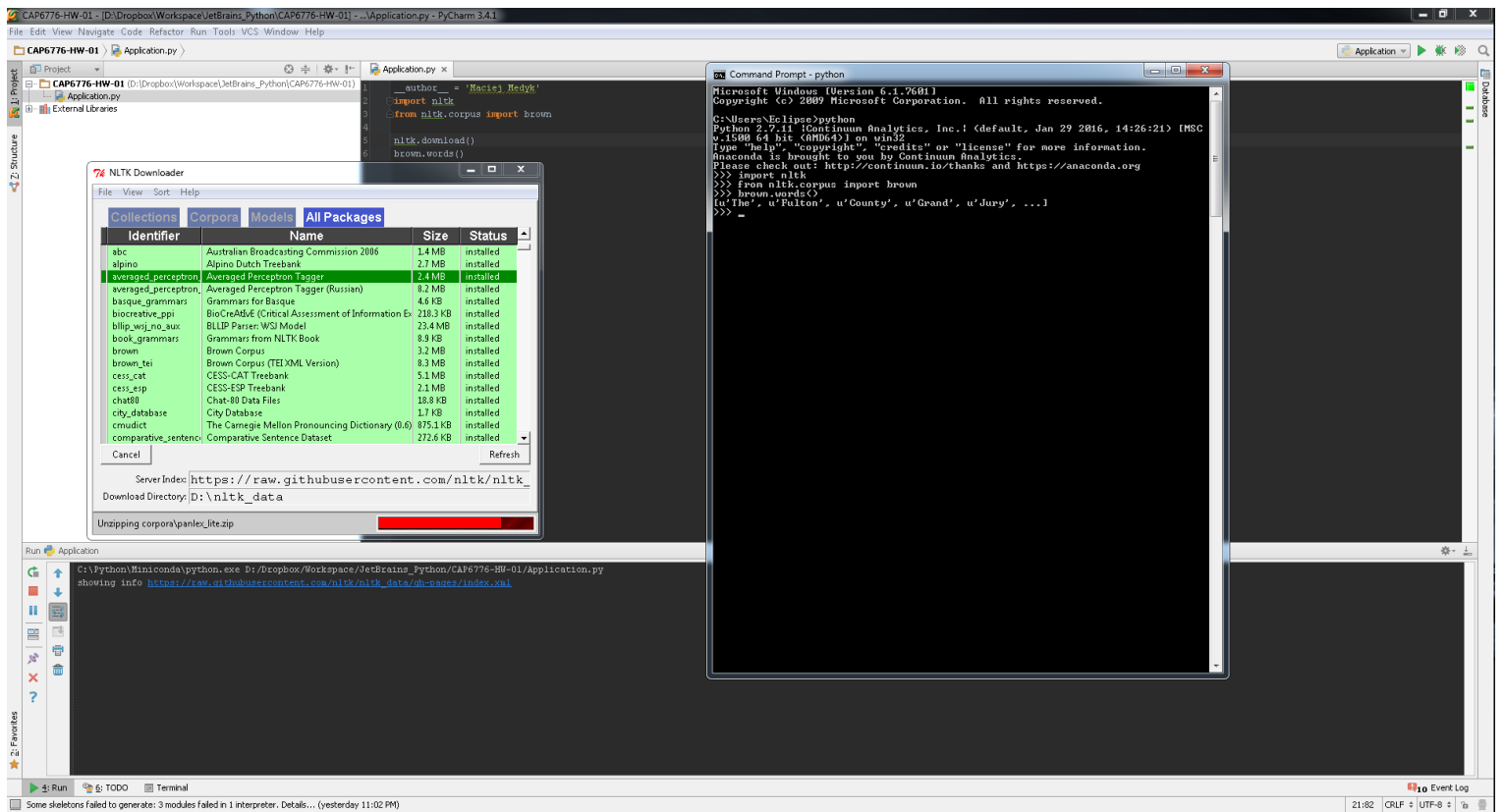
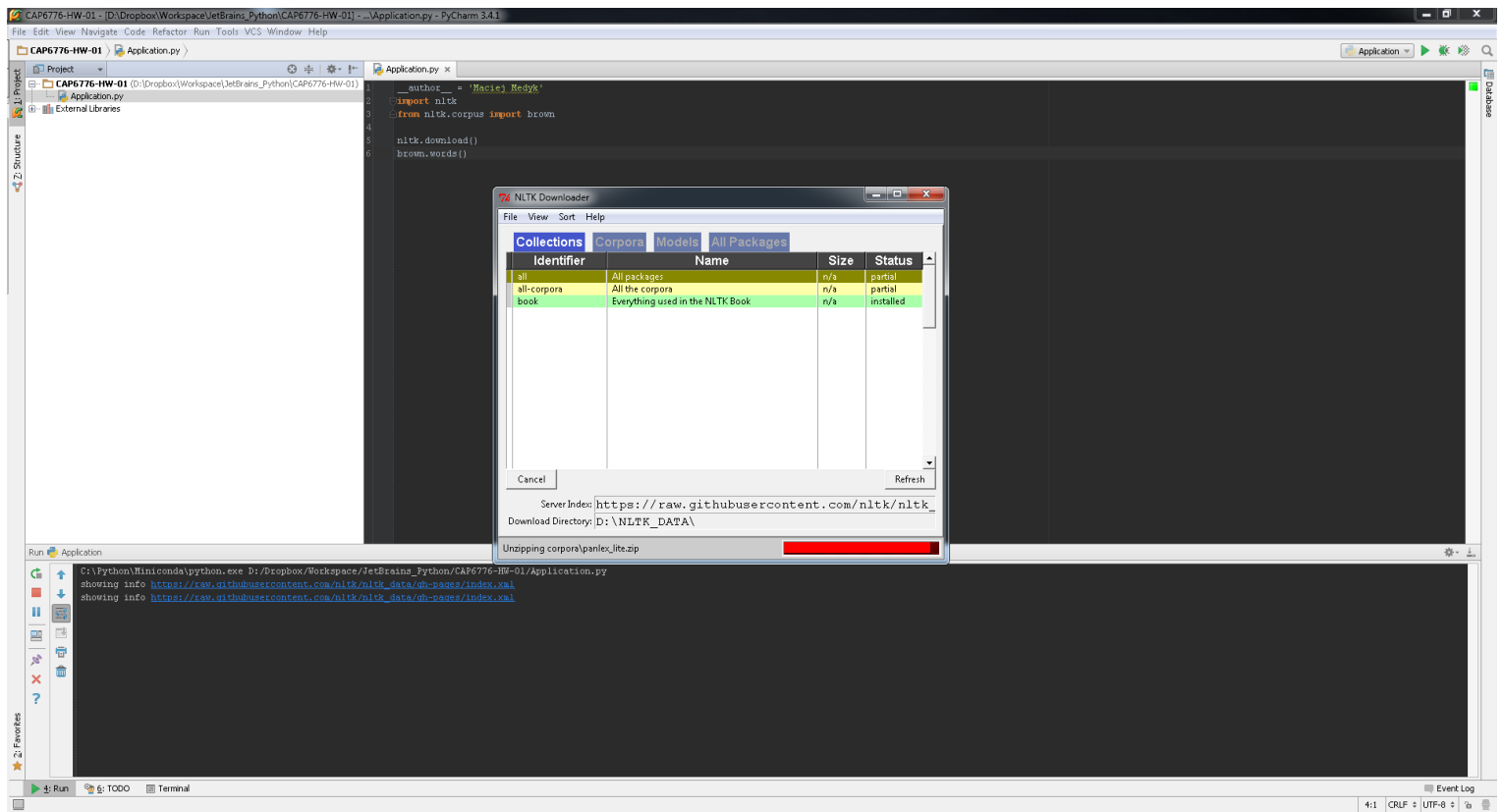


Maciej Medyk – CAP6776 – Information Retrieval

Question 1 – [3.00pt] – Install Python II and NLTK



PyCharm 3.4.1 interface showing a Python script for TF-IDF calculation. The script imports necessary libraries and processes a document from '137871newsML.txt'. It uses a TfidfVectorizer to create a document-term matrix and calculates TF-IDF scores for each word in the document. The output is a list of words and their corresponding TF-IDF scores, sorted by score in descending order.

