

File:../data/experiments/base\_cased/training\_folder/  
no\_attention/predictions\_strong\_matching/ace2004  
checkpoint=54, opt\_thr=0.03, model=local:  
pem\_similarityglobal: local

Sample: b'20001115\_AFP\_ARB\_0093\_eng'

[1 1 Bandar Seri Begawan 1] 11-15 -LRB- [2 2 AFP 2] -RRB- -  
The [3 3 United States 3] today Wednesday deemed the order  
issued by [1 Palestinian 1] President [2 Yasser Arafat 2] for  
a ceasefire in territories under [3 Palestinian Authority 3]  
control as a positive gesture but considered that it does not  
release constitute a release form the terms of the [4 Sharm 4]  
el-Sheikh agreement . James Stewart , the [4 4 White House 4]  
's spokesman in [5 Bandar Seri Begawan 5] , the capital of  
the [6 Sultanate of Brunei 6] which [7 American 7] President [8  
8 Bill Clinton 8] is visiting , said `` of course we  
positively welcome the announcement aimed at stopping the  
violence . But the important point is that [9 Palestinian 9]  
and [10 Israeli 10] officials take the right steps to provide  
a safer climate in the whole region . '' [11 Clinton 11] is  
to take part in the [12 Asia-Pacific Economic Cooperation 12]  
forum -LRB- [13 APEC 13] -RRB- in [14 Brunei 14] . The [15  
American 15] spokesman said President [16 Bill Clinton 16]  
spoke to [17 Arafat 17] on the telephone after arriving in [18  
Brunei 18] yesterday evening Tuesday . He said `` their  
conversation focused on ways to stop the violence because we  
believe that it is extremely important for the sides to take  
immediate measures to end the tension and reduce the  
violence . '' A high-level [19 Palestinian 19] source said  
today that the 1 Palestinian Superior Security Council led by  
President [20 Yasser Arafat 20] issued orders to stop  
shooting inside area `` A '' under [21 Palestinian 21]  
control . The source who wished to remain anonymous said that  
`` the Palestinian Superior Security Council led by President  
[22 Yasser Arafat 22] issued orders to stop unknown armed  
persons from shooting inside zone `` A '' under complete [23  
Palestinian 23] control and from between buildings and  
civilians in the [24 Palestinian Territories 24] . '' Zone ``  
A '' covers areas controlled completely by the [25  
Palestinian Authority 25] in the [26 West Bank 26] and [27  
Gaza Strip 27] .

1: 56611 Bandar\_Seri\_Begawan

2: 471981 Agence\_France-Presse

3: 3434750 United\_States

4: 33057 White\_House

1: 41207485 Palestinian\_National\_Council

[winner: 241405 State\_of\_Palestine, pem\_pos=9,  
final\_scores=-0.9339542, final\_local\_scores=-1.0924225,  
pem=-2.8508017, similarity=-13.007247]

RECALL MISS: 41207485 Palestinian\_National\_Council]

[winner: 253407 Council, pem\_pos=0,

final\_scores=-0.9413133, final\_local\_scores=-0.95077366,  
pem=-1.7500058, similarity=-14.766465]  
RECALL MISS: 41207485 Palestinian\_National\_Council]  
[winner: 27890274 Superior, West\_Virginia, pem\_pos=5,  
final\_scores=-1.3416647, final\_local\_scores=-1.2932777,  
pem=-2.619499, similarity=-20.25903]  
RECALL MISS: 41207485 Palestinian\_National\_Council]  
[winner: 11221224 Adventures\_in\_Freestyle, pem\_pos=6,  
final\_scores=-1.602793, final\_local\_scores=-1.5644922,  
pem=-2.4593651, similarity=-29.223446]  
RECALL MISS: 41207485 Palestinian\_National\_Council]  
1: 56611 Bandar\_Seri\_Begawan  
[pem\_pos=0, final\_scores=0.25304186,  
final\_local\_scores=0.30817372, pem=0.0, similarity=13.286231]  
2: 471981 Agence\_France\_Presse  
[pem\_pos=0, final\_scores=0.10071181,  
final\_local\_scores=0.11393696, pem=-0.6605333,  
similarity=11.081902]  
3: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.43231076,  
final\_local\_scores=0.43853378, pem=-0.073457584,  
similarity=17.57591]  
4: 33057 White\_House  
[pem\_pos=0, final\_scores=0.4231022,  
final\_local\_scores=0.42864668, pem=-0.0056634876,  
similarity=16.912184]  
1: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.7267313,  
final\_local\_scores=0.7253932, pem=-2.8508017,  
similarity=41.239994]  
2: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.6060758,  
final\_local\_scores=0.5930039, pem=0.0, similarity=21.786133]  
3: 24093 Palestinian\_National\_Authority  
[pem\_pos=0, final\_scores=0.4471243,  
final\_local\_scores=0.41463107, pem=0.0, similarity=16.46313]  
4: 419335 Sharm\_el-Sheikh  
[pem\_pos=0, final\_scores=0.14679442,  
final\_local\_scores=0.15007901, pem=0.0, similarity=8.56837]  
5: 56611 Bandar\_Seri\_Begawan  
[pem\_pos=0, final\_scores=0.21440138,  
final\_local\_scores=0.2659649, pem=0.0, similarity=12.026636]  
6: 3466 Brunei  
[pem\_pos=0, final\_scores=0.18645881,  
final\_local\_scores=0.18190506, pem=0.0, similarity=9.518123]  
7: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.5544532,  
final\_local\_scores=0.5719558, pem=-0.12994486,  
similarity=21.864676]  
8: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.54541594,  
final\_local\_scores=0.54993296, pem=0.0, similarity=20.500809]

9: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.5250942,  
final\_local\_scores=0.5051353, pem=-2.8508017,  
similarity=34.66706]  
10: 9282173 Israel  
[pem\_pos=0, final\_scores=0.49228317,  
final\_local\_scores=0.4726085, pem=-0.34790626,  
similarity=20.08526]  
11: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.83591825,  
final\_local\_scores=0.8727538, pem=0.0, similarity=30.134426]  
12: 135976 Asia-Pacific\_Economic\_Cooperation  
[pem\_pos=0, final\_scores=0.41221386,  
final\_local\_scores=0.4345218, pem=0.0, similarity=17.05671]  
13: 135976 Asia-Pacific\_Economic\_Cooperation  
[pem\_pos=0, final\_scores=0.05531381,  
final\_local\_scores=0.04213603, pem=0.0, similarity=5.347136]  
14: 3466 Brunei  
[pem\_pos=0, final\_scores=0.27488977,  
final\_local\_scores=0.27850243, pem=-0.386843,  
similarity=14.504494]  
15: 3434750 United\_States  
[pem\_pos=1, final\_scores=0.4563015,  
final\_local\_scores=0.46541625, pem=-0.12994486,  
similarity=18.685324]  
16: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.59433204,  
final\_local\_scores=0.6033663, pem=0.0, similarity=22.095367]  
17: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.57457244,  
final\_local\_scores=0.56268305, pem=0.0, similarity=20.881298]  
18: 3466 Brunei  
[pem\_pos=0, final\_scores=0.2838853,  
final\_local\_scores=0.28832868, pem=-0.386843,  
similarity=14.797729]  
19: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.52387136,  
final\_local\_scores=0.50379956, pem=-2.8508017,  
similarity=34.6272]  
20: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.61114293,  
final\_local\_scores=0.59853894, pem=0.0, similarity=21.95131]  
21: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.5382374,  
final\_local\_scores=0.51949227, pem=-2.8508017,  
similarity=35.0955]  
22: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.6152013,  
final\_local\_scores=0.6029721, pem=0.0, similarity=22.083603]  
23: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.15776522,  
final\_local\_scores=0.10447365, pem=-2.8508017,

similarity=22.710518]  
24: 96665 Palestinian\_territories  
[pem\_pos=0, final\_scores=0.06998612,  
final\_local\_scores=0.020381153, pem=0.0, similarity=4.697927]  
25: 24093 Palestinian\_National\_Authority  
[pem\_pos=0, final\_scores=0.30428898,  
final\_local\_scores=0.25860518, pem=0.0, similarity=11.807008]  
26: 96665 Palestinian\_territories  
[pem\_pos=1, final\_scores=0.031510457,  
final\_local\_scores=-0.02612567, pem=-1.929408,  
similarity=13.802491]  
27: 12047 Gaza\_Strip  
[pem\_pos=0, final\_scores=0.34355348,  
final\_local\_scores=0.33000073, pem=0.0, similarity=13.9375925]

Sample: b'PRI20001128\_2000\_0055'

Documented . From [1 1 NPR 1] news in [1 1 Washington 1] , I  
'm Corey Flintoff . A Florida State judge has ordered local  
election officials to ship thousands of ballots to the state  
capital . The move comes as the judge decides whether to  
order a recount of those ballots in the presidential  
election . [2 NPR 2] 's Steve Inski was in the courtroom .  
Steve , what are the issues that came up in this hearing  
tonight ? Corey judge N. Sander Sauls is deciding whether to  
order a recount as you said of those ballots , 9000 or more  
actually from [2 2 Miami Dade County 2] that apparently were  
never counted at all , at least according to the [3 3 Democra  
ts 3] in the presidential election and about 3300 ballots  
from [4 4 Palm Beach County 4] that were considered by local  
officials but then thrown out , disallowed . Judge Sauls is  
deciding whether to count those ballots and go through them  
and perhaps add some votes to [5 5 Al Gore 5] 's final  
total , which is exactly what the [3 Democrats 3] want . He  
is not ready to do that yet . He did not agree to the [4  
Democratic 4] proposal to begin counting tomorrow morning but  
he has ordered that the ballots be shipped up . Later this  
week , Miami Police apparently are going to escort them upto [5  
Tallahassee 5] and they will be ready to be counted if the  
judge decides at a later time to do so . Now time is a big  
consideration from here . What 's the judge 's timetable ?  
Not as fast as the [6 Democrats 6] would like , not as slowly  
as [7 Republicans 7] would prefer . He is going to order a  
major hearing on Saturday to review all the legal and  
evidentiary issues , look at evidence , perhaps hear from  
witnesses and after that the judge will decide if there are  
any ballots to count . It might take another five days or so  
to do it . Neither side seemed particularly pleased by the  
timetable but the judge said , well , I am going to make  
everybody equally unhappy . Thank you . [8 NPR 8] 's Steven  
Inski reporting from [9 Tallahassee 9] , [10 Florida 10] . As  
the battle over votes goes on in [11 Florida 11] , 2 Washingt  
on State is in the middle of recounting the votes cast in its

very tight US Senate race . Jennifer Nesan reports from member station KPLU in [6 6 Seattle 6] . The recount of the votes cast for [12 Democrat 12] Maria Catwell and [13 Republican 13] Senator Slate Gordon is calm and organized . The State 's Elections Director Garry Macintosh says the process is about as exciting as watching paint dry . There are a lot of things about the election 's process I think that are fairly complex . This is n't probably one of them . Catwell leads [14 Republican 14] incumbent Slate Gordon by fewer then 2000 votes . At this point neither of the candidates plan to contest any of the votes . State law requires a recount when the margin of victory is less than one half of 1 percent . The recount is expected to be completed by the end of the week . No statewide recount in [15 Washington 15] has ever changed the initial outcome . For [16 NPR 16] news , I 'm Jennifer Nesan in [17 Seattle 17] .

1: 108956 Washington,\_D.C.

2: 27905225 Miami-Dade\_County,\_Florida

3: 5043544 Democratic\_Party\_(United\_States)

4: 73647 Palm\_Beach\_County,\_Florida

5: 5042706 Al\_Gore

6: 11388236 Seattle

1: 29697232 NPR

[winner: 29697232 NPR, pem\_pos=0,  
final\_scores=0.19449793, final\_local\_scores=0.18908387,  
pem=0.0, similarity=9.732353]

gt: 29697232 NPR, pem\_pos=0, final\_scores=0.19449793,  
final\_local\_scores=0.18908387, pem=0.0, similarity=9.732353 ]

[winner: 961853 Newsreader\_(Usenet), pem\_pos=4,  
final\_scores=-1.5506979, final\_local\_scores=-1.5400653,  
pem=-2.1872969, similarity=-29.974043]

RECALL MISS: 29697232 NPR]

2: 13015878 Washington\_(state)

[winner: 228600 Washington\_State\_University, pem\_pos=0,  
final\_scores=0.008222505, final\_local\_scores=0.007250622,  
pem=-1.1284691, similarity=10.442877]

gt: 13015878 Washington\_(state), pem\_pos=2,  
final\_scores=-0.15067299, final\_local\_scores=-0.20107716,  
pem=-1.9198062, similarity=8.529374 ]

1: 108956 Washington,\_D.C.

[pem\_pos=0, final\_scores=0.57117635,  
final\_local\_scores=0.5431863, pem=-1.7496481,  
similarity=29.814333]

2: 27905225 Miami-Dade\_County,\_Florida

[pem\_pos=0, final\_scores=0.3340724,  
final\_local\_scores=0.3286753, pem=0.0, similarity=13.89804]

3: 5043544 Democratic\_Party\_(United\_States)

[pem\_pos=0, final\_scores=0.48706847,  
final\_local\_scores=0.46068645, pem=-0.18007085,  
similarity=18.816769]

4: 73647 Palm\_Beach\_County,\_Florida

[pem\_pos=0, final\_scores=0.3042606,

final\_local\_scores=0.3164174, pem=0.0, similarity=13.532239]  
5: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.3555675,  
final\_local\_scores=0.35372487, pem=0.0, similarity=14.645569]  
6: 11388236 Seattle  
[pem\_pos=0, final\_scores=0.51493496,  
final\_local\_scores=0.5108702, pem=-0.1054736,  
similarity=19.908682]  
1: 29697232 NPR  
[pem\_pos=0, final\_scores=0.19449793,  
final\_local\_scores=0.18908387, pem=0.0, similarity=9.732353]  
2: 29697232 NPR  
[pem\_pos=0, final\_scores=0.2312461,  
final\_local\_scores=0.22922564, pem=0.0, similarity=10.9302635]  
3: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.50164765,  
final\_local\_scores=0.4766119, pem=-0.18007085,  
similarity=19.292017]  
4: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.4730062,  
final\_local\_scores=0.4453255, pem=-0.076093495,  
similarity=17.792923]  
5: 57700 Tallahassee,\_Florida  
[pem\_pos=0, final\_scores=0.3318503,  
final\_local\_scores=0.30137986, pem=-0.024764989,  
similarity=13.218165]  
6: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.48799354,  
final\_local\_scores=0.46169692, pem=-0.18007085,  
similarity=18.846924]  
7: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.47259957,  
final\_local\_scores=0.44599694, pem=-0.11427376,  
similarity=18.02059]  
8: 29697232 NPR  
[pem\_pos=0, final\_scores=0.20955713,  
final\_local\_scores=0.20553374, pem=0.0, similarity=10.223249]  
9: 57700 Tallahassee,\_Florida  
[pem\_pos=0, final\_scores=0.4511674,  
final\_local\_scores=0.43171573, pem=-0.024764989,  
similarity=17.107647]  
10: 18933066 Florida  
[pem\_pos=0, final\_scores=0.18844123,  
final\_local\_scores=0.15200049, pem=-0.15405534,  
similarity=9.463488]  
11: 18933066 Florida  
[pem\_pos=0, final\_scores=0.12475465,  
final\_local\_scores=0.08243257, pem=-0.15405534,  
similarity=7.387442]  
12: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.33819872,  
final\_local\_scores=0.2980688, pem=-0.061652232,

```

similarity=13.319955]
13: 32070 Republican_Party_(United_States)
    [pem_pos=0, final_scores=0.36693054,
final_local_scores=0.33056965, pem=-0.1310191,
similarity=14.667073]
14: 32070 Republican_Party_(United_States)
    [pem_pos=0, final_scores=0.41853118,
final_local_scores=0.38693547, pem=-0.1310191,
similarity=16.349142]
15: 108956 Washington,_D.C.
    [pem_pos=0, final_scores=0.4271124,
final_local_scores=0.38581836, pem=-1.7496481,
similarity=25.11816]
16: 29697232 NPR
    [pem_pos=0, final_scores=0.13164799,
final_local_scores=0.12042984, pem=0.0, similarity=7.6835794]
17: 11388236 Seattle
    [pem_pos=0, final_scores=0.3586554,
final_local_scores=0.34015858, pem=-0.1054736,
similarity=14.814305]

```

Sample: b'20001115\_AFP\_ARB\_0072\_eng'

[1 1 Baghdad 1] 11-15 -LRB- 1 AFP -RRB- - The head of the French Interests Section in [1 Baghdad 1] , Ambassador Andre Janier , today Wednesday welcomed [2 2 Iraq 2] 's decision to deal in euros and said this would encourage relations between the two countries . In an interview in today 's edition of the official newspaper `` al-Jamhouriya , ' ' Janier said `` we [2 French 2] believe this decision is positive for the international economy , and this decision encourages economic relations , especially in terms of oil , between [3 Iraq 3] and [4 France 4] . ' ' The [5 French 5] diplomat carried on to say `` we have a vested interest in this matter as [6 Europeans 6] and [7 French 7] because all trade payments and exchanges are paid in euros and there is no danger to [8 European 8] and [9 French 9] companies because of fluctuations in the price of the dollar on the stock exchange . This compensates for any loss in buying and selling in euros rather than dollars . ' ' [10 Iraq 10] decided to deal in euros , rather than the dollar , as of the sixth of this month for its oil sales and outside trade deals .

1: 4492 Baghdad

2: 7515928 Iraq

1: 471981 Agence\_France-Presse

```

    [winner: 471981 Agence_France-Presse, pem_pos=0,
final_scores=-0.02723211, final_local_scores=-0.012810551,
pem=-0.6605333, similarity=7.2995033]

```

```

    gt: 471981 Agence_France-Presse, pem_pos=0,
final_scores=-0.02723211, final_local_scores=-0.012810551,
pem=-0.6605333, similarity=7.2995033 ]

```

1: 4492 Baghdad

[pem\_pos=0, final\_scores=0.20703246,  
final\_local\_scores=0.17424792, pem=0.0, similarity=9.2896185]  
2: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.60147613,  
final\_local\_scores=0.5804516, pem=-0.25379938,  
similarity=22.791748]  
1: 4492 Baghdad  
[pem\_pos=0, final\_scores=0.33362722,  
final\_local\_scores=0.31253347, pem=0.0, similarity=13.416335]  
2: 5843419 France  
[pem\_pos=0, final\_scores=0.51553243,  
final\_local\_scores=0.4720605, pem=-0.80223274,  
similarity=22.539606]  
3: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.6188285,  
final\_local\_scores=0.5994064, pem=-0.25379938,  
similarity=23.357397]  
4: 5843419 France  
[pem\_pos=0, final\_scores=0.5369143,  
final\_local\_scores=0.49541694, pem=-0.16240834,  
similarity=19.757143]  
5: 5843419 France  
[pem\_pos=0, final\_scores=0.7015657,  
final\_local\_scores=0.6752736, pem=-0.80223274,  
similarity=28.603891]  
6: 9239 Europe  
[pem\_pos=0, final\_scores=0.5837738,  
final\_local\_scores=0.573416, pem=-0.23495062,  
similarity=22.479288]  
7: 5843419 France  
[pem\_pos=0, final\_scores=0.589631,  
final\_local\_scores=0.55300194, pem=-0.80223274,  
similarity=24.955061]  
8: 9239 Europe  
[pem\_pos=0, final\_scores=0.45057315,  
final\_local\_scores=0.4279145, pem=-0.25557294,  
similarity=18.24938]  
9: 5843419 France  
[pem\_pos=0, final\_scores=0.70463747,  
final\_local\_scores=0.6786291, pem=-0.80223274,  
similarity=28.704025]  
10: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.4982851,  
final\_local\_scores=0.46773112, pem=-0.25379938,  
similarity=19.427942]

Sample: b'20001115\_AFP\_ARB\_0210\_eng'

[1 1 Tallahassee 1] -LRB- [2 2 United States 2] -RRB- 11-15 -  
LRB- 1 AFP -RRB- - The 2 Supreme Court in 3 Florida today  
Wednesday refused the application by the state 's authorities  
to stop the new hand count of votes in some counties .  
1: 57700 Tallahassee, Florida



2: 3434750 United\_States

1: 471981 Agence\_France-Presse

[winner: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=-0.017736569, final\_local\_scores=0.05808711,  
pem=-0.6605333, similarity=9.415231]

gt: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=-0.017736569, final\_local\_scores=0.05808711,  
pem=-0.6605333, similarity=9.415231 ]

2: 1635683 Supreme\_Court\_of\_Florida

[winner: 31737 Supreme\_Court\_of\_the\_United\_States,  
pem\_pos=0, final\_scores=-0.03512071,  
final\_local\_scores=-0.018858, pem=-0.9940822,  
similarity=8.932926]

RECALL MISS: 1635683 Supreme\_Court\_of\_Florida]

3: 18933066 Florida

[winner: 18933066 Florida, pem\_pos=0,  
final\_scores=-0.014775202, final\_local\_scores=-0.0651268,  
pem=-0.15405534, similarity=2.9839764]

gt: 18933066 Florida, pem\_pos=0,  
final\_scores=-0.014775202, final\_local\_scores=-0.0651268,  
pem=-0.15405534, similarity=2.9839764 ]

1: 57700 Tallahassee, \_Florida

[pem\_pos=0, final\_scores=0.05096735,  
final\_local\_scores=0.0722851, pem=-0.024764989,  
similarity=6.3815203]

2: 3434750 United\_States

[pem\_pos=0, final\_scores=0.28136724,  
final\_local\_scores=0.2830722, pem=-0.073457584,  
similarity=12.936626]

Sample: b'20000815\_AFP\_ARB\_0071\_eng'

[1 1 Beirut 1] 8-15 -LRB- [2 2 AFP 2] -RRB- - [3 3 Lebanon 3]  
has asked the international peacekeeping force operating in  
the south to investigate reports indicating that [4 4 Israel 4  
] is penetrating the shared border between the two countries  
and to prevent this from happening . According to official  
sources , the [1 Lebanese 1] authorities contacted the  
peacekeeping high command and asked it to put an end to  
infringements by [2 Israel 2] , which is carrying out  
operations inside [3 Lebanese 3] territory , in particular  
dumping sand and building slopes . The source added that if  
this is confirmed , then such acts constitute a provocation  
and an infringement of [4 Lebanese 4] sovereignty . Witnesses  
said two [5 Israeli 5] bulldozers started work today Tuesday  
on a slope in the Abbad sector inside [6 Lebanese 6]  
territory on the border between [7 Lebanon 7] and [8 Israel 8]  
. These witnesses told [9 Agence France Presse 9] that an [10  
Israeli 10] army patrol equipped with ultra-modern binoculars  
penetrated a few meters into [11 Lebanese 11] territory but  
withdrew back to [12 Israeli 12] territory as soon as a  
peacekeeping patrol arrived . Families in southern [13  
Lebanon 13] visit this area where , according to tradition ,

the burial site of Sheikh Abbad , who died 500 years ago , is located . For their part , [14 Israelis 14] also regard the site as a shrine because , according to them , [15 Rabbi Ashi 15] lived there in the 5th century . The [16 Israeli 16] army had previously carried out operations inside [17 Lebanese 17] territory in the same border region on August 11 . Last Friday , an [18 Israeli 18] bulldozer crossed the wire fence in Wadi Marjayoun near the [19 Israeli 19] village of Matla and spread sand over a length of hundreds of meters and a width of six meters in order to expose the footprints of people crossing the border or tire prints of cars from the [20 Lebanese 20] side .

1: 37428 Beirut

2: 471981 Agence\_France-Presse

3: 17771 Lebanon

4: 9282173 Israel

1: 37428 Beirut

[pem\_pos=0, final\_scores=0.4092031,  
final\_local\_scores=0.37020463, pem=-0.046614014,  
similarity=15.390852]

2: 471981 Agence\_France-Presse

[pem\_pos=0, final\_scores=0.082000926,  
final\_local\_scores=0.0977194, pem=-0.6605333,  
similarity=10.597938]

3: 17771 Lebanon

[pem\_pos=0, final\_scores=0.55716044,  
final\_local\_scores=0.496422, pem=-0.4097839,  
similarity=21.132402]

4: 9282173 Israel

[pem\_pos=0, final\_scores=0.68402666,  
final\_local\_scores=0.63868874, pem=-0.21279564,  
similarity=24.306675]

1: 17771 Lebanon

[pem\_pos=0, final\_scores=0.59200287,  
final\_local\_scores=0.53448206, pem=-0.55808944,  
similarity=23.074701]

2: 9282173 Israel

[pem\_pos=0, final\_scores=0.6804989,  
final\_local\_scores=0.6348352, pem=-0.21279564,  
similarity=24.191675]

3: 17771 Lebanon

[pem\_pos=0, final\_scores=0.62932855,  
final\_local\_scores=0.57525474, pem=-0.55808944,  
similarity=24.29144]

4: 17771 Lebanon

[pem\_pos=0, final\_scores=0.6428148,  
final\_local\_scores=0.5899864, pem=-0.55808944,  
similarity=24.73106]

5: 9282173 Israel

[pem\_pos=0, final\_scores=0.61437386,  
final\_local\_scores=0.56260365, pem=-0.34790626,  
similarity=22.770897]

6: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.5838223,  
final\_local\_scores=0.525546, pem=-0.55808944,  
similarity=22.808031]  
7: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.56363386,  
final\_local\_scores=0.50349325, pem=-0.4097839,  
similarity=21.343424]  
8: 9282173 Israel  
[pem\_pos=0, final\_scores=0.55397195,  
final\_local\_scores=0.4966237, pem=-0.21279564,  
similarity=20.06717]  
9: 471981 Agence\_France\_Presse  
[pem\_pos=0, final\_scores=0.46152377,  
final\_local\_scores=0.5122904, pem=0.0, similarity=19.377481]  
10: 9282173 Israel  
[pem\_pos=0, final\_scores=0.58150476,  
final\_local\_scores=0.5266991, pem=-0.34790626,  
similarity=21.699434]  
11: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.6240862,  
final\_local\_scores=0.5695282, pem=-0.55808944,  
similarity=24.120548]  
12: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5398342,  
final\_local\_scores=0.48118037, pem=-0.34790626,  
similarity=20.341063]  
13: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.53772324,  
final\_local\_scores=0.47518986, pem=-0.4097839,  
similarity=20.498795]  
14: 9282173 Israel  
[pem\_pos=1, final\_scores=0.3341152,  
final\_local\_scores=0.25646365, pem=-0.5460739,  
similarity=14.712735]  
15: 59044 Rav\_Ashi  
[pem\_pos=0, final\_scores=0.17260896,  
final\_local\_scores=0.29923192, pem=0.0, similarity=13.01939]  
16: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5805647,  
final\_local\_scores=0.52567226, pem=-0.34790626,  
similarity=21.668789]  
17: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.57237965,  
final\_local\_scores=0.5130467, pem=-0.55808944,  
similarity=22.435026]  
18: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5203204,  
final\_local\_scores=0.45986456, pem=-0.34790626,  
similarity=19.704958]  
19: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5594003,

final\_local\_scores=0.5025534, pem=-0.34790626,  
similarity=20.978878]  
20: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.5581704,  
final\_local\_scores=0.49752522, pem=-0.55808944,  
similarity=21.971834]

Sample: b'VOA20001020\_2100\_1853'

The [1 US 1] 1 navy now says the [2 2 USS Cole 2] was being  
refueled when an explosion ripped through it in [1 1 Yemen 1]  
last week , killing 17 . The revised accounting of the  
incident was given in a navy statement Friday raising new  
questions about how the small boat carrying the explosives  
was able to get near the ship and set off the blast .

1: 350939 Yemen

1: 20518076 United\_States\_Navy

[winner: 20518076 United\_States\_Navy, pem\_pos=0,  
final\_scores=-1.2856207, final\_local\_scores=-1.3994505,  
pem=-1.0602508, similarity=-31.906872]

gt: 20518076 United\_States\_Navy, pem\_pos=0,  
final\_scores=-1.2856207, final\_local\_scores=-1.3994505,  
pem=-1.0602508, similarity=-31.906872 ]

2: 197334 USS\_Cole\_(DDG-67)

[winner: 31981 USS\_Cole\_bombing, pem\_pos=0,  
final\_scores=0.05776559, final\_local\_scores=0.10422009,  
pem=-0.15458126, similarity=8.040485]

gt: 197334 USS\_Cole\_(DDG-67), pem\_pos=1,  
final\_scores=-0.1329229, final\_local\_scores=-0.04624909,  
pem=-0.39582554, similarity=4.8621087 ]

1: 350939 Yemen

[pem\_pos=0, final\_scores=0.4772607,  
final\_local\_scores=0.4940638, pem=-0.27215755,  
similarity=20.313597]

1: 3434750 United\_States

[pem\_pos=0, final\_scores=0.11615594,  
final\_local\_scores=0.14977562, pem=-0.48755783,  
similarity=11.2107315]

2: 31981 USS\_Cole\_bombing

[pem\_pos=0, final\_scores=0.05776559,  
final\_local\_scores=0.10422009, pem=-0.15458126,  
similarity=8.040485]

Sample: b'20001115\_AFP\_ARB\_0184\_eng'

[1 1 Jerusalem 1] 11-15 -LRB- [2 2 AFP 2] -RRB- - Official [1  
Israeli 1] radio reported that the lesser security cabinet  
which met this evening Wednesday has made a series of  
practical decisions pointing to its desire to avoid  
confrontations with the [2 Palestinians 2] . The radio  
reported `` senior political officials '' as saying at the  
close of the security cabinet 's meeting , chaired by Prime  
Minister [3 Ehud Barak 3] , that the army would continue with  
the procedure it has followed over the past days and that

there would be no escalation . Military officials who also attended the meeting that [4 Palestinian 4] President [5 Yasser Arafat 5] `` wants to drag [6 Israel 6] into an escalation which would lead to international intervention . '' The head of the prime minister 's office , Jalaad Shir , had previously said on official television that `` a compromise must be found between a studied response not leading to escalation and the need to pave the way for political openness , on the condition that the [7 Palestinians 7] resolutely stop the violence and incitement to violence . '' Shir added `` only then will it be possible to think about resuming the political process and on what basis . '' The security cabinet was convened to a session with the aim of examining new measures to deal with the escalation in attacks against [8 Israelis 8] in the [9 West Bank 9] and [10 Gaza Strip 10] . Last week , [11 Barak 11] gave his consent , for the first time since the [12 Intifada 12] began , to elimination of Hussein Abiyat the military leader of [13 Fatah 13] movement led by [14 Arafat 14] killed when a missile fired from a combat helicopter targeted his car . [15 Israeli 15] security forces arrested 15 [16 Fatah 16] officials in the [17 1 West Bank 17] on the night of Tuesday to Wednesday . In his statement , [18 Barak 18] praised this operation and the elimination of Abiyat , stressing that this policy would continue . He said `` the long arm of our forces will reach those who attack our civilians and soldiers . ''

1: 16043 Jerusalem

2: 471981 Agence\_France-Presse

1: 33209 West\_Bank

[winner: 96665 Palestinian\_territories, pem\_pos=1, final\_scores=0.09886728, final\_local\_scores=0.04294163, pem=-1.929408, similarity=15.863597]

gt: 33209 West\_Bank, pem\_pos=0, final\_scores=-0.22582509, final\_local\_scores=-0.3332244, pem=-0.053944394, similarity=-5.5610094 ]

1: 16043 Jerusalem

[pem\_pos=0, final\_scores=0.27571386, final\_local\_scores=0.2542159, pem=-0.13239029, similarity=12.395981]

2: 471981 Agence\_France-Presse

[pem\_pos=0, final\_scores=0.14966382, final\_local\_scores=0.16402277, pem=-0.6605333, similarity=12.576562]

1: 9282173 Israel

[pem\_pos=0, final\_scores=0.4039325, final\_local\_scores=0.3601024, pem=-0.34790626, similarity=16.727854]

2: 241405 State\_of\_Palestine

[pem\_pos=5, final\_scores=0.37120527, final\_local\_scores=0.34686178, pem=-1.7909594, similarity=24.180277]

3: 168212 Ehud\_Barak  
[pem\_pos=0, final\_scores=0.6696437,  
final\_local\_scores=0.6570201, pem=0.0, similarity=23.696503]  
4: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.40067738,  
final\_local\_scores=0.37905562, pem=-2.8508017,  
similarity=30.904593]  
5: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.6629655,  
final\_local\_scores=0.6339933, pem=0.0, similarity=23.009338]  
6: 9282173 Israel  
[pem\_pos=0, final\_scores=0.63348836,  
final\_local\_scores=0.6108573, pem=-0.21279564,  
similarity=23.476128]  
7: 241405 State\_of\_Palestine  
[pem\_pos=5, final\_scores=0.43165857,  
final\_local\_scores=0.41289788, pem=-1.7909594,  
similarity=26.150925]  
8: 9282173 Israel  
[pem\_pos=1, final\_scores=0.38273764,  
final\_local\_scores=0.34004715, pem=-0.5460739,  
similarity=17.207033]  
9: 96665 Palestinian\_territories  
[pem\_pos=1, final\_scores=0.030070469,  
final\_local\_scores=-0.03754288, pem=-1.929408,  
similarity=13.461779]  
10: 12047 Gaza\_Strip  
[pem\_pos=0, final\_scores=0.464024,  
final\_local\_scores=0.42801845, pem=0.0, similarity=16.862637]  
11: 168212 Ehud\_Barak  
[pem\_pos=0, final\_scores=0.54343116,  
final\_local\_scores=0.5243674, pem=0.0, similarity=19.737883]  
12: 71712 First\_Intifada  
[pem\_pos=0, final\_scores=0.27021807,  
final\_local\_scores=0.23132113, pem=-0.066496804,  
similarity=11.354417]  
13: 11104 Fatah  
[pem\_pos=0, final\_scores=0.39298886,  
final\_local\_scores=0.3540807, pem=0.0, similarity=14.656187]  
14: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.60134137,  
final\_local\_scores=0.5704605, pem=0.0, similarity=21.113392]  
15: 9282173 Israel  
[pem\_pos=0, final\_scores=0.55611306,  
final\_local\_scores=0.5263365, pem=-0.34790626,  
similarity=21.688612]  
16: 11104 Fatah  
[pem\_pos=0, final\_scores=0.4083886,  
final\_local\_scores=0.37090254, pem=0.0, similarity=15.158184]  
17: 96665 Palestinian\_territories  
[pem\_pos=1, final\_scores=0.09886728,  
final\_local\_scores=0.04294163, pem=-1.929408,

similarity=15.863597]  
18: 168212 Ehud\_Barak  
[pem\_pos=0, final\_scores=0.59116286,  
final\_local\_scores=0.576507, pem=0.0, similarity=21.29383]

Sample: b'PRI20001122\_2000\_0320'

The [1 1 Yasser Arafat 1]  
said today that , it will avenge the deaths of four of its  
members who were killed today by [2 Israeli 2] soldiers .  
Later in the day , a car bombing in the Northern Israeli town  
of [3 Hadera 3] , killed two people and injured at least 50 .  
[4 Israeli 4] Prime Minister [5 Ehud Barak 5] said he would  
retaliate . 1 Corey Flintoff , [6 2 NPR 6] news , [2 2 Washin  
gton 2] .

1: 34289 Yasser\_Arafat  
2: 108956 Washington,\_D.C.