```
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.corpus import stopwords

author = 'Mario Nedyk'
import nltk, string
from sklearn.feature_extraction.text import TfidfVectorizer
from nltk.corpus import stopwords
```

Counter({'u'said': 9, 'u'appeal': 8, 'u'italian': 7, 'u'berlusconi': 7, 'u'hear': 5, 'u'fraud': 5, 'u'brown': 5, 'u'sto': 5, 'u'the': 5, 'u'paper': 4, 'u'office': 4, 'u'lord': 4, 'u'company': 4, 'u'ground': 3, 'u'ainist': 3, 'u'prime': 3, 'u'ailan': 3, 'u'investig': 3, 'u'friday': 3, 'u'transfer': 3, 'u'use': 3, 'u'appeal': 3, 'u'he': 3, 'u'house': 3, 'u'jud': 3, 'u'fals': 2, 'u'number': 2, 'u'judgment': 2, 'u'layer': 2, 'u'illan': 2, 'u'although': 2, 'u'justic': 2, 'u'although': 2, 'u'forward': 2, 'u'invest': 2, 'u'forward': 2, 'u'account': 2, 'u'ground': 2, 'u'britain': 2, 'u'court': 2, 'u'itali': 2, 'u'channel': 2, 'u'16260': 1, 'u'mone': 1, 'u'mamb': 1, 'u'char': 1, 'u'ail': 1, 'u'illicit': 1, 'u'vet': 1, 'u'involv': 1, 'u'presumpt': 1, 'u'football': 1, 'u'ainan': 1, 'u'utv': 1, 'u'april': 1, 'u'late': 1, 'u'hus': 1, 'u'decid': 1, 'u'ainan': 1, 'u'sent': 1, 'u'sourc': 1, 'u'wide': 1, 'u'donat': 1, 'u'next': 1, 'u'dickson': 1, 'u'politician': 1, 'u'look': 1, 'u'lost': 1, 'u'ayth': 1, 'u'level': 1, 'u'51': 1, 'u'wednesday': 1, 'u'bag': 1, 'u'larg': 1, 'u'leav': 1, 'u'fish': 1, 'u'surreptit': 1, 'u'82': 1, 'u'home': 1, 'u'even': 1, 'u'wear': 1, 'u'given': 1, 'u'lead': 1, 'u'legal': 1, 'u'page': 1, 'u'senior': 1, 'u'aspect': 1, 'u'recognis': 1, 'u'convict': 1, 'u'current': 1, 'u'fidel': 1, 'u'fight': 1, 'u'we': 1, 'u'prosecut': 1, 'u'stem': 1, 'u'job': 1, 'u'succeed': 1, 'u'put': 1, 'u'bas': 1, 'u'search': 1, 'u'ped': 1, 'u'english': 1, 'u'ould': 1, 'u'larger': 1, 'u'british': 1, 'u'david': 1, 'u'trial': 1, 'u'mount': 1, 'u'statement': 1, 'u'place': 1, 'u'decid': 1, 'u'presid': 1, 'u'sean': 1, 'u'mong': 1, 'u'bettino': 1, 'u'om': 1, 'u'ac': 1, 'u'multifac': 1, 'u'one': 1, 'u'expect': 1, 'u'mink': 1, 'u'leadi': 1, 'u'crack': 1, 'u'upr': 1, 'u'ailiv': 1, 'u'describ': 1, 'u'quest': 1, 'u'three': 1, 'u'leat': 1, 'u'happ': 1, 'u'reven': 1, 'u'om': 1, 'u'insere': 1, 'u'illag': 1, 'u'masid': 1, 'u'individa': 1, 'u'look': 1, 'u'offshor': 1, 'u'underst': 1, 'u'underst': 1, 'u'welcom': 1, 'u'judgment': 1, 'u'said': 1, 'u'farrack': 1, 'u'he': 1, 'u'56': 1, 'u'said': 1, 'u'jud': 1, 'u'ject': 1, 'u'ainan': 1, 'u'agament': 1, 'u'p': 1, 'u'forward': 1, 'u'apell': 1, 'u'we': 1, 'u'look': 1, 'u'forward': 1, 'u'speed': 1, 'u'transfer': 1, 'u'document': 1, 'u'ailan': 1, 'u'investig': 1, 'u'get': 1, 'u'job': 1, 'u'said': 1, 'u'hear': 1, 'u'friday': 1, 'u'decid': 1, 'u'whether': 1, 'u'berlusconi': 1, 'u'appeal': 1, 'u'today': 1, 'u'rule': 1, 'u'british': 1, 'u'house': 1, 'u'lord': 1, 'u'fo': 1, 'u'senior': 1, 'u'assist': 1, 'u'director': 1, 'u'chri': 1, 'u'dickson': 1, 'u'said': 1, 'u'judg': 1, 'u'current': 1, 'u'hear': 1, 'u'war': 1, 'u'house': 1, 'u'lord': 1, 'u'hear': 1, 'u'take': 1, 'u'year': 1, 'u'would': 1, 'u'sean': 1, 'u'paper': 1, 'u'would': 1, 'u'arriv': 1, 'u'late': 1, 'u'use': 1, 'u'italian': 1, 'u'proceed': 1, 'u'he': 1, 'u'said': 1, 'u'apell': 1, 'u'given': 1, 'u'leav': 1, 'u'appeal': 1, 'u'friday': 1, 'u'expect': 1, 'u'paper': 1, 'u'transfer': 1, 'u'masid': 1, 'u'although': 1, 'u'decid': 1, 'u'home': 1, 'u'office': 1, 'u'16260': 1, 'u'pound': 1})

Question 3 – [7.00pt] – Calculate TF-IDF for each word in each document and generate document-word matrix

PyCharm 3.4.1 interface showing a Python script for TF-IDF calculation. The script imports necessary libraries and processes a document from '137871newsML.txt'. It uses a TfidfVectorizer to create a document-term matrix and calculates TF-IDF scores for each word in the document. The output is a list of words and their corresponding TF-IDF scores, sorted by score in descending order.