1: 10448282 Corey\_Flintoff

[winner: 10448282 Corey\_Flintoff, pem\_pos=0,  
final\_scores=-0.00792481, final\_local\_scores=0.15937367,  
pem=0.0, similarity=8.845741]

gt: 10448282 Corey\_Flintoff, pem\_pos=0,  
final\_scores=-0.00792481, final\_local\_scores=0.15937367,  
pem=0.0, similarity=8.845741 ]

2: 29697232 NPR

[winner: 29697232 NPR, pem\_pos=0,  
final\_scores=0.06592266, final\_local\_scores=0.12179935,  
pem=0.0, similarity=7.724448]

gt: 29697232 NPR, pem\_pos=0, final\_scores=0.06592266,  
final\_local\_scores=0.12179935, pem=0.0, similarity=7.724448 ]

[winner: 961853 Newsreader\_(Usenet), pem\_pos=4,  
final\_scores=-1.6245427, final\_local\_scores=-1.6141624,  
pem=-2.1872969, similarity=-32.18525]

RECALL MISS: 29697232 NPR]

1: 34289 Yasser\_Arafat

[pem\_pos=0, final\_scores=0.5807961,  
final\_local\_scores=0.58720654, pem=0.0, similarity=21.613127]

2: 108956 Washington,\_D.C.

[pem\_pos=0, final\_scores=0.2912479,  
final\_local\_scores=0.30225736, pem=-1.7496481,  
similarity=22.624533]

1: 241405 State\_of\_Palestine

[pem\_pos=9, final\_scores=0.50084716,  
final\_local\_scores=0.5218458, pem=-2.8508017,  
similarity=35.165737]

2: 9282173 Israel

[pem\_pos=1, final\_scores=0.55618423,  
final\_local\_scores=0.5265754, pem=-0.34790626,  
similarity=21.69574]

3: 323304 Hadera

[pem\_pos=0, final\_scores=0.24268423,  
final\_local\_scores=0.28094772, pem=0.0, similarity=12.473753]

4: 9282173 Israel

[pem\_pos=0, final\_scores=0.5457534,  
final\_local\_scores=0.51518124, pem=-0.34790626,  
similarity=21.355719]  
5: 168212 Ehud\_Barak  
[pem\_pos=0, final\_scores=0.5800683,  
final\_local\_scores=0.6109032, pem=0.0, similarity=22.320284]  
6: 29697232 NPR  
[pem\_pos=0, final\_scores=0.06592266,  
final\_local\_scores=0.12179935, pem=0.0, similarity=7.724448]

Sample: b'20001115\_AFP\_ARB\_0061\_eng'

[1 1 Gaza 1] 11-15 -LRB- [1 1 AFP 1] -RRB- - A high-level [2  
Palestinian 2] source said today Wednesday that the  
Palestinian Superior Security Council led by President [3  
Yasser Arafat 3] issued orders to stop shooting inside area  
`` A '' under [4 Palestinian 4] control . The source said  
that `` the Palestinian Superior Security Council led by  
President [5 Yasser Arafat 5] issued orders to stop unknown  
armed persons from shooting inside zone `` A '' under  
complete [6 Palestinian 6] control and between buildings and  
civilians in the [7 Palestinian Territories 7] . '' The [8  
Palestinian 8] decision comes a few hours before discussions  
between [9 Israeli 9] Prime Minister [10 Ehud Barak 10] and  
his top security advisors to explore the measures to be taken  
against [11 Palestinians 11] after four [12 Israeli 12]  
soldiers were shot in [13 Palestinian 13] areas last Monday .  
1: 471981 Agence\_France-Presse

1: 55387 Gaza

[winner: 12047 Gaza\_Strip, pem\_pos=1,  
final\_scores=0.056817934, final\_local\_scores=-0.00673151,  
pem=-1.374985, similarity=11.366214]

gt: 55387 Gaza, pem\_pos=0, final\_scores=-0.10510437,  
final\_local\_scores=-0.17772657, pem=-0.30319342,  
similarity=0.43481058 ]

1: 471981 Agence\_France-Presse

[pem\_pos=0, final\_scores=0.089916036,  
final\_local\_scores=0.11688453, pem=-0.6605333,  
similarity=11.169863]

1: 12047 Gaza\_Strip

[pem\_pos=1, final\_scores=0.056817934,  
final\_local\_scores=-0.00673151, pem=-1.374985,  
similarity=11.366214]

2: 241405 State\_of\_Palestine

[pem\_pos=9, final\_scores=0.5773781,  
final\_local\_scores=0.5297878, pem=-2.8508017,  
similarity=35.40274]

3: 34289 Yasser\_Arafat

[pem\_pos=0, final\_scores=0.6073585,  
final\_local\_scores=0.5913901, pem=0.0, similarity=21.737972]

4: 241405 State\_of\_Palestine

[pem\_pos=9, final\_scores=0.565121,  
final\_local\_scores=0.51639867, pem=-2.8508017,



similarity=35.00318]  
 5: 34289 Yasser\_Arafat  
     [pem\_pos=0, final\_scores=0.6216095,  
 final\_local\_scores=0.60695714, pem=0.0, similarity=22.202524]  
 6: 241405 State\_of\_Palestine  
     [pem\_pos=9, final\_scores=0.38815963,  
 final\_local\_scores=0.32309532, pem=-2.8508017,  
 similarity=29.234625]  
 7: 24093 Palestinian\_National\_Authority  
     [pem\_pos=2, final\_scores=0.10534988,  
 final\_local\_scores=0.024991632, pem=-1.3862944,  
 similarity=12.374397]  
 8: 241405 State\_of\_Palestine  
     [pem\_pos=9, final\_scores=0.5484697,  
 final\_local\_scores=0.4982097, pem=-2.8508017,  
 similarity=34.46039]  
 9: 9282173 Israel  
     [pem\_pos=0, final\_scores=0.6014314,  
 final\_local\_scores=0.5836188, pem=-0.34790626,  
 similarity=23.39803]  
 10: 168212 Ehud\_Barak  
     [pem\_pos=0, final\_scores=0.6817664,  
 final\_local\_scores=0.71302634, pem=0.0, similarity=25.367842]  
 11: 241405 State\_of\_Palestine  
     [pem\_pos=5, final\_scores=0.12866633,  
 final\_local\_scores=0.037748516, pem=-1.7909594,  
 similarity=14.955719]  
 12: 9282173 Israel  
     [pem\_pos=1, final\_scores=0.5336433,  
 final\_local\_scores=0.50957054, pem=-0.34790626,  
 similarity=21.188282]  
 13: 241405 State\_of\_Palestine  
     [pem\_pos=9, final\_scores=0.31175327,  
 final\_local\_scores=0.23963296, pem=-2.8508017,  
 similarity=26.743942]

Sample: b'20000715\_AFP\_ARB\_0072\_eng'

[1 1 Jerusalem 1] 7-15 -LRB- [2 2 AFP 2] -RRB- - A high  
 level [1 Israeli 1] army official has said today Saturday  
 that [2 Israel 2] believes [3 3 Iran 3] is set to begin  
 acquiring nuclear capability for military purposes from 2005  
 and will be in a position to equip missiles with nuclear  
 warheads , capable of reaching [3 Israel 3] , within ten  
 years . The official , who wished to remain anonymous , told  
 journalists : `` [4 Iran 4] will start accumulating nuclear  
 capability for military purposes from 2005 and will be in a  
 position , within the space of 10 years , to produce nuclear  
 warheads that can be fitted to missiles capable of striking [5  
 Israel 5] . '' The official made these statements after [6  
 Tehran 6] announced today Saturday that it had ``  
 successfully '' completed a new test on the Shahab 3 ground-  
 ground missile with a range of 1300 kilometers . The official

further said : `` [7 Iran 7] is not capable of developing a nuclear program without 1 Moscow 's help . '' He said that according to information obtained by [8 Israel 8] , tests on the Shahab 3 ground-ground missile were in fact `` successful and reached their maximum range of 1300 kilometers which allows them to reach strategic centers in [9 Israel 9] . '' The official said that the design of the [10 Shahab 10] missile is based on a [11 North Korean 11] missile called Nodong , itself developed from the [12 Soviet 12] Scud missile . The official also stressed that [13 Iran 13] regards the `` destruction of [14 Israel 14] as legitimate , '' saying that `` [15 Tehran 15] is embroiled in strategic terrorist activities targeting [16 Israel 16] , especially alongside [17 Palestinian 17] and [18 Lebanese 18] organizations . '' The official further considered that `` the [19 Shahab 19] missile will not come into operation for two years and that an improved range of 2000 kilometers would only be achieved after this period . '' [20 Israel 20] 's Deputy Defense Minister , [21 Ephraim Sneh 21] , said today Saturday that [22 Israel 22] would `` know how to find a suitable response '' to the [23 Iranian 23] Shahab 3 ground-ground missile with a 1300 kilometer range . [24 Sneh 24] said on army radio that `` [25 Israel 25] will prepare itself to find a suitable response to this threat , '' without giving any further details . The [26 Iranian 26] army announced today Saturday that it had `` successfully '' completed a new test on Shahab 3 ground-ground missiles . An official at the [27 Iranian 27] 2 Ministry of Defense told Tehran Radio that the `` [28 Shahab 28] 3 defense missile was tested a second time to ensure it conforms to international standards . '' In July 1998 , the first test on this missile was announced . It is capable of reaching the majority of countries in the region , including [29 Israel 29] .

1: 16043 Jerusalem

2: 471981 Agence\_France-Presse

3: 14653 Iran

1: 19004 Moscow

[winner: 19004 Moscow, pem\_pos=0,  
final\_scores=-0.0730136, final\_local\_scores=-0.090761304,  
pem=-0.106582455, similarity=1.9608266]

gt: 19004 Moscow, pem\_pos=0, final\_scores=-0.0730136,  
final\_local\_scores=-0.090761304, pem=-0.106582455,  
similarity=1.9608266 ]

2: 18379483

Ministry\_of\_Defense\_and\_Armed\_Forces\_Logistics\_(Iran)

[winner: 3226446 Ministry\_of\_Defense\_(Israel), pem\_pos=0,  
final\_scores=-0.5855336, final\_local\_scores=-0.6427183,  
pem=-1.6469514, similarity=-6.1339064]

RECALL MISS: 18379483

Ministry\_of\_Defense\_and\_Armed\_Forces\_Logistics\_(Iran)]

1: 16043 Jerusalem

[pem\_pos=0, final\_scores=0.16049342,

final\_local\_scores=0.13449743, pem=-0.13239029,  
similarity=8.823343]  
2: 471981 Agence\_France-Presse  
[pem\_pos=0, final\_scores=0.032201275,  
final\_local\_scores=0.06123966, pem=-0.6605333,  
similarity=9.509309]  
3: 14653 Iran  
[pem\_pos=0, final\_scores=0.62536556,  
final\_local\_scores=0.6491537, pem=-0.13856095,  
similarity=24.215271]  
1: 9282173 Israel  
[pem\_pos=0, final\_scores=0.54559034,  
final\_local\_scores=0.5046386, pem=-0.34790626,  
similarity=21.041103]  
2: 9282173 Israel  
[pem\_pos=0, final\_scores=0.62107176,  
final\_local\_scores=0.58709055, pem=-0.21279564,  
similarity=22.76688]  
3: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5395083,  
final\_local\_scores=0.49799484, pem=-0.21279564,  
similarity=20.10809]  
4: 14653 Iran  
[pem\_pos=0, final\_scores=0.69485617,  
final\_local\_scores=0.72506166, pem=-0.13856095,  
similarity=26.480515]  
5: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5246845,  
final\_local\_scores=0.48180217, pem=-0.21279564,  
similarity=19.624866]  
6: 14653 Iran  
[pem\_pos=5, final\_scores=0.26816243,  
final\_local\_scores=0.25896356, pem=-2.6998966,  
similarity=26.50016]  
7: 14653 Iran  
[pem\_pos=0, final\_scores=0.7096095,  
final\_local\_scores=0.74117744, pem=-0.13856095,  
similarity=26.961443]  
8: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5672632,  
final\_local\_scores=0.52831286, pem=-0.21279564,  
similarity=21.012838]  
9: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5236421,  
final\_local\_scores=0.48066348, pem=-0.21279564,  
similarity=19.590885]  
10: 262773 Shahab-3  
[pem\_pos=0, final\_scores=0.07446031,  
final\_local\_scores=0.0829919, pem=-0.24850939,  
similarity=7.917789]  
11: 21255 North\_Korea  
[pem\_pos=0, final\_scores=0.46710438,

final\_local\_scores=0.47838962, pem=0.0, similarity=18.365814]  
12: 26779 Soviet\_Union  
[pem\_pos=0, final\_scores=0.12947969,  
final\_local\_scores=0.10831533, pem=-0.097171314,  
similarity=7.8504915]  
13: 14653 Iran  
[pem\_pos=0, final\_scores=0.6463728,  
final\_local\_scores=0.67210096, pem=-0.13856095,  
similarity=24.900063]  
14: 9282173 Israel  
[pem\_pos=0, final\_scores=0.60158014,  
final\_local\_scores=0.56579894, pem=-0.21279564,  
similarity=22.131496]  
15: 14653 Iran  
[pem\_pos=5, final\_scores=0.27429146,  
final\_local\_scores=0.2656586, pem=-2.6998966,  
similarity=26.699953]  
16: 9282173 Israel  
[pem\_pos=0, final\_scores=0.6136579,  
final\_local\_scores=0.57899207, pem=-0.21279564,  
similarity=22.525206]  
17: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.45158857,  
final\_local\_scores=0.47940588, pem=-2.8508017,  
similarity=33.899242]  
18: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.48701507,  
final\_local\_scores=0.46015984, pem=-0.55808944,  
similarity=20.856777]  
19: 262773 Shahab-3  
[pem\_pos=0, final\_scores=0.09128149,  
final\_local\_scores=0.10136646, pem=-0.24850939,  
similarity=8.466123]  
20: 9282173 Israel  
[pem\_pos=0, final\_scores=0.6558999,  
final\_local\_scores=0.625135, pem=-0.21279564,  
similarity=23.902203]  
21: 6728098 Efraim\_Sneh  
[pem\_pos=0, final\_scores=0.17140104,  
final\_local\_scores=0.2185533, pem=0.0, similarity=10.611779]  
22: 9282173 Israel  
[pem\_pos=0, final\_scores=0.7057038,  
final\_local\_scores=0.6795382, pem=-0.21279564,  
similarity=25.525703]  
23: 14653 Iran  
[pem\_pos=0, final\_scores=0.10409464,  
final\_local\_scores=0.07974444, pem=-0.57066226,  
similarity=9.572796]  
24: 6728098 Efraim\_Sneh  
[pem\_pos=0, final\_scores=0.1734642,  
final\_local\_scores=0.22080699, pem=0.0, similarity=10.679034]  
25: 9282173 Israel

[pem\_pos=0, final\_scores=0.60266393,  
final\_local\_scores=0.5669828, pem=-0.21279564,  
similarity=22.166828]  
26: 14653 Iran  
[pem\_pos=0, final\_scores=0.53803164,  
final\_local\_scores=0.5537547, pem=-0.57066226,  
similarity=23.718206]  
27: 14653 Iran  
[pem\_pos=0, final\_scores=0.32186872,  
final\_local\_scores=0.31762952, pem=-0.57066226,  
similarity=16.67176]  
28: 262773 Shahab-3  
[pem\_pos=0, final\_scores=0.041606233,  
final\_local\_scores=0.047103792, pem=-0.24850939,  
similarity=6.8468165]  
29: 9282173 Israel  
[pem\_pos=0, final\_scores=0.56271046,  
final\_local\_scores=0.5233397, pem=-0.21279564,  
similarity=20.864428]

Sample: b'20001015\_AFP\_ARB\_0053\_eng'

1 Beirut 10-15 -LRB- 2 AFP -RRB- - The Saudi Soccer Federation sacked the team 's [1 Czech 1] coach Milan Machala from his post today Sunday and appointed his assistant , Nasser Jawhar , in his place . The title-holding [2 Saudi 2] team suffered a bitter defeat -LRB- 1-4 -RRB- against its [3 Japanese 3] counterpart in its first match in the championship yesterday Saturday . The Youth and Sports Bureau in [4 Saudi Arabia 4] released a statement saying : `` Prince Sultan Ben Fahd , president of the Youth and Sports Bureau and president of the Saudi Soccer Federation , has issued a decision by order of which [5 Czech 5] coach Machala has been discharged and Nasser Jawhar has been placed in charge of training the team during the championship . ' ' He carried on : `` This decision is the result of the Saudi Federation ensuring the best technical conditions for the team in order to reflect the true level of [6 Saudi 6] soccer . ' ' In a telephone call with [7 Agence France Presse 7] at his hotel , [8 Machala 8] refused to give his opinion regarding the decision and said `` no comment . ' ' [9 Machala 9] is the first victim of the 12th [10 Asian Nations Cup 10] finals hosted by [1 1 Lebanon 1] until this October 29 as the [11 Japanese 11] blow was fatal , especially since the [12 Saudi 1 2] team holds the title from the previous series in the [13 3 Emirates 13] in 1996 . Machala took over training the [14 Saudi 14] team in February 1999 and led it in the Continents Championship hosted by [2 2 Mexico 2] in July of that year . Machala is considered one of the best-known coaches in the Gulf region and has achieved remarkable results with the [15 Kuwaiti 15] team , which he led to win the [16 Gulf Cup 16] on two occasions and to take fourth place in the last Asian Nations Championship in the [17 Emirates 17] , as well as

second place in the Asian Games in [3 3 Bangkok 3] in late 1998 .

1: 17771 Lebanon

2: 3966054 Mexico

3: 56615 Bangkok

1: 37428 Beirut

[winner: 37428 Beirut, pem\_pos=0,  
final\_scores=0.020781055, final\_local\_scores=-0.0179305,  
pem=-0.046614014, similarity=3.8081255]

gt: 37428 Beirut, pem\_pos=0, final\_scores=0.020781055,  
final\_local\_scores=-0.0179305, pem=-0.046614014,  
similarity=3.8081255 ]

2: 471981 Agence\_France-Presse

[winner: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=-0.10040212, final\_local\_scores=-0.09362685,  
pem=-0.6605333, similarity=4.887784]

gt: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=-0.10040212, final\_local\_scores=-0.09362685,  
pem=-0.6605333, similarity=4.887784 ]

3: 69328 United\_Arab\_Emirates

[winner: 142715 Emirates\_(airline), pem\_pos=0,  
final\_scores=0.1332423, final\_local\_scores=0.15235484,  
pem=-0.28409576, similarity=10.181242]

gt: 69328 United\_Arab\_Emirates, pem\_pos=3,  
final\_scores=-0.043442085, final\_local\_scores=-0.13926676,  
pem=-2.7219982, similarity=14.736364 ]

1: 17771 Lebanon

[pem\_pos=0, final\_scores=0.40322024,  
final\_local\_scores=0.37575108, pem=-0.4097839,  
similarity=17.531342]

2: 3966054 Mexico

[pem\_pos=0, final\_scores=0.31975013,  
final\_local\_scores=0.30953205, pem=-0.18258883,  
similarity=14.319713]

3: 56615 Bangkok

[pem\_pos=0, final\_scores=0.58713794,  
final\_local\_scores=0.6038151, pem=-0.0005963434,  
similarity=22.112001]

1: 5321 Czech\_Republic

[pem\_pos=1, final\_scores=0.29585803,  
final\_local\_scores=0.29886788, pem=-1.208016,  
similarity=19.577906]

2: 349303 Saudi\_Arabia

[pem\_pos=0, final\_scores=0.68214923,  
final\_local\_scores=0.6575692, pem=0.0, similarity=23.71289]

3: 15573 Japan

[pem\_pos=0, final\_scores=0.7237804,  
final\_local\_scores=0.74530977, pem=-0.99076414,  
similarity=31.719173]

4: 349303 Saudi\_Arabia

[pem\_pos=0, final\_scores=0.40368897,  
final\_local\_scores=0.35339358, pem=-0.25454172,

similarity=16.01992]  
5: 5321 Czech\_Republic  
[pem\_pos=1, final\_scores=0.23915364,  
final\_local\_scores=0.23692697, pem=-1.208016,  
similarity=17.729465]  
6: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.6227892,  
final\_local\_scores=0.5927273, pem=0.0, similarity=21.777878]  
7: 471981 Agence\_France-Presse  
[pem\_pos=0, final\_scores=0.3584255,  
final\_local\_scores=0.4215213, pem=0.0, similarity=16.668749]  
8: 441979 Machala  
[pem\_pos=0, final\_scores=0.053571507,  
final\_local\_scores=0.09075691, pem=0.0, similarity=6.79808]  
9: 441979 Machala  
[pem\_pos=0, final\_scores=0.038071975,  
final\_local\_scores=0.07382606, pem=0.0, similarity=6.29283]  
10: 250683 AFC\_Asian\_Cup  
[pem\_pos=0, final\_scores=0.12301175,  
final\_local\_scores=0.14730504, pem=0.0, similarity=8.485589]  
11: 15573 Japan  
[pem\_pos=0, final\_scores=0.7406147,  
final\_local\_scores=0.75520045, pem=-0.99076414,  
similarity=32.014328]  
12: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.5267903,  
final\_local\_scores=0.48786318, pem=0.0, similarity=18.648523]  
13: 142715 Emirates\_(airline)  
[pem\_pos=0, final\_scores=0.1332423,  
final\_local\_scores=0.15235484, pem=-0.28409576,  
similarity=10.181242]  
14: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.7166484,  
final\_local\_scores=0.6952542, pem=0.0, similarity=24.837484]  
15: 7515890 Kuwait  
[pem\_pos=0, final\_scores=0.5526547,  
final\_local\_scores=0.5214036, pem=-0.041184552,  
similarity=19.873404]  
16: 3728043 Gulf\_Cup\_of\_Nations  
[pem\_pos=0, final\_scores=0.1330068,  
final\_local\_scores=0.17670587, pem=-0.12871213,  
similarity=10.062924]  
17: 142715 Emirates\_(airline)  
[pem\_pos=0, final\_scores=0.08308928,  
final\_local\_scores=0.097570285, pem=-0.28409576,  
similarity=8.546362]

Sample: b'VOA20001220\_2000\_0060'

I am David Coler , [1 1 VOA News 1] . President Elect [1  
George W Bush 1] made more cabinet appointments Wednesday .  
He named Al Koha Chairman Paul O ' Neil as Treasury Secretary  
and former California Food and Agriculture Secretary , Anne

Venemon as Secretary of Agriculture . He also named an old friend , an oil business associate , [2 Don Evans 2] as Secretary of Commerce . The promotion of free enterprise first in [2 2 America 2] and then abroad will be our first priority . Free flow of capital , free and open competition . We will strive to be an advocate for [3 US 3] businesses first in [4 America 4] and also those wading into the waters of the global market place . Mr. [5 Bush 5] also named former [6 Cuban 6] refugee [7 Mel Martinez 7] as Secretary of Housing and Urban Development . Mr. Martinez is a close political ally of the President Elect 's brother , [8 Florida 8] Governor [9 Jeb Bush 9] . Mr. [10 Bush 10] says he will step down Thursday as Governor of [3 3 Texas 3] .

1: 103191 Voice\_of\_America

2: 3434750 United\_States

3: 29810 Texas

1: 103191 Voice\_of\_America

[pem\_pos=0, final\_scores=0.13933642,  
final\_local\_scores=0.170896, pem=0.0, similarity=9.18959]

2: 3434750 United\_States

[pem\_pos=0, final\_scores=0.43047756,  
final\_local\_scores=0.43509597, pem=-0.78483206,  
similarity=21.341883]

3: 29810 Texas

[pem\_pos=0, final\_scores=0.7322361,  
final\_local\_scores=0.7667841, pem=-0.16191302,  
similarity=27.852587]

1: 3414021 George\_W.\_Bush

[pem\_pos=0, final\_scores=0.33848864,  
final\_local\_scores=0.26661193, pem=-0.017176678,  
similarity=12.139355]

2: 285101 Donald\_Evans

[pem\_pos=1, final\_scores=0.27485996,  
final\_local\_scores=0.31206012, pem=-0.042925045,  
similarity=13.635642]

3: 3434750 United\_States

[pem\_pos=0, final\_scores=0.035180494,  
final\_local\_scores=-0.007293433, pem=-0.48755783,  
similarity=6.5234776]

4: 3434750 United\_States

[pem\_pos=0, final\_scores=0.3848558,  
final\_local\_scores=0.38526118, pem=-0.78483206,  
similarity=19.854713]

5: 3414021 George\_W.\_Bush

[pem\_pos=0, final\_scores=0.33136725,  
final\_local\_scores=0.26664206, pem=0.0, similarity=12.046844]

6: 5042481 Cuba

[pem\_pos=0, final\_scores=0.43047172,  
final\_local\_scores=0.46766812, pem=-0.37537703,  
similarity=20.087221]

7: 296686 Mel\_Martínez

[pem\_pos=0, final\_scores=0.42160672,



final\_local\_scores=0.4367813, pem=0.0, similarity=17.124138]  
8: 18933066 Florida  
[pem\_pos=0, final\_scores=0.257847,  
final\_local\_scores=0.2601929, pem=-0.15405534,  
similarity=12.692165]  
9: 189322 Jeb\_Bush  
[pem\_pos=0, final\_scores=0.4763183,  
final\_local\_scores=0.46512508, pem=0.0, similarity=17.969973]  
10: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.37747878,  
final\_local\_scores=0.31701192, pem=0.0, similarity=13.549981]

Sample: b'chthb\_165\_eng'

[1 1 Xinhua News Agency 1] , [2 2 Shanghai 2] , April 3rd ,  
by reporter Jierong Zhou Recently , [3 3 HSBC 3] has moved  
its [1 Shanghai 1] branch to the 1 China Shipping Mansion in  
the Pudong Lujiazui financial trading district , becoming the  
third foreign capital bank to be approved to operate [2 RMB 2]  
business and shift to [4 4 Pudong 4] . At the moment , [3  
Shanghai 3] already has 8 foreign capital banks approved to  
operate [4 RMB 4] business . They are the [5 5 US 5] 's [6 6 C  
itibank 6] , [7 7 Hong Kong 7] 's [5 HSBC 5] , [8 8 Japan 8]  
's [9 9 Tokyo Mitsubishi Bank 9] , [6 Japan 6] 's [7  
Industrial Bank 7] , the Shanghai Branch of the Standard and  
Chartered Bank , [8 Shanghai 8] 's BNP Paris Bank , the [9  
Shanghai Branch 9] of [10 Japan 10] 's [11 Dai-Ichi Kangyo  
Bank 11] and the Shanghai Branch of [12 Japan 12] 's [13  
Sanwa Bank 13] . According to regulations , these 8 banks  
will all move to [14 Pudong 14] soon . After the policy was  
formally announced for the Pudong Lujiazui Financial Trade  
District to allow foreign capital banks to operate [15 RMB 15]  
business , more than 40 foreign capital banks and financial  
institutions have submitted applications to establish  
branches in [16 Pudong 16] . When selecting locations in [17  
Lujiazui 17] , some foreign financial institutions do not  
mind the real estate prices , but focus more on whether the  
class of the building is compatible with the standing of the  
company . The Franklin Templeton Company , whose headquarters  
are in [18 California 18] , the [19 US 19] , and has  
registered capital of 130 billion US dollars all around the  
world , decided to establish its local [20 China 20]  
headquarters in [21 Pudong 21] . The president of the company  
and the [22 Asian 22] regional president personally surveyed [23  
Pudong 23] and decided to choose office locations in the  
Shanghai Securities Building , which has a floor rate of more  
than 2700 US dollars per square meter . Now , the Templeton  
Company has formally signed a contract , and bought the floor  
area of the entire 18th floor of the Securities Building .  
Currently , foreign-capital banks and financial institutions  
such as [24 Hong Kong 24] 's [25 HSBC 25] , the Shanghai  
Branch of the [26 Japan Industrial Bank 26] , the Holland Co-  
operation Bank , Shanghai Branch of the Belgium Credit Bank ,

etc. , have successively settled down in the Lujiazui Financial Trading District . Soon , a large number of domestic and foreign financial institutions , and large companies will also come to successively settle down , including the finance street which has already formed in the [ 27 Waitan 27] district , and the [28 Shanghai 28] central commercial district framework that is forming . -LRB- End - RRB-

1: 263168 Xinhua\_News\_Agency

2: 27643 Shanghai

3: 322572 HSBC

4: 402737 Pudong

5: 3434750 United\_States

6: 231026 Citibank

7: 13404 Hong\_Kong

8: 15573 Japan

9: 2823064 The\_Bank\_of\_Tokyo-Mitsubishi\_UFJ

1: 13926709 International\_Ocean\_Shipping\_Building

[winner: 17576047 China\_Merchants\_Energy\_Shipping,  
pem\_pos=1, final\_scores=-0.6726252,  
final\_local\_scores=-0.59187865, pem=-0.6931472,  
similarity=-9.803684]

RECALL MISS: 13926709

International\_Ocean\_Shipping\_Building]

[winner: 557698 Mansion, pem\_pos=0,  
final\_scores=-0.88047516, final\_local\_scores=-0.90092987,  
pem=-0.09583251, similarity=-22.27468]

RECALL MISS: 13926709

International\_Ocean\_Shipping\_Building]

1: 263168 Xinhua\_News\_Agency

[pem\_pos=0, final\_scores=0.42345232,  
final\_local\_scores=0.45586604, pem=0.0, similarity=17.693665]

2: 27643 Shanghai

[pem\_pos=0, final\_scores=0.4649455,  
final\_local\_scores=0.42622203, pem=-0.10072064,  
similarity=17.356762]

3: 322572 HSBC

[pem\_pos=0, final\_scores=0.34508616,  
final\_local\_scores=0.35390884, pem=0.0, similarity=14.651058]

4: 402737 Pudong

[pem\_pos=0, final\_scores=0.29184598,  
final\_local\_scores=0.26971343, pem=0.0, similarity=12.138499]

5: 3434750 United\_States

[pem\_pos=0, final\_scores=0.2426322,  
final\_local\_scores=0.24385142, pem=-0.48755783,  
similarity=14.018141]

6: 231026 Citibank

[pem\_pos=0, final\_scores=0.29678816,  
final\_local\_scores=0.34032935, pem=0.0, similarity=14.245819]

7: 13404 Hong\_Kong

[pem\_pos=0, final\_scores=0.61815625,  
final\_local\_scores=0.5919483, pem=-0.22402541,

similarity=22.972916]  
8: 15573 Japan  
[pem\_pos=0, final\_scores=0.78545326,  
final\_local\_scores=0.82038933, pem=-0.13549899,  
similarity=29.30863]  
9: 2823064 The\_Bank\_of\_Tokyo-Mitsubishi\_UFJ  
[pem\_pos=0, final\_scores=0.3739406,  
final\_local\_scores=0.44036764, pem=0.0, similarity=17.231161]  
1: 27643 Shanghai  
[pem\_pos=0, final\_scores=0.43549317,  
final\_local\_scores=0.39404988, pem=-0.10072064,  
similarity=16.39668]  
2: 66958 Renminbi  
[pem\_pos=0, final\_scores=0.16858934,  
final\_local\_scores=0.14823928, pem=0.0, similarity=8.513469]  
3: 27643 Shanghai  
[pem\_pos=0, final\_scores=0.3781994,  
final\_local\_scores=0.3314651, pem=-0.10072064,  
similarity=14.529027]  
4: 66958 Renminbi  
[pem\_pos=0, final\_scores=0.14208539,  
final\_local\_scores=0.11928773, pem=0.0, similarity=7.6494966]  
5: 322572 HSBC  
[pem\_pos=0, final\_scores=0.30749935,  
final\_local\_scores=0.31107497, pem=0.0, similarity=13.37281]  
6: 15573 Japan  
[pem\_pos=0, final\_scores=0.6086171,  
final\_local\_scores=0.6272227, pem=-0.13549899,  
similarity=23.544153]  
7: 3522901 Industrial\_Bank\_Co.  
[pem\_pos=2, final\_scores=0.109676376,  
final\_local\_scores=0.15255994, pem=-0.6931472,  
similarity=12.411848]  
8: 27643 Shanghai  
[pem\_pos=0, final\_scores=0.45045727,  
final\_local\_scores=0.41039586, pem=-0.10072064,  
similarity=16.884478]  
9: 27643 Shanghai  
[pem\_pos=0, final\_scores=0.03183578,  
final\_local\_scores=-0.049561664, pem=0.0, similarity=2.610694]  
10: 15573 Japan  
[pem\_pos=0, final\_scores=0.75822717,  
final\_local\_scores=0.79064894, pem=-0.13549899,  
similarity=28.421118]  
11: 384768 Dai-Ichi\_Kangyo\_Bank  
[pem\_pos=0, final\_scores=0.11084585,  
final\_local\_scores=0.15925565, pem=0.0, similarity=8.842219]  
12: 15573 Japan  
[pem\_pos=0, final\_scores=0.7621624,  
final\_local\_scores=0.79494756, pem=-0.13549899,  
similarity=28.549398]  
13: 2823064 The\_Bank\_of\_Tokyo-Mitsubishi\_UFJ

[pem\_pos=0, final\_scores=0.3545817,  
final\_local\_scores=0.41922098, pem=0.0, similarity=16.600103]  
14: 402737 Pudong  
[pem\_pos=0, final\_scores=0.26995915,  
final\_local\_scores=0.24580544, pem=0.0, similarity=11.425037]  
15: 66958 Renminbi  
[pem\_pos=0, final\_scores=0.20116974,  
final\_local\_scores=0.18382841, pem=0.0, similarity=9.57552]  
16: 402737 Pudong  
[pem\_pos=0, final\_scores=0.30792212,  
final\_local\_scores=0.2872742, pem=0.0, similarity=12.662548]  
17: 2437737 Lujiazui  
[pem\_pos=0, final\_scores=0.22823267,  
final\_local\_scores=0.2410717, pem=0.0, similarity=11.283773]  
18: 5407 California  
[pem\_pos=0, final\_scores=0.48041588,  
final\_local\_scores=0.53371435, pem=-0.08662026,  
similarity=20.48787]  
19: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.11492102,  
final\_local\_scores=0.10434635, pem=-0.48755783,  
similarity=9.855031]  
20: 5405 China  
[pem\_pos=0, final\_scores=0.53023976,  
final\_local\_scores=0.52320963, pem=-0.15669827,  
similarity=20.555481]  
21: 402737 Pudong  
[pem\_pos=0, final\_scores=0.23462938,  
final\_local\_scores=0.20721301, pem=0.0, similarity=10.273362]  
22: 689 Asia  
[pem\_pos=1, final\_scores=0.35774827,  
final\_local\_scores=0.37293434, pem=-1.9623682,  
similarity=25.890482]  
23: 402737 Pudong  
[pem\_pos=0, final\_scores=0.10267575,  
final\_local\_scores=0.062141865, pem=0.0, similarity=5.94415]  
24: 13404 Hong\_Kong  
[pem\_pos=0, final\_scores=0.63463324,  
final\_local\_scores=0.60994685, pem=-0.22402541,  
similarity=23.510029]  
25: 322572 HSBC  
[pem\_pos=0, final\_scores=0.31229955,  
final\_local\_scores=0.3163185, pem=0.0, similarity=13.529288]  
26: 1490243 Industrial\_Bank\_of\_Japan  
[pem\_pos=0, final\_scores=0.1932256,  
final\_local\_scores=0.25605342, pem=0.0, similarity=11.730858]  
27: 453167 The\_Bund  
[pem\_pos=0, final\_scores=0.18358211,  
final\_local\_scores=0.17234072, pem=0.0, similarity=9.232704]  
28: 27643 Shanghai  
[pem\_pos=0, final\_scores=0.58655125,  
final\_local\_scores=0.55905783, pem=-0.10072064,

similarity=21.320848]

Sample: b'20001115\_AFP\_ARB\_0089\_eng'

Albus -LRB- [1 1 Lebanon 1] -RRB- 11-15 -LRB- 1 AFP -RRB- - Thousands of [1 Palestinian 1] refugees in [2 Lebanon 2] commemorated in their camps today Wednesday the twelfth anniversary of the symbolic announcement of independence by the [2 2 Palestinian National Council 2] in [3 3 Algeria 3] in 1988 . Around five thousand [3 Palestinians 3] form the Albus Camp near 2 Souer -LRB- 83 km south of [4 4 Beirut 4] - RRB- gathered and followed a march by children dressed in military clothing who held stones and raised pictures of [5 5 Mohamed al-Dura 5] , the 12-year-old [4 Palestinian 4] child killed by [5 Israeli 5] army bullets in the [6 Palestinian Territories 6] in early October . A France Presse correspondent reported that around 200 [7 Palestinians 7] marched along the streets of Ayn Alhilwa refugee camp near [6 6 Saida 6] , one of southern [7 7 Lebanon 7] 's main cities , toward the headquarters of the [8 Fatah 8] movement led by [9 Palestinian 9] President [10 Yasser Arafat 10] carrying the [1 1 Palestinian 11] flag . A group of young people burned the [1 2 Israeli 12] and [13 American 13] flags outside the headquarters . [14 Khalid Arif 14] , am official with the [15 Fatah 15] movement in [16 Saida 16] , said the [17 Intifada 17] against [18 Israel 18] which started late last September and has resulted in over two hundred deaths would `` continue until the [19 Israeli 19] occupiers and settlers were chased out of the [20 Palestinian Territories 20] . '' He added , `` The [21 Intifada 21] will continue until a [22 Palestinian 22] state with holy [8 8 Jerusalem 8] as its capital is created and the refugees return to their homes . '' In the nearby Ayn Alhilwa refugee camp , students between ages fifteen and sixteen marched through the camp 's alleyways , chanting `` [2 3 Jerusalem 23] is ours '' then the burned [24 Israeli 24] and [25 American 25] flags . [26 Fatah 26] movement called on [27 Palestinians 27] to declare today Wednesday as a day `` to embody national independence '' and to stop `` [28 Israeli 28] occupying forces and settler bandits from entering occupied [29 Palestinian 29] territory , '' as a commemoration of the twelfth anniversary of the symbolic announcement of [30 Palestinian 30] independence . [31 Palestinian 31] President [32 Yasser Arafat 32] had symbolically announced a [33 Palestinian 33] state on November 15 , 1988 in [34 Algeria 34] when the first [35 Palestinian Intifada 35] was at its height and which continued for seven years until the [36 Oslo 36] peace accords were signed in 1993 .