```
tf-idf for each word
```

100554newsML.txt	100594newsML.txt	100614newsML.txt	130040newsML.txt	137871newsML.txt
0.064077798184	0.0790335897544	0.0825906290022	0.0	0.0
100	0.0	0.0	0.0470401010308	0.0
1040	0.032038989092	0.0395167948772	0.0412953145011	0.0
1135	0.032038989092	0.0395167948772	0.0412953145011	0.0
130	0.032038989092	0.0395167948772	0.0412953145011	0.0
136	0.0	0.0590056849221	0.0	0.0
141	0.0	0.0590056849221	0.0	0.0
15	0.0	0.0	0.0	0.0394572624027
150	0.0385969909435	0.0	0.0	0.0379516573231
156	0.0	0.0952108301304	0.0497479766054	0.0
16	0.0478399754196	0.0	0.0	0.0
160	0.032038989092	0.0395167948772	0.0412953145011	0.0
16260	0.0	0.0	0.0	0.0394572624027
16270	0.0	0.0	0.0	0.0470401010308
16393	0.0385969909435	0.0	0.0497479766054	0.0
185	0.0478399754196	0.0	0.0	0.0
1994	0.0	0.0	0.0	0.0394572624027
1997	0.032038989092	0.0395167948772	0.0412953145011	0.0
200	0.0478399754196	0.0	0.0	0.0
2003	0.032038989092	0.0395167948772	0.0412953145011	0.0
2004	0.0478399754196	0.0	0.0	0.0
225	0.032038989092	0.0790335897544	0.0825906290022	0.0
23	0.0	0.0	0.0	0.0470401010308
2392	0.0	0.0	0.0	0.0470401010308
24	0.0	0.0476054130652	0.0497479766054	0.0
30	0.0	0.0	0.0	0.0470401010308
350	0.0	0.0	0.0	0.0470401010308
37	0.032038989092	0.0395167948772	0.0412953145011	0.0
39	0.032038989092	0.0395167948772	0.0412953145011	0.0
40	0.0771939818871	0.0	0.0	0.0379516573231
455	0.032038989092	0.0395167948772	0.0412953145011	0.0
51	0.0	0.0	0.0	0.0394572624027
545	0.032038989092	0.0395167948772	0.0412953145011	0.0
58	0.0	0.0590056849221	0.0	0.0
82	0.0	0.0	0.0	0.0394572624027
87	0.0	0.0476054130652	0.0497479766054	0.0
ab11	0.0478399754196	0.0	0.0	0.0
ab1	0.064077798184	0.0395167948772	0.0412953145011	0.0
ac	0.0	0.0	0.0	0.0394572624027
account	0.0	0.0476054130652	0.0497479766054	0.0
achiev	0.0	0.0	0.0	0.0789145248053
achiev	0.0	0.0476054130652	0.0497479766054	0.0
ad	0.039044153324	0.0332427736336	0.0347389204097	0.026501572388
admit	0.0478399754196	0.0	0.0	0.0
afford	0.0478399754196	0.0	0.0	0.0
agre	0.039044153324	0.0332427736336	0.0347389204097	0.0222295333743
ahead	0.032038989092	0.0395167948772	0.0412953145011	0.0

Question 4 – [5.00pt] – Calculate pairwise cosine similarity for the documents

```
CAP6776-HW-01 - [D:\Dropbox\Workspace\UetBrains_Python\CAP6776-HW-01] - ...CAP6776-HW-01.py - PyCharm 3.4.1
File Edit View Navigate Code Refactor Run Tools VCS Window Help

CAP6776-HW-01 | CAP6776-HW-01.py >

Project | files | CAP6776-HW-01 (D:\Dropbox\Workspace\UetBrains_Python\CAP6776-HW-01)
  10055newsML.txt
  10059newsML.txt
  10061newsML.txt
  13004newsML.txt
  13787newsML.txt
  fi.txt
  fz.txt
  fz.txt

Run | HW-01

cosine similarity for documents
=====
10055newsML.txt 10059newsML.txt 10061newsML.txt 13004newsML.txt 13787newsML.txt

10055newsML.txt 1.0 0.747076672441 0.772366997261 0.166756995412 0.109543969878
10059newsML.txt 0.0269522076662 1.0 0.9746691127 0.122279601282 0.0838630885123
10061newsML.txt 0.772366997261 0.9746691127 1.0 0.123710873913 0.0898335710639
13004newsML.txt 0.166756995412 0.122279601282 0.123710873913 1.0 0.0688468107264
13787newsML.txt 0.109543969878 0.0838630885123 0.0898335710639 0.0688468107264 1.0

Process finished with exit code 0
```

Addendum – Full program output

```
file : 10055newsML.txt
=====