1: 17771 Lebanon

2: 41207485 Palestinian\_National\_Council

3: 358 Algeria

4: 37428 Beirut

5: 1091190 Muhammad\_al-Durrah\_incident

6: 28366 Sidon  
7: 17771 Lebanon  
8: 16043 Jerusalem  
1: 471981 Agence\_France-Presse  
[winner: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=0.024735644, final\_local\_scores=0.028136343,  
pem=-0.6605333, similarity=8.5214405|  
gt: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=0.024735644, final\_local\_scores=0.028136343,  
pem=-0.6605333, similarity=8.5214405 ]  
2: 30796 Tyre, Lebanon  
[winner: 14646364 Piet\_Souer, pem\_pos=1,  
final\_scores=-0.20786105, final\_local\_scores=-0.08292982,  
pem=0.0, similarity=1.6149217|  
RECALL MISS: 30796 Tyre, Lebanon]  
1: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.31168962,  
final\_local\_scores=0.25524107, pem=-0.4097839,  
similarity=13.935084]  
2: 41207485 Palestinian\_National\_Council  
[pem\_pos=0, final\_scores=0.17267688,  
final\_local\_scores=0.18261996, pem=-0.013409173,  
similarity=9.612378]  
3: 358 Algeria  
[pem\_pos=0, final\_scores=0.4725135,  
final\_local\_scores=0.48587668, pem=-0.33694586,  
similarity=20.421606]  
4: 37428 Beirut  
[pem\_pos=0, final\_scores=0.5266289,  
final\_local\_scores=0.50482684, pem=-0.046614014,  
similarity=19.408247]  
5: 1091190 Muhammad\_al-Durrah\_incident  
[pem\_pos=0, final\_scores=0.35973418,  
final\_local\_scores=0.41645432, pem=0.0, similarity=16.51754]  
6: 28366 Sidon  
[pem\_pos=0, final\_scores=0.36638856,  
final\_local\_scores=0.36959255, pem=-0.02797111,  
similarity=15.271203]  
7: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.5552884,  
final\_local\_scores=0.52133566, pem=-0.4097839,  
similarity=21.875877]  
8: 16043 Jerusalem  
[pem\_pos=0, final\_scores=0.51640445,  
final\_local\_scores=0.49600166, pem=-0.13239029,  
similarity=19.611351]  
1: 1114732 Palestine  
[pem\_pos=1, final\_scores=0.20822476,  
final\_local\_scores=0.13285595, pem=-1.9100745,  
similarity=18.44168]  
2: 17771 Lebanon  
[pem\_pos=0, final\_scores=0.53974634,

final\_local\_scores=0.50435835, pem=-0.4097839,  
similarity=21.36924]  
3: 1114732 Palestine  
[pem\_pos=1, final\_scores=0.12542294,  
final\_local\_scores=0.04180652, pem=-1.6411982,  
similarity=14.262394]  
4: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.19770543,  
final\_local\_scores=0.14451832, pem=-2.8508017,  
similarity=23.90553]  
5: 9282173 Israel  
[pem\_pos=0, final\_scores=0.54771507,  
final\_local\_scores=0.5016696, pem=-0.34790626,  
similarity=20.952501]  
6: 96665 Palestinian\_territories  
[pem\_pos=0, final\_scores=0.10755585,  
final\_local\_scores=0.03695318, pem=0.0, similarity=5.1924696]  
7: 241405 State\_of\_Palestine  
[pem\_pos=5, final\_scores=0.23459713,  
final\_local\_scores=0.1848169, pem=-1.7909594,  
similarity=19.344534]  
8: 11104 Fatah  
[pem\_pos=0, final\_scores=0.44351608,  
final\_local\_scores=0.42379653, pem=0.0, similarity=16.736647]  
9: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.44520146,  
final\_local\_scores=0.41487014, pem=-2.8508017,  
similarity=31.97337]  
10: 34289 Yasser\_Arafat  
[pem\_pos=0, final\_scores=0.5955458,  
final\_local\_scores=0.58339024, pem=0.0, similarity=21.49924]  
11: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.27478898,  
final\_local\_scores=0.22872037, pem=-2.8508017,  
similarity=26.41829]  
12: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5385097,  
final\_local\_scores=0.49161416, pem=-0.34790626,  
similarity=20.65243]  
13: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.4585716,  
final\_local\_scores=0.47985077, pem=-0.12994486,  
similarity=19.116077]  
14: 13639042 Khalid\_Mahmud\_Arif  
[pem\_pos=0, final\_scores=0.24499683,  
final\_local\_scores=0.34908256, pem=0.0, similarity=14.507032]  
15: 11104 Fatah  
[pem\_pos=0, final\_scores=0.37032527,  
final\_local\_scores=0.34384668, pem=0.0, similarity=14.350783]  
16: 28366 Sidon  
[pem\_pos=0, final\_scores=0.2870955,  
final\_local\_scores=0.28297696, pem=-0.02797111,

similarity=12.68642]  
17: 71712 First\_Intifada  
[pem\_pos=0, final\_scores=0.24032356,  
final\_local\_scores=0.20349649, pem=-0.066496804,  
similarity=10.524075]  
18: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5557955,  
final\_local\_scores=0.5104962, pem=-0.21279564,  
similarity=20.481153]  
19: 9282173 Israel  
[pem\_pos=0, final\_scores=0.53112894,  
final\_local\_scores=0.4835518, pem=-0.34790626,  
similarity=20.411833]  
20: 96665 Palestinian\_territories  
[pem\_pos=0, final\_scores=0.042562082,  
final\_local\_scores=-0.036167897, pem=0.0,  
similarity=3.0103908]  
21: 71712 First\_Intifada  
[pem\_pos=0, final\_scores=0.27215987,  
final\_local\_scores=0.23940438, pem=-0.066496804,  
similarity=11.595637]  
22: 1114732 Palestine  
[pem\_pos=1, final\_scores=0.28868788,  
final\_local\_scores=0.22074968, pem=-1.9100745,  
similarity=21.064606]  
23: 16043 Jerusalem  
[pem\_pos=0, final\_scores=0.58772403,  
final\_local\_scores=0.57390743, pem=-0.13239029,  
similarity=21.936214]  
24: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5238777,  
final\_local\_scores=0.47563082, pem=-0.34790626,  
similarity=20.175453]  
25: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.43801856,  
final\_local\_scores=0.45739967, pem=-0.12994486,  
similarity=18.446093]  
26: 11104 Fatah  
[pem\_pos=0, final\_scores=0.36172956,  
final\_local\_scores=0.3344572, pem=0.0, similarity=14.070582]  
27: 1114732 Palestine  
[pem\_pos=1, final\_scores=0.24243046,  
final\_local\_scores=0.16961944, pem=-1.6411982,  
similarity=18.076586]  
28: 9282173 Israel  
[pem\_pos=0, final\_scores=0.56043094,  
final\_local\_scores=0.51555973, pem=-0.34790626,  
similarity=21.367012]  
29: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.1611109,  
final\_local\_scores=0.10454434, pem=-2.8508017,  
similarity=22.712627]



30: 241405 State\_of\_Palestine  
 [pem\_pos=9, final\_scores=0.1903417,  
 final\_local\_scores=0.13588697, pem=-2.8508017,  
 similarity=23.647955]  
 31: 241405 State\_of\_Palestine  
 [pem\_pos=9, final\_scores=0.41316557,  
 final\_local\_scores=0.37987578, pem=-2.8508017,  
 similarity=30.929066]  
 32: 34289 Yasser\_Arafat  
 [pem\_pos=0, final\_scores=0.66020864,  
 final\_local\_scores=0.6540245, pem=0.0, similarity=23.607107]  
 33: 1114732 Palestine  
 [pem\_pos=1, final\_scores=0.26085263,  
 final\_local\_scores=0.19034392, pem=-1.9100745,  
 similarity=20.157236]  
 34: 358 Algeria  
 [pem\_pos=0, final\_scores=0.5163155,  
 final\_local\_scores=0.5337238, pem=-0.33694586,  
 similarity=21.849459]  
 35: 71712 First\_Intifada  
 [pem\_pos=0, final\_scores=0.10773216,  
 final\_local\_scores=0.056748882, pem=0.0, similarity=5.7832127]  
 36: 22309 Oslo  
 [pem\_pos=0, final\_scores=0.31985587,  
 final\_local\_scores=0.3985945, pem=-0.035297178,  
 similarity=16.17652]

Sample: b'PRI20001201\_2000\_1828'

From [1 1 BBC News 1] in [2 2 London 2] , I am Gregor Cragy for The World . The [3 3 US Supreme Court 3] has held a hearing to decide whether to intervene in the dispute over the presidential election results . Judith Spencer reports on the legal wrangling that is underway . Inside the stately [1 Supreme Court 1] building , lawyers for [2 Republican 2] [3 George W. Bush 3] and [4 Democrat 4] [5 AL Gore 5] argued over the validity of hand re-counts in the crucial state of [6 Florida 6] , just as 1 Ruth Bater Ginsburg pressed [7 Bush 7] 's lawyer to explain why the [8 Federal High Court 8] should intervene in the [4 4 Florida Supreme Court 4] 's move to extend the deadline for finishing the re-counts . When we read a State court decision , we should read it in the light most favorable to the integrity of the [9 State Supreme Court 9] . But later on , Justice [10 2 Sandra Day 10] O'Connor suggested the State court might have been out of line . The legislature had very clearly said , you know , 7 days after , that 's the date and it just does look like very dramatic change made by the [11 Florida 11] court . The [12 US Supreme Court 12] could rule on the matter as early as tomorrow . For The World , I am Judith Spencer in [5 5 Washington 5] . Meanwhile , the [13 Florida Supreme Court 13] has rejected an emergency appeal from [14 Al Gore 14] for an immediate re-count of thousands of disputed ballots from [15 Palm Beach 15]

and [16 Miami Dade 16] counties .

1: 1139893 BBC\_News

2: 17867 London

3: 31737 Supreme\_Court\_of\_the\_United\_States

4: 1635683 Supreme\_Court\_of\_Florida

5: 108956 Washington,\_D.C.

1: 185232 Ruth\_Bader\_Ginsburg

[winner: 185232 Ruth\_Bader\_Ginsburg, pem\_pos=0,  
final\_scores=-0.25985867, final\_local\_scores=-0.31032443,  
pem=-0.5024102, similarity=-2.4388022]

gt: 185232 Ruth\_Bader\_Ginsburg, pem\_pos=0,  
final\_scores=-0.25985867, final\_local\_scores=-0.31032443,  
pem=-0.5024102, similarity=-2.4388022 ]

[winner: 30219226 Ruth,\_West\_Virginia, pem\_pos=4,  
final\_scores=-1.2380666, final\_local\_scores=-1.1923814,  
pem=-3.2352626, similarity=-13.899464]

RECALL MISS: 185232 Ruth\_Bader\_Ginsburg]

[winner: 20623593 Ryan\_Bater, pem\_pos=0,  
final\_scores=-1.8181325, final\_local\_scores=-1.8655677,  
pem=-0.009302417, similarity=-51.53196]

RECALL MISS: 185232 Ruth\_Bader\_Ginsburg]

2: 42640 Sandra\_Day\_O'Connor

[winner: 42640 Sandra\_Day\_O'Connor, pem\_pos=0,  
final\_scores=0.2719984, final\_local\_scores=0.25943598,  
pem=0.0, similarity=11.8318]

gt: 42640 Sandra\_Day\_O'Connor, pem\_pos=0,  
final\_scores=0.2719984, final\_local\_scores=0.25943598,  
pem=0.0, similarity=11.8318 ]

[winner: 4463625 James\_O'Connor\_(Louisiana\_politician),  
pem\_pos=7, final\_scores=-1.7651623,  
final\_local\_scores=-1.805063, pem=-3.3581038,  
similarity=-31.515078]

gt: 42640 Sandra\_Day\_O'Connor, pem\_pos=3,  
final\_scores=-2.200557, final\_local\_scores=-2.4497056,  
pem=-2.721405, similarity=-54.214962 ]

1: 1139893 BBC\_News

[pem\_pos=0, final\_scores=0.13022079,  
final\_local\_scores=0.1658574, pem=0.0, similarity=9.039228]

2: 17867 London

[pem\_pos=0, final\_scores=0.5960172,  
final\_local\_scores=0.64963424, pem=-0.09086737,  
similarity=23.970245]

3: 31737 Supreme\_Court\_of\_the\_United\_States

[pem\_pos=0, final\_scores=0.28000146,  
final\_local\_scores=0.24153528, pem=0.0, similarity=11.297607]

4: 1635683 Supreme\_Court\_of\_Florida

[pem\_pos=0, final\_scores=0.24585636,  
final\_local\_scores=0.23020831, pem=0.0, similarity=10.959588]

5: 108956 Washington,\_D.C.

[pem\_pos=0, final\_scores=0.42663974,  
final\_local\_scores=0.39714402, pem=-1.7496481,  
similarity=25.45614]

1: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.25986516,  
final\_local\_scores=0.21953946, pem=-0.9940822,  
similarity=16.047182]  
2: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.33543396,  
final\_local\_scores=0.31169906, pem=-0.1310191,  
similarity=14.103937]  
3: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.29790777,  
final\_local\_scores=0.24830627, pem=0.0, similarity=11.499667]  
4: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.35722023,  
final\_local\_scores=0.33721924, pem=-0.061652232,  
similarity=14.488282]  
5: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.5522954,  
final\_local\_scores=0.5589735, pem=0.0, similarity=20.770597]  
6: 18933066 Florida  
[pem\_pos=0, final\_scores=0.2629453,  
final\_local\_scores=0.2541981, pem=-0.15405534,  
similarity=12.513268]  
7: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.2845189,  
final\_local\_scores=0.23933738, pem=0.0, similarity=11.2320175]  
8: 15668259 Islamabad\_High\_Court  
[pem\_pos=0, final\_scores=0.050155506,  
final\_local\_scores=0.13720039, pem=0.0, similarity=8.184046]  
9: 28945 State\_supreme\_court  
[pem\_pos=0, final\_scores=0.053232267,  
final\_local\_scores=0.049159855, pem=-0.2950351,  
similarity=7.1611876]  
10: 42640 Sandra\_Day\_O'Connor  
[pem\_pos=0, final\_scores=0.2719984,  
final\_local\_scores=0.25943598, pem=0.0, similarity=11.8318]  
11: 18933066 Florida  
[pem\_pos=0, final\_scores=0.15717484,  
final\_local\_scores=0.13865998, pem=-0.15405534,  
similarity=9.06538]  
12: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.23913126,  
final\_local\_scores=0.19689083, pem=0.0, similarity=9.965328]  
13: 1635683 Supreme\_Court\_of\_Florida  
[pem\_pos=0, final\_scores=0.19307949,  
final\_local\_scores=0.17255759, pem=0.0, similarity=9.239176]  
14: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.36115426,  
final\_local\_scores=0.35018083, pem=0.0, similarity=14.539807]  
15: 109617 Palm\_Beach,\_Florida  
[pem\_pos=0, final\_scores=0.1376443,  
final\_local\_scores=0.14986292, pem=-0.6411934,  
similarity=12.048831]

16: 27905225 Miami-Dade\_County,\_Florida  
[pem\_pos=0, final\_scores=0.2941305,  
final\_local\_scores=0.29169205, pem=0.0, similarity=12.794386]

Sample: b'20001115\_AFP\_ARB\_0060\_eng'

[1 1 Bandar Seri Begawan 1] 11-15 -LRB- 1 AFP -RRB- - A high-level [1 American 1] official announced today Wednesday in the wake of the meeting between [2 American 2] President [3 Bill Clinton 3] and his [4 Russian 4] counterpart [5 Vladimir Putin 5] that [6 Putin 6] 's proposals on reducing the [7 Russian 7] and [8 American 8] nuclear capability to under 1500 nuclear heads does not include `` many new elements . '' The official who wished to remain anonymous confirmed `` we have paid a great deal of attention to President [9 Putin 9] 's announcement but it does not contain many new elements , in fact just a few new disparities which require further examination and discussion by experts . '' [10 Clinton 10] and [11 Putin 11] are taking part in the [12 Asia-Pacific Economic Cooperation 12] -LRB- APEC -RRB- forum in [13 Bandar Seri Begawan 13] the capital of the 2 Sultanate of [14 Brunei 14] . The presidents met for a working lunch of around 75 minutes . The [15 American 15] official stressed that what is important in [16 Putin 16] 's proposal is that it falls within the `` framework adopted by the [17 United States 17] and [18 Russia 18] to discuss strategic issues from a defensive and offensive point of view . '' [19 Putin 19] proposed limiting the [20 Russian 20] and [21 American 21] nuclear capabilities by reducing the number of nuclear heads to under 1500 each , a smaller number than that in the Russian-American Start-3 treaty currently being negotiated . He proposed that the current reduction should be between 2500 to 2000 nuclear heads . [22 Putin 22] said this reduction would depend on compliance with the Russian-American anti-ABM missile treaty signed in 1972 , which the [23 Americans 23] are attempting to amend to set up anti- missile defenses - LRB- NMD -RRB- . [24 Washington 24] has so far confirmed agreeing to the 2000 to 2500 nuclear warhead ceiling . The [25 United States 25] has not taken a final position on the [26 NMD 26] system because [27 Clinton 27] , whose term ends in January , is leaving the decision in this matter to his successor . In response to a question about [28 Republican 28] presidential candidate [29 George Bush Junior 29] 's approach to deploying the [30 NMD 30] system , more ambitious than the one proposed by the [31 Clinton 31] administration , the [32 American 32] official said the [33 United States 33] would continue `` complying with the present framework '' in these strategic discussions . He added `` while this framework does not reflect the basic interests of the [34 United States 34] , we feel it is still a valid approach in our discussions with [35 Russia 35] on this matter . '' He also said that [36 Clinton 36] raised the issue of the transfer of sensitive [37 Russian 37] technology to [2 2 Iran 2] , adding `` President [

38 Putin 38] has expressed his agreement that these are real threats and is strongly committed to limiting this , but there is still a lot of work to do '' in this area . [39 Clinton 39] also touched on the matter of [40 American 40] Edmond Bob who is being tried in a closed court in [41 Russia 41] on charges of spying . The [42 United States 42] believes he is innocent of these charges and is demanding his release on humanitarian grounds . The official said [43 Putin 43] understands our concern and we hope he will do something as quickly as possible . ''

1: 56611 Bandar\_Seri\_Begawan

2: 14653 Iran

1: 471981 Agence\_France-Presse

[winner: 471981 Agence\_France-Presse, pem\_pos=0, final\_scores=0.025662467, final\_local\_scores=0.055996835, pem=-0.6605333, similarity=9.352853]

gt: 471981 Agence\_France-Presse, pem\_pos=0, final\_scores=0.025662467, final\_local\_scores=0.055996835, pem=-0.6605333, similarity=9.352853 ]

2: 3466 Brunei

[winner: 3466 Brunei, pem\_pos=0, final\_scores=0.24005102, final\_local\_scores=0.26080537, pem=-0.386843, similarity=13.9763775]

gt: 3466 Brunei, pem\_pos=0, final\_scores=0.24005102, final\_local\_scores=0.26080537, pem=-0.386843, similarity=13.9763775 ]

[winner: 4280342 Sultanate\_of\_Maguindanao, pem\_pos=3, final\_scores=-1.0819365, final\_local\_scores=-1.067621, pem=-1.807297, similarity=-17.941864]

RECALL MISS: 3466 Brunei]

1: 56611 Bandar\_Seri\_Begawan

[pem\_pos=0, final\_scores=0.19522034, final\_local\_scores=0.26714152, pem=0.0, similarity=12.0617485]

2: 14653 Iran

[pem\_pos=0, final\_scores=0.48352873, final\_local\_scores=0.5814333, pem=-0.13856095, similarity=22.194359]

1: 3434750 United\_States

[pem\_pos=1, final\_scores=0.5359146, final\_local\_scores=0.5104943, pem=-0.12994486, similarity=20.030542]

2: 3434750 United\_States

[pem\_pos=0, final\_scores=0.5842634, final\_local\_scores=0.5642292, pem=-0.12994486, similarity=21.634098]

3: 3356 Bill\_Clinton

[pem\_pos=0, final\_scores=0.62014467, final\_local\_scores=0.59921974, pem=0.0, similarity=21.971624]

4: 25391 Russia

[pem\_pos=1, final\_scores=0.73145694, final\_local\_scores=0.76557124, pem=-1.0515022, similarity=32.654118]

5: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.553857,  
final\_local\_scores=0.55378044, pem=0.0, similarity=20.615625]  
6: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.29766792,  
final\_local\_scores=0.27393273, pem=0.0, similarity=12.264412]  
7: 25391 Russia  
[pem\_pos=1, final\_scores=0.8470134,  
final\_local\_scores=0.89179915, pem=-1.0515022,  
similarity=36.421013]  
8: 3434750 United\_States  
[pem\_pos=1, final\_scores=0.62008667,  
final\_local\_scores=0.6024395, pem=-0.12994486,  
similarity=22.77437]  
9: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.36255318,  
final\_local\_scores=0.34481004, pem=0.0, similarity=14.379532]  
10: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.8409886,  
final\_local\_scores=0.8425461, pem=0.0, similarity=29.23297]  
11: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.3997746,  
final\_local\_scores=0.38546878, pem=0.0, similarity=15.59287]  
12: 135976 Asia-Pacific\_Economic\_Cooperation  
[pem\_pos=0, final\_scores=0.31167847,  
final\_local\_scores=0.33596253, pem=0.0, similarity=14.115505]  
13: 56611 Bandar\_Seri\_Begawan  
[pem\_pos=0, final\_scores=0.22522517,  
final\_local\_scores=0.29991725, pem=0.0, similarity=13.039842]  
14: 3466 Brunei  
[pem\_pos=0, final\_scores=0.24005102,  
final\_local\_scores=0.26080537, pem=-0.386843,  
similarity=13.9763775]  
15: 3434750 United\_States  
[pem\_pos=1, final\_scores=0.53723353,  
final\_local\_scores=0.51193506, pem=-0.12994486,  
similarity=20.073538]  
16: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.36229026,  
final\_local\_scores=0.34452283, pem=0.0, similarity=14.370961]  
17: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.5749522,  
final\_local\_scores=0.5540581, pem=-0.073457584,  
similarity=21.023386]  
18: 25391 Russia  
[pem\_pos=0, final\_scores=0.7869895,  
final\_local\_scores=0.82623214, pem=-0.20299491,  
similarity=29.850046]  
19: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.3948025,  
final\_local\_scores=0.3800375, pem=0.0, similarity=15.43079]  
20: 25391 Russia

[pem\_pos=1, final\_scores=0.7901642,  
final\_local\_scores=0.8297, pem=-1.0515022,  
similarity=34.56785]  
21: 3434750 United\_States  
[pem\_pos=1, final\_scores=0.5982527,  
final\_local\_scores=0.5785892, pem=-0.12994486,  
similarity=22.06263]  
22: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.50659686,  
final\_local\_scores=0.50215584, pem=0.0, similarity=19.075045]  
23: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.4614253,  
final\_local\_scores=0.4300472, pem=-0.5581965,  
similarity=19.958738]  
24: 108956 Washington,\_D.C.  
[pem\_pos=0, final\_scores=0.07942413,  
final\_local\_scores=0.022393107, pem=-1.7496481,  
similarity=14.272826]  
25: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.5451418,  
final\_local\_scores=0.5214948, pem=-0.073457584,  
similarity=20.051632]  
26: 36647151 Missile\_defense\_systems\_of\_various\_nations  
[pem\_pos=0, final\_scores=0.124196395,  
final\_local\_scores=0.1576834, pem=-0.014117905,  
similarity=8.872075]  
27: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.79693323,  
final\_local\_scores=0.79442227, pem=0.0, similarity=27.796858]  
28: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.36830634,  
final\_local\_scores=0.38845122, pem=-0.1310191,  
similarity=16.394375]  
29: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.33358115,  
final\_local\_scores=0.2899942, pem=0.0, similarity=12.743718]  
30: 36647151 Missile\_defense\_systems\_of\_various\_nations  
[pem\_pos=0, final\_scores=0.16410069,  
final\_local\_scores=0.20127279, pem=-0.014117905,  
similarity=10.17287]  
31: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.80203825,  
final\_local\_scores=0.7999988, pem=0.0, similarity=27.963274]  
32: 3434750 United\_States  
[pem\_pos=1, final\_scores=0.5671497,  
final\_local\_scores=0.5446139, pem=-0.12994486,  
similarity=21.048738]  
33: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.519669,  
final\_local\_scores=0.49366963, pem=-0.073457584,  
similarity=19.221273]  
34: 3434750 United\_States

[pem\_pos=0, final\_scores=0.53037864,  
 final\_local\_scores=0.5053683, pem=-0.073457584,  
 similarity=19.570385]  
 35: 25391 Russia  
 [pem\_pos=0, final\_scores=0.73656136,  
 final\_local\_scores=0.7711471, pem=-0.20299491,  
 similarity=28.206198]  
 36: 3356 Bill\_Clinton  
 [pem\_pos=0, final\_scores=0.8581274,  
 final\_local\_scores=0.86126757, pem=0.0, similarity=29.791655]  
 37: 25391 Russia  
 [pem\_pos=1, final\_scores=0.75628895,  
 final\_local\_scores=0.7926965, pem=-1.0515022,  
 similarity=33.46359]  
 38: 32817 Vladimir\_Putin  
 [pem\_pos=0, final\_scores=0.4012639,  
 final\_local\_scores=0.38709563, pem=0.0, similarity=15.641419]  
 39: 3356 Bill\_Clinton  
 [pem\_pos=0, final\_scores=0.78140956,  
 final\_local\_scores=0.77746505, pem=0.0, similarity=27.290821]  
 40: 3434750 United\_States  
 [pem\_pos=1, final\_scores=0.524978,  
 final\_local\_scores=0.49946886, pem=-0.12994486,  
 similarity=19.70152]  
 41: 25391 Russia  
 [pem\_pos=0, final\_scores=0.7865427,  
 final\_local\_scores=0.8257441, pem=-0.20299491,  
 similarity=29.835482]  
 42: 3434750 United\_States  
 [pem\_pos=0, final\_scores=0.49408442,  
 final\_local\_scores=0.46572238, pem=-0.073457584,  
 similarity=18.387272]  
 43: 32817 Vladimir\_Putin  
 [pem\_pos=0, final\_scores=0.42662138,  
 final\_local\_scores=0.41479474, pem=0.0, similarity=16.468016]

Sample: b'NYT20001124\_2050\_0257'

ECONOMISTS &LR; &QL; &UR; By MARK LANDLER &LR; &QC; &QL;  
 &UR; c. 2000 N.Y. Times News Service &LR; &QC; &QL; [1 1 HONG  
 KONG 1] \_ For most of this year , [2 2 Asia 2] 's economies  
 have performed a sort of levitating act \_ growing by double-  
 digit rates month after month even as economists predicted  
 they would fall to earth . That dichotomy was on vivid  
 display here Friday : [1 Hong Kong 1] reported that its  
 economic output surged 10.4 percent in the third quarter ,  
 and it raised its projection for full-year growth to 10  
 percent from 8.5 percent . Yet a block away , the [3 3 Intern  
 ational Monetary Fund 3] warned in a news conference that [2  
 Asian 2] economies could be hit hard by soaring oil prices  
 and turmoil in the financial markets . The [3 IMF 3] said a \$  
 5 increase in the price of a barrel of oil could slow the  
 growth rates of [4 Asian 4] countries next year . `` The



prospects for growth in 2001 have weakened , '' said David Robinson , assistant director of research at the fund . `` There has also been increased political instability across the world , not least in [5 Asia 5] . '' Several [6 Asian 6] countries \_ notably the [4 4 Philippines 4] , [5 5 Taiwan 5] , and [6 6 Indonesia 6] \_ have been gripped by political tumult in recent weeks . The [7 Philippines 7] president , [8 Joseph Estrada 8] , was impeached last week , while [7 7 Chen Shui-bian 7] , the president of [9 Taiwan 9] , is trying to fend off a recall vote . Even in more stable countries , like [8 8 Malaysia 8] and [9 9 South Korea 9] , Robinson said the Fund was concerned by the haphazard pace of economic reforms . In many countries , the banks and corporations remain financially shaky . `` We 're still seeing chaebols in deep trouble , '' he said , referring to [10 Korea 10] 's heavily indebted conglomerates . But Robinson acknowledged that these weaknesses had not yet been reflected in the economic performance of the countries . On Tuesday , [11 South Korea 11] reported a 9.2 percent increase in its third-quarter output \_ beating the estimates of most analysts and extending a string of powerful quarters . [12 Indonesia 12] , [13 Malaysia 13] , and [10 10 Singapore 10] have also beaten expectations , while [14 Hong Kong 14] 's growth rate was nearly three percentage points above a median forecast of 12 economists surveyed by [15 1 Bloomberg News 15] . Like other [16 Asian 16] economies , [17 Hong Kong 17] has been buoyed by an unquenchable appetite for its exports , principally in the [11 11 United States 11] . Exports surged 17.7 percent in the third quarter . [18 Hong Kong 18] has done a particularly booming business in repackaging and shipping goods from mainland [12 12 China 12] . Consumer spending rose by a much less impressive 5.6 percent , as retail sales slackened from the second quarter . And consumer prices fell 2 percent , extending a decline that began during the crisis . Ian Perkin , chief economist at the [13 13 Hong Kong General Chamber of Commerce 13] said the double-digit growth in [19 Hong Kong 19] and other [20 Asian 20] countries masked their underlying condition . While the region 's growth in exports has been torrid , domestic demand has remained tenuous . `` We have n't built up our domestic economies yet , '' he said . `` The recovery has relied almost completely on export growth . `` If consumer spending slows down in the [21 States 21] , it will have a devastating effect on [22 Asia 22] , '' said Perkin . `` 2001 could be a crunch year . '' Robinson of the [23 IMF 23] was similarly wary about next year . In October , the [24 IMF 24] projected that the world economy would grow 4.2 percent in 2001 , while [25 Asia 25] as a region would grow 6.6 percent . On Friday , Robinson said the growth rate could be lower , though he would not say by how much . A principal culprit of possibly slower growth is the dependence of many [26 Asian 26] countries on imported oil . Oil prices have risen \$ 5 per barrel since the [27 IMF 27]

issued its semi-annual economic outlook in October . [28 Asia 28] has little control over oil , but Robinson said the region could improve its economic prospects by calming its political waters . One bright spot in the [29 Asian 29] landscape is [30 China 30] , which is on track to record 8 percent growth in 2000 . Robinson said the main question facing [31 Chinese 31] leaders was whether they need to spend billions of dollars in public funds to further stimulate an economy that was already expanding . Like Perkin , Robinson said [32 Asia 32] 's future depended in large part on the health of the [33 United States 33] . But he disputed those who predict dramatically slowing growth \_ the so-called hard landing \_ for the [34 American 34] economy . `` The chances of that happening are less than 40 percent , '' he said , `` substantially less . ''

1: 13404 Hong\_Kong

2: 689 Asia

3: 15251 International\_Monetary\_Fund

4: 23440 Philippines

5: 25734 Taiwan

6: 14579 Indonesia

7: 55077 Chen\_Shui-bian

8: 3607937 Malaysia

9: 27019 South\_Korea

10: 27318 Singapore

11: 3434750 United\_States

12: 5405 China

13: 18058884 Hong\_Kong\_General\_Chamber\_of\_Commerce

1: 3057255 Bloomberg\_News

[winner: 1974777 Bloomberg\_Television, pem\_pos=0, final\_scores=0.2526068, final\_local\_scores=0.3292271, pem=0.0, similarity=13.914506]

gt: 3057255 Bloomberg\_News, pem\_pos=2, final\_scores=-0.08100924, final\_local\_scores=-0.039545655, pem=-2.0322127, similarity=13.961081 ]

1: 13404 Hong\_Kong

[pem\_pos=0, final\_scores=0.107858464, final\_local\_scores=0.027755171, pem=-0.22402541, similarity=6.1362677]

2: 689 Asia

[pem\_pos=0, final\_scores=0.6468355, final\_local\_scores=0.6168384, pem=-0.1494309, similarity=23.31003]

3: 15251 International\_Monetary\_Fund

[pem\_pos=0, final\_scores=0.4093879, final\_local\_scores=0.47171706, pem=0.0, similarity=18.16669]

4: 23440 Philippines

[pem\_pos=0, final\_scores=0.38958198, final\_local\_scores=0.3533911, pem=-0.18084323, similarity=15.619061]

5: 25734 Taiwan

[pem\_pos=0, final\_scores=0.4399401,

final\_local\_scores=0.394094, pem=-0.077886656,  
similarity=16.273823]  
6: 14579 Indonesia  
[pem\_pos=0, final\_scores=0.5266091,  
final\_local\_scores=0.49707502, pem=-0.23151328,  
similarity=20.182428]  
7: 55077 Chen\_Shui-bian  
[pem\_pos=0, final\_scores=0.6246091,  
final\_local\_scores=0.73399204, pem=0.0, similarity=25.9935]  
8: 3607937 Malaysia  
[pem\_pos=0, final\_scores=0.6035941,  
final\_local\_scores=0.58344597, pem=-0.20223404,  
similarity=22.600685]  
9: 27019 South\_Korea  
[pem\_pos=0, final\_scores=0.4532469,  
final\_local\_scores=0.4344625, pem=-0.22449642,  
similarity=18.275787]  
10: 27318 Singapore  
[pem\_pos=0, final\_scores=0.5053318,  
final\_local\_scores=0.46333128, pem=-0.1922004,  
similarity=18.961658]  
11: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.5316796,  
final\_local\_scores=0.5178029, pem=-0.073457584,  
similarity=19.941458]  
12: 5405 China  
[pem\_pos=0, final\_scores=0.5321843,  
final\_local\_scores=0.49093425, pem=-0.15669827,  
similarity=19.59232]  
13: 18058884 Hong\_Kong\_General\_Chamber\_of\_Commerce  
[pem\_pos=0, final\_scores=0.12614374,  
final\_local\_scores=0.20570076, pem=0.0, similarity=10.228233]  
1: 13404 Hong\_Kong  
[pem\_pos=0, final\_scores=0.5237066,  
final\_local\_scores=0.48200607, pem=-0.22402541,  
similarity=19.69202]  
2: 689 Asia  
[pem\_pos=1, final\_scores=0.49135584,  
final\_local\_scores=0.44700038, pem=-1.9623682,  
similarity=28.10076]  
3: 15251 International\_Monetary\_Fund  
[pem\_pos=0, final\_scores=0.31267977,  
final\_local\_scores=0.36607814, pem=0.0, similarity=15.0142145]  
4: 689 Asia  
[pem\_pos=1, final\_scores=0.5593584,  
final\_local\_scores=0.5215661, pem=-1.9623682,  
similarity=30.32595]  
5: 689 Asia  
[pem\_pos=0, final\_scores=0.59520596,  
final\_local\_scores=0.56044084, pem=-0.1494309,  
similarity=21.627014]  
6: 689 Asia

[pem\_pos=1, final\_scores=0.42462987,  
final\_local\_scores=0.37411243, pem=-1.9623682,  
similarity=25.925638]  
7: 23440 Philippines  
[pem\_pos=0, final\_scores=0.43492067,  
final\_local\_scores=0.40291673, pem=-0.18084323,  
similarity=17.097006]  
8: 279629 Joseph\_Estrada  
[pem\_pos=0, final\_scores=0.3799414,  
final\_local\_scores=0.4913966, pem=0.0, similarity=18.753967]  
9: 25734 Taiwan  
[pem\_pos=0, final\_scores=0.4739309,  
final\_local\_scores=0.4312238, pem=-0.077886656,  
similarity=17.381851]  
10: 16749 Korea  
[pem\_pos=0, final\_scores=0.43479413,  
final\_local\_scores=0.4030425, pem=-0.27902022,  
similarity=17.63466]  
11: 27019 South\_Korea  
[pem\_pos=0, final\_scores=0.4219551,  
final\_local\_scores=0.400281, pem=-0.22449642,  
similarity=17.255743]  
12: 14579 Indonesia  
[pem\_pos=0, final\_scores=0.61001265,  
final\_local\_scores=0.5881807, pem=-0.23151328,  
similarity=22.901205]  
13: 3607937 Malaysia  
[pem\_pos=0, final\_scores=0.6503075,  
final\_local\_scores=0.63447315, pem=-0.20223404,  
similarity=24.123438]  
14: 13404 Hong\_Kong  
[pem\_pos=0, final\_scores=0.47259772,  
final\_local\_scores=0.42617738, pem=-0.22402541,  
similarity=18.025982]  
15: 1974777 Bloomberg\_Television  
[pem\_pos=0, final\_scores=0.2526068,  
final\_local\_scores=0.3292271, pem=0.0, similarity=13.914506]  
16: 689 Asia  
[pem\_pos=1, final\_scores=0.51389974,  
final\_local\_scores=0.47190946, pem=-1.9623682,  
similarity=28.844097]  
17: 13404 Hong\_Kong  
[pem\_pos=0, final\_scores=0.48868436,  
final\_local\_scores=0.4437496, pem=-0.22402541,  
similarity=18.550371]  
18: 13404 Hong\_Kong  
[pem\_pos=0, final\_scores=0.5676977,  
final\_local\_scores=0.5300597, pem=-0.22402541,  
similarity=21.126038]  
19: 13404 Hong\_Kong  
[pem\_pos=0, final\_scores=0.472524,  
final\_local\_scores=0.4260968, pem=-0.22402541,

similarity=18.023577]  
20: 689 Asia  
[pem\_pos=1, final\_scores=0.46761686,  
final\_local\_scores=0.4210692, pem=-1.9623682,  
similarity=27.326921]  
21: 3434750 United\_States  
[pem\_pos=2, final\_scores=0.04130064,  
final\_local\_scores=-0.020647615, pem=-2.0821934,  
similarity=14.796838]  
22: 689 Asia  
[pem\_pos=0, final\_scores=0.533978,  
final\_local\_scores=0.49355865, pem=-0.1494309,  
similarity=19.631115]  
23: 15251 International\_Monetary\_Fund  
[pem\_pos=0, final\_scores=0.47938174,  
final\_local\_scores=0.5505839, pem=0.0, similarity=20.520235]  
24: 15251 International\_Monetary\_Fund  
[pem\_pos=0, final\_scores=0.3222453,  
final\_local\_scores=0.37652707, pem=0.0, similarity=15.326033]  
25: 689 Asia  
[pem\_pos=0, final\_scores=0.5971815,  
final\_local\_scores=0.5625989, pem=-0.1494309,  
similarity=21.691414]  
26: 689 Asia  
[pem\_pos=1, final\_scores=0.5007779,  
final\_local\_scores=0.45757586, pem=-1.9623682,  
similarity=28.416353]  
27: 15251 International\_Monetary\_Fund  
[pem\_pos=0, final\_scores=0.36396933,  
final\_local\_scores=0.42210418, pem=0.0, similarity=16.686144]  
28: 689 Asia  
[pem\_pos=0, final\_scores=0.622866,  
final\_local\_scores=0.59065527, pem=-0.1494309,  
similarity=22.528673]  
29: 689 Asia  
[pem\_pos=1, final\_scores=0.5039223,  
final\_local\_scores=0.46072733, pem=-1.9623682,  
similarity=28.510399]  
30: 5405 China  
[pem\_pos=0, final\_scores=0.57196194,  
final\_local\_scores=0.5343853, pem=-0.15669827,  
similarity=20.888987]  
31: 5405 China  
[pem\_pos=1, final\_scores=0.6537663,  
final\_local\_scores=0.6237441, pem=-1.4196125,  
similarity=30.423553]  
32: 689 Asia  
[pem\_pos=0, final\_scores=0.5638228,  
final\_local\_scores=0.5261596, pem=-0.1494309,  
similarity=20.603992]  
33: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.52152085,

final\_local\_scores=0.50670606, pem=-0.073457584,  
similarity=19.610306]  
34: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.5665481,  
final\_local\_scores=0.5558915, pem=-0.12994486,  
similarity=21.385286]

Sample: b'20001115\_AFP\_ARB\_0217\_eng'

[1 1 Bandar Seri Begawan 1] 11-15 -LRB- [2 2 AFP 2] -RRB- -  
A [1 Japanese 1] official announced today Wednesday that [2  
Japan 2] and [3 3 Russia 3] are to hold a summit before the  
end of the current year in an attempt to find a solution to  
their dispute about the Coral Islands , which has hindered  
the signing of a peace treaty between the two countries since  
the end of [3 World War II 3] . The decision to hold the  
summit was made during the meeting between [4 Russian 4]  
President [5 Vladimir Putin 5] and [6 Japanese 6] Prime  
Minister [7 Yoshiro Mori 7] on the periphery of the [8 Asia-  
Pacific Economic Cooperation 8] forum in [4 4 Brunei 4] . The  
[9 Japanese 9] official said that those responsible had ``  
agreed to meet at 1 Irkotsik -LRB- east [10 Russia 10] -RRB-  
before the end of the year . In 1997 , [11 Russian 11]  
President [12 Boris Yeltsin 12] and [13 Japanese 13] Prime  
Minister Rirotaro Hashimoto agreed to sign the peace treaty  
before the end of 2000 , but negotiations have not progressed  
much since then . [14 Russia 14] and [15 Japan 15] have been  
unable to sign a peace agreement since 1945 because of their  
dispute concerning four of the Coral islands located between [  
16 Russia 16] and [17 Japan 17] and which were invaded by [18  
Russian 18] forces in the last few days of [19 World War II 19  
] . [20 Putin 20] visited [21 Japan 21] last September but  
the visit did not produce any result on the issue of the  
islands .

1: 56611 Bandar\_Seri\_Begawan  
2: 471981 Agence\_France-Presse  
3: 25391 Russia  
4: 3466 Brunei

1: 56872 Irkutsk no overlap with any filtered span

1: 56611 Bandar\_Seri\_Begawan  
[pem\_pos=0, final\_scores=0.27966404,  
final\_local\_scores=0.34509513, pem=0.0, similarity=14.38804]  
2: 471981 Agence\_France-Presse  
[pem\_pos=0, final\_scores=0.04461299,  
final\_local\_scores=0.09806129, pem=-0.6605333,  
similarity=10.60814]  
3: 25391 Russia  
[pem\_pos=0, final\_scores=0.68318623,  
final\_local\_scores=0.6677543, pem=-0.20299491,  
similarity=25.12075]  
4: 3466 Brunei  
[pem\_pos=0, final\_scores=0.27817303,  
final\_local\_scores=0.269157, pem=-0.386843,

similarity=14.225607]  
1: 15573 Japan  
[pem\_pos=0, final\_scores=0.698295,  
final\_local\_scores=0.6771792, pem=-0.99076414,  
similarity=29.68602]  
2: 15573 Japan  
[pem\_pos=0, final\_scores=0.7222782,  
final\_local\_scores=0.7033772, pem=-0.13549899,  
similarity=25.816755]  
3: 32927 World\_War\_II  
[pem\_pos=0, final\_scores=0.43250477,  
final\_local\_scores=0.44018018, pem=0.0, similarity=17.225567]  
4: 25391 Russia  
[pem\_pos=2, final\_scores=0.8045389,  
final\_local\_scores=0.8003137, pem=-1.0515022,  
similarity=33.690903]  
5: 32817 Vladimir\_Putin  
[pem\_pos=0, final\_scores=0.50207466,  
final\_local\_scores=0.53513205, pem=0.0, similarity=20.05912]  
6: 15573 Japan  
[pem\_pos=0, final\_scores=0.887098,  
final\_local\_scores=0.8834178, pem=-0.99076414,  
similarity=35.84059]  
7: 352104 Yoshirō\_Mori  
[pem\_pos=0, final\_scores=0.56097984,  
final\_local\_scores=0.674694, pem=0.0, similarity=24.223928]  
8: 135976 Asia-Pacific\_Economic\_Cooperation  
[pem\_pos=0, final\_scores=0.3278548,  
final\_local\_scores=0.3704533, pem=0.0, similarity=15.144778]  
9: 15573 Japan  
[pem\_pos=0, final\_scores=0.7419883,  
final\_local\_scores=0.72490746, pem=-0.99076414,  
similarity=31.110325]  
10: 25391 Russia  
[pem\_pos=0, final\_scores=0.70001,  
final\_local\_scores=0.6861318, pem=-0.20299491,  
similarity=25.66917]  
11: 25391 Russia  
[pem\_pos=2, final\_scores=0.8991545,  
final\_local\_scores=0.9036668, pem=-1.0515022,  
similarity=36.775166]  
12: 23278520 Boris\_Yeltsin  
[pem\_pos=0, final\_scores=0.61316067,  
final\_local\_scores=0.6716319, pem=0.0, similarity=24.132547]  
13: 15573 Japan  
[pem\_pos=0, final\_scores=0.91047484,  
final\_local\_scores=0.90895337, pem=-0.99076414,  
similarity=36.602623]  
14: 25391 Russia  
[pem\_pos=0, final\_scores=0.75853986,  
final\_local\_scores=0.75006664, pem=-0.20299491,  
similarity=27.577114]

15: 15573 Japan  
 [pem\_pos=0, final\_scores=0.6114383,  
 final\_local\_scores=0.58230144, pem=-0.13549899,  
 similarity=22.203613]  
 16: 25391 Russia  
 [pem\_pos=0, final\_scores=0.921048,  
 final\_local\_scores=0.9275821, pem=-0.20299491,  
 similarity=32.87453]  
 17: 15573 Japan  
 [pem\_pos=0, final\_scores=0.640061,  
 final\_local\_scores=0.6135674, pem=-0.13549899,  
 similarity=23.136652]  
 18: 25391 Russia  
 [pem\_pos=3, final\_scores=0.90477365,  
 final\_local\_scores=0.9098049, pem=-1.0515022,  
 similarity=36.95834]  
 19: 32927 World\_War\_II  
 [pem\_pos=0, final\_scores=0.4434812,  
 final\_local\_scores=0.45217025, pem=0.0, similarity=17.583376]  
 20: 32817 Vladimir\_Putin  
 [pem\_pos=0, final\_scores=0.4025281,  
 final\_local\_scores=0.42639256, pem=0.0, similarity=16.814117]  
 21: 15573 Japan  
 [pem\_pos=0, final\_scores=0.6538184,  
 final\_local\_scores=0.6285953, pem=-0.13549899,  
 similarity=23.585114]

Sample: b'chtb\_267\_eng'

[1 1 Xinhua News Agency 1] , [2 2 Beijing 2] , September 1st , by reporter Jingcai Wu The [3 3 Weir Group 3] of the [4 4 US 4] recently signed an implementation agreement in [1 Beijing 1] with [5 5 Jiangsu Province 5] 's Electric Power Department to build [6 6 China 6] 's first large-capacity power plant with liquefied natural gas as fuel . Aixing Dan , director of the International Co-operation Department of the Electric Power Industry Ministry , on this matter , expressed to [2 Xinhua News Agency 2] reporters today that if this power plant is successful , it will be a model for similar plants to be built along coastal regions . Coal , water , oil , etc. , are normally used for traditional electricity generation . However , using liquefied natural gas as fuel for joint circulatory electricity generation has advantages of higher power generation efficiency and lower total cost , is beneficial to the environment , etc. . According to presentations , with this type of electricity generation method , a single engine with liquefied natural gas as fuel can produce waste gas with temperatures being as high as 500 to 600 degrees centigrade . This gas in turn can heat a boiler for the production of steam to turn a steam turbine which generates electricity . Because of the efficiency of this construction method , the electricity generated is relatively competitive in price . At the same time , using



natural gas as fuel will also reduce environmental pollution along coastal regions . The [3 Weir Group 3] , whose headquarters is in the [4 US 4] , is a large , specialized corporation investing in the area of electricity generation . The installed capacity of electricity it has invested in and developed globally has reached 15 million kilowatts . Through the contact of the Yizhong Company of the [5 US 5] , the [6 Weir Group 6] and the Jiangsu Electric Power Department reached an initial agreement on a plan to build a joint circulatory electricity generation plant . This power plant , which will be situated in [7 7 Rudong 7] , [7 Jiangsu 7] , has an annual generation capacity of 2.4 million kilowatts . The investment amount is 2.4 billion US dollars . The [8 Chinese 8] side holds 30 % of the shares while the [9 Weir Group 9] holds 70 % . According to the agreement , the [10 Weir Group 10] will be responsible for the supply of natural gas from the international market for use by the power plant . Renzu Luo , managing director of the [11 Weir Group 11] , is optimistic over the prospect of using liquefied natural gas as fuel for electricity generation along [12 China 12] 's coastal regions . He holds that according to the experience of [13 Asian 13] energy market development , liquefied natural gas has become the main material of clean fuel used for electricity generation . When the gross national product per capita reached 3000 US dollars in [8 8 Japan 8] , [9 9 South Korea 9] and [10 10 Taiwan 10] , etc. , environmental protection awareness generally rose , and all began to obviously increase usage of electricity generated by liquefied natural gas . As the gross national product in [14 China 14] 's coastal regions begins to approach this benchmark , it can be said to have had the conditions for electricity generation using natural gas in terms of financial resources . John Weir , president of the [15 Weir Group 15] who visited [16 China 16] with [17 US 17] Minister Brown and signed this agreement for electricity generation plant construction in [18 Jiangsu 18] , held that now is the best time to develop a project of joint circulatory electricity generation plant with natural gas as fuel in [19 China 19] . Both parties , the [20 US 20] and [21 China 21] , will invest more time , money and resources to make this power plant project a successful model . -LRB- End -RRB-

1: 263168 Xinhua\_News\_Agency

2: 18603746 Beijing

3: 3762139 Weir\_Group

4: 3434750 United\_States

5: 166412 Jiangsu

6: 5405 China

7: 8094498 Rudong\_County

8: 15573 Japan

9: 27019 South\_Korea

10: 25734 Taiwan

1: 263168 Xinhua\_News\_Agency

[pem\_pos=0, final\_scores=0.42652917,  
final\_local\_scores=0.4661516, pem=0.0, similarity=18.000607]  
2: 18603746 Beijing  
[pem\_pos=0, final\_scores=0.36944634,  
final\_local\_scores=0.35182628, pem=-0.042161036,  
similarity=14.818189]  
3: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.03443624,  
final\_local\_scores=0.07556935, pem=0.0, similarity=6.344853]  
4: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.12268554,  
final\_local\_scores=0.07511519, pem=-0.48755783,  
similarity=8.982716]  
5: 166412 Jiangsu  
[pem\_pos=0, final\_scores=0.43441266,  
final\_local\_scores=0.4383132, pem=0.0, similarity=17.169853]  
6: 5405 China  
[pem\_pos=0, final\_scores=0.47099483,  
final\_local\_scores=0.42459887, pem=-0.15669827,  
similarity=17.61274]  
7: 8094498 Rudong\_County  
[pem\_pos=0, final\_scores=0.037438974,  
final\_local\_scores=0.13028064, pem=0.0, similarity=7.977547]  
8: 15573 Japan  
[pem\_pos=0, final\_scores=0.6883177,  
final\_local\_scores=0.7053793, pem=-0.13549899,  
similarity=25.876503]  
9: 27019 South\_Korea  
[pem\_pos=0, final\_scores=0.43431228,  
final\_local\_scores=0.42994446, pem=-0.22449642,  
similarity=18.14096]  
10: 25734 Taiwan  
[pem\_pos=0, final\_scores=0.37724334,  
final\_local\_scores=0.34026673, pem=-0.077886656,  
similarity=14.667511]  
1: 18603746 Beijing  
[pem\_pos=0, final\_scores=0.490138,  
final\_local\_scores=0.48366362, pem=-0.042161036,  
similarity=18.75248]  
2: 263168 Xinhua\_News\_Agency  
[pem\_pos=0, final\_scores=0.48230022,  
final\_local\_scores=0.527073, pem=0.0, similarity=19.818623]  
3: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.26284385,  
final\_local\_scores=0.32506993, pem=0.0, similarity=13.790448]  
4: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.1895477,  
final\_local\_scores=0.14815193, pem=-0.48755783,  
similarity=11.162277]  
5: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.13500391,  
final\_local\_scores=0.08857112, pem=-0.48755783,

similarity=9.384267]  
6: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.18722962,  
final\_local\_scores=0.24247289, pem=0.0, similarity=11.325587]  
7: 166412 Jiangsu  
[pem\_pos=0, final\_scores=0.3624162,  
final\_local\_scores=0.35966796, pem=0.0, similarity=14.822922]  
8: 5405 China  
[pem\_pos=1, final\_scores=0.54688996,  
final\_local\_scores=0.50750273, pem=-1.4196125,  
similarity=26.954681]  
9: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.3018586,  
final\_local\_scores=0.36768764, pem=0.0, similarity=15.062246]  
10: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.3131292,  
final\_local\_scores=0.37999904, pem=0.0, similarity=15.429642]  
11: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.2610495,  
final\_local\_scores=0.3231099, pem=0.0, similarity=13.7319565]  
12: 5405 China  
[pem\_pos=0, final\_scores=0.48934597,  
final\_local\_scores=0.4446447, pem=-0.15669827,  
similarity=18.210947]  
13: 689 Asia  
[pem\_pos=1, final\_scores=0.44321436,  
final\_local\_scores=0.44141978, pem=-1.9623682,  
similarity=27.934223]  
14: 5405 China  
[pem\_pos=0, final\_scores=0.48364443,  
final\_local\_scores=0.4384166, pem=-0.15669827,  
similarity=18.02509]  
15: 3762139 Weir\_Group  
[pem\_pos=0, final\_scores=0.24570523,  
final\_local\_scores=0.30634865, pem=0.0, similarity=13.231768]  
16: 5405 China  
[pem\_pos=0, final\_scores=0.5756379,  
final\_local\_scores=0.5389055, pem=-0.15669827,  
similarity=21.023878]  
17: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.2090541,  
final\_local\_scores=0.1694597, pem=-0.48755783,  
similarity=11.798144]  
18: 166412 Jiangsu  
[pem\_pos=0, final\_scores=0.3206777,  
final\_local\_scores=0.314075, pem=0.0, similarity=13.4623375]  
19: 5405 China  
[pem\_pos=0, final\_scores=0.554838,  
final\_local\_scores=0.5161848, pem=-0.15669827,  
similarity=20.345846]  
20: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.22382103,

final\_local\_scores=0.18559033, pem=-0.48755783,  
similarity=12.279513]  
21: 5405 China  
[pem\_pos=0, final\_scores=0.5199551,  
final\_local\_scores=0.4780805, pem=-0.15669827,  
similarity=19.208738]

Sample: b'20001015\_AFP\_ARB\_0229\_eng'

[1 1 London 1] 10-16 -LRB- 1 AFP -RRB- - A Saudi Airways  
spokesman in [1 London 1] announced that passengers on the [2  
Saudi 2] 777 [3 Boeing 3] , hijacked to [2 2 Baghdad 2] on  
Saturday and which arrived in [3 3 Riyadh 3] yesterday  
evening Sunday , left the [4 Saudi 4] capital for [5 London 5]  
at 2h43 local time -LRB- GMT 22h43 -RRB- , approximately 30  
hours after the hijacking operation by two [6 Saudi 6]  
hijackers began . The [7 Boeing 7] arrived at 2 Riyadh Intern  
ational Airport at 21h30 yesterday Sunday from [8 Baghdad 8]  
through [9 Jordanian 9] air space with 104 passengers on  
board , including 14 crew members . Saudi Airways said  
Saturday that 40 [10 British 10] , 14 [11 Saudis 11] , 15 [12  
Pakistanis 12] , 4 [13 Yemenis 13] and 4 South Africans , 2 [1  
4 Kenyans 14] and an [15 American 15] , [16 Frenchman 16] , [1  
7 Spaniard 17] , [18 Swiss 18] , [19 Swede 19] , [20 Nigerian  
20] , [21 Indian 21] , [22 Omani 22] , [23 Palestinian 23]  
and [24 Lebanese 24] were on board the plane .

1: 17867 London

2: 4492 Baghdad

3: 51554 Riyadh

1: 471981 Agence\_France-Presse

[winner: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=-0.039916888, final\_local\_scores=-0.028028838,  
pem=-0.6605333, similarity=6.8453593]

gt: 471981 Agence\_France-Presse, pem\_pos=0,  
final\_scores=-0.039916888, final\_local\_scores=-0.028028838,  
pem=-0.6605333, similarity=6.8453593 ]

2: 471186 King\_Khalid\_International\_Airport

[winner: 51554 Riyadh, pem\_pos=0,  
final\_scores=-0.3747313, final\_local\_scores=-0.46484807,  
pem=0.0, similarity=-9.782281]

gt: 471186 King\_Khalid\_International\_Airport, pem\_pos=2,  
final\_scores=-0.667352, final\_local\_scores=-0.6744125,  
pem=-1.7747052, similarity=-6.3849792 ]

[winner: 150179

International\_Air\_Transport\_Association\_airport\_code,  
pem\_pos=1, final\_scores=-0.73735106,  
final\_local\_scores=-0.72118855, pem=-1.6488862,  
similarity=-8.465092]

RECALL MISS: 471186 King\_Khalid\_International\_Airport]

1: 17867 London

[pem\_pos=0, final\_scores=0.49446756,  
final\_local\_scores=0.47702777, pem=-0.09086737,  
similarity=18.819323]

2: 4492 Baghdad  
[pem\_pos=0, final\_scores=0.30960983,  
final\_local\_scores=0.2879748, pem=0.0, similarity=12.683455]  
3: 51554 Riyadh  
[pem\_pos=0, final\_scores=0.39345628,  
final\_local\_scores=0.38037592, pem=0.0, similarity=15.440888]  
1: 17867 London  
[pem\_pos=0, final\_scores=0.5897775,  
final\_local\_scores=0.5811395, pem=-0.09086737,  
similarity=21.926226]  
2: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.5210035,  
final\_local\_scores=0.47513413, pem=0.0, similarity=18.268663]  
3: 18933266 Boeing  
[pem\_pos=0, final\_scores=0.11168726,  
final\_local\_scores=0.15804121, pem=0.0, similarity=8.805978]  
4: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.5525088,  
final\_local\_scores=0.5095489, pem=0.0, similarity=19.295668]  
5: 17867 London  
[pem\_pos=0, final\_scores=0.5836026,  
final\_local\_scores=0.5743943, pem=-0.09086737,  
similarity=21.724934]  
6: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.5521783,  
final\_local\_scores=0.5091879, pem=0.0, similarity=19.284895]  
7: 18933266 Boeing  
[pem\_pos=0, final\_scores=0.11682229,  
final\_local\_scores=0.16365045, pem=0.0, similarity=8.973369]  
8: 4492 Baghdad  
[pem\_pos=0, final\_scores=0.35546392,  
final\_local\_scores=0.33806345, pem=0.0, similarity=14.178201]  
9: 7515964 Jordan  
[pem\_pos=0, final\_scores=0.5449013,  
final\_local\_scores=0.5062353, pem=-0.18608934,  
similarity=20.208767]  
10: 31717 United\_Kingdom  
[pem\_pos=0, final\_scores=0.5027266,  
final\_local\_scores=0.47149992, pem=-0.48936236,  
similarity=20.82144]  
11: 349303 Saudi\_Arabia  
[pem\_pos=0, final\_scores=0.34224457,  
final\_local\_scores=0.27986714, pem=0.0, similarity=12.441506]  
12: 23235 Pakistan  
[pem\_pos=6, final\_scores=0.038958922,  
final\_local\_scores=-0.02418974, pem=-1.993598,  
similarity=14.209339]  
13: 350939 Yemen  
[pem\_pos=0, final\_scores=0.38391858,  
final\_local\_scores=0.35256836, pem=-0.30390495,  
similarity=16.263739]  
14: 188171 Kenya

[pem\_pos=0, final\_scores=0.19169803,  
 final\_local\_scores=0.17747179, pem=0.0, similarity=9.385825]  
 15: 3434750 United\_States  
 [pem\_pos=0, final\_scores=0.3864196,  
 final\_local\_scores=0.34841815, pem=-0.12994486,  
 similarity=15.193866]  
 16: 5843419 France  
 [pem\_pos=1, final\_scores=0.16625924,  
 final\_local\_scores=0.12063441, pem=-0.7184992,  
 similarity=11.596994]  
 17: 26667 Spain  
 [pem\_pos=1, final\_scores=0.12441988,  
 final\_local\_scores=0.088844284, pem=-0.7552963,  
 similarity=10.848422]  
 18: 26748 Switzerland  
 [pem\_pos=0, final\_scores=0.4700442,  
 final\_local\_scores=0.46701646, pem=-0.1392413,  
 similarity=18.783632]  
 19: 5058739 Sweden  
 [pem\_pos=0, final\_scores=0.2750855,  
 final\_local\_scores=0.26096043, pem=-0.6464619,  
 similarity=15.392853]  
 20: 21383 Nigeria  
 [pem\_pos=0, final\_scores=0.47114503,  
 final\_local\_scores=0.47515798, pem=-0.075027294,  
 similarity=18.677385]  
 21: 14533 India  
 [pem\_pos=0, final\_scores=0.34534484,  
 final\_local\_scores=0.3253664, pem=-0.618294,  
 similarity=17.161674]  
 22: 22316 Oman  
 [pem\_pos=0, final\_scores=0.36200112,  
 final\_local\_scores=0.32472238, pem=-0.16408218,  
 similarity=14.672381]  
 23: 241405 State\_of\_Palestine  
 [pem\_pos=9, final\_scores=0.22990267,  
 final\_local\_scores=0.24610919, pem=-2.8508017,  
 similarity=26.937206]  
 24: 17771 Lebanon  
 [pem\_pos=0, final\_scores=0.501653,  
 final\_local\_scores=0.46409285, pem=-0.55808944,  
 similarity=20.974146]

Sample: b'chthb\_227\_eng'

[1 Xinhua News Agency 1] , [2 Moscow 2] , August 31st , by  
 reporter Rong Xie Presidents of 3 [3 Baltic Sea 3] countries  
 issued a joint statement on August 31st , pointing out that [4  
 Russia 4] 's withdrawal of troops from the 3 [5 Baltic Sea 5]  
 countries would `` strengthen [6 European 6] security '' .  
 The statement emphasized that [7 Russia 7] 's completing its  
 troop withdrawal should promote the development of relations  
 between [8 Russian 8] and [9 Baltic Sea 9] countries . This

joint statement was issued on the occasion of [10 Russia 10] entirely completing its troop withdrawal from the three [11 Baltic Sea 11] countries . [12 Russian 12] troops stationed in [13 Lithuania 13] had withdrawn to [14 Russia 14] one year earlier . On August 31st of this year , [15 Russian 15] troops stationed in [16 Latvia 16] withdrew completely and [17 Russian 17] troops stationed in [18 Estonia 18] also completed withdrawal on August 29th of this year . In an interview with International Correspondence Communication News Agency on the 31st , Klelov , vice-minister of [19 Russia 19] 's Foreign Affairs , pointed out that , although [20 Russia 20] has withdrawn its troops from the [21 Baltic Sea 21] countries , if the [22 Baltic Sea 22] countries infringe upon the rights of local Russian-speaking residents , the [23 Russian 23] government still has measures for exerting influence over these countries , which include political , economical and humanitarian measures . -LRB- End -RRB-

1: 263168 Xinhua\_News\_Agency

[pem\_pos=0, final\_scores=0.37059236, final\_local\_scores=0.4725926, pem=0.0, similarity=18.192818]

2: 19004 Moscow

[pem\_pos=0, final\_scores=0.28391737, final\_local\_scores=0.26476973, pem=-0.106582455, similarity=12.570581]

3: 3335 Baltic\_Sea

[pem\_pos=0, final\_scores=0.1360998, final\_local\_scores=0.08481252, pem=0.0, similarity=6.620688]

4: 25391 Russia

[pem\_pos=0, final\_scores=0.804578, final\_local\_scores=0.78352296, pem=-0.20299491, similarity=28.575518]

5: 3335 Baltic\_Sea

[pem\_pos=0, final\_scores=0.24660696, final\_local\_scores=0.20823404, pem=0.0, similarity=10.303832]

6: 9239 Europe

[pem\_pos=0, final\_scores=0.45980698, final\_local\_scores=0.45574725, pem=-0.25557294, similarity=19.079966]

7: 25391 Russia

[pem\_pos=0, final\_scores=0.84238654, final\_local\_scores=0.82482314, pem=-0.20299491, similarity=29.807999]

8: 25391 Russia

[pem\_pos=1, final\_scores=0.7988983, final\_local\_scores=0.77731884, pem=-1.0515022, similarity=33.00469]

9: 3335 Baltic\_Sea

[pem\_pos=0, final\_scores=0.12922122, final\_local\_scores=0.0772987, pem=0.0, similarity=6.3964605]

10: 25391 Russia

[pem\_pos=0, final\_scores=0.78953105, final\_local\_scores=0.7670864, pem=-0.20299491,

similarity=28.085018]  
11: 3335 Baltic\_Sea  
[pem\_pos=0, final\_scores=0.2619692,  
final\_local\_scores=0.22501495, pem=0.0, similarity=10.804608]  
12: 25391 Russia  
[pem\_pos=1, final\_scores=0.8967307,  
final\_local\_scores=0.88418585, pem=-1.0515022,  
similarity=36.193817]  
13: 17675 Lithuania  
[pem\_pos=0, final\_scores=0.582248,  
final\_local\_scores=0.57382977, pem=-0.39766282,  
similarity=23.37649]  
14: 25391 Russia  
[pem\_pos=0, final\_scores=0.7658649,  
final\_local\_scores=0.7412349, pem=-0.20299491,  
similarity=27.313559]  
15: 25391 Russia  
[pem\_pos=1, final\_scores=0.8271757,  
final\_local\_scores=0.8082075, pem=-1.0515022,  
similarity=33.926468]  
16: 17514 Latvia  
[pem\_pos=0, final\_scores=0.5596369,  
final\_local\_scores=0.5477748, pem=-0.35144013,  
similarity=22.347591]  
17: 25391 Russia  
[pem\_pos=1, final\_scores=0.83787376,  
final\_local\_scores=0.8198936, pem=-1.0515022,  
similarity=34.275204]  
18: 28222445 Estonia  
[pem\_pos=0, final\_scores=0.5699429,  
final\_local\_scores=0.57024443, pem=-0.4106629,  
similarity=23.340193]  
19: 25391 Russia  
[pem\_pos=0, final\_scores=0.9372147,  
final\_local\_scores=0.9284083, pem=-0.20299491,  
similarity=32.899185]  
20: 25391 Russia  
[pem\_pos=0, final\_scores=0.8370243,  
final\_local\_scores=0.8189656, pem=-0.20299491,  
similarity=29.633198]  
21: 3335 Baltic\_Sea  
[pem\_pos=0, final\_scores=0.22127675,  
final\_local\_scores=0.18056464, pem=0.0, similarity=9.478122]  
22: 3335 Baltic\_Sea  
[pem\_pos=0, final\_scores=0.20385383,  
final\_local\_scores=0.16153276, pem=0.0, similarity=8.910172]  
23: 25391 Russia  
[pem\_pos=1, final\_scores=0.8771396,  
final\_local\_scores=0.86278546, pem=-1.0515022,  
similarity=35.555183]

Sample: b'NYT20001129\_2040\_0383'



QUICKLY &LR; &QL; -LRB- Repeating ; this is now a [1 New York Times 1] front-page story -RRB- &QL; -LRB- ART ADV : Graphic is being sent to [2 NYT 2] graphic clients . Nonsubscribers can purchase one-time rights by calling 888-603-1036 or 888-346-9867 . -RRB- &QL; &UR; By DAVID FIRESTONE &LR; &QC; &QL; &UR; c. 2000 N.Y. Times News Service &LR; &QC; &QL; [1 1 TALLAHASSEE 1] , 1 Fla. . \_ Lawyers for Vice President [3 Al Gore 3] said Wednesday their only chance for victory in his contest of the Florida election would be shattered if they have to wait until Saturday to begin counting disputed ballots , and they began an appeal to the [4 Florida 2 Supreme Court 4] to do the counting itself , immediately . The appeal was filed with the [2 2 District Court of Appeal 2] late Wednesday afternoon , and will be submitted to the state 's highest court early Thursday , [5 Gore 5] 's lawyers said . It essentially asks the [6 Supreme Court 6] to take over [7 Gore 7] 's contest lawsuit from the slow-moving trial court , to count the ballots itself , and eventually to declare [8 Gore 8] the winner if he wins the recount . But lawyers for Gov. [9 George W. Bush 9] said they had no intention of allowing any hand counting to begin without a vigorous legal fight . To that end , they persuaded Circuit Judge N. Sanders Sauls to order 1.1 million ballots \_ every vote cast in [10 Miami-Dade 10] and [11 Palm Beach 11] counties \_ brought up from south Florida and placed in the court 's custody . That represents 83 times the number of disputed ballots that the judge had agreed on Tuesday to transport to [12 Tallahassee 12] at the [13 Gore 13] team 's request . And it could take weeks if not months to count them all , the point that Barry Richard , [14 Bush 14] 's chief trial lawyer here , said he was trying to make in requesting that the ballots be transported here . `` We did it to make a point , '' Richard said in an interview . `` If you recount any ballots , you have to recount all of them . Now , we do n't think any ballots should be counted , because the manual recount of ballots is finished . But in the event the court determines that some ballots should be counted , we wanted them here so they could all be counted . '' Aware that they could not stop the movement of all the ballots , [15 Gore 15] 's lawyers barely objected to the move , asking only that the disputed ballots that might contain hidden [16 Gore 16] votes be separated from the mass . `` I know when it 's futile , '' said [17 David Boies 17] , [18 Gore 18] 's chief trial lawyer , to Sauls . Nonetheless , [19 Gore 19] 's lawyers were furious at the move , aware that it represents yet another enormous barrier to their hopes for an immediate count of the 14,000 disputed ballots in the two counties and a quick decision from the judge that [20 Gore 20] really won the election . Even if the [21 Florida Supreme Court 21] should swiftly agree to their request and orders the counting to begin by Friday , that would still put them three days behind their original timetable . And facing a Dec. 12

deadline for the state to name electors for the [3 3 Electoral College 3] , [22 Gore 22] has absolutely no days to spare .  
`` It 's a stunt , '' said Kendall Coffey , one of [23 Gore 23] 's election lawyers , as he stormed out of Sauls ' courtroom Wednesday , referring to the request for all the ballots . `` They 're trying to bog down the case , the system , the clerk 's office and everywhere else they can .  
'' The decision by [24 Gore 24] 's lawyers to appeal the case shows how perilous their situation is . By appealing , the lawyers were forced to put on paper their fear that their case could be over if they can not begin counting ballots until Saturday . The decision to wait until a Saturday hearing `` is effectively a final order denying all relief sought by plaintiffs because the action now pending below must be completed before Dec. 12 , 2000 , in order to offer any relief , '' said the notice filed with the Court of Appeal . It adds , `` the court must allow sufficient time to complete an accurate and fair count of the contested ballots , which is essential for the proper resolution of the contest action . '' But the decision to bring up all the ballots means that they will not be assembled in [25 Tallahassee 25] until late Friday , a day later than planned . And even then , the hurdles are enormous . There will have to be another hearing before the [26 Florida Supreme Court 26] , and [27 Gore 27] 's lawyers will have to refute three arguments to be raised by the [28 Bush 28] team : There should be no more counting at all ; if there is any counting , it will have to include all 1.16 million ballots ; and if there is any counting of disputed ballots , dimples can not be counted as votes . If these issues are heard before Sanders , there will have to be briefs , witnesses , and evidence on each point , which could take days . If the [29 Supreme Court 29] agrees to short-circuit the process , there will still have to be lengthy briefs filed , and the court may need to wait several days before finding time for yet another momentous hearing in its schedule .