Channel tunnel operator Eurotunnel on Monday announced details of a deal giving bank creditors 45.5 percent of the company in return for wiping out 1.0 billion pounds ($1.6 billion) of its massive debts. The long-awaited but highly complex restructuring of nearly nearly nine billion pounds of debt and unpaid interest throws the company a lifeline which could secure what is still likely to be a difficult future. The deal, announced simultaneously in Paris and London, brings the company back from the brink of bankruptcy but leaves current shareholders, who have already seen their investment dwindle, owning only 54.5 percent of the company. "We have fixed and capped the interest payments and arranged only to pay what is available in cash," Eurotunnel co-chairman Alastair Morton told reporters at a news conference. "Avoiding having to do this again is the name of the game." Morton said the plan provides the Anglo-French company with the medium term financial stability to consolidate its commercial position and develop its operations, adding that the firm was now making a profit before interest. Although shareholders will see their holdings diluted, they were offered the prospect of a brighter future and urged to be patient after months of uncertainty while Eurotunnel wrestled to reduce the crippling interest payments negotiated during the tunnel's construction. Eurotunnel, which has taken around half of the market in the busiest cross-Channel route from the European ferry companies, said a strong operating performance could allow it to pay its first dividend within the next 10 years. French co-chairman Patrick Ponsolle told reporters at a Paris news conference that the dividend could come as early as 2004 if the company performed "very well". Eurotunnel and the banks have come up with an ingenious formula to help the company get over the early years of the deal when, despite the swaps of debt for equity and bonds, it will still not be able to afford the annual interest bill of 400 million pounds. If its revenue, after costs and depreciation, is less than 400 million pounds, then the company will issue "Stabilisation notes" to a maximum of 1.85 billion pounds to the banks. Eurotunnel would not pay interest on these notes (which would constitute a debt issue) for ten years. Analysts said that under the deal, Eurotunnel's ability to finance its debt would become sustainable, at least for a few years. "If you look at the current cash flow of between 150 and 200 million pounds a year, what they can't find (to meet the bill) they will roll forward into the stabilisation notes, and they can keep that going for seven, eight, nine years," said an analyst at one major investment bank. "So they are here for that time," he added. The company said in a statement there was still considerable work to be done to finalise and agree the details of the plan before it can be submitted to shareholders and the bank group for approval, probably early in the Spring of 1997. Eurotunnel said the debt-for-equity swap would be at 130 pence, or 10.40 francs, per share -- considerably below the level of 160 pence widely reported in the run up to the deal. The company said a further 3.7 billion pounds of debt would be converted into new financial instruments and existing shareholders would be able to participate in this issue. If they choose not to take up free warrants entitling them to subscribe to this, Eurotunnel said shareholders' interests may be reduced further to just over 39 percent of the company by the end of December 2003. Eurotunnel's shares, which were suspended last week at 113.5 pence ahead of Monday's announcement, will resume trading on Tuesday. Shareholders and all 225 creditor banks have to agree the deal. "I'm hopeful but I'm not taking it (approval) for granted," Morton admitted, "Shareholders are pretty angry in France." Asked what would happen if the banks reject the deal, Morton said, "Nobody wants a collapse, nobody wants a doomsday scenario." ($1=.6393 Pound)