1: 57700 Tallahassee,Florida

2: 1634101 Florida\_District\_Courts\_of\_Appeal

3: 85533 Electoral\_College\_(United\_States)

1: 18933066 Florida

[winner: 18933066 Florida, pem\_pos=0, final\_scores=-0.1576716, final\_local\_scores=-0.17619869, pem=0.0, similarity=-1.1684074]

gt: 18933066 Florida, pem\_pos=0, final\_scores=-0.1576716, final\_local\_scores=-0.17619869, pem=0.0, similarity=-1.1684074 ]

[winner: 4721653 Zyphe\_language, pem\_pos=12, final\_scores=-1.4598804, final\_local\_scores=-1.4195514, pem=-4.4188404, similarity=-14.24219]

RECALL MISS: 18933066 Florida]

2: 1635683 Supreme\_Court\_of\_Florida

[winner: 1635683 Supreme\_Court\_of\_Florida, pem\_pos=0,  
final\_scores=0.1778159, final\_local\_scores=0.2045126,  
pem=0.0, similarity=10.192777]  
gt: 1635683 Supreme\_Court\_of\_Florida, pem\_pos=0,  
final\_scores=0.1778159, final\_local\_scores=0.2045126,  
pem=0.0, similarity=10.192777 ]  
1: 57700 Tallahassee, Florida  
[pem\_pos=0, final\_scores=0.35513592,  
final\_local\_scores=0.37426198, pem=-0.024764989,  
similarity=15.393113]  
2: 1634101 Florida\_District\_Courts\_of\_Appeal  
[pem\_pos=0, final\_scores=0.034099624,  
final\_local\_scores=0.099673465, pem=0.0, similarity=7.064168]  
3: 85533 Electoral\_College\_(United\_States)  
[pem\_pos=0, final\_scores=0.10736643,  
final\_local\_scores=0.10289797, pem=-0.16437481,  
similarity=8.054289]  
1: 30680 The\_New\_York\_Times  
[pem\_pos=0, final\_scores=0.37412488,  
final\_local\_scores=0.40594822, pem=0.0, similarity=16.204018]  
2: 30680 The\_New\_York\_Times  
[pem\_pos=0, final\_scores=0.16090296,  
final\_local\_scores=0.17036706, pem=0.0, similarity=9.173806]  
3: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.42688626,  
final\_local\_scores=0.35378593, pem=0.0, similarity=14.64739]  
4: 1635683 Supreme\_Court\_of\_Florida  
[pem\_pos=0, final\_scores=0.1778159,  
final\_local\_scores=0.2045126, pem=0.0, similarity=10.192777]  
5: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.78020257,  
final\_local\_scores=0.74046856, pem=0.0, similarity=26.186771]  
6: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.10081945,  
final\_local\_scores=0.09499748, pem=-0.9940822,  
similarity=12.330601]  
7: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.64579576,  
final\_local\_scores=0.59364957, pem=0.0, similarity=21.8054]  
8: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.6483187,  
final\_local\_scores=0.59640545, pem=0.0, similarity=21.887642]  
9: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.27702028,  
final\_local\_scores=0.22291726, pem=0.0, similarity=10.742008]  
10: 27905225 Miami-Dade\_County, Florida  
[pem\_pos=0, final\_scores=0.18348153,  
final\_local\_scores=0.20782185, pem=0.0, similarity=10.291532]  
11: 109617 Palm\_Beach, Florida  
[pem\_pos=0, final\_scores=0.08675943,  
final\_local\_scores=0.113586366, pem=-0.6411934,  
similarity=10.966267]

12: 57700 Tallahassee, Florida  
[pem\_pos=0, final\_scores=0.30458802,  
final\_local\_scores=0.31904605, pem=-0.024764989,  
similarity=13.745359]  
13: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.70201176,  
final\_local\_scores=0.655057, pem=0.0, similarity=23.63792]  
14: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.37350637,  
final\_local\_scores=0.3305024, pem=0.0, similarity=13.952563]  
15: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.78265065,  
final\_local\_scores=0.7431428, pem=0.0, similarity=26.266577]  
16: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.7581852,  
final\_local\_scores=0.71641797, pem=0.0, similarity=25.469053]  
17: 38783 David\_Boies  
[pem\_pos=0, final\_scores=0.16043653,  
final\_local\_scores=0.20395088, pem=0.0, similarity=10.176014]  
18: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.7877905,  
final\_local\_scores=0.7487572, pem=0.0, similarity=26.43412]  
19: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.71364206,  
final\_local\_scores=0.6677613, pem=0.0, similarity=24.017042]  
20: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.69547325,  
final\_local\_scores=0.6479147, pem=0.0, similarity=23.42478]  
21: 1635683 Supreme\_Court\_of\_Florida  
[pem\_pos=0, final\_scores=0.20282348,  
final\_local\_scores=0.23182958, pem=0.0, similarity=11.00797]  
22: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.74408394,  
final\_local\_scores=0.70101446, pem=0.0, similarity=25.009382]  
23: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.79666847,  
final\_local\_scores=0.75845504, pem=0.0, similarity=26.723524]  
24: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.9160457,  
final\_local\_scores=0.8888566, pem=0.0, similarity=30.614967]  
25: 57700 Tallahassee, Florida  
[pem\_pos=0, final\_scores=0.26108795,  
final\_local\_scores=0.27152887, pem=-0.024764989,  
similarity=12.327352]  
26: 1635683 Supreme\_Court\_of\_Florida  
[pem\_pos=0, final\_scores=0.17798452,  
final\_local\_scores=0.2046968, pem=0.0, similarity=10.198274]  
27: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.7563891,  
final\_local\_scores=0.714456, pem=0.0, similarity=25.410505]  
28: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.34266138,

final\_local\_scores=0.29680896, pem=0.0, similarity=12.947084]  
29: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.05672522,  
final\_local\_scores=0.046831235, pem=-0.9940822,  
similarity=10.893224]

Sample: b'NYT20001217\_2241\_0165'

&LR; -LRB- ART ADV : Photo \_ NYT8 \_ is being sent to [1 NYT 1]  
] photo clients . Graphic is being sent to [2 NYT 2] graphics  
clients . Nonsubscribers can make one-time purchases by  
calling 888-603-1036 or 888-346-9867 . -RRB- &QL; &UR; By  
LAURENCE ZUCKERMAN &QC; &LR; &QL; &UR; c. 2000 N.Y. Times  
News Service &QC; &LR; &QL; Booking Your Ticket : \_ Fly  
nonstop . The fewer connections you make , the lower the  
chance that you will suffer a delay . \_ Pick airports with  
care . If you have to connect , choose the least-congested  
airport possible ; some airports , like [3 1 Chicago 3] O'Hare  
and [4 2 La Guardia 4] in [1 1 New York 1] , experience  
chronic delays . Take into consideration the weather and the  
season ; an airport in the 3 South might have fewer winter  
snowstorms but more spring and summer thunderstorms . \_ Check  
times between connecting flights . Airlines often book tight  
connections and do not take responsibility if you miss  
yours . Assume each leg of your flight will arrive late , and  
book connections that allow ample time . \_ Check on-time  
records . The 4 Transportation Department requires major  
domestic airlines to make the on-time record for every flight  
available to consumers . You can check online at [www.dot.gov/airconsumer](http://www.dot.gov/airconsumer) . That site also reports each airline 's on-time  
record for the previous month . \_ Travel on off-peak days and  
at off-peak times . Saturday is usually the least-traveled  
day of the week ; so are major holidays . Monday mornings and  
Friday afternoons are peak travel times for business  
travelers . Flights leaving earlier in the day are less  
likely to be delayed , because of the cascading effect that  
disruptions have on airline schedules . \_ Use smaller  
airports . Instead of [2 2 Los Angeles International 2] , for  
example , consider flying into [5 5 Burbank 5] or [3 3 John Wa  
yne Airport 3] in [4 4 Orange County 4] , [5 5 Calif 5] . ,  
or use [6 6 Westchester County Airport 6] instead of [6 6 JFK  
6] in [7 New York 7] . \_ Avoid electronic tickets . Though  
they are convenient and more secure than an easily lost paper  
ticket , they are not transferable between most airlines if  
you need to switch carriers because of delays . \_ Watch for  
labor disruptions . Check newspapers and Web sites like  
[www.thenewfoundlandgroup.com](http://www.thenewfoundlandgroup.com) to identify airlines  
experiencing labor strife , which often increases  
cancellations and delays . \_ Leave a number . Make sure the  
airline you book with has a phone number at which you can be  
notified of delays or cancellations . \_ Use a travel agent .  
A good one can be a great ally in the event of a problem \_  
and may save you time and money . Day of Departure : \_ Be

informed . Check your flight 's status before heading to the airport . Automated telephone services or airline [8 Web 8] sites often provide more up-to-date information than do ticket agents . The [7 7 Federal Aviation Administration 7] 's Web site , [www.fly.faa.gov](http://www.fly.faa.gov) , has real-time information on operations at the biggest airports . \_ Pack defensively . Take along a change of clothes and a toothbrush in your carry-on bag . If traveling with an infant , take 24 hours ' worth of diapers and extra formula . When possible , avoid checking luggage , in case you need to switch flights or carriers . Ship your gifts ahead . \_ Educate yourself . Find out what your airline promises to do for passengers in the event of a delay . Print out a copy of the policy from the airline 's Web site , and carry it with you \_ in case the airline representatives need a reminder . \_ If a Flight Is Delayed : \_ Keep your cool . Becoming angry at airline personnel is counterproductive . Sympathy for their plight will sometimes produce results . \_ Do n't stand in line . To get on another flight , it is often more effective to call the airline from the airport than to wait at the ticket counter . Carry a cell phone ; if delays are widespread , public telephones may be in demand . \_ Check the competition . If another airline has seats to your destination , ask your airline to endorse your ticket to the other line . \_ Get what you paid for . If you paid full fare or are a high-mileage frequent flier , let the airline know . Your odds of getting a seat on the next flight out will increase , or the airline might pay for your hotel , if you are delayed overnight . \_ Consider driving . If the last leg of your flight is a 200-mile hop , renting a car might be a better choice than waiting out a delay . Getting Even : \_ Write a letter . The airlines have pledged to respond to consumer complaints within 60 days . They almost always offer a refund , reimbursement of expenses , a discount on your next ticket \_ or all three . Be sure to save the relevant paperwork . \_ Make it a federal case . Send a copy of your letter to [8 8 Washington 8] . Complaints are charged against each airline in the Transportation Department 's monthly Air Travel Consumer Report and serve as a basis for rule making and enforcement action . Write to : 7 Aviation Consumer Protection Div . -LRB- C-75 -RRB- Department of Transportation 400 Seventh St. S.W. Washington , DC 20590

1: 645042 New\_York\_City

2: 18131 Los\_Angelès\_International\_Airport

3: 250456 John\_Wayne\_Airport

4: 45330 Orange\_County,\_California

5: 5407 California

6: 249690 Westchester\_County\_Airport

7: 11186 Federal\_Aviation\_Administration

8: 108956 Washington,\_D.C.

1: 248449 O'Hare\_International\_Airport

[winner: 6886 Chicago, pem\_pos=0,  
final\_scores=0.13269188, final\_local\_scores=0.06804888,

pem=-0.20658566, similarity=7.243872|  
 RECALL MISS: 248449 O'Hare\_International\_Airport]  
 [winner: 14691719 Mick\_O'Hare, pem\_pos=16,  
 final\_scores=-1.3870419, final\_local\_scores=-1.3111973,  
 pem=-3.4965076, similarity=-16.024475|  
 gt: 248449 O'Hare\_International\_Airport, pem\_pos=0,  
 final\_scores=-1.6381879, final\_local\_scores=-1.8245692,  
 pem=-0.43603384, similarity=-47.987854 ]  
 2: 248547 LaGuardia\_Airport  
 [winner: 11418 Fiorello\_H.\_La\_Guardia, pem\_pos=1,  
 final\_scores=0.09429078, final\_local\_scores=0.098837584,  
 pem=-0.67818034, similarity=10.727274|  
 gt: 248547 LaGuardia\_Airport, pem\_pos=2,  
 final\_scores=0.027354285, final\_local\_scores=0.0077116042,  
 pem=-1.3025954, similarity=11.403559 ]  
 3: 179553 Southern\_United\_States  
 [winner: 179553 Southern\_United\_States, pem\_pos=0,  
 final\_scores=-0.16341725, final\_local\_scores=-0.19670103,  
 pem=-1.3313899, similarity=5.4600673|  
 gt: 179553 Southern\_United\_States, pem\_pos=0,  
 final\_scores=-0.16341725, final\_local\_scores=-0.19670103,  
 pem=-1.3313899, similarity=5.4600673 ]  
 4: 58235 United\_States\_Department\_of\_Transportation  
 [winner: 58235  
 United\_States\_Department\_of\_Transportation, pem\_pos=0,  
 final\_scores=-0.10392597, final\_local\_scores=-0.13516852,  
 pem=0.0, similarity=0.056014422|  
 gt: 58235 United\_States\_Department\_of\_Transportation,  
 pem\_pos=0, final\_scores=-0.10392597,  
 final\_local\_scores=-0.13516852, pem=0.0,  
 similarity=0.056014422 ]  
 5: 250473 Bob\_Hope\_Airport  
 [winner: 107608 Burbank,California, pem\_pos=0,  
 final\_scores=0.43797666, final\_local\_scores=0.424133,  
 pem=-0.24073757, similarity=18.055855|  
 gt: 250473 Bob\_Hope\_Airport, pem\_pos=5,  
 final\_scores=-0.19202408, final\_local\_scores=-0.2174297,  
 pem=-2.7235656, similarity=12.41235 ]  
 6: 97218 John\_F.\_Kennedy\_International\_Airport  
 [winner: 5119376 John\_F.\_Kennedy, pem\_pos=1,  
 final\_scores=0.15325512, final\_local\_scores=0.13095057,  
 pem=-0.86228985, similarity=12.686806|  
 gt: 97218 John\_F.\_Kennedy\_International\_Airport,  
 pem\_pos=2, final\_scores=-0.00040875375,  
 final\_local\_scores=-0.035227165, pem=-1.2221509,  
 similarity=9.6847105 ]  
 7: 13025240 Airline\_complaints  
 [winner: 3617770 Div\_Sultan\_Rumlu, pem\_pos=2,  
 final\_scores=-0.7186469, final\_local\_scores=-0.56888354,  
 pem=-1.3862944, similarity=-5.3480215|  
 RECALL MISS: 13025240 Airline\_complaints]  
 [winner: 1370999 Australian\_referendum,\_1937\_(Aviation),

pem\_pos=5, final\_scores=-1.0154802,  
final\_local\_scores=-0.92688113, pem=-2.2853432,  
similarity=-11.142218|  
RECALL MISS: 13025240 Airline\_complaints]  
[winner: 23867559 Protection\_New\_York, pem\_pos=1,  
final\_scores=-1.5013769, final\_local\_scores=-1.4601991,  
pem=-1.1594036, similarity=-33.180527|  
RECALL MISS: 13025240 Airline\_complaints]  
[winner: 286317 Consumer\_Reports, pem\_pos=1,  
final\_scores=-2.1476543, final\_local\_scores=-2.2729557,  
pem=-0.6776367, similarity=-60.054726|  
RECALL MISS: 13025240 Airline\_complaints]  
1: 645042 New\_York\_City  
[pem\_pos=1, final\_scores=0.46773726,  
final\_local\_scores=0.42471087, pem=-1.7906687,  
similarity=26.501867]  
2: 18131 Los\_Angeles\_International\_Airport  
[pem\_pos=0, final\_scores=0.2963627,  
final\_local\_scores=0.27926362, pem=0.0, similarity=12.423496]  
3: 250456 John\_Wayne\_Airport  
[pem\_pos=0, final\_scores=0.30320042,  
final\_local\_scores=0.32557252, pem=0.0, similarity=13.805447]  
4: 45330 Orange\_County\_California  
[pem\_pos=0, final\_scores=0.3119833,  
final\_local\_scores=0.29586673, pem=-0.449498,  
similarity=15.363406]  
5: 5407 California  
[pem\_pos=0, final\_scores=0.70970327,  
final\_local\_scores=0.7215297, pem=0.0, similarity=25.6216]  
6: 249690 Westchester\_County\_Airport  
[pem\_pos=0, final\_scores=0.19461213,  
final\_local\_scores=0.23821348, pem=0.0, similarity=11.198479]  
7: 11186 Federal\_Aviation\_Administration  
[pem\_pos=0, final\_scores=0.3319831,  
final\_local\_scores=0.37048936, pem=0.0, similarity=15.145855]  
8: 108956 Washington\_D.C.  
[pem\_pos=0, final\_scores=0.3086601,  
final\_local\_scores=0.2623701, pem=-1.7496481,  
similarity=21.434217]  
1: 30680 The\_New\_York\_Times  
[pem\_pos=0, final\_scores=0.0633996,  
final\_local\_scores=0.0456516, pem=0.0, similarity=5.452048]  
2: 30680 The\_New\_York\_Times  
[pem\_pos=0, final\_scores=0.13390295,  
final\_local\_scores=0.12266582, pem=0.0, similarity=7.7503057]  
3: 6886 Chicago  
[pem\_pos=0, final\_scores=0.13269188,  
final\_local\_scores=0.06804888, pem=-0.20658566,  
similarity=7.243872]  
4: 11418 Fiorello\_H\_La\_Guardia  
[pem\_pos=1, final\_scores=0.09429078,  
final\_local\_scores=0.098837584, pem=-0.67818034,



similarity=10.727274]  
 5: 107608 Burbank, California  
 [pem\_pos=0, final\_scores=0.43797666,  
 final\_local\_scores=0.424133, pem=-0.24073757,  
 similarity=18.055855]  
 6: 5119376 John\_F.\_Kennedy  
 [pem\_pos=1, final\_scores=0.15325512,  
 final\_local\_scores=0.13095057, pem=-0.86228985,  
 similarity=12.686806]  
 7: 645042 New\_York\_City  
 [pem\_pos=1, final\_scores=0.36642295,  
 final\_local\_scores=0.31404042, pem=-1.7906687,  
 similarity=23.19924]  
 8: 33139 World\_Wide\_Web  
 [pem\_pos=0, final\_scores=0.07234819,  
 final\_local\_scores=0.12491226, pem=-0.34117427,  
 similarity=9.672703]

Sample: b'APW20001001\_2021\_0521'

[1 1 CAIRO 1] , [2 2 Egypt 2] -LRB- 1 AP -RRB- \_ In his  
 foreign policy debut as [3 3 Syria 3] 's president , [1  
 Bashar Assad 1] met Sunday with [2 Egyptian 2] 2 President [3  
 Hosni Mubarak 3] in talks on [4 Mideast 4] peace and the  
 escalating violence in the [5 Palestinian 5] territories . [6  
 Assad 6] , on his first trip abroad since becoming president  
 in July , also dealt with the issue of [4 4 Iraq 4] in the  
 talks with [7 Mubarak 7] , [8 Egyptian 8] Foreign Minister [9  
 Amr Mussa 9] said . The [10 Syrian 10] leader 's visit came  
 on a fourth day of clashes that have raged in the [11 3 West B  
 ank 11] , [5 5 Gaza Strip 5] and [6 6 Jerusalem 6] , leaving  
 more than two dozen [12 Palestinians 12] dead . `` There is  
 complete agreement -LRB- between [13 Mubarak 13] and [14  
 Assad 14] -RRB- as to the seriousness of the situation and  
 that the negative repercussions would adversely effect the  
 peace process , which is in itself in a very precarious  
 state , ' ' [15 Moussa 15] said . [16 Assad 16] , 35 , has  
 focused on domestic issues since his father , [17 Hafez Assad  
 17] , died in June after 30 years in power . He has taken  
 steps to modernize his nation 's bureaucracy and revive its  
 stagnant economy . `` This was the first meeting of the first  
 visit of President [18 Bashar Assad 18] after the funeral of  
 his father , so it was only normal for the discussions to  
 have a general tone ... in addition to certain specifics of  
 special importance to both countries and the [19 Arab 19]  
 world , ' ' [20 Moussa 20] said . The late [21 Syrian 21]  
 president rarely traveled abroad . In his last few years ,  
 however , he had sent [22 Bashar 22] , educated as an eye  
 doctor , on foreign trips as part of his grooming to become  
 president . Prospects appear dim for progress on either the [2  
 3 Syrian 23] or [24 Palestinian 24] peace track . The  
 deadlock in [25 Palestinian-Israeli 25] talks is exacerbated  
 by the worst violence in four years . [26 Syria 26] 's

government-run media said last week that no progress could be expected now because of the [27 U.S. 27] elections and [28 Israeli 28] Prime Minister [29 Ehud Barak 29] 's precarious situation as leader of a minority government . Syrian-Israeli peace negotiations broke down in January , with [30 Syria 30] demanding a full withdrawal from the [31 Golan Heights 31] , a strategic plateau captured by [32 Israel 32] in 1967 . [33 Israel 33] insists on waiting to settle borders until it knows what it will gain by returning [34 Golan 34] land . [35 Assad 35] has made it clear he intends to follow his father 's course in any negotiations with [36 Israel 36] . [37 Mubarak 37] and [38 Assad 38] met after a week when [39 Russia 39] , [40 France 40] , [41 Jordan 41] and [42 Yemen 42] sent planes to [43 Baghdad 43] , challenging [44 U.N. 44] sanctions imposed on [45 Iraq 45] after its 1990 invasion of [46 Kuwait 46] . [47 Iraq 47] is hoping other [48 Arab 48] countries will follow , despite [49 U.S. 49] complaints , easing its isolation . [50 Moussa 50] said the presidents discussed the situation in [51 Iraq 51] and would return to it Monday . However , he would not comment on whether [52 Syria 52] and [53 Egypt 53] were planning to send planes to [54 Baghdad 54] . [55 Egypt 55] and [56 Syria 56] are on good terms , but they have had their differences . [57 Hafez Assad 57] 's attempts to rally [58 Arab 58] leaders to present a unified front against [59 Israel 59] gained little ground . He derided [60 Egypt 60] for signing a peace treaty with [61 Israel 61] in 1979 .

1: 6293 Cairo

2: 8087628 Egypt

3: 7515849 Syria

4: 7515928 Iraq

5: 12047 Gaza\_Strip

6: 16043 Jerusalem

1: 18935732 Associated\_Press

[winner: 18935732 Associated\_Press, pem\_pos=1, final\_scores=-0.2176148, final\_local\_scores=-0.21635607, pem=-0.68645424, similarity=1.3662612]

gt: 18935732 Associated\_Press, pem\_pos=1, final\_scores=-0.2176148, final\_local\_scores=-0.21635607, pem=-0.68645424, similarity=1.3662612 ]

2: 335341 President\_of\_Egypt

[winner: 364813 Bashar\_al-Assad, pem\_pos=0, final\_scores=-1.341855, final\_local\_scores=-1.5554006, pem=0.0, similarity=-42.32654]

RECALL MISS: 335341 President\_of\_Egypt]

3: 33209 West\_Bank

[winner: 96665 Palestinian\_territories, pem\_pos=1, final\_scores=0.05233313, final\_local\_scores=-0.010123998, pem=-1.929408, similarity=14.280013]

gt: 33209 West\_Bank, pem\_pos=0, final\_scores=-0.2670529, final\_local\_scores=-0.35428184, pem=-0.053944394, similarity=-6.1894054 ]

1: 6293 Cairo  
[pem\_pos=0, final\_scores=0.34315854,  
final\_local\_scores=0.31013262, pem=-0.15211283,  
similarity=14.171902]  
2: 8087628 Egypt  
[pem\_pos=0, final\_scores=0.50845176,  
final\_local\_scores=0.47297436, pem=-0.27480945,  
similarity=19.698668]  
3: 7515849 Syria  
[pem\_pos=0, final\_scores=0.6011165,  
final\_local\_scores=0.57307357, pem=-0.26162508,  
similarity=22.61413]  
4: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.70375,  
final\_local\_scores=0.6840075, pem=-0.25379938,  
similarity=25.882063]  
5: 12047 Gaza Strip  
[pem\_pos=0, final\_scores=0.38914675,  
final\_local\_scores=0.36852366, pem=0.0, similarity=15.087194]  
6: 16043 Jerusalem  
[pem\_pos=0, final\_scores=0.4102581,  
final\_local\_scores=0.40289855, pem=-0.13239029,  
similarity=16.832968]  
1: 364813 Bashar\_al-Assad  
[pem\_pos=0, final\_scores=0.4768017,  
final\_local\_scores=0.43223417, pem=0.0, similarity=16.988443]  
2: 8087628 Egypt  
[pem\_pos=0, final\_scores=0.4078859,  
final\_local\_scores=0.3631214, pem=-1.052751,  
similarity=20.651003]  
3: 51884 Hosni\_Mubarak  
[pem\_pos=0, final\_scores=0.6034507,  
final\_local\_scores=0.587891, pem=0.0, similarity=21.633553]  
4: 19323 Middle\_East  
[pem\_pos=0, final\_scores=0.5639774,  
final\_local\_scores=0.5636235, pem=0.0, similarity=20.909363]  
5: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.39861083,  
final\_local\_scores=0.3906641, pem=-2.8508017,  
similarity=31.251013]  
6: 364813 Bashar\_al-Assad  
[pem\_pos=1, final\_scores=0.40901875,  
final\_local\_scores=0.36010206, pem=0.0, similarity=14.835876]  
7: 51884 Hosni\_Mubarak  
[pem\_pos=0, final\_scores=0.46051824,  
final\_local\_scores=0.4332583, pem=0.0, similarity=17.019005]  
8: 8087628 Egypt  
[pem\_pos=1, final\_scores=0.47059256,  
final\_local\_scores=0.43161887, pem=-1.052751,  
similarity=22.695105]  
9: 201993 Amr\_Moussa  
[pem\_pos=0, final\_scores=0.5883916,

final\_local\_scores=0.6112062, pem=0.0, similarity=22.329325]  
10: 7515849 Syria  
[pem\_pos=0, final\_scores=0.6107028,  
final\_local\_scores=0.5835452, pem=-0.60890055,  
similarity=24.815163]  
11: 96665 Palestinian\_territories  
[pem\_pos=1, final\_scores=0.05233313,  
final\_local\_scores=-0.010123998, pem=-1.929408,  
similarity=14.280013]  
12: 241405 State\_of\_Palestine  
[pem\_pos=5, final\_scores=0.4571032,  
final\_local\_scores=0.45455813, pem=-1.7909594,  
similarity=27.39415]  
13: 51884 Hosni\_Mubarak  
[pem\_pos=0, final\_scores=0.49459857,  
final\_local\_scores=0.47048587, pem=0.0, similarity=18.12995]  
14: 364813 Bashar\_al-Assad  
[pem\_pos=1, final\_scores=0.40282613,  
final\_local\_scores=0.35333762, pem=0.0, similarity=14.634012]  
15: 201993 Amr\_Moussa  
[pem\_pos=0, final\_scores=0.32275456,  
final\_local\_scores=0.32103807, pem=-0.9024974,  
similarity=18.57805]  
16: 70843 Hafez\_al-Assad  
[pem\_pos=0, final\_scores=0.40616804,  
final\_local\_scores=0.36922145, pem=0.0, similarity=15.108018]  
17: 70843 Hafez\_al-Assad  
[pem\_pos=0, final\_scores=0.5209567,  
final\_local\_scores=0.49239272, pem=0.0, similarity=18.783693]  
18: 364813 Bashar\_al-Assad  
[pem\_pos=0, final\_scores=0.44711024,  
final\_local\_scores=0.39980078, pem=0.0, similarity=16.020565]  
19: 159433 Arab\_world  
[pem\_pos=14, final\_scores=0.48430026,  
final\_local\_scores=0.46503425, pem=-3.8052368,  
similarity=38.660732]  
20: 201993 Amr\_Moussa  
[pem\_pos=0, final\_scores=0.35626984,  
final\_local\_scores=0.35764837, pem=-0.9024974,  
similarity=19.670574]  
21: 7515849 Syria  
[pem\_pos=0, final\_scores=0.6422671,  
final\_local\_scores=0.6180244, pem=-0.60890055,  
similarity=25.84409]  
22: 364813 Bashar\_al-Assad  
[pem\_pos=0, final\_scores=0.5578929,  
final\_local\_scores=0.52188236, pem=0.0, similarity=19.663723]  
23: 7515849 Syria  
[pem\_pos=0, final\_scores=0.54323304,  
final\_local\_scores=0.5098447, pem=-0.60890055,  
similarity=22.615793]  
24: 241405 State\_of\_Palestine

[pem\_pos=9, final\_scores=0.34354955,  
final\_local\_scores=0.330518, pem=-2.8508017,  
similarity=29.456133]  
25: 492331 Arab\_citizens\_of\_Israel  
[pem\_pos=0, final\_scores=0.03350167,  
final\_local\_scores=-0.0059438646, pem=0.0,  
similarity=3.9123363]  
26: 7515849 Syria  
[pem\_pos=0, final\_scores=0.6348013,  
final\_local\_scores=0.6098692, pem=-0.26162508,  
similarity=23.712185]  
27: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.44236964,  
final\_local\_scores=0.47470564, pem=-0.14075686,  
similarity=19.021334]  
28: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5882711,  
final\_local\_scores=0.5703955, pem=-0.34790626,  
similarity=23.00342]  
29: 168212 Ehud\_Barak  
[pem\_pos=0, final\_scores=0.7062873,  
final\_local\_scores=0.74855715, pem=0.0, similarity=26.428152]  
30: 7515849 Syria  
[pem\_pos=0, final\_scores=0.65849954,  
final\_local\_scores=0.63575584, pem=-0.26162508,  
similarity=24.484694]  
31: 7515865 Golan\_Heights  
[pem\_pos=0, final\_scores=0.43748218,  
final\_local\_scores=0.4319188, pem=0.0, similarity=16.97903]  
32: 9282173 Israel  
[pem\_pos=0, final\_scores=0.56344146,  
final\_local\_scores=0.5432729, pem=-0.21279564,  
similarity=21.459274]  
33: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5842077,  
final\_local\_scores=0.5659569, pem=-0.21279564,  
similarity=22.13621]  
34: 7515865 Golan\_Heights  
[pem\_pos=1, final\_scores=0.37143117,  
final\_local\_scores=0.35976815, pem=-1.3369088,  
similarity=22.09623]  
35: 364813 Bashar\_al-Assad  
[pem\_pos=1, final\_scores=0.45689106,  
final\_local\_scores=0.4123953, pem=0.0, similarity=16.396412]  
36: 9282173 Israel  
[pem\_pos=0, final\_scores=0.58959645,  
final\_local\_scores=0.5718433, pem=-0.21279564,  
similarity=22.311872]  
37: 51884 Hosni\_Mubarak  
[pem\_pos=0, final\_scores=0.50132984,  
final\_local\_scores=0.4778388, pem=0.0, similarity=18.349377]  
38: 364813 Bashar\_al-Assad

[pem\_pos=1, final\_scores=0.44336492,  
final\_local\_scores=0.39762002, pem=0.0, similarity=15.955488]  
39: 25391 Russia  
[pem\_pos=0, final\_scores=0.6811345,  
final\_local\_scores=0.76117855, pem=-0.20299491,  
similarity=27.908716]  
40: 5843419 France  
[pem\_pos=0, final\_scores=0.6263167,  
final\_local\_scores=0.6791112, pem=-0.16240834,  
similarity=25.238947]  
41: 7515964 Jordan  
[pem\_pos=0, final\_scores=0.6977279,  
final\_local\_scores=0.67378783, pem=-0.53647494,  
similarity=27.11432]  
42: 350939 Yemen  
[pem\_pos=0, final\_scores=0.6530455,  
final\_local\_scores=0.65595627, pem=-0.27215755,  
similarity=25.14479]  
43: 4492 Baghdad  
[pem\_pos=0, final\_scores=0.35329443,  
final\_local\_scores=0.31713414, pem=0.0, similarity=13.553628]  
44: 31769 United\_Nations  
[pem\_pos=0, final\_scores=0.51685524,  
final\_local\_scores=0.53647494, pem=0.0, similarity=20.099195]  
45: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.5853357,  
final\_local\_scores=0.5546578, pem=-0.25379938,  
similarity=22.022009]  
46: 7515890 Kuwait  
[pem\_pos=0, final\_scores=0.44416398,  
final\_local\_scores=0.41698104, pem=-0.3372397,  
similarity=18.36722]  
47: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.5463988,  
final\_local\_scores=0.51212525, pem=-0.25379938,  
similarity=20.752752]  
48: 159433 Arab\_world  
[pem\_pos=14, final\_scores=0.47575122,  
final\_local\_scores=0.45569575, pem=-3.8052368,  
similarity=38.382053]  
49: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.45506215,  
final\_local\_scores=0.48857027, pem=-0.14075686,  
similarity=19.435081]  
50: 201993 Amr\_Moussa  
[pem\_pos=0, final\_scores=0.41694087,  
final\_local\_scores=0.42392224, pem=-0.9024974,  
similarity=21.64832]  
51: 7515928 Iraq  
[pem\_pos=0, final\_scores=0.6331355,  
final\_local\_scores=0.60687184, pem=-0.25379938,  
similarity=23.58018]

52: 7515849 Syria  
     [pem\_pos=0, final\_scores=0.67044294,  
 final\_local\_scores=0.6488022, pem=-0.26162508,  
 similarity=24.874023]  
 53: 8087628 Egypt  
     [pem\_pos=0, final\_scores=0.6031384,  
 final\_local\_scores=0.5764051, pem=-0.27480945,  
 similarity=22.785248]  
 54: 4492 Baghdad  
     [pem\_pos=0, final\_scores=0.3260889,  
 final\_local\_scores=0.28741622, pem=0.0, similarity=12.666786]  
 55: 8087628 Egypt  
     [pem\_pos=0, final\_scores=0.5792006,  
 final\_local\_scores=0.5502567, pem=-0.27480945,  
 similarity=22.004929]  
 56: 7515849 Syria  
     [pem\_pos=0, final\_scores=0.71049255,  
 final\_local\_scores=0.6925503, pem=-0.26162508,  
 similarity=26.179554]  
 57: 70843 Hafez\_al-Assad  
     [pem\_pos=0, final\_scores=0.48749477,  
 final\_local\_scores=0.4558407, pem=0.0, similarity=17.69291]  
 58: 159433 Arab\_world  
     [pem\_pos=14, final\_scores=0.27385992,  
 final\_local\_scores=0.23516011, pem=-3.8052368,  
 similarity=31.80083]  
 59: 9282173 Israel  
     [pem\_pos=0, final\_scores=0.5787452,  
 final\_local\_scores=0.55999, pem=-0.21279564,  
 similarity=21.958145]  
 60: 8087628 Egypt  
     [pem\_pos=0, final\_scores=0.5971195,  
 final\_local\_scores=0.56983036, pem=-0.27480945,  
 similarity=22.589045]  
 61: 9282173 Israel  
     [pem\_pos=0, final\_scores=0.54381746,  
 final\_local\_scores=0.5218367, pem=-0.21279564,  
 similarity=20.819578]

Sample: b'20001115\_AFP\_ARB\_0212\_eng'

[1 Washington 1] 11-15 -LRB- [2 AFP 2] -RRB- - Stewart  
 Talbot , [1 1 America 1] 's Assistant Secretary of State and  
 `` second in charge `` at the [3 State Department 3] after [2  
 2 Madeleine Albright 2] , has decided to resign from the  
 administration to head one of the departments at [3 3 Yale Un  
 iversity 3] in the state of [4 Connecticut 4] -LRB- northeast  
 -RRB- . [4 4 Richard Levin 4] , the Chancellor of this  
 prestigious university , said Talbot would head the 1 Globali  
 zation Studies Center as of next July and would also teach at  
 the university . Talbot is the second top-ranking [5 American  
 5] diplomat in a few days to announce his intention of  
 resigning at the end of President [6 Bill Clinton 6] 's

second term next January 20 . The special envoy to the [5 5 Middle East 5] , [7 Dennis Ross 7] , announced last week that he intends to resign from his post , for personal and family reasons , once the current administration 's term comes to an end . Talbot , a [8 Yale 8] graduate and former correspondent with [9 2 Time 9] magazine , was appointed assistant to the Secretary of State in 1994 after acting as roving ambassador and special advisor on former [10 Soviet Union 10] Affairs .