[u'channel', u'tunnel', u'oper', u'eurotunnel', u'monday', u'announc', u'detail', u'deal', u'give', u'bank', u'creditor', u'455', u'percent', u'compani', u'return', u'wipe', u'10', u'billion', u'pound', u'16', u'billion', u'massiv', u'debt', u'the', u'longawait', u'highli', u'complex', u'restructur', u'nearli', u'nearlii', u'nine', u'billion', u'pound', u'debt', u'unpaid', u'interest', u'throw', u'compani', u'lifelin', u'could', u'secur', u'still', u'like', u'difficult', u'futur', u'the', u'deal', u'announc', u'simultan', u'pari', u'london', u'bring', u'compani', u'back', u'brink', u'bankruptci', u'leav', u'current', u'sharehold', u'alreadi', u'seen', u'invest', u'dwindl', u'own', u'545', u'percent', u'compani', u'we', u'fix', u'cap', u'interest', u'payment', u'arrang', u'pay', u'avail', u'cash', u'eurotunnel', u'cochairman', u'alastair', u'morton', u'told', u'report', u'news', u'confer', u'avoid', u'name', u'game', u'morton', u'said', u'plan', u'provid', u'anglofrench', u'compani', u'medium', u'term', u'financi', u'stabil', u'consolid', u'commerci', u'posit', u'develop', u'oper', u'ad', u'firm', u'make', u'profit', u'interest', u'although', u'sharehold', u'see', u'hold', u'dilut', u'offer', u'prospect', u'brighter', u'futur', u'urg', u'patient', u'month', u'uncertain', u'eurotunnel', u'wrestl', u'reduc', u'crippi', u'interest', u'payment', u'negoti', u'tunnel', u'construct', u'eurotunnel', u'taken', u'around', u'half', u'market', u'busiest', u'crosschannel', u'rout', u'european', u'ferri', u'compani', u'said', u'strong', u'oper', u'perform', u'could', u'allow', u'pay', u'first', u'dividend', u'within', u'next', u'10', u'year', u'french', u'cochairman', u'patrick', u'ponsol', u'told', u'report', u'pari', u'news', u'confer', u'dividend', u'could', u'come', u'earli', u'2004', u'compani', u'perform', u'well', u'eurotunnel', u'bank', u'come', u'ingeni', u'formula', u'help', u'compani', u'get', u'earlii', u'year', u'deal', u'despit', u'swap', u'debt', u'equiti', u'bond', u'still', u'abl', u'afford', u'annual', u'interest', u'bill', u'400', u'million', u'pound', u'if', u'revenu', u'cost', u'depreci', u'less', u'400', u'million', u'pound', u'compani', u'issu', u'stabil', u'note', u'maximum', u'185', u'billion', u'pound', u'bank', u'eurotunnel', u'would', u'pay', u'interest', u'note', u'would', u'constitut', u'debt', u'issu', u'ten', u'year', u'analyst', u'said', u'deal', u'eurotunnel', u'abil', u'financ', u'debt', u'would', u'becom', u'sustain', u'least', u'year', u'if', u'look', u'current', u'cash', u'flow', u'150', u'200', u'million', u'pound', u'year', u'cant', u'find', u'meet', u'bill', u'roll', u'forward', u'stabil', u'note', u'keep', u'go', u'seven', u'eight', u'nine', u'year', u'said', u'analyst', u'one', u'major', u'invest', u'bank', u'so', u'time', u'ad', u'the', u'compani', u'said', u'statement', u'still', u'consider', u'work', u'done', u'finalis', u'agre', u'detail', u'plan', u'submit', u'sharehold', u'bank', u'group', u'approv', u'probabl', u'earli', u'spring', u'1997', u'eurotunnel', u'said', u'debtforequ',