1: 3434750 United\_States

2: 133249 Madeleine\_Albright

3: 34273 Yale\_University

4: 457646 Rick\_Levin

5: 19323 Middle\_East

1: 2802942 Yale\_Center\_for\_the\_Study\_of\_Globalization

[winner: 666100 Glocalization, pem\_pos=1,  
final\_scores=-1.0541284, final\_local\_scores=-1.043776,  
pem=-1.9338181, similarity=-16.54224]

RECALL MISS: 2802942

Yale\_Center\_for\_the\_Study\_of\_Globalization]

[winner: 13749514

Center\_Township,\_LaPorte\_County,\_Indiana, pem\_pos=11,  
final\_scores=-1.0990006, final\_local\_scores=-1.0223355,  
pem=-4.030373, similarity=-4.501021]

RECALL MISS: 2802942

Yale\_Center\_for\_the\_Study\_of\_Globalization]

[winner: 9318072 Science\_fiction\_studies, pem\_pos=5,  
final\_scores=-1.6951096, final\_local\_scores=-1.7660244,  
pem=-1.7503334, similarity=-39.09339]

RECALL MISS: 2802942

Yale\_Center\_for\_the\_Study\_of\_Globalization]

2: 31600 Time\_(magazine)

[winner: 31600 Time\_(magazine), pem\_pos=0,  
final\_scores=0.25067163, final\_local\_scores=0.24848482,  
pem=-0.31107646, similarity=13.196678]

gt: 31600 Time\_(magazine), pem\_pos=0,  
final\_scores=0.25067163, final\_local\_scores=0.24848482,  
pem=-0.31107646, similarity=13.196678 ]

[winner: 896497 Magazine\_(firearms), pem\_pos=0,  
final\_scores=-1.0191278, final\_local\_scores=-0.9967933,  
pem=-0.62660754, similarity=-22.248999]

gt: 31600 Time\_(magazine), pem\_pos=16,  
final\_scores=-1.9581869, final\_local\_scores=-2.1788378,  
pem=-5.21196, similarity=-32.5877 ]

1: 3434750 United\_States

[pem\_pos=0, final\_scores=0.6067086,  
final\_local\_scores=0.5938429, pem=-0.78483206,  
similarity=26.079208]

2: 133249 Madeleine\_Albright

[pem\_pos=0, final\_scores=0.63246024,  
final\_local\_scores=0.66284955, pem=0.0, similarity=23.870466]

3: 34273 Yale\_University

[pem\_pos=0, final\_scores=0.36325216,



final\_local\_scores=0.3507073, pem=0.0, similarity=14.555518]  
 4: 457646 Rick\_Levin  
 [pem\_pos=0, final\_scores=0.25442278,  
 final\_local\_scores=0.30285916, pem=0.0, similarity=13.127634]  
 5: 19323 Middle\_East  
 [pem\_pos=0, final\_scores=0.3172925,  
 final\_local\_scores=0.3180875, pem=0.0, similarity=13.582078]  
 1: 108956 Washington,\_D.C.  
 [pem\_pos=0, final\_scores=0.3361314,  
 final\_local\_scores=0.28633237, pem=-1.7496481,  
 similarity=22.1493]  
 2: 471981 Agence\_France-Presse  
 [pem\_pos=0, final\_scores=0.08987428,  
 final\_local\_scores=0.12137306, pem=-0.6605333,  
 similarity=11.30381]  
 3: 31975 United\_States\_Department\_of\_State  
 [pem\_pos=0, final\_scores=0.3347215,  
 final\_local\_scores=0.29545847, pem=0.0, similarity=12.906783]  
 4: 6466 Connecticut  
 [pem\_pos=0, final\_scores=0.21026723,  
 final\_local\_scores=0.19446075, pem=-0.24788867,  
 similarity=11.240867]  
 5: 3434750 United\_States  
 [pem\_pos=0, final\_scores=0.623065,  
 final\_local\_scores=0.6117098, pem=-0.12994486,  
 similarity=23.051014]  
 6: 3356 Bill\_Clinton  
 [pem\_pos=0, final\_scores=0.51303035,  
 final\_local\_scores=0.5062914, pem=0.0, similarity=19.198458]  
 7: 1399016 Dennis\_Ross  
 [pem\_pos=0, final\_scores=0.5080742,  
 final\_local\_scores=0.55403715, pem=0.0, similarity=20.623287]  
 8: 34273 Yale\_University  
 [pem\_pos=0, final\_scores=0.29335004,  
 final\_local\_scores=0.2743498, pem=-0.3216871,  
 similarity=14.026242]  
 9: 31600 Time\_(magazine)  
 [pem\_pos=0, final\_scores=0.25067163,  
 final\_local\_scores=0.24848482, pem=-0.31107646,  
 similarity=13.196678]  
 10: 26779 Soviet\_Union  
 [pem\_pos=0, final\_scores=0.24986778,  
 final\_local\_scores=0.2605568, pem=-0.075323656,  
 similarity=12.274869]

Sample: b'20001115\_AFP\_ARB\_0065\_eng'

1 Camp Zeist -LRB- 2 Holland -RRB- 11-15 -LRB- 3 AFP -RRB- -  
 Today Wednesday , the judges in charge of looking into the [1  
 Lockerbie 1] case in Camp Zeist -LRB- Holland -RRB- were  
 faced with an awkward problem : how to deal with a witness  
 who is refusing to answer questions and who is at the same  
 time serving a life sentence and who is not subject to [2

Scottish [2] law ? The problem arose after [3 Palestinian 3] Mahmood Abu Talib , whose testimony the court has been hearing since Friday , refused to continue answering a question by defense lawyer Richard Keen about the detailed reasons for his having lived in the former [1 1 Soviet Union 1] for a period of 18 months in the 70s . The lawyer asked the judges to force [4 Abu Talib 4] to answer the question aimed at demonstrating the witness 's `` professional terrorism ' precedents . 46-year-old [5 Abu Talib 5] was sentenced to life imprisonment in 1990 in [2 2 Sweden 2] for terrorist acts in [3 3 Amsterdam 3] , [4 4 Copenhagen 4] and [5 5 Stockholm 5] between 1985 and 1986 . The defense council in the [6 Lockerbie 6] court is attempting to show his involvement in the [7 Lockerbie 7] attack . Based on [8 Scottish 8] law , followed in this court , if [9 Abu Talib 9] insists on not responding to the question , this is considered to be `` contempt of court '' and means he may be punished in accordance with [10 Scottish 10] law covering such incidents . The problem is that the judges may not punish [11 Abu Talib 11] unless [12 Sweden 12] hands him over to [13 Scotland 13] after the court case ends to apply the punishment for `` contempt of court . '' So far , the judges have not asked [14 Abu Talib 14] to answer the defense 's question in what appears to be a move to avoid a tricky situation if the witness refuses to answer . The court president , Ronald Sutherland , who adjourned the session for a short while in an attempt to find a solution to this dilemma , said `` this raises an important question of principle . '' The two [15 Libyans 15] , 4 Alameen Khalifa Fakhima and 5 Abdulbasit al-Magrahi , are suspected of sending a bomb from [6 6 Malta 6] which was used to explode the [16 Pan Am 16] plane over the [17 Scottish 17] village of [18 Lockerbie 18] on December 21 , 1988 killing 270 people .

1: 26779 Soviet\_Union

2: 5058739 Sweden

3: 844 Amsterdam

4: 5166 Copenhagen

5: 26741 Stockholm

6: 19137 Malta

1: 118688 Zeist

[winner: 377762 Scottish\_Court\_in\_the\_Netherlands,  
pem\_pos=0, final\_scores=-0.2328065,  
final\_local\_scores=-0.19006355, pem=0.0,  
similarity=-1.5821624|

gt: 118688 Zeist, pem\_pos=1, final\_scores=-0.66416895,  
final\_local\_scores=-0.6886048, pem=-2.4079456,  
similarity=-3.3648438 ]

2: 13288 Holland

[winner: 13288 Holland, pem\_pos=0,  
final\_scores=-0.90594065, final\_local\_scores=-0.9945092,  
pem=-0.6316271, similarity=-22.15354|  
gt: 13288 Holland, pem\_pos=0, final\_scores=-0.90594065,

```

final_local_scores=-0.9945092, pem=-0.6316271,
similarity=-22.15354 ]
3: 471981 Agence_France-Presse
  [winner: 471981 Agence_France-Presse, pem_pos=0,
final_scores=-0.07617855, final_local_scores=-0.04580573,
pem=-0.6605333, similarity=6.3148613]
  gt: 471981 Agence_France-Presse, pem_pos=0,
final_scores=-0.07617855, final_local_scores=-0.04580573,
pem=-0.6605333, similarity=6.3148613 ]
4: 2060944 Lamin_Khalifah_Fhimah
  [winner: 2441231 Tali_Fahima, pem_pos=1,
final_scores=-0.87218076, final_local_scores=-0.86746246,
pem=-0.6931472, similarity=-18.027655]
  RECALL MISS: 2060944 Lamin_Khalifah_Fhimah]
  [winner: 12526424 Saaed_Khalifa, pem_pos=18,
final_scores=-1.4213768, final_local_scores=-1.3716366,
pem=-3.6109178, similarity=-17.205927]
  RECALL MISS: 2060944 Lamin_Khalifah_Fhimah]
5: 2028438 Abdelbaset_al-Megrahi
  [winner: 5140095 Abdul_Basit 'Abd_us-Samad, pem_pos=0,
final_scores=-0.88997513, final_local_scores=-0.8470445,
pem=0.0, similarity=-21.187786]
  RECALL MISS: 2028438 Abdelbaset_al-Megrahi]
1: 26779 Soviet_Union
  [pem_pos=0, final_scores=0.2890315,
final_local_scores=0.31639957, pem=-0.075323656,
similarity=13.941329]
2: 5058739 Sweden
  [pem_pos=0, final_scores=0.6786671,
final_local_scores=0.70009786, pem=-0.2135508,
similarity=26.14335]
3: 844 Amsterdam
  [pem_pos=0, final_scores=0.695259,
final_local_scores=0.7461705, pem=-0.0642803,
similarity=26.706497]
4: 5166 Copenhagen
  [pem_pos=0, final_scores=0.553683,
final_local_scores=0.5728658, pem=-0.11866472,
similarity=21.830486]
5: 26741 Stockholm
  [pem_pos=0, final_scores=0.5994557,
final_local_scores=0.6197214, pem=-0.1047564,
similarity=23.153118]
6: 19137 Malta
  [pem_pos=0, final_scores=0.43200696,
final_local_scores=0.44079196, pem=-0.39222723,
similarity=19.376818]
1: 14410202 Lockerbie
  [pem_pos=0, final_scores=0.030835912,
final_local_scores=0.011685282, pem=0.0, similarity=4.438425]
2: 26994 Scotland
  [pem_pos=0, final_scores=0.44735855,

```

final\_local\_scores=0.42709607, pem=-0.47951394,  
similarity=19.442783]  
3: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.0920582,  
final\_local\_scores=0.12561625, pem=-2.8508017,  
similarity=23.341455]  
4: 23569239 Abu\_Talib\_ibn\_Abdul-Muttalib  
[pem\_pos=0, final\_scores=0.104036555,  
final\_local\_scores=0.17268392, pem=0.0, similarity=9.242946]  
5: 23569239 Abu\_Talib\_ibn\_Abdul-Muttalib  
[pem\_pos=0, final\_scores=0.1534336,  
final\_local\_scores=0.2266427, pem=0.0, similarity=10.853183]  
6: 14410202 Lockerbie  
[pem\_pos=0, final\_scores=0.103546366,  
final\_local\_scores=0.09111041, pem=0.0, similarity=6.808629]  
7: 14410202 Lockerbie  
[pem\_pos=0, final\_scores=0.115032986,  
final\_local\_scores=0.1036578, pem=0.0, similarity=7.1830683]  
8: 26994 Scotland  
[pem\_pos=0, final\_scores=0.31884402,  
final\_local\_scores=0.28671342, pem=-0.47951394,  
similarity=15.253485]  
9: 23569239 Abu\_Talib\_ibn\_Abdul-Muttalib  
[pem\_pos=0, final\_scores=0.103750154,  
final\_local\_scores=0.17237106, pem=0.0, similarity=9.233609]  
10: 26994 Scotland  
[pem\_pos=0, final\_scores=0.36658448,  
final\_local\_scores=0.3388626, pem=-0.47951394,  
similarity=16.809721]  
11: 23569239 Abu\_Talib\_ibn\_Abdul-Muttalib  
[pem\_pos=0, final\_scores=0.11702715,  
final\_local\_scores=0.18727615, pem=0.0, similarity=9.678407]  
12: 5058739 Sweden  
[pem\_pos=0, final\_scores=0.63562965,  
final\_local\_scores=0.653086, pem=-0.2135508,  
similarity=24.740423]  
13: 26994 Scotland  
[pem\_pos=0, final\_scores=0.3875968,  
final\_local\_scores=0.3618154, pem=-0.22905664,  
similarity=16.132652]  
14: 23569239 Abu\_Talib\_ibn\_Abdul-Muttalib  
[pem\_pos=0, final\_scores=0.0934857,  
final\_local\_scores=0.16115871, pem=0.0, similarity=8.899011]  
15: 17633 Libya  
[pem\_pos=4, final\_scores=0.0800194,  
final\_local\_scores=0.06353569, pem=-1.6365398,  
similarity=14.885503]  
16: 81024 Pan\_American\_World\_Airways  
[pem\_pos=0, final\_scores=0.29027957,  
final\_local\_scores=0.3467067, pem=0.0, similarity=14.436132]  
17: 26994 Scotland  
[pem\_pos=0, final\_scores=0.47840816,

final\_local\_scores=0.46101302, pem=-0.47951394,  
similarity=20.454933]  
18: 14410202 Lockerbie  
[pem\_pos=0, final\_scores=0.25028223,  
final\_local\_scores=0.25139704, pem=0.0, similarity=11.591902]

Sample: b'NYT20001125\_1558\_0117'

&UR; c. 2000 N.Y. Times News Service &LR; &QC; &QL; [1 1 GRO  
SSE POINTE PARK 1] , [1 1 Mich. 1] . \_ It is deer hunting  
season in [2 Michigan 2] , and hunters who climb trees with  
loaded shotguns have state officials redoubling safety  
education efforts . Two men have been killed , and a third  
has been injured since the hunting season opened on Nov. 15 .  
All three accidents involved shotguns , elevated platforms  
and a practice relatively new to [3 Michigan 3] hunters \_  
shooting at deer from trees . The [4 Michigan 4] season , one  
of the busiest , continues through Nov. 30 . Firearm hunting  
from elevated platforms has long been legal in other states  
with heavy hunting , like [5 2 New York 5] and [2 2 Pennsylv  
nia 2] . But the practice is only in its third season in [6  
Michigan 6] , and officials say it appears that hunters still  
have a lot to learn about safety . Lt. Suzanne Koppelo ,  
hunter safety administrator of the law enforcement division  
of the Michigan 3 Department of Natural Resources &UR; , &LR;  
said that since firearm platform hunting began in [7 Michigan  
7] there had been seven shooting accidents linked to the  
practice . Three were fatal . Two hunters shot themselves on  
platforms and another was killed on the ground by a hunter  
shooting from a platform . `` If hunters exercised the kind  
of caution one would expect at heights of 20-some feet , it  
would be almost a nonissue , ' ' said Koppelo . `` But given  
human nature , that 's not the case . ' ' More than 750,000  
licensed hunters are taking to the [8 Michigan 8] woods this  
year , where they are expected to kill about 500,000 of the  
state 's 1.9 million deer . The numbers reinforce the need to  
improve education about platform hunting with firearms ,  
Koppelo said . The practice is more commonly associated with  
hunters using bows and arrows . Instead of hiding on the  
ground or walking through the woods , they wait quietly ,  
typically at a height of 15 to 20 feet , and fire down at  
deer that wander below . Despite the hunter killed from the  
platform , Sgt. Larry D. Sargent , hunter education field  
coordinator for the law enforcement division , says the  
practice does offer one safety advantage : errant shots  
typically thump harmlessly to the ground , rather than  
whistling through the woods and striking other people . The  
problems primarily arise when hunters get in and out of  
position . `` It 's kind of hard climbing a tree with a gun ,  
' ' Rob Williams said , as he and a hunting partner , Art  
Bush , waited on a recent morning in a conventional  
earthbound hunting blind about 20 miles west of [3 3 Detroit 3  
] . For Williams , Rule No. 1 is not to climb a tree with a

loaded gun . He says that when he hunts from a platform he climbs up with a rope dangling from his belt to his gun , which he leaves unloaded on the ground . When he is secure in his stand , he pulls his gun up after him and loads it . Sargent says this is proper procedure . One accident this year , however , illustrates the importance of following each step of the safety procedures . A 16-year-old hunter used a haul line , but it was tied to the trigger guard and he left the gun loaded , Koppelo said . The teen-ager 's gun went off and he was shot in the foot . Others are hurt falling out of trees . Koppelo said that state officials were trying to keep count of these injuries but that it was difficult , since they relied on hunters reporting minor accidents . `` A lot of guys are embarrassed , '' Sargent said . `` They 'll say they fell working in the yard or something . '' Others , Koppelo says , do not want to report accidents for fear the law will be changed . And she adds that though anecdotal reports show [9 Michigan 9] 's large hunting population to be taking slowly to firearm deer hunting from platforms , most expect the practice to catch on as it has in other states . As Peter Thompson , a 16-year-old hunter , said , `` Usually the deer do n't look up . ''

1: 119261 Grosse\_Pointe\_Park,\_Michigan

2: 23332 Pennsylvania

3: 8687 Detroit

1: 18859 Michigan

[winner: 18859 Michigan, pem\_pos=0,  
final\_scores=0.6055379, final\_local\_scores=0.53799015,  
pem=0.0, similarity=20.144413]

gt: 18859 Michigan, pem\_pos=0, final\_scores=0.6055379,  
final\_local\_scores=0.53799015, pem=0.0,  
similarity=20.144413 ]

[winner: 4721653 Zyphe\_language, pem\_pos=12,  
final\_scores=-1.3736796, final\_local\_scores=-1.3028207,  
pem=-4.4188404, similarity=-10.758715]

RECALL MISS: 18859 Michigan]

2: 8210131 New\_York

[winner: 645042 New\_York\_City, pem\_pos=1,  
final\_scores=0.3574925, final\_local\_scores=0.33284193,  
pem=-1.7906687, similarity=23.760315]

gt: 8210131 New\_York, pem\_pos=0,  
final\_scores=0.024271712, final\_local\_scores=-0.046278484,  
pem=-0.2889045, similarity=4.279778 ]

3: 6480872 Michigan\_Department\_of\_Natural\_Resources

[winner: 6480872

Michigan\_Department\_of\_Natural\_Resources, pem\_pos=0,  
final\_scores=-0.5637547, final\_local\_scores=-0.5819928,  
pem=0.0, similarity=-13.278113]

gt: 6480872 Michigan\_Department\_of\_Natural\_Resources,  
pem\_pos=0, final\_scores=-0.5637547,  
final\_local\_scores=-0.5819928, pem=0.0,  
similarity=-13.278113 ]

1: 119261 Grosse\_Pointe\_Park,\_Michigan  
 [pem\_pos=0, final\_scores=0.11210142,  
 final\_local\_scores=0.13253987, pem=0.0, similarity=8.044967]  
 2: 23332 Pennsylvania  
 [pem\_pos=0, final\_scores=0.46186537,  
 final\_local\_scores=0.42399615, pem=-0.11810702,  
 similarity=17.384888]  
 3: 8687 Detroit  
 [pem\_pos=0, final\_scores=0.56153554,  
 final\_local\_scores=0.5261439, pem=-0.21700948,  
 similarity=20.97103]  
 1: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.6055379,  
 final\_local\_scores=0.53799015, pem=0.0, similarity=20.144413]  
 2: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.6496467,  
 final\_local\_scores=0.58711267, pem=-0.24788691,  
 similarity=22.958372]  
 3: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.71115714,  
 final\_local\_scores=0.65430343, pem=-0.24788691,  
 similarity=24.96348]  
 4: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.58049107,  
 final\_local\_scores=0.51157063, pem=-0.24788691,  
 similarity=20.704048]  
 5: 645042 New\_York\_City  
 [pem\_pos=1, final\_scores=0.3574925,  
 final\_local\_scores=0.33284193, pem=-1.7906687,  
 similarity=23.760315]  
 6: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.68440217,  
 final\_local\_scores=0.6250777, pem=-0.24788691,  
 similarity=24.091326]  
 7: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.6172689,  
 final\_local\_scores=0.5517448, pem=-0.24788691,  
 similarity=21.902925]  
 8: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.6111698,  
 final\_local\_scores=0.5450825, pem=-0.24788691,  
 similarity=21.70411]  
 9: 18859 Michigan  
 [pem\_pos=0, final\_scores=0.7399808,  
 final\_local\_scores=0.6857889, pem=-0.24788691,  
 similarity=25.903069]

Sample: b'VOA20001129\_2000\_0364'

Welcome to this hour of VOA News Now . I 'm Erin Brumett in [ 1 1 Washington 1] . The incredibly close Florida vote count for President is in the state courts and heading for the [1 US 1 Supreme Court 1] at the end of the week . That 's

virtually unprecedented , as the [2 Federal Courts 2] have tended to stay out of elections , which are the responsibility of the states . 2 Joe O'Grossman is a professor of constitutional law at [2 2 Baltimore 2] 's [3 3 J ohns Hopkins University 3] . He says that in this instance the Supreme Court probably felt the circumstances were such that they had no real alternative but to take the case . Well , I think they took the case because it seemed like it was such a important issue that it would be kind of unseemly for them to try to reject it out of hand . In fact , I think that [4 4 Lawrence Tribe 4] , the [3 Harvard Law School 3] Professor who is arguing for [5 5 Gore 5] tomorrow , is going to ask the court to hold the case , moot and dismiss it . And the [4 Bush 4] campaign has brought the case , but recognizes that the particular issues that they brought to the court last week are really not very relevant anymore . They are going to probably , well , I know what they are going to do . They are going to try to broaden the case to make it a kind of a general referendum on election procedures . So , they are looking for a coronation . And the [5 Gore 5] people at this point just want to get rid of this case , because they think they have a better chance in the [6 Florida 6] courts . [7 Republicans 7] generally support states ' rights . So , does it appear somewhat ironic that they are going to the [8 Supreme Court 8] in this instance ? No , I think first of all the notion that [9 Republicans 9] favor states ' rights and [10 Democrats 10] oppose it , is much too simplistic . That 's generally true , but it 's only true on some issues . And in this case , you could argue that the [11 Bush 11] people are going to the [12 Supreme Court 12] to preserve the right of [13 Florida 13] , to determine its own electors . So , there is a kind of states ' rights issue there . But they need , or at least they once thought they needed the [14 Supreme Court 14] to endorse it . So it goes both ways . What are the justices likely looking for ? Well , I do n't know . But I think that - and I think they will be divided , I mean I do n't think they will be unanimous . Obviously , they would like to be able to achieve some kind of resolution of the case . [15 American 15] election laws are often challenged at the local level and sometimes in the [16 Federal 16] courts , but this is certainly the first time we have had a [17 Supreme Court 17] challenge over presidential election . So , there is a lot of new law here to be made or at least law that has n't been made yet . It 's possible that if the court accepts the invitation of the [18 Bush 18] campaign , at least as it is expressed in the briefs they submitted yesterday , that the court will in the sense , take the entire issue , even - not just the specific questions that it was asked last week and try to resolve the question . For example , if the court says it is clearly the fact that the 3 legislature of the state of [19 Florida 19] can determine for itself who its electors will be , that will pretty much end the case , because that



's what the legislature already plans to do . And that does n't mean the [20 Gore 20] people wo n't keep trying , because ultimately [6 6 Congress 6] has to decide and only [21 Congress 21] can decide if there are competing slates of electors submitted , which slate to accept . The [22 Johns Hopkins University 22] constitutional law professor Joe O'Grossman says , it 's still not certain the [23 Supreme Court 23] will actually hear the Florida case . As their first order , business Friday , attorneys for Vice President [ 24 Gore 24] will move to have the court dismiss the case brought by [25 Republican 25] [26 George W Bush 26] .

1: 108956 Washington,\_D.C.

2: 26997138 Baltimore

3: 38420 Johns\_Hopkins\_University

4: 758546 Laurence\_Tribe

5: 5042706 Al\_Gore

6: 31756 United\_States\_Congress

1: 31737 Supreme\_Court\_of\_the\_United\_States

[winner: 31737 Supreme\_Court\_of\_the\_United\_States,

pem\_pos=0, final\_scores=0.31373423,

final\_local\_scores=0.2822363, pem=0.0, similarity=12.512207|

gt: 31737 Supreme\_Court\_of\_the\_United\_States, pem\_pos=0,

final\_scores=0.31373423, final\_local\_scores=0.2822363,

pem=0.0, similarity=12.512207 ]

2: 5022991 Joel\_Grossman

[winner: 145422 Joe\_Biden, pem\_pos=9,

final\_scores=-0.8793642, final\_local\_scores=-1.0081649,

pem=-4.1998224, similarity=-3.1566489|

RECALL MISS: 5022991 Joel\_Grossman]

3: 429346 Florida\_Legislature

[winner: 4503828 Parliament\_of\_Saint\_Lucia, pem\_pos=12,

final\_scores=-1.3400279, final\_local\_scores=-1.2866693,

pem=-3.2829933, similarity=-16.453634|

RECALL MISS: 429346 Florida\_Legislature]

1: 108956 Washington,\_D.C.

[pem\_pos=0, final\_scores=0.4394098,

final\_local\_scores=0.40726823, pem=-1.7496481,

similarity=25.758266]

2: 26997138 Baltimore

[pem\_pos=0, final\_scores=0.52001166,

final\_local\_scores=0.54298306, pem=-0.19949137,

similarity=21.378275]

3: 38420 Johns\_Hopkins\_University

[pem\_pos=0, final\_scores=0.33488125,

final\_local\_scores=0.3601384, pem=0.0, similarity=14.836961]

4: 758546 Laurence\_Tribe

[pem\_pos=0, final\_scores=0.4785729,

final\_local\_scores=0.5252773, pem=0.0, similarity=19.765036]

5: 5042706 Al\_Gore

[pem\_pos=0, final\_scores=0.64990544,

final\_local\_scores=0.64397436, pem=0.0, similarity=23.307192]

6: 31756 United\_States\_Congress

[pem\_pos=0, final\_scores=0.25536913,  
final\_local\_scores=0.20149106, pem=-0.3349529,  
similarity=11.924133]  
1: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.31373423,  
final\_local\_scores=0.2822363, pem=0.0, similarity=12.512207]  
2: 1636627 Federal\_judiciary\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.13907908,  
final\_local\_scores=0.12923011, pem=0.0, similarity=7.9461975]  
3: 158984 Harvard\_Law\_School  
[pem\_pos=0, final\_scores=0.22473793,  
final\_local\_scores=0.23553163, pem=0.0, similarity=11.118446]  
4: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.3059337,  
final\_local\_scores=0.24730054, pem=0.0, similarity=11.469654]  
5: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.7213745,  
final\_local\_scores=0.72204345, pem=0.0, similarity=25.63693]  
6: 18933066 Florida  
[pem\_pos=0, final\_scores=0.085966334,  
final\_local\_scores=0.07510051, pem=-0.15405534,  
similarity=7.1686387]  
7: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.4314028,  
final\_local\_scores=0.40899038, pem=-0.11427376,  
similarity=16.91624]  
8: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.2589032,  
final\_local\_scores=0.22234178, pem=-0.9940822,  
similarity=16.13081]  
9: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.41705006,  
final\_local\_scores=0.39331216, pem=-0.11427376,  
similarity=16.448372]  
10: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.41297626,  
final\_local\_scores=0.39059365, pem=-0.18007085,  
similarity=16.72506]  
11: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.22016893,  
final\_local\_scores=0.15126663, pem=0.0, similarity=8.603811]  
12: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.18669377,  
final\_local\_scores=0.14346391, pem=-0.9940822,  
similarity=13.776937]  
13: 18933066 Florida  
[pem\_pos=0, final\_scores=0.060784176,  
final\_local\_scores=0.04759285, pem=-0.15405534,  
similarity=6.3477554]  
14: 31737 Supreme\_Court\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.11116205,  
final\_local\_scores=0.06095697, pem=-0.9940822,

similarity=11.314764]  
 15: 3434750 United\_States  
     [pem\_pos=0, final\_scores=0.518627,  
 final\_local\_scores=0.5298783, pem=-0.12994486,  
 similarity=20.609]  
 16: 195149 Federal\_government\_of\_the\_United\_States  
     [pem\_pos=1, final\_scores=0.032831743,  
 final\_local\_scores=-0.014860153, pem=-1.574776,  
 similarity=12.210134]  
 17: 31737 Supreme\_Court\_of\_the\_United\_States  
     [pem\_pos=0, final\_scores=0.12829365,  
 final\_local\_scores=0.07967065, pem=-0.9940822,  
 similarity=11.873218]  
 18: 3414021 George\_W.\_Bush  
     [pem\_pos=0, final\_scores=0.30039108,  
 final\_local\_scores=0.24124604, pem=0.0, similarity=11.288976]  
 19: 18933066 Florida  
     [pem\_pos=0, final\_scores=0.23152183,  
 final\_local\_scores=0.23409781, pem=-0.15405534,  
 similarity=11.913435]  
 20: 5042706 Al\_Gore  
     [pem\_pos=0, final\_scores=0.67413986,  
 final\_local\_scores=0.6704468, pem=0.0, similarity=24.097183]  
 21: 31756 United\_States\_Congress  
     [pem\_pos=0, final\_scores=0.2646821,  
 final\_local\_scores=0.21296558, pem=-0.3349529,  
 similarity=12.266556]  
 22: 38420 Johns\_Hopkins\_University  
     [pem\_pos=0, final\_scores=0.31399536,  
 final\_local\_scores=0.33317763, pem=0.0, similarity=14.032398]  
 23: 31737 Supreme\_Court\_of\_the\_United\_States  
     [pem\_pos=0, final\_scores=0.1439778,  
 final\_local\_scores=0.0968032, pem=-0.9940822,  
 similarity=12.384487]  
 24: 5042706 Al\_Gore  
     [pem\_pos=0, final\_scores=0.4757098,  
 final\_local\_scores=0.4503411, pem=-1.2035761,  
 similarity=24.074024]  
 25: 32070 Republican\_Party\_(United\_States)  
     [pem\_pos=0, final\_scores=0.35961205,  
 final\_local\_scores=0.3305699, pem=-0.1310191,  
 similarity=14.66708]  
 26: 3414021 George\_W.\_Bush  
     [pem\_pos=0, final\_scores=0.1837657,  
 final\_local\_scores=0.109338805, pem=-0.017176678,  
 similarity=7.4460106]

Sample: b'20001115\_AFP\_ARB\_0030\_eng'

[1 1 Gaza 1] 11-15 -LRB- [1 1 AFP 1] -RRB- - A [2  
 Palestinian 2] medical source reported today Wednesday that  
 more than fifteen [3 Palestinians 3] were injured this  
 morning in the town of [4 Rafah 4] in the southern [5 Gaza

Strip 5] in clashes with the [6 Israeli 6] army . Medical sources said `` at least 15 [7 Palestinians 7] were injured by live bullets fired by [8 Israeli 8] soldiers , including six children all shot in the head . ' ' Eyewitnesses said the clashes broke out when the [9 Israeli 9] army began bulldozing [10 Palestinian 10] land near the border with [2 2 Egypt 2] and opened fire randomly on [11 Palestinian 11] citizens when they threw stones at the soldiers in protest . An [12 Israeli 12] tank also fired a rocket . Elsewhere , [13 Israeli 13] bulldozers dug up dozens of hectares of [14 Palestinian 14] farmland near the village of Alkarara -LRB- near the Kosovim checkpoint -RRB- last night for the second day in a row , destroying two houses and a well , where the army erected a new military camp instead . The town 's authorities said last night the [15 Israeli 15] army dug up several hectares planted with orange trees in southern [16 Gaza 16] -LRB- 2 Nitsarim -RRB- . The bulldozing of land began the night before last when an [17 Israeli 17] soldier was killed in a shooting incident in the area .

1: 471981 Agence\_France-Presse

2: 8087628 Egypt

1: 55387 Gaza

[winner: 12047 Gaza\_Strip, pem\_pos=1, final\_scores=0.09009029, final\_local\_scores=0.014540076, pem=-1.374985, similarity=12.001]

gt: 55387 Gaza, pem\_pos=0, final\_scores=-0.06647116, final\_local\_scores=-0.15296096, pem=-0.30319342, similarity=1.1738658 ]

2: 2509332 Netzarim\_(settlement) no overlap with any filtered span

1: 471981 Agence\_France-Presse

[pem\_pos=0, final\_scores=0.108233616, final\_local\_scores=0.14519522, pem=-0.6605333, similarity=12.01471]

2: 8087628 Egypt

[pem\_pos=0, final\_scores=0.5051058, final\_local\_scores=0.4940418, pem=-0.27480945, similarity=20.327362]

1: 12047 Gaza\_Strip

[pem\_pos=1, final\_scores=0.09009029, final\_local\_scores=0.014540076, pem=-1.374985, similarity=12.001]

2: 241405 State\_of\_Palestine

[pem\_pos=9, final\_scores=0.35237497, final\_local\_scores=0.302822, pem=-2.8508017, similarity=28.62963]

3: 241405 State\_of\_Palestine

[pem\_pos=5, final\_scores=0.30883396, final\_local\_scores=0.25526005, pem=-1.7909594, similarity=21.446697]

4: 584373 Rafah

[pem\_pos=0, final\_scores=0.31213337,

final\_local\_scores=0.32141745, pem=0.0, similarity=13.681451]  
5: 12047 Gaza\_Strip  
[pem\_pos=0, final\_scores=0.42020816,  
final\_local\_scores=0.37514365, pem=0.0, similarity=15.284748]  
6: 9282173 Israel  
[pem\_pos=1, final\_scores=0.6064795,  
final\_local\_scores=0.5603408, pem=-0.34790626,  
similarity=22.703367]  
7: 241405 State\_of\_Palestine  
[pem\_pos=5, final\_scores=0.2804417,  
final\_local\_scores=0.22424579, pem=-1.7909594,  
similarity=20.52117]  
8: 9282173 Israel  
[pem\_pos=1, final\_scores=0.5853661,  
final\_local\_scores=0.5372775, pem=-0.34790626,  
similarity=22.015116]  
9: 9282173 Israel  
[pem\_pos=1, final\_scores=0.5574489,  
final\_local\_scores=0.50678235, pem=-0.34790626,  
similarity=21.105076]  
10: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.2220098,  
final\_local\_scores=0.16041785, pem=-2.8508017,  
similarity=24.380005]  
11: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.24107863,  
final\_local\_scores=0.18124765, pem=-2.8508017,  
similarity=25.001608]  
12: 9282173 Israel  
[pem\_pos=1, final\_scores=0.5686211,  
final\_local\_scores=0.51898617, pem=-0.34790626,  
similarity=21.469263]  
13: 9282173 Israel  
[pem\_pos=1, final\_scores=0.52575547,  
final\_local\_scores=0.47216207, pem=-0.34790626,  
similarity=20.07194]  
14: 241405 State\_of\_Palestine  
[pem\_pos=9, final\_scores=0.12210481,  
final\_local\_scores=0.051286817, pem=-2.8508017,  
similarity=21.123318]  
15: 9282173 Israel  
[pem\_pos=1, final\_scores=0.5878369,  
final\_local\_scores=0.53997654, pem=-0.34790626,  
similarity=22.09566]  
16: 12047 Gaza\_Strip  
[pem\_pos=1, final\_scores=0.0912215,  
final\_local\_scores=0.018456727, pem=-1.374985,  
similarity=12.117881]  
17: 9282173 Israel  
[pem\_pos=0, final\_scores=0.5688117,  
final\_local\_scores=0.5191944, pem=-0.34790626,  
similarity=21.47548]

Sample: b'NYT20001101\_2212\_0429'

MERGER &QL; By ALEX BERENSON &QC; &LR; &QL; &UR; c. 2000  
N.Y. Times News Service &QC; &LR; &QL; &UR; &LR; Back in  
January , when [1 Internet 1] stocks were hot , [2 America  
Online 2] agreed to buy [3 Time Warner 3] with [4 AOL 4]  
stock . The deal , if approved by regulators , would leave [5  
AOL 5] shareholders with 55 percent of the merged company ,  
even though [6 Time Warner 6] 's revenue and cash flow are  
far greater . Since then , [7 AOL 7] 's stock has fallen by  
almost a third , destroying most of the premium the company  
had offered for control of [1 1 Time Warner 1] . So a [8 Time  
Warner 8] shareholder might conclude that [9 America Online 9]  
's chairman , [10 Stephen Case 10] , had bamboozled his [11  
Time Warner 11] counterpart , [12 Gerald Levin 12] , into  
taking [13 AOL 13] 's shares at the worst possible time . ``  
Short term , anyone objective would say probably Steve got  
the better part of the deal , ' ' said Larry Haverty , a  
senior vice president at 1 State Street Research in [2 2 Bost  
on 2] and a longtime media investor . But Haverty , like many  
other [14 Time Warner 14] shareholders and industry  
analysts , hardly cares . With the merger likely to be  
completed later this month , he predicts that the merged  
company will grow quickly by combining [15 AOL 15] 's  
marketing skills and captive base of online users with [16  
Time Warner 16] 's huge library of films , magazines and the  
like and its relationships with major advertisers . `` The  
deal strategically and financially makes all the sense in the  
world , ' ' Haverty said . `` Five years out , if you 're a  
shareholder in this company , you 're going to be a happy  
camper . ' ' That view is widely shared . `` Strategically ,  
it makes a lot of sense , ' ' said Tom Wolzien of [3 3 Sanford  
C. Bernstein 3] . `` These companies fill the holes of the  
other one . They 're subscription-advertising combo models .  
The reality is the pieces are very compatible . ' ' To be  
sure , the recent history of big media mergers is hardly  
comforting for shareholders in either [17 AOL 17] or [18 Time  
Warner 18] . [4 4 Disney 4] stumbled badly after it bought [5  
5 Capital Cities/ABC 5] in 1996 , and [19 Time Warner 19]  
stagnated for five years after it was created in 1990 by the  
merger of [20 2 Time 20] and 3 Warner . `` Huge deals are  
really hard to do and create a lot of uncertainty , ' ' said [6  
6 Roger McNamee 6] of Integral Capital , which sold its  
404,000 [21 AOL 21] shares the day the merger was announced .  
But if the merged company can overcome that hurdle , it will  
have big advantages over [22 Disney 22] and other traditional  
media companies , which are still struggling to figure out  
how they can profit from the [23 Internet 23] , said Jessica  
Reif of [7 7 Merrill Lynch 7] . `` Given the multiple  
distribution platforms that they have , it 's obviously a  
very powerful consumer engine , ' ' Reif said . `` There 's a  
reason that you see [24 Disney 24] and to a lesser extent [8 8

NBC 8] screaming in [9 9 Washington 9] . '' John Schreiber , an assistant portfolio manager at 4 Janus Capital , the mutual fund company that is [25 Time Warner 25] 's largest shareholder , with more than 120 million shares , said the companies had already begun to demonstrate how they would work together . For example , [26 AOL 26] has pitched [27 Time Warner 27] 's magazines to its users , resulting in hundreds of thousands of new subscriptions . And the companies are trying to sell big advertisers on deals that will reach both [28 America Online 28] users and the people who watch [29 Time Warner 29] cable channels and read its magazines , Schreiber said . `` Once the merger closes , I think you 'll see some very powerful cross-promotional deals across all of [30 Time Warner 30] 's advertisers , '' he said . Schreiber also says the risk of cultural clashes between the companies is limited , because the companies disclosed publicly where top executives would stand in the combined company only a few weeks after the merger was announced . So people who feel their responsibilities have been diminished have had months to look for new jobs , and some have already left , he said . Given the strength he sees in the combined company , Schreiber said the question of whether Levin could have negotiated a better deal with [31 America Online 31] is moot . Shares in cable companies like [10 10 Comcast 10] have fallen 15 percent or more this year , while big media companies have been flat or up slightly . So `` it 's not as if [32 Time Warner 32] shareholders have lost out on a great deal of upside , '' Schreiber said .

1: 83045 Time\_Warner

2: 24437894 Boston

3: 20833274 Sanford\_Bernstein

4: 37398 The\_Walt\_Disney\_Company

5: 2359720 Capital\_Cities\_Communications

6: 16355166 Roger\_McNamee

7: 33731299 Merrill\_Lynch

8: 21780 NBC

9: 108956 Washington,\_D.C.

10: 303749 Comcast

1: 26824 State\_Street\_Corporation

[winner: 1401256 State\_Street\_(Madison), pem\_pos=0, final\_scores=-1.0196317, final\_local\_scores=-1.007041, pem=-0.92191434, similarity=-20.948885]

gt: 26824 State\_Street\_Corporation, pem\_pos=2, final\_scores=-1.2356641, final\_local\_scores=-1.2925177, pem=-1.4717546, similarity=-26.477964 ]

[winner: 7015339 Carmellon\_Research, pem\_pos=4, final\_scores=-1.1316285, final\_local\_scores=-1.0496423, pem=-2.3978953, similarity=-14.193581]

RECALL MISS: 26824 State\_Street\_Corporation]

2: 6696324 Time\_Inc.

[winner: 31600 Time\_(magazine), pem\_pos=0, final\_scores=0.22710602, final\_local\_scores=0.2286005,

pem=-0.31107646, similarity=12.603291|  
gt: 6696324 Time\_Inc., pem\_pos=11,  
final\_scores=-0.5144954, final\_local\_scores=-0.6130364,  
pem=-4.2874045, similarity=9.111055 ]  
3: 389886 Warner\_Communications  
[winner: 454544 Warner\_Music\_Group, pem\_pos=0,  
final\_scores=-0.023783848, final\_local\_scores=-0.07453349,  
pem=-1.3944875, similarity=9.448924|  
RECALL MISS: 389886 Warner\_Communications]  
4: 1310841 Janus\_Capital\_Group  
[winner: 10829920 Capitals\_(typeface), pem\_pos=7,  
final\_scores=-0.99746746, final\_local\_scores=-0.8937132,  
pem=-3.250049, similarity=-4.9061985|  
RECALL MISS: 1310841 Janus\_Capital\_Group]  
[winner: 159852 Janus, pem\_pos=0,  
final\_scores=-1.2758934, final\_local\_scores=-1.277694,  
pem=-0.2638665, similarity=-32.60428|  
gt: 1310841 Janus\_Capital\_Group, pem\_pos=13,  
final\_scores=-1.3547115, final\_local\_scores=-1.3771112,  
pem=-3.6898952, similarity=-16.939806 ]  
1: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.25618315,  
final\_local\_scores=0.18541485, pem=0.0, similarity=9.622862]  
2: 24437894 Boston  
[pem\_pos=0, final\_scores=0.51058066,  
final\_local\_scores=0.563107, pem=-0.23164093,  
similarity=22.153648]  
3: 20833274 Sanford\_Bernstein  
[pem\_pos=0, final\_scores=0.19041727,  
final\_local\_scores=0.30535465, pem=0.0, similarity=13.202105]  
4: 37398 The\_Walt\_Disney\_Company  
[pem\_pos=0, final\_scores=0.43526775,  
final\_local\_scores=0.40516025, pem=-0.24767768,  
similarity=17.527412]  
5: 2359720 Capital\_Cities\_Communications  
[pem\_pos=0, final\_scores=0.36594182,  
final\_local\_scores=0.37099296, pem=0.0, similarity=15.160882]  
6: 16355166 Roger\_McNamee  
[pem\_pos=0, final\_scores=0.04422097,  
final\_local\_scores=0.118258804, pem=0.0, similarity=7.618791]  
7: 33731299 Merrill\_Lynch  
[pem\_pos=0, final\_scores=0.3863706,  
final\_local\_scores=0.42420113, pem=0.0, similarity=16.74872]  
8: 21780 NBC  
[pem\_pos=0, final\_scores=0.2419943,  
final\_local\_scores=0.20859194, pem=-0.06403453,  
similarity=10.662743]  
9: 108956 Washington,\_D.C.  
[pem\_pos=0, final\_scores=0.32942402,  
final\_local\_scores=0.36721182, pem=-1.7496481,  
similarity=24.562904]  
10: 303749 Comcast



[pem\_pos=0, final\_scores=0.35138863,  
final\_local\_scores=0.31601563, pem=0.0, similarity=13.520249]  
1: 33139 World\_Wide\_Web  
[pem\_pos=1, final\_scores=0.0334502,  
final\_local\_scores=0.045417607, pem=-3.3911073,  
similarity=23.886433]  
2: 1397 AOL  
[pem\_pos=0, final\_scores=0.5791753,  
final\_local\_scores=0.5459073, pem=0.0, similarity=20.380676]  
3: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.3932491,  
final\_local\_scores=0.33440983, pem=0.0, similarity=14.069169]  
4: 1397 AOL  
[pem\_pos=0, final\_scores=0.6354022,  
final\_local\_scores=0.6073267, pem=0.0, similarity=22.213552]  
5: 1397 AOL  
[pem\_pos=0, final\_scores=0.61913294,  
final\_local\_scores=0.589555, pem=0.0, similarity=21.68321]  
6: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.31293464,  
final\_local\_scores=0.2466785, pem=0.0, similarity=11.451092]  
7: 1397 AOL  
[pem\_pos=0, final\_scores=0.5841723,  
final\_local\_scores=0.55136585, pem=0.0, similarity=20.54357]  
8: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.32760137,  
final\_local\_scores=0.26342842, pem=0.0, similarity=11.950942]  
9: 1397 AOL  
[pem\_pos=0, final\_scores=0.5812906,  
final\_local\_scores=0.548218, pem=0.0, similarity=20.449633]  
10: 285956 Steve\_Case  
[pem\_pos=0, final\_scores=0.16781168,  
final\_local\_scores=0.19261268, pem=0.0, similarity=9.83766]  
11: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.36034864,  
final\_local\_scores=0.29865128, pem=0.0, similarity=13.002063]  
12: 2020385 Gerald\_M.\_Levin  
[pem\_pos=0, final\_scores=0.12621532,  
final\_local\_scores=0.18752605, pem=0.0, similarity=9.685864]  
13: 1397 AOL  
[pem\_pos=0, final\_scores=0.51190025,  
final\_local\_scores=0.47241956, pem=0.0, similarity=18.187654]  
14: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.33372217,  
final\_local\_scores=0.2703197, pem=0.0, similarity=12.156592]  
15: 1397 AOL  
[pem\_pos=0, final\_scores=0.5941573,  
final\_local\_scores=0.5622729, pem=0.0, similarity=20.869059]  
16: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.34541482,  
final\_local\_scores=0.28233832, pem=0.0, similarity=12.515251]  
17: 1397 AOL

[pem\_pos=0, final\_scores=0.60278034,  
final\_local\_scores=0.5716923, pem=0.0, similarity=21.15015]  
18: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.2978831,  
final\_local\_scores=0.23023695, pem=0.0, similarity=10.960443]  
19: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.32658,  
final\_local\_scores=0.26158392, pem=0.0, similarity=11.895899]  
20: 31600 Time\_(magazine)  
[pem\_pos=0, final\_scores=0.22710602,  
final\_local\_scores=0.2286005, pem=-0.31107646,  
similarity=12.603291]  
21: 1397 AOL  
[pem\_pos=0, final\_scores=0.50517255,  
final\_local\_scores=0.4650706, pem=0.0, similarity=17.968348]  
22: 37398 The\_Walt\_Disney\_Company  
[pem\_pos=0, final\_scores=0.44344133,  
final\_local\_scores=0.41408867, pem=-0.24767768,  
similarity=17.793856]  
23: 33139 World\_Wide\_Web  
[pem\_pos=1, final\_scores=0.19988744,  
final\_local\_scores=0.227225, pem=-3.3911073,  
similarity=29.311928]  
24: 37398 The\_Walt\_Disney\_Company  
[pem\_pos=0, final\_scores=0.3503526,  
final\_local\_scores=0.31240338, pem=-0.24767768,  
similarity=14.759362]  
25: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.35332245,  
final\_local\_scores=0.290796, pem=0.0, similarity=12.767646]  
26: 1397 AOL  
[pem\_pos=0, final\_scores=0.5757176,  
final\_local\_scores=0.54213035, pem=0.0, similarity=20.267963]  
27: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.2964629,  
final\_local\_scores=0.22868559, pem=0.0, similarity=10.914147]  
28: 1397 AOL  
[pem\_pos=0, final\_scores=0.56535983,  
final\_local\_scores=0.5308161, pem=0.0, similarity=19.930323]  
29: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.33590847,  
final\_local\_scores=0.2719541, pem=0.0, similarity=12.205365]  
30: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.3914432,  
final\_local\_scores=0.33261728, pem=0.0, similarity=14.0156765]  
31: 1397 AOL  
[pem\_pos=0, final\_scores=0.554792,  
final\_local\_scores=0.5192723, pem=0.0, similarity=19.585835]  
32: 83045 Time\_Warner  
[pem\_pos=0, final\_scores=0.28174978,  
final\_local\_scores=0.21334246, pem=0.0, similarity=10.456278]

Sample: b'20001115\_AFP\_ARB\_0013\_eng'

[1 1 Bandar Seri Begawan 1] 11-15 -LRB- [2 2 AFP 2] -RRB- - [1 American 1] President [2 Bill Clinton 2] attempted today Wednesday to reassure money markets about doubts caused by the battle between the [3 Republicans 3] and [4 Democrats 4] on the result of the presidential elections in the [5 United States 5] , which is still unknown and dependent on the results of the recount in the 1 State of Florida . [6 Clinton 6] told a group of businessmen on the periphery of the [7 Asia-Pacific Economic Cooperation 7] forum in [3 3 Brunei 3] that `` one of the issues both Vice-President [8 Al Gore 8] and Governor [9 George Bush 9] agree on in these elections is the need to maintain a strong and pioneering role by the [10 United States 10] in order to achieve flourishing trade and greater integration for the world 's economy -LSB- ... -RSB- and the world should not be concerned because the two candidates are strongly committed to following this path . '' The [11 American 11] stock market saw an upheaval as a result of the delay in announcing the final results of the presidential elections in the [12 United States 12] but it improved on Tuesday . Initial results in [13 Florida 13] on Tuesday evening showed [14 Republican 14] candidate [15 George Bush 15] to be 300 votes ahead of [16 Democrat 16] candidate [17 Al Gore 17] but some voting stations were still recounting votes today Wednesday . In [18 Brunei 18] , President [19 Clinton 19] refused to make any predications about who might win the presidency in this exceptional election battle between the [20 Democratic 20] and [21 Republican 21] candidates . He said `` we must be careful when we talk about predictions ; the only thing I can predict with confidence is that I will not be taking part in the next [22 Apec 22] summit next year and at present , I ca n't say who will be representing the [23 United States 23] . ''

1: 56611 Bandar\_Seri\_Begawan

2: 471981 Agence\_France-Presse

3: 3466 Brunei

1: 18933066 Florida

[winner: 18933066 Florida, pem\_pos=0, final\_scores=0.005418375, final\_local\_scores=-0.023443326, pem=0.0, similarity=3.3901174]

gt: 18933066 Florida, pem\_pos=0, final\_scores=0.005418375, final\_local\_scores=-0.023443326, pem=0.0, similarity=3.3901174 ]

1: 56611 Bandar\_Seri\_Begawan

[pem\_pos=0, final\_scores=0.18205576, final\_local\_scores=0.26753107, pem=0.0, similarity=12.073374]

2: 471981 Agence\_France-Presse

[pem\_pos=0, final\_scores=0.032648757, final\_local\_scores=0.07958904, pem=-0.6605333, similarity=10.056891]

3: 3466 Brunei

[pem\_pos=0, final\_scores=0.31451368,

final\_local\_scores=0.35443166, pem=-0.386843,  
similarity=16.770372]  
1: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.50481707,  
final\_local\_scores=0.48676968, pem=-0.12994486,  
similarity=19.322552]  
2: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.7408264,  
final\_local\_scores=0.7184739, pem=0.0, similarity=25.530407]  
3: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.43078083,  
final\_local\_scores=0.41028744, pem=-0.11427376,  
similarity=16.954948]  
4: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.43008965,  
final\_local\_scores=0.41146708, pem=-0.18007085,  
similarity=17.347965]  
5: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.52569765,  
final\_local\_scores=0.5095786, pem=-0.073457584,  
similarity=19.696028]  
6: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.8286608,  
final\_local\_scores=0.8173504, pem=0.0, similarity=28.48108]  
7: 135976 Asia-Pacific\_Economic\_Cooperation  
[pem\_pos=0, final\_scores=0.37040925,  
final\_local\_scores=0.40794402, pem=0.0, similarity=16.263577]  
8: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.367172,  
final\_local\_scores=0.34271047, pem=0.0, similarity=14.316876]  
9: 11955 George\_H.\_W.\_Bush  
[pem\_pos=1, final\_scores=0.27818507,  
final\_local\_scores=0.22297403, pem=-0.80923575,  
similarity=15.144452]  
10: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.5155222,  
final\_local\_scores=0.4984634, pem=-0.073457584,  
similarity=19.364328]  
11: 3434750 United\_States  
[pem\_pos=1, final\_scores=0.57516843,  
final\_local\_scores=0.5617969, pem=-0.12994486,  
similarity=21.561514]  
12: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.53334373,  
final\_local\_scores=0.5179308, pem=-0.073457584,  
similarity=19.945276]  
13: 18933066 Florida  
[pem\_pos=0, final\_scores=0.08870821,  
final\_local\_scores=0.07434712, pem=-0.15405534,  
similarity=7.1461563]  
14: 32070 Republican\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.4244163,

final\_local\_scores=0.40333515, pem=-0.1310191,  
 similarity=16.838541]  
 15: 11955 George\_H.\_W.\_Bush  
     [pem\_pos=1, final\_scores=0.25209278,  
 final\_local\_scores=0.19447213, pem=-0.80923575,  
 similarity=14.2939]  
 16: 5043544 Democratic\_Party\_(United\_States)  
     [pem\_pos=0, final\_scores=0.42116255,  
 final\_local\_scores=0.40171552, pem=-0.061652232,  
 similarity=16.41298]  
 17: 5042706 Al\_Gore  
     [pem\_pos=0, final\_scores=0.36339462,  
 final\_local\_scores=0.33858424, pem=0.0, similarity=14.193742]  
 18: 3466 Brunei  
     [pem\_pos=0, final\_scores=0.32341963,  
 final\_local\_scores=0.36416006, pem=-0.386843,  
 similarity=17.060688]  
 19: 3356 Bill\_Clinton  
     [pem\_pos=0, final\_scores=0.8764927,  
 final\_local\_scores=0.8695995, pem=0.0, similarity=30.040297]  
 20: 5043544 Democratic\_Party\_(United\_States)  
     [pem\_pos=0, final\_scores=0.4683817,  
 final\_local\_scores=0.45329535, pem=-0.076093495,  
 similarity=18.03076]  
 21: 32070 Republican\_Party\_(United\_States)  
     [pem\_pos=0, final\_scores=0.39823395,  
 final\_local\_scores=0.37473494, pem=-0.1310191,  
 similarity=15.985053]  
 22: 135976 Asia-Pacific\_Economic\_Cooperation  
     [pem\_pos=0, final\_scores=0.23273899,  
 final\_local\_scores=0.2537982, pem=0.0, similarity=11.663557]  
 23: 3434750 United\_States  
     [pem\_pos=0, final\_scores=0.4082026,  
 final\_local\_scores=0.3812331, pem=-0.073457584,  
 similarity=15.865943]

Sample: b'PRI20001031\_2000\_1824'

From [1 1 BBC News 1] in [2 2 London 2] , I am Gregor Craigy  
 for The World . A [1 Singapore Airlines 1] 747 flying from [3  
 3 Taiwan 3] to [4 4 Los Angeles 4] has crashed . Flight SQ006  
 was carrying 179 passengers and crew when it crashed in  
 flames shortly after take off . At least 65 people are known  
 to have been killed and around 30 passengers remain  
 unaccounted for . David Chezan reports . Survivors from the [2  
 Singapore Airlines 2] flight described how the plane was  
 taxing down the runway when there was a loud noise and the  
 aircraft split in two . An [3 American 3] passenger said he  
 had almost decided not to board the plane because of what he  
 called unbelievable wind and rain . Survivors described how  
 the tail section turned over several times , leaving  
 passengers hanging by their safety belts , some were quickly  
 engulfed by flames . One survivor described saying a man , as

he put it , `` lit up like a torch . '' `` Flames shot up right next to me and there was jet fuel all over the place , '' he said . David Chezan , [4 BBC News 4] .

1: 1139893 BBC\_News

2: 17867 London

3: 25734 Taiwan

4: 18110 Los\_Angeles

1: 1139893 BBC\_News

[pem\_pos=0, final\_scores=0.17339374, final\_local\_scores=0.16421449, pem=0.0, similarity=8.990201]

2: 17867 London

[pem\_pos=0, final\_scores=0.63263386, final\_local\_scores=0.63868386, pem=-0.09086737, similarity=23.643463]

3: 25734 Taiwan

[pem\_pos=0, final\_scores=0.38166863, final\_local\_scores=0.40370613, pem=-0.077886656, similarity=16.560669]

4: 18110 Los\_Angeles

[pem\_pos=0, final\_scores=0.5179978, final\_local\_scores=0.5308947, pem=-0.100652285, similarity=20.480032]

1: 161591 Singapore\_Airlines

[pem\_pos=0, final\_scores=0.26653856, final\_local\_scores=0.2864743, pem=0.0, similarity=12.638677]

2: 161591 Singapore\_Airlines

[pem\_pos=0, final\_scores=0.26924247, final\_local\_scores=0.28942785, pem=0.0, similarity=12.726817]

3: 3434750 United\_States

[pem\_pos=0, final\_scores=0.37316918, final\_local\_scores=0.35957327, pem=-0.12994486, similarity=15.526756]

4: 1139893 BBC\_News

[pem\_pos=0, final\_scores=0.074922785, final\_local\_scores=0.05664994, pem=0.0, similarity=5.78026]

Sample: b'NYT20001106\_1705\_0187'

-LRB- EDS : This early version will be updated from [1 Davenport 1] , [2 Iowa 2] . -RRB- &QL; -LRB- ART ADV : Photos NYT11 and 25 are being sent to [3 NYT 3] photo clients . Nonsubscribers can purchase one-time rights by calling 888-603-1036 or 888-346-9867 . -RRB- &QL; &UR; By FRANK BRUNI &QC; &LR; &QL; &UR; c. 2000 N.Y. Times News Service &QC; &LR; &QL; [1 1 CHATTANOOGA 1] , [4 1 Tenn. 4] . \_ Nearly 17 months after he first issued his call for a `` fresh start after a season of cynicism , '' Gov. [5 George W. Bush 5] ended his quest for the presidency Monday on a nearly identical note , pledging to purge [2 2 Washington 2] of what he cast as a crippling discord . The course that [6 Bush 6] charted over his final hours on the campaign trail , plotted with as much concern for symbolism as for raw electoral arithmetic , left no doubt about where he placed the blame for that

atmosphere . His first major rally was in [7 Tennessee 7] and his last was planned for [3 3 Arkansas 3] , the home states , respectively , of Vice President [8 Al Gore 8] and President [9 Clinton 9] . The [10 Texas 10] governor 's words made equally clear that he saw himself as the country 's best hope for bridging ideological divides , healing partisan wounds and making sure that [11 Americans 11] could gaze upon the [4 4 White House 4] with unfettered respect . `` There 's a better day ahead if we have a leader who 's willing to unite the country , to bring people together , to elevate the people 's business above everyday politics , '' [12 Bush 12] said at a rally inside a hangar at the airport here , where supporters ' chants evoked the nearness of the moment of decision . `` One more day ! '' they roared . `` One more day ! '' What the nation needed , [13 Bush 13] told them , was `` a president who can unite this nation , a president who puts aside the endless partisan bickering that seems to gridlock our nation 's capitol , a president who puts the people first , a president who lifts this nation 's spirits . `` I 'll be that president , '' he added , a sentence that was equal parts promise and prediction . It underscored the optimism that [14 Bush 14] has , from the very beginning , sought to project to voters . His schedule over the course of a 16-hour day was also an expression of that confidence . In between [15 Tennessee 15] and [16 Arkansas 16] , his campaign plane touched down in [5 5 Wisconsin 5] and [6 6 Iowa 6] , and while any one of those four states could wind up providing the key to a [17 Bush 17] victory , none have as many electoral votes as such battlegrounds as [7 7 Michigan 7] and [8 8 Pennsylvania 8] . But they are genuine tossups that were initially predicted to be much more hospitable to [18 Gore 18] . By visiting them Monday , [19 Bush 19] emphasized the breadth of his appeal and the sometimes surprising vulnerability of the vice president . He did not let his assertion of strength and confidence distract him from the important business of beseeching supporters to grant him victory . With just 24 hours before [20 Election Day 20] , [21 Bush 21] in many ways stripped campaigning to its essence : coaxing and cajoling , pleading and persuading . `` I want you to understand that I ca n't win without you , '' [22 Bush 22] told a crowd of more than 1,000 people at a convention center in [9 9 Green Bay 9] , [23 Wis. 23] `` I hope you redouble your efforts to make sure people get out to the polls . '' Minutes later , he added : `` When you go out there and tell the folks where we stand on the issues and where we stand when it comes to bringing people together to get things done , and you tell them that the core of this campaign is the inherent trust in the [24 American 24] people , I believe it does n't matter what political party they 're in . They 're going to come our way . '' In [25 Chattanooga 25] , [26 Bush 26] repeatedly needled [27 Gore 27] , which was much of the point of going to [28 Tennessee 28] ,

where [29 Gore 29] is not at all certain to prevail on Tuesday . The [30 Texas 30] governor claimed that [31 Gore 31] 's decades of experience in [32 Washington 32] had estranged him from the rest of the country by making him too trusting of federal government and too fond of federal spending . `` My opponent vows to carry his home state , '' [33 Bush 33] said . `` He may win [34 Washington 34] , [10 10 D.C. 10] , but he 's not going to win [35 Tennessee 35] . `` He forgot his roots , '' [36 Bush 36] added . `` He forgot where he 's from . He trusts Washington . We trust the people . ''

1: 102797 Chattanooga, Tennessee

2: 108956 Washington, D.C.

3: 1930 Arkansas

4: 33057 White House

5: 33127 Wisconsin

6: 26810748 Iowa

7: 18859 Michigan

8: 23332 Pennsylvania

9: 217033 Green Bay, Wisconsin

10: 108956 Washington, D.C.

1: 30395 Tennessee

[winner: 30395 Tennessee, pem\_pos=0,  
final\_scores=0.3963527, final\_local\_scores=0.40095085,  
pem=0.0, similarity=16.054886]

gt: 30395 Tennessee, pem\_pos=0, final\_scores=0.3963527,  
final\_local\_scores=0.40095085, pem=0.0,  
similarity=16.054886 ]

[winner: 4721653 Zyphe language, pem\_pos=12,  
final\_scores=-1.3504664, final\_local\_scores=-1.3113806,  
pem=-4.4188404, similarity=-11.014157]

RECALL MISS: 30395 Tennessee]

1: 102797 Chattanooga, Tennessee

[pem\_pos=0, final\_scores=0.39962387,  
final\_local\_scores=0.44035923, pem=-0.16817571,  
similarity=18.145475]

2: 108956 Washington, D.C.

[pem\_pos=0, final\_scores=0.33345705,  
final\_local\_scores=0.29051661, pem=-1.7496481,  
similarity=22.274166]

3: 1930 Arkansas

[pem\_pos=0, final\_scores=0.5043716,  
final\_local\_scores=0.5053317, pem=-0.32007957,  
similarity=20.910461]

4: 33057 White House

[pem\_pos=0, final\_scores=0.50881505,  
final\_local\_scores=0.4764877, pem=-0.0056634876,  
similarity=18.339855]

5: 33127 Wisconsin

[pem\_pos=0, final\_scores=0.49140418,  
final\_local\_scores=0.5132793, pem=-0.21268336,  
similarity=20.563597]

6: 26810748 Iowa



[pem\_pos=0, final\_scores=0.16093479,  
final\_local\_scores=0.13943583, pem=-0.3546882,  
similarity=10.1796055]  
7: 18859 Michigan  
[pem\_pos=0, final\_scores=0.6002708,  
final\_local\_scores=0.6236512, pem=-0.24788691,  
similarity=24.048756]  
8: 23332 Pennsylvania  
[pem\_pos=0, final\_scores=0.4688937,  
final\_local\_scores=0.47907937, pem=-0.11810702,  
similarity=19.02868]  
9: 217033 Green\_Bay,\_Wisconsin  
[pem\_pos=0, final\_scores=0.3172453,  
final\_local\_scores=0.3640536, pem=-0.59786385,  
similarity=18.205076]  
10: 108956 Washington,\_D.C.  
[pem\_pos=0, final\_scores=0.09786455,  
final\_local\_scores=0.03316757, pem=-0.15729709,  
similarity=5.934905]  
1: 8599 Davenport,\_Iowa  
[pem\_pos=0, final\_scores=0.10300873,  
final\_local\_scores=0.12874293, pem=-1.1472282,  
similarity=14.1704645]  
2: 26810748 Iowa  
[pem\_pos=0, final\_scores=0.11060445,  
final\_local\_scores=0.08445759, pem=-0.3546882,  
similarity=8.538945]  
3: 30680 The\_New\_York\_Times  
[pem\_pos=0, final\_scores=0.06049429,  
final\_local\_scores=0.05691944, pem=0.0, similarity=5.7883024]  
4: 30395 Tennessee  
[pem\_pos=0, final\_scores=0.3963527,  
final\_local\_scores=0.40095085, pem=0.0, similarity=16.054886]  
5: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.23068182,  
final\_local\_scores=0.14583358, pem=0.0, similarity=8.441678]  
6: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.3783744,  
final\_local\_scores=0.3078669, pem=0.0, similarity=13.277076]  
7: 30395 Tennessee  
[pem\_pos=0, final\_scores=0.38332635,  
final\_local\_scores=0.38672155, pem=-0.27977526,  
similarity=17.151716]  
8: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.41526586,  
final\_local\_scores=0.38021398, pem=0.0, similarity=15.436057]  
9: 3356 Bill\_Clinton  
[pem\_pos=0, final\_scores=0.5301667,  
final\_local\_scores=0.48102885, pem=-1.920079,  
similarity=28.886263]  
10: 29810 Texas  
[pem\_pos=0, final\_scores=0.7646516,

final\_local\_scores=0.79878956, pem=-0.16191302,  
similarity=28.807695]  
11: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.4281451,  
final\_local\_scores=0.42415583, pem=-0.5581965,  
similarity=19.782928]  
12: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.3562618,  
final\_local\_scores=0.28371218, pem=0.0, similarity=12.556251]  
13: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.49201804,  
final\_local\_scores=0.43319935, pem=0.0, similarity=17.017246]  
14: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.40154463,  
final\_local\_scores=0.33317682, pem=0.0, similarity=14.032373]  
15: 30395 Tennessee  
[pem\_pos=0, final\_scores=0.56431633,  
final\_local\_scores=0.5844256, pem=-0.27977526,  
similarity=23.051601]  
16: 1930 Arkansas  
[pem\_pos=0, final\_scores=0.51667863,  
final\_local\_scores=0.5187752, pem=-0.32007957,  
similarity=21.311644]  
17: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.33248967,  
final\_local\_scores=0.25774476, pem=0.0, similarity=11.781331]  
18: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.70123684,  
final\_local\_scores=0.6942636, pem=0.0, similarity=24.807922]  
19: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.45452684,  
final\_local\_scores=0.39137208, pem=0.0, similarity=15.769037]  
20: 1050340 Election\_Day\_(United\_States)  
[pem\_pos=0, final\_scores=0.108695015,  
final\_local\_scores=0.14681658, pem=-0.07676029,  
similarity=8.888448]  
21: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.41467482,  
final\_local\_scores=0.34751955, pem=0.0, similarity=14.460389]  
22: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.47215134,  
final\_local\_scores=0.41086608, pem=0.0, similarity=16.350777]  
23: 33127 Wisconsin  
[pem\_pos=0, final\_scores=0.42750496,  
final\_local\_scores=0.44347912, pem=0.0, similarity=17.324015]  
24: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.52403456,  
final\_local\_scores=0.5289005, pem=-0.12994486,  
similarity=20.57982]  
25: 102797 Chattanooga,\_Tennessee  
[pem\_pos=0, final\_scores=0.3917786,  
final\_local\_scores=0.4317894, pem=-0.16817571,

similarity=17.889734]  
 26: 3414021 George\_W.\_Bush  
     [pem\_pos=0, final\_scores=0.41979414,  
 final\_local\_scores=0.35343185, pem=0.0, similarity=14.636825]  
 27: 5042706 Al\_Gore  
     [pem\_pos=0, final\_scores=0.74785084,  
 final\_local\_scores=0.7451823, pem=0.0, similarity=26.327438]  
 28: 30395 Tennessee  
     [pem\_pos=0, final\_scores=0.4632553,  
 final\_local\_scores=0.47403175, pem=-0.27977526,  
 similarity=19.757227]  
 29: 5042706 Al\_Gore  
     [pem\_pos=0, final\_scores=0.72322184,  
 final\_local\_scores=0.7182789, pem=0.0, similarity=25.524588]  
 30: 29810 Texas  
     [pem\_pos=0, final\_scores=0.7506667,  
 final\_local\_scores=0.78351307, pem=-0.16191302,  
 similarity=28.351814]  
 31: 5042706 Al\_Gore  
     [pem\_pos=0, final\_scores=0.8186639,  
 final\_local\_scores=0.82253486, pem=0.0, similarity=28.635794]  
 32: 108956 Washington,\_D.C.  
     [pem\_pos=0, final\_scores=0.45814002,  
 final\_local\_scores=0.42671376, pem=-1.7496481,  
 similarity=26.33856]  
 33: 3414021 George\_W.\_Bush  
     [pem\_pos=0, final\_scores=0.4493882,  
 final\_local\_scores=0.38663274, pem=0.0, similarity=15.6276045]  
 34: 108956 Washington,\_D.C.  
     [pem\_pos=0, final\_scores=0.043105975,  
 final\_local\_scores=-0.028240561, pem=-1.7496481,  
 similarity=12.761816]  
 35: 30395 Tennessee  
     [pem\_pos=0, final\_scores=0.3880399,  
 final\_local\_scores=0.39187032, pem=-0.27977526,  
 similarity=17.305367]  
 36: 3414021 George\_W.\_Bush  
     [pem\_pos=0, final\_scores=0.52018076,  
 final\_local\_scores=0.46432227, pem=0.0, similarity=17.946016]

Sample: b'NYT20001109\_1946\_0315'

&UR; By JAMES DAO &LR; &QC; &QL; &UR; c. 2000 N.Y. Times  
 News Service &LR; &QC; &QL; [1 1 WASHINGTON 1] \_ [2 2 Ralph Na  
 der 2] may not be feeling any regrets about his [1 Green  
 Party 1] presidential campaign , but some visitors to the  
 chat room on Nader2000.org , [2 Nader 2] 's official campaign  
 website , sounded as if a bad hangover had come crashing down  
 on them . There was this entry , for instance , posted  
 Thursday morning by a writer named `` wreckz '' : `` I ,  
 too , feel I should apologize to [3 3 America 3] . I was  
 duped . [3 Ralph Nader 3] 's enjoyment of all of this is an  
 indication of a man more concerned with vindictive politics

than the future of our country and planet . I am no longer a member of the [4 Green Party 4] . Instead , I 'm a member of the green movement \_ a movement without [5 Ralph Nader 5] .

'' Or this one , also posted Thursday morning , from `` Noni '' : `` I saw an interview with [6 Ralph Nader 6] today and he seemed to be enjoying what has happened ! I 've never seen him smile so broadly . I regret my support . [7 Nader 7] , you 're a well-intending man and I agree with most of your positions . But the symbolic run for the [4 4 White House 4] has destroyed our Party and our cause . Very bad things are in our future . You should not be smiling . And we , the misled , should feel shame for what we 've done . '' The remarks expressed fears that votes for [8 Nader 8] may have put Gov. [9 George W. Bush 9] in a position to defeat Vice President [10 Al Gore 10] . In [5 5 Florida 5] , where [11 Gore 11] trailed [12 Bush 12] by only a few hundred votes Thursday , a shift of just a fraction of [13 Nader 13] 's 95,000 votes could have given the state , and the [14 White House 14] , to [15 Gore 15] , the most likely major-party alternative for [16 Nader 16] 's supporters . John Ruth , for example , wrote on the Nader2000 site Thursday morning that he voted for [17 Nader 17] because he thought they `` shared a common goal '' but that he was now regretting it . `` [18 Gore 18] -LRB- despite what you have said -RRB- is NOT an environmental clone of [19 G.W. Bush 19] . Because of your actions and statements prior to Nov. 7th , we are on the brink of a [20 Bush 20] presidency . No money , no support , no respect for you , [21 Ralph 21] , ever again ! '' And in [6 6 Oregon 6] , a state that was still too close to call Thursday afternoon , [22 Nader 22] won about 5 percent of the vote \_ enough to cause some [23 Nader 23] voters to feel morning-after qualms . Julie Quastler , 28 , who voted for [24 Nader 24] in [7 7 Portland 7] , said on Wednesday that she was feeling `` a little alarmed and disappointed '' that [25 Bush 25] seemed on the brink of winning the presidency \_ suggesting that she did not accept [26 Nader 26] 's assertion that the two parties were equally bad . `` I 'm just so sad , '' she said . `` [27 Bush 27] is probably going to be our president and [28 Nader 28] did n't get his 5 percent . It seems like a lose-lose end result in some ways . '' Winning 5 percent of the vote nationwide would have qualified the [29 Green Party 29] for millions of dollars in federal campaign funds in 2004 . But Quastler , the development director for a nonprofit cafe , said she did not feel guilty about her vote . `` I voted for [30 Nader 30] because he was most aligned with my values , '' she said . Of course , there were also countless verbal high-fives among [31 Nader 31] supporters Thursday who felt the [32 Green Party 32] had sent a bracing wake-up call to the two major parties . `` Stand tall and proud , fellow [33 Nader 33] supporters , '' said one message on the Nader2000 site signed GO LOOK ELSEWHERE FOR SCAPEGOATS . `` We fought the good fight for what we

believed in . This does not make us villains . '' Such sentiments seemed particularly strong in Florida . Dorothy Byrne , a state coordinator for the Florida 1 Green Party , said she had been inundated with angry phone calls and e-mails from [34 Democrats 34] , but has yet to receive one regretful note from a [35 Nader 35] voter . `` I can see positives for the Greens with either [36 Bush 36] or [37 Gore 37] being president , '' she said . `` If [38 Gore 38] was in , we would n't have to worry about women 's reproductive rights . But his record on the environment is so abysmal I do n't see how [39 Bush 39] can be any worse . '' Barbara Lange , the [40 Everglades 40] chairwoman for the Florida chapter of the [8 8 Sierra Club 8] , said she and other environmentalists voted for [41 Nader 41] because [42 Gore 42] had not taken a stand against the development of a commercial airport near the [43 Everglades 43] . `` I voted my conscience , '' she said . `` And I 'm okay with that . '' Patricia Newell , an organizer for [44 Nader 44] at the [9 9 University of Florida 9] in [10 10 Gainesville 10] , said that [45 Nader 45] had won far fewer votes in Florida than his supporters had expected \_ due , she said , to `` scare tactics '' by the [46 Democrats 46] . `` Anyone who voted with [47 Nader 47] on Tuesday knew exactly what they were doing , '' she said . `` On the contrary to feeling guilty , they feel very good about sticking with the candidate that they felt was best . ''

1: 108956 Washington,\_D.C.

2: 25671 Ralph\_Nader

3: 3434750 United\_States

4: 33057 White\_House

5: 18933066 Florida

6: 26811621 Oregon

7: 23503 Portland,\_Oregon

8: 242098 Sierra\_Club

9: 60611 University\_of\_Florida

10: 108959 Gainesville,\_Florida

1: 20478485 Green\_Party\_of\_Florida

[winner: 30696 Green\_Party\_of\_the\_United\_States,  
pem\_pos=0, final\_scores=-0.055461183,  
final\_local\_scores=-0.111825496, pem=-1.1386364,  
similarity=6.944699]

RECALL MISS: 20478485 Green\_Party\_of\_Florida]

1: 108956 Washington,\_D.C.

[pem\_pos=0, final\_scores=0.23294918,  
final\_local\_scores=0.22610897, pem=-1.7496481,  
similarity=20.352114]

2: 25671 Ralph\_Nader

[pem\_pos=0, final\_scores=0.45520514,  
final\_local\_scores=0.3874532, pem=0.0, similarity=15.652089]

3: 3434750 United\_States

[pem\_pos=0, final\_scores=0.5661629,  
final\_local\_scores=0.61331505, pem=-0.78483206,

similarity=26.660295]  
4: 33057 White\_House  
[pem\_pos=0, final\_scores=0.41999525,  
final\_local\_scores=0.41386795, pem=-0.0056634876,  
similarity=16.471157]  
5: 18933066 Florida  
[pem\_pos=0, final\_scores=0.070871934,  
final\_local\_scores=0.100621134, pem=-0.15405534,  
similarity=7.9302254]  
6: 26811621 Oregon  
[pem\_pos=0, final\_scores=0.357866,  
final\_local\_scores=0.43335384, pem=-0.28686622,  
similarity=18.581879]  
7: 23503 Portland,\_Oregon  
[pem\_pos=0, final\_scores=0.30552906,  
final\_local\_scores=0.35404158, pem=-0.66320467,  
similarity=18.261631]  
8: 242098 Sierra\_Club  
[pem\_pos=0, final\_scores=0.2765203,  
final\_local\_scores=0.27514166, pem=0.0, similarity=12.300488]  
9: 60611 University\_of\_Florida  
[pem\_pos=0, final\_scores=0.22818552,  
final\_local\_scores=0.28540143, pem=0.0, similarity=12.606661]  
10: 108959 Gainesville,\_Florida  
[pem\_pos=0, final\_scores=0.08411397,  
final\_local\_scores=0.13200161, pem=-0.6585603,  
similarity=11.610258]  
1: 30696 Green\_Party\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.1933146,  
final\_local\_scores=0.16169602, pem=-1.1386364,  
similarity=15.107126]  
2: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.46323198,  
final\_local\_scores=0.3966843, pem=0.0, similarity=15.927563]  
3: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.5447711,  
final\_local\_scores=0.4852904, pem=0.0, similarity=18.571747]  
4: 30696 Green\_Party\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.05959101,  
final\_local\_scores=0.015623331, pem=-1.1386364,  
similarity=10.748026]  
5: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.44286674,  
final\_local\_scores=0.3739754, pem=0.0, similarity=15.249886]  
6: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.5128259,  
final\_local\_scores=0.45039517, pem=0.0, similarity=17.530403]  
7: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.4859349,  
final\_local\_scores=0.42148376, pem=0.0, similarity=16.66763]  
8: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.40660995,

final\_local\_scores=0.33483326, pem=0.0, similarity=14.081805]  
9: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.21147917,  
final\_local\_scores=0.15810663, pem=0.0, similarity=8.80793]  
10: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.38199663,  
final\_local\_scores=0.33138287, pem=0.0, similarity=13.978839]  
11: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.64394826,  
final\_local\_scores=0.6185581, pem=0.0, similarity=22.548721]  
12: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.32053477,  
final\_local\_scores=0.27855963, pem=0.0, similarity=12.402488]  
13: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.3642711,  
final\_local\_scores=0.28858453, pem=0.0, similarity=12.701651]  
14: 33057 White\_House  
[pem\_pos=0, final\_scores=0.4578783,  
final\_local\_scores=0.4552495, pem=-0.0056634876,  
similarity=17.706064]  
15: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.66270846,  
final\_local\_scores=0.6390508, pem=0.0, similarity=23.160263]  
16: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.40890795,  
final\_local\_scores=0.33734354, pem=0.0, similarity=14.156717]  
17: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.40361577,  
final\_local\_scores=0.33156258, pem=0.0, similarity=13.984201]  
18: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.78463656,  
final\_local\_scores=0.7722386, pem=0.0, similarity=27.134853]  
19: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.33073765,  
final\_local\_scores=0.2883784, pem=0.0, similarity=12.695499]  
20: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.20494331,  
final\_local\_scores=0.15096721, pem=0.0, similarity=8.594875]  
21: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.18482946,  
final\_local\_scores=0.092108876, pem=0.0, similarity=6.8384256]  
22: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.40707988,  
final\_local\_scores=0.33534658, pem=0.0, similarity=14.097124]  
23: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.40472227,  
final\_local\_scores=0.3327713, pem=0.0, similarity=14.020272]  
24: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.36445796,  
final\_local\_scores=0.28878865, pem=0.0, similarity=12.707742]  
25: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.41733813,

final\_local\_scores=0.38430262, pem=0.0, similarity=15.55807]  
26: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.3862725,  
final\_local\_scores=0.31261766, pem=0.0, similarity=13.418847]  
27: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.35031605,  
final\_local\_scores=0.31204873, pem=0.0, similarity=13.40187]  
28: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.4090696,  
final\_local\_scores=0.33752012, pem=0.0, similarity=14.161986]  
29: 30696 Green\_Party\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.1917731,  
final\_local\_scores=0.16001219, pem=-1.1386364,  
similarity=15.056878]  
30: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.4614399,  
final\_local\_scores=0.3947267, pem=0.0, similarity=15.869144]  
31: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.39332372,  
final\_local\_scores=0.3203201, pem=0.0, similarity=13.648704]  
32: 30696 Green\_Party\_of\_the\_United\_States  
[pem\_pos=0, final\_scores=0.1406353,  
final\_local\_scores=0.104151905, pem=-1.1386364,  
similarity=13.389895]  
33: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.4147349,  
final\_local\_scores=0.3437086, pem=0.0, similarity=14.346663]  
34: 5043544 Democratic\_Party\_(United\_States)  
[pem\_pos=0, final\_scores=0.4497102,  
final\_local\_scores=0.4486605, pem=-0.18007085,  
similarity=18.457891]  
35: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.40161425,  
final\_local\_scores=0.32937628, pem=0.0, similarity=13.918958]  
36: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.28636885,  
final\_local\_scores=0.24123856, pem=0.0, similarity=11.288753]  
37: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.6091823,  
final\_local\_scores=0.58058155, pem=0.0, similarity=21.415424]  
38: 5042706 Al\_Gore  
[pem\_pos=0, final\_scores=0.77687293,  
final\_local\_scores=0.76375806, pem=0.0, similarity=26.881777]  
39: 3414021 George\_W.\_Bush  
[pem\_pos=0, final\_scores=0.48046488,  
final\_local\_scores=0.45421654, pem=0.0, similarity=17.64444]  
40: 87003 Everglades  
[pem\_pos=0, final\_scores=0.058179483,  
final\_local\_scores=0.13054189, pem=0.0, similarity=7.9853435]  
41: 25671 Ralph\_Nader  
[pem\_pos=0, final\_scores=0.38997674,  
final\_local\_scores=0.31666404, pem=0.0, similarity=13.539599]



42: 5042706 Al\_Gore  
 [pem\_pos=0, final\_scores=0.7076634,  
 final\_local\_scores=0.6881572, pem=0.0, similarity=24.625696]  
 43: 87003 Everglades  
 [pem\_pos=0, final\_scores=0.12770377,  
 final\_local\_scores=0.20648661, pem=0.0, similarity=10.251685]  
 44: 25671 Ralph\_Nader  
 [pem\_pos=0, final\_scores=0.38249797,  
 final\_local\_scores=0.30849466, pem=0.0, similarity=13.295809]  
 45: 25671 Ralph\_Nader  
 [pem\_pos=0, final\_scores=0.3228693,  
 final\_local\_scores=0.24335936, pem=0.0, similarity=11.352041]  
 46: 5043544 Democratic\_Party\_(United\_States)  
 [pem\_pos=0, final\_scores=0.48097914,  
 final\_local\_scores=0.4828171, pem=-0.18007085,  
 similarity=19.477192]  
 47: 25671 Ralph\_Nader  
 [pem\_pos=0, final\_scores=0.43784213,  
 final\_local\_scores=0.3689497, pem=0.0, similarity=15.099909]

Sample: b'NYT20001123\_1511\_0062'

WORLD ' &LR; &QL; &UR; By STUART ELLIOTT &LR; &QC; &QL; &UR;  
 c. 2000 N.Y. Times News Service &LR; &QC; &QL; It 's time  
 again to ask 20 questions about advertising , marketing and  
 the media . \_ Will [1 Americans 1] ever refer to the day  
 before [2 Thanksgiving 2] as `` Thanksgiving Wednesday , ''  
 as [1 1 ABC 1] did repeatedly this week in commercials  
 promoting its Wednesday sitcoms `` The Drew Carey Show '' and  
 `` [3 Spin City 3] '' ? \_ If a [4 Hollywood 4] studio makes a  
 movie about the Florida election standoff , would the title  
 be `` It 's a Chad , Chad , Chad , Chad World '' ? \_ How many  
 readers of [5 The Wall Street Journal 5] noticed that an  
 article on Wednesday carrying the headline `` Genetically  
 Altered Protein Is Found in Still More Corn '' ran adjacent  
 to an advertisement for a golf tournament on [2 2 Fox Sports 2  
 ] sponsored by [3 3 Sun Microsystems 3] that featured a  
 fanciful photograph of a gigantic ear of corn as big as a [6  
 Buick 6] ? \_ Does n't a new television commercial for [4 4 Fi  
 delity Investments 4] , centered on a telephone conversation  
 between a female employee and a male customer , resemble a [7  
 Citibank 7] campaign from the late 1980s that was so  
 brilliantly parodied as the `` First Citiwide Change Bank ''  
 commercials on `` [8 Saturday Night Live 8] '' ? \_ Did  
 shoppers at the [9 New York 9] supermarket chain Gristede 's  
 rush to take advantage of a recent sale on the `` BIG 150-  
 count box '' of Scotties facial tissues , or did they recall  
 that Scotties has usually come in a 175-count box ? \_ Can 1 Oc  
 cean Spray Cranberries get a refund on a full-page ad that ran  
 in newspapers on Nov. 8 and carried the headline `` Now That  
 the Election Is Over , You Can Focus on Some Real News '' ? \_  
 Was it a coincidence that an ad for the Carnation Coffee-mate  
 line of coffee creamers in the recent issue of 2 In Style Ent

entaining 2000 magazine ran opposite an article about `` great trends '' in party-giving that began by describing carnations as `` the chic flower of the moment '' ? \_ Should an [10 English 10] teacher have been consulted before the Tropicana division of [5 5 PepsiCo 5] produced posters that described its Pure Premium orange juice as being `` as packed with Vitamin C as a bus on a rainy day '' ? \_ Will [6 6 Jack Nicholson 6] , [7 7 Helen Hunt 7] , [8 8 Greg Kinnear 8] and 3 James L. Brooks complain about an ad from 4 Campbell Soup Co . for Campbell 's Cream of Mushroom soup that carried the slogan `` As m 'm ! m 'm ! good ! as it gets '' ? \_ Now that the [9 9 Roundabout Theater Company 9] is calling its new facility in [10 10 Times Square 10] the [11 American Airlines Theater 11] , will the ushers give out peanuts instead of [12 Playbills 12] ? \_ Why did the 5 Reader 's Digest Association go to the trouble of airbrushing the identifying information from a label on a pair of jeans worn by a man in a photograph featured in an ad for 6 Reader 's Digest magazine , but leave the red tab on the back pocket that tells everyone the pants are 7 Levi 's ? \_ Do television commercials for prescription drugs disclose that side effects in tests were no worse `` than a sugar pill '' because research found that consumers were n't familiar with the word `` placebo '' ? \_ What was in the zeitgeist that led two disparate advertisers \_ 8 Empire Blue Cross Blue Shield and the 9 MSN network of [11 11 Microsoft 11] \_ to introduce about the same time campaigns centered on butterflies ? \_ And what was in the zeitgeist that led two automotive marketers \_ [12 12 Land Rover 12] and [13 13 Volvo 13] \_ to introduce about the same time commercials centered on drivers rescuing animals on rainy roads ? \_ Are folks old enough to remember the heyday of the radio and television personality [13 Arthur Godfrey 13] reminded of his longtime greeting , `` How are ya ? , '' every time they hear the name of the new [14 Lucent Technologies 14] spinoff company , Avaya Communication ? \_ How appealing was a recent offer made by Sprint Communications Co . to give buyers of a Sprint PCS telephone and a calling plan a free `` pre-owned phone '' ? \_ Was it an inside joke that the name of the character played by 10 John Corbett on `` [15 Sex and the City 15] , '' Aidan Shaw , was almost identical to the name of a longtime star of gay sex films , Aiden Shaw ? \_ How many tickets did the [14 14 Houston Grand Opera 14] sell after mailing out a promotional brochure featuring photographs of cattle with headlines like `` It 's udderly fantastic entertainment ! '' and `` Great opera . No bull . '' ? \_ If [15 15 CBS 15] shows the first episode of `` Survivor II '' after its broadcast of [16 Super Bowl XXXV 16] , will the network someday show the first episode of `` Survivor XXXV '' after its broadcast of Super Bowl LXVIII ? \_ Will cheerful chorus members dressed in V-neck sweaters interrupt their serenades to online shopping long enough to tell an advertising columnist , `` You ask a lot of questions for someone from [16 16 Brooklyn 16] '' ?

1: 62027 American\_Broadcasting\_Company  
2: 1699539 Fox\_Sports\_(United\_States)  
3: 26980 Sun\_Microsystems  
4: 439738 Fidelity\_Investments  
5: 1213210 PepsiCo  
6: 18617829 Jack\_Nicholson  
7: 44569 Helen\_Hunt  
8: 598459 Greg\_Kinnear  
9: 6418322 Roundabout\_Theatre\_Company  
10: 31389 Times\_Square  
11: 19001 Microsoft  
12: 18724 Land\_Rover  
13: 32412 Volvo  
14: 389898 Houston\_Grand\_Opera  
15: 37653 CBS  
16: 47384 Brooklyn

1: 1136380 Ocean\_Spray\_(cooperative)  
[winner: 7025 Cranberry, pem\_pos=1,  
final\_scores=-0.81836414, final\_local\_scores=-0.77171624,  
pem=-0.29774684, similarity=-17.320646]  
RECALL MISS: 1136380 Ocean\_Spray\_(cooperative)]  
[winner: 18842359 Ocean, pem\_pos=0,  
final\_scores=-0.9573421, final\_local\_scores=-0.9286402,  
pem=-1.0522059, similarity=-17.900703]  
RECALL MISS: 1136380 Ocean\_Spray\_(cooperative)]  
[winner: 21785342 Snow\_spray, pem\_pos=12,  
final\_scores=-1.2347496, final\_local\_scores=-1.1540041,  
pem=-2.7725887, similarity=-15.270302]  
RECALL MISS: 1136380 Ocean\_Spray\_(cooperative)]  
2: 6349753 InStyle  
[winner: 6349753 InStyle, pem\_pos=0,  
final\_scores=-0.86317575, final\_local\_scores=-0.9384386,  
pem=0.0, similarity=-23.915167]  
gt: 6349753 InStyle, pem\_pos=0,  
final\_scores=-0.86317575, final\_local\_scores=-0.9384386,  
pem=0.0, similarity=-23.915167 ]  
[winner: 34548 2000, pem\_pos=0, final\_scores=-0.9375043,  
final\_local\_scores=-0.9069182, pem=-0.8862115,  
similarity=-18.15518]  
RECALL MISS: 6349753 InStyle]  
[winner: 896497 Magazine\_(firearms), pem\_pos=0,  
final\_scores=-1.2395424, final\_local\_scores=-1.2410066,  
pem=-0.62660754, similarity=-29.536808]  
RECALL MISS: 6349753 InStyle]  
[winner: 9262 Entertainment, pem\_pos=0,  
final\_scores=-1.623925, final\_local\_scores=-1.7972462,  
pem=0.0, similarity=-49.543697]  
RECALL MISS: 6349753 InStyle]  
3: 432249 James\_L.\_Brooks  
[winner: 432249 James\_L.\_Brooks, pem\_pos=0,  
final\_scores=-0.057996973, final\_local\_scores=-0.07336374,  
pem=0.0, similarity=1.9003924]

gt: 432249 James\_L.\_Brooks, pem\_pos=0,  
final\_scores=-0.057996973, final\_local\_scores=-0.07336374,  
pem=0.0, similarity=1.9003924 ]  
4: 25125130 Campbell\_Soup\_Company  
[winner: 6744197 Atoy\_Co, pem\_pos=1,  
final\_scores=-0.42584884, final\_local\_scores=-0.3649251,  
pem=-2.1566277, similarity=4.9276967]  
RECALL MISS: 25125130 Campbell\_Soup\_Company]  
[winner: 108150 Campbell,\_California, pem\_pos=0,  
final\_scores=-0.56159353, final\_local\_scores=-0.59059995,  
pem=-2.019329, similarity=-2.5535424]  
RECALL MISS: 25125130 Campbell\_Soup\_Company]  
[winner: 12927133 Soup\_(Apple), pem\_pos=4,  
final\_scores=-1.349763, final\_local\_scores=-1.3051057,  
pem=-2.1241436, similarity=-23.305817]  
gt: 25125130 Campbell\_Soup\_Company, pem\_pos=11,  
final\_scores=-2.8667219, final\_local\_scores=-3.1572788,  
pem=-4.7194905, similarity=-64.46446 ]  
5: 2362549 Reader's\_Digest\_Association  
[winner: 2362549 Reader's\_Digest\_Association, pem\_pos=0,  
final\_scores=-0.23503946, final\_local\_scores=-0.21226907,  
pem=-0.095310144, similarity=-1.7265084]  
gt: 2362549 Reader's\_Digest\_Association, pem\_pos=0,  
final\_scores=-0.23503946, final\_local\_scores=-0.21226907,  
pem=-0.095310144, similarity=-1.7265084 ]  
6: 465434 Reader's\_Digest  
[winner: 465434 Reader's\_Digest, pem\_pos=0,  
final\_scores=-0.2181512, final\_local\_scores=-0.28512043,  
pem=0.0, similarity=-4.4188495]  
gt: 465434 Reader's\_Digest, pem\_pos=0,  
final\_scores=-0.2181512, final\_local\_scores=-0.28512043,  
pem=0.0, similarity=-4.4188495 ]  
[winner: 896497 Magazine\_(firearms), pem\_pos=0,  
final\_scores=-1.28705, final\_local\_scores=-1.2929015,  
pem=-0.62660754, similarity=-31.08546]  
RECALL MISS: 465434 Reader's\_Digest]  
7: 503153 Levi\_Strauss\_&\_Co.  
[winner: 246921 Levi, pem\_pos=0,  
final\_scores=-0.074271604, final\_local\_scores=0.048242196,  
pem=-0.49640274, similarity=8.228871]  
gt: 503153 Levi\_Strauss\_&\_Co., pem\_pos=7,  
final\_scores=-0.41818625, final\_local\_scores=-0.49409327,  
pem=-3.878242, similarity=10.435468 ]  
8: 3462395 Blue\_Cross\_Blue\_Shield\_Association  
[winner: 3462395 Blue\_Cross\_Blue\_Shield\_Association,  
pem\_pos=0, final\_scores=0.02014874,  
final\_local\_scores=0.020621866, pem=-0.080042705,  
similarity=5.140395]  
gt: 3462395 Blue\_Cross\_Blue\_Shield\_Association,  
pem\_pos=0, final\_scores=0.02014874,  
final\_local\_scores=0.020621866, pem=-0.080042705,  
similarity=5.140395 ]

9: 185041 MSN  
[winner: 185041 MSN, pem\_pos=0, final\_scores=-0.13671337, final\_local\_scores=-0.16522878, pem=-0.10645999, similarity=-0.2620974]  
gt: 185041 MSN, pem\_pos=0, final\_scores=-0.13671337, final\_local\_scores=-0.16522878, pem=-0.10645999, similarity=-0.2620974 ]  
[winner: 85024 Network\_layer, pem\_pos=12, final\_scores=-1.5811054, final\_local\_scores=-1.5851848, pem=-3.1110392, similarity=-26.297045]  
RECALL MISS: 185041 MSN]

10: 678304 John\_Corbett\_(actor)  
[winner: 678304 John\_Corbett\_(actor), pem\_pos=0, final\_scores=-0.13024588, final\_local\_scores=-0.13892919, pem=-0.07093271, similarity=0.32953173]  
gt: 678304 John\_Corbett\_(actor), pem\_pos=0, final\_scores=-0.13024588, final\_local\_scores=-0.13892919, pem=-0.07093271, similarity=0.32953173 ]

1: 62027 American\_Broadcasting\_Company  
[pem\_pos=0, final\_scores=0.47764963, final\_local\_scores=0.45682538, pem=-0.3336378, similarity=19.536667]

2: 1699539 Fox\_Sports\_(United\_States)  
[pem\_pos=0, final\_scores=0.18839781, final\_local\_scores=0.19974375, pem=-0.1659384, similarity=10.952864]

3: 26980 Sun\_Microsystems  
[pem\_pos=0, final\_scores=0.22898178, final\_local\_scores=0.28343466, pem=0.0, similarity=12.547969]

4: 439738 Fidelity\_Investments  
[pem\_pos=0, final\_scores=0.26396173, final\_local\_scores=0.29337773, pem=0.0, similarity=12.844689]

5: 1213210 PepsiCo  
[pem\_pos=0, final\_scores=0.04012783, final\_local\_scores=0.019183457, pem=0.0, similarity=4.6621857]

6: 18617829 Jack\_Nicholson  
[pem\_pos=0, final\_scores=0.13974993, final\_local\_scores=0.1368959, pem=0.0, similarity=8.174959]

7: 44569 Helen\_Hunt  
[pem\_pos=0, final\_scores=0.17898087, final\_local\_scores=0.17961419, pem=0.0, similarity=9.449759]

8: 598459 Greg\_Kinnear  
[pem\_pos=0, final\_scores=0.23808907, final\_local\_scores=0.27286142, pem=0.0, similarity=12.232442]

9: 6418322 Roundabout\_Theatre\_Company  
[pem\_pos=0, final\_scores=0.13752244, final\_local\_scores=0.15304878, pem=0.0, similarity=8.656994]

10: 31389 Times\_Square  
[pem\_pos=0, final\_scores=0.3401912, final\_local\_scores=0.33299044, pem=0.0, similarity=14.026812]

11: 19001 Microsoft  
[pem\_pos=0, final\_scores=0.19025163,

final\_local\_scores=0.20085403, pem=-0.006739087,  
similarity=10.120246]  
12: 18724 Land\_Rover  
[pem\_pos=0, final\_scores=0.14166467,  
final\_local\_scores=0.1782045, pem=0.0, similarity=9.407691]  
13: 32412 Volvo  
[pem\_pos=0, final\_scores=0.11578627,  
final\_local\_scores=0.13620391, pem=-0.15991406,  
similarity=9.023947]  
14: 389898 Houston\_Grand\_Opera  
[pem\_pos=0, final\_scores=0.26617038,  
final\_local\_scores=0.31173506, pem=0.0, similarity=13.392509]  
15: 37653 CBS  
[pem\_pos=0, final\_scores=0.43104684,  
final\_local\_scores=0.40305585, pem=-0.10726234,  
similarity=16.701012]  
16: 47384 Brooklyn  
[pem\_pos=0, final\_scores=0.30838537,  
final\_local\_scores=0.30188748, pem=-0.15890065,  
similarity=13.962764]  
1: 3434750 United\_States  
[pem\_pos=0, final\_scores=0.32638705,  
final\_local\_scores=0.33176172, pem=-0.5581965,  
similarity=17.025703]  
2: 8948452 Thanksgiving  
[pem\_pos=0, final\_scores=0.15521772,  
final\_local\_scores=0.14861307, pem=-0.18263912,  
similarity=9.517843]  
3: 255339 Spin\_City  
[pem\_pos=0, final\_scores=0.13591115,  
final\_local\_scores=0.13636845, pem=0.0, similarity=8.15922]  
4: 53849 Hollywood  
[pem\_pos=1, final\_scores=0.24720778,  
final\_local\_scores=0.20499781, pem=-2.0106025,  
similarity=21.141226]  
5: 173070 The\_Wall\_Street\_Journal  
[pem\_pos=0, final\_scores=0.4833098,  
final\_local\_scores=0.4974029, pem=0.0, similarity=18.933208]  
6: 42254 Buick  
[pem\_pos=0, final\_scores=0.24463294,  
final\_local\_scores=0.26772732, pem=-0.04999765,  
similarity=12.351126]  
7: 231026 Citibank  
[pem\_pos=0, final\_scores=0.30506325,  
final\_local\_scores=0.34463343, pem=0.0, similarity=14.374262]  
8: 763013 Saturday\_Night\_Live  
[pem\_pos=0, final\_scores=0.279742,  
final\_local\_scores=0.2566532, pem=-0.003355763,  
similarity=11.767005]  
9: 645042 New\_York\_City  
[pem\_pos=1, final\_scores=0.45217025,  
final\_local\_scores=0.4311431, pem=-1.7906687,

similarity=26.693817]  
 10: 9316 England  
     [pem\_pos=1, final\_scores=0.047663108,  
 final\_local\_scores=0.06392187, pem=-1.3527626,  
 similarity=13.353803]  
 11: 4401391 American\_Airlines\_Theatre  
     [pem\_pos=0, final\_scores=0.19728123,  
 final\_local\_scores=0.24433851, pem=0.0, similarity=11.381262]  
 12: 877312 Playbill  
     [pem\_pos=0, final\_scores=0.043137476,  
 final\_local\_scores=0.05119902, pem=0.0, similarity=5.617594]  
 13: 418568 Arthur\_Godfrey  
     [pem\_pos=0, final\_scores=0.3401811,  
 final\_local\_scores=0.35191894, pem=0.0, similarity=14.591676]  
 14: 18157 Lucent  
     [pem\_pos=0, final\_scores=0.17504577,  
 final\_local\_scores=0.22608197, pem=0.0, similarity=10.83645]  
 15: 82842 Sex\_and\_the\_City  
     [pem\_pos=0, final\_scores=0.12301542,  
 final\_local\_scores=0.09754932, pem=-0.11152261,  
 similarity=7.6072564]  
 16: 29160 Super\_Bowl\_XXXV  
     [pem\_pos=0, final\_scores=0.0308422,  
 final\_local\_scores=0.08623229, pem=0.0, similarity=6.6630564]

Sample: b'VOA20001223\_2000\_0139'

Reformist allies of [1 Yugoslav 1] President [2 Vojislav Kostunica 2] have scored a decisive victory in today 's [3 Serbian 3] parliamentary elections . The 1 Democratic opposition of [4 Serbia 4] , which supports the President , estimates it has won 65 % of the vote . The coalition took power after a popular uprising ousted long time [5 Yugoslav 5] President [1 1 Slobodan Milosevic 1] in October . Mr. [6 Milosevic 6] 's [2 2 Socialist Party 2] and its allies won about 1/5th of the vote .

1: 25862556 Slobodan\_Milošević

2: 417429 Socialist\_Party\_of\_Serbia

1: 424665 Democratic\_Opposition\_of\_Serbia

[winner: 5043544 Democratic\_Party\_(United\_States),  
 pem\_pos=0, final\_scores=-0.03013198,  
 final\_local\_scores=0.018891707, pem=-0.076093495,  
 similarity=5.0672874]

RECALL MISS: 424665 Democratic\_Opposition\_of\_Serbia]  
 [winner: 2667603 Controversy\_over\_Cantor's\_theory,  
 pem\_pos=6, final\_scores=-1.2427417,  
 final\_local\_scores=-1.180685, pem=-2.472125,  
 similarity=-17.700483]

RECALL MISS: 424665 Democratic\_Opposition\_of\_Serbia]

1: 25862556 Slobodan\_Milošević

[pem\_pos=0, final\_scores=0.61048096,  
 final\_local\_scores=0.5968015, pem=0.0, similarity=21.89946]  
 2: 417429 Socialist\_Party\_of\_Serbia

[pem\_pos=19, final\_scores=0.21962844,  
final\_local\_scores=0.20866376, pem=-4.5017786,  
similarity=34.798027]  
1: 34244 Yugoslavia  
[pem\_pos=1, final\_scores=0.34179354,  
final\_local\_scores=0.29407924, pem=-0.7596219,  
similarity=16.996567]  
2: 424674 Vojislav\_Koštunica  
[pem\_pos=0, final\_scores=0.7662656,  
final\_local\_scores=0.8169474, pem=0.0, similarity=28.469053]  
3: 29265 Serbia  
[pem\_pos=1, final\_scores=0.528584,  
final\_local\_scores=0.5336491, pem=-1.2504034,  
similarity=26.814754]  
4: 29265 Serbia  
[pem\_pos=0, final\_scores=0.06728722,  
final\_local\_scores=0.029752433, pem=-0.34077588,  
similarity=6.830777]  
5: 34244 Yugoslavia  
[pem\_pos=1, final\_scores=0.34668475,  
final\_local\_scores=0.29942214, pem=-0.7596219,  
similarity=17.15601]  
6: 25862556 Slobodan\_Milošević  
[pem\_pos=0, final\_scores=0.212144,  
final\_local\_scores=0.16167894, pem=0.0, similarity=8.914535]