```



```
u'swap', u'would', u'130', u'penc', u'1040', u'franc', u'per', u'share', u'consider', u'level', u'160', u'penc', u'wide', u'report', u'run', u'deal', u'the', u'compani', u'said', u'37', u'billion', u'pound', u'debt', u'would', u'convert', u'new', u'financi', u'instrument', u'exist', u'sharehold', u'would', u'abl', u'particip', u'issu', u'if', u'choos', u'take', u'free', u'warrant', u'entitl', u'subscrib', u'eurotunnel', u'said', u'sharehold', u'interest', u'may', u'reduc', u'39', u'percent', u'compani', u'end', u'decemb', u'2003', u'eurotunnel', u'share', u'suspend', u'last', u'week', u'1135', u'penc', u'ahead', u'monday', u'announc', u'resum', u'trade', u'tuesday', u'sharehold', u'225', u'creditor', u'bank', u'agre', u'deal', u'im', u'hope', u'im', u'take', u'approx', u'grant', u'morton', u'admit', u'sharehold', u'pretti', u'angri', u'franc', u'ask', u'would', u'happen', u'bank', u'reject', u'deal', u'morton', u'said', u'nobodi', u'want', u'collaps', u'nobodi', u'want', u'doomsday', u'scenario', u'16393', u'pound']

Counter({u'compani': 12, u'eurotunnel': 10, u'said': 9, u'pound': 8, u'bank': 7, u'deal': 7, u'sharehold': 7, u'would': 7, u'interest': 7, u'year': 6, u'debt': 6, u'billion': 5, u'morton': 4, u'the': 4, u'penc': 3, u'earli': 3, u'report': 3, u'still': 3, u'announc': 3, u'million': 3, u'could': 3, u'oper': 3, u'pay': 3, u'percent': 3, u'note': 3, u'if': 3, u'issu': 3, u'agre': 2, u'approx': 2, u'consider': 2, u'stabilis': 2, u'swap': 2, u'take': 2, u'nearli': 2, u'financi': 2, u'tunnel': 2, u'bill': 2, u'creditor': 2, u'abl': 2, u'current': 2, u'told': 2, u'franc': 2, u'confer': 2, u'news': 2, u'come': 2, u'dividend': 2, u'reduc': 2, u'nobodi': 2, u'400': 2, u'pari': 2, u'analyst': 2, u'10': 2, u'cash': 2, u'nine': 2, u'ad': 2, u'im': 2, u'share': 2, u'want': 2, u'perform': 2, u'detail': 2, u'futur': 2, u'cochairman': 2, u'monday': 2, u'plan': 2, u'payment': 2, u'invest': 2, u'help': 1, u'less': 1, u'particip': 1, u'month': 1, u'bring': 1, u'130': 1, u'london': 1, u'abil': 1, u'go': 1, u'entitl': 1, u'simultan': 1, u'brink': 1, u'consolid': 1, u'submit': 1, u'group': 1, u'pretti': 1, u'equiti': 1, u'anglofrench': 1, u'fix': 1, u'brighter': 1, u'545': 1, u'sustain': 1, u'warrant': 1, u'collaps': 1, u'maximum': 1, u'1135': 1, u'return': 1, u'get': 1, u'financ': 1, u'game': 1, u'half': 1, u'one': 1, u'wipe': 1, u'like': 1, u'longawait': 1, u'term': 1, u'name': 1, u'level': 1, u'patrick': 1, u'busiest': 1, u'probabl': 1, u'roll': 1, u'european': 1, u'crosschannel': 1, u'bond': 1, u'spring': 1, u'back': 1, u'commerci': 1, u'see': 1, u'cost': 1, u'run': 1, u'happen': 1, u'constitut': 1, u'profit': 1, u'avoid': 1, u'per': 1, u'find': 1, u'restructur': 1, u'leav': 1, u'wide': 1, u'formula': 1, u'we': 1, u'patient': 1, u'despit': 1, u'225': 1, u'french': 1, u'debtforequ': 1, u'lifelin': 1, u'free': 1, u'although': 1, u'meet': 1, u'ask': 1, u'end': 1, u'throw': 1, u'convert': 1, u'last': 1, u'ahead': 1, u'resum': 1, u'annual': 1, u'37': 1, u'keep': 1, u'unpaid': 1, u'admit': 1, u'ponsol': 1, u'uncertain': 1, u'first': 1, u'major': 1, u'own': 1, u'tuesday': 1, u'within': 1, u'455': 1, u'bankruptci': 1, u'alreadi': 1, u'done': 1, u'least': 1, u'massiv': 1, u'market': 1, u'highli': 1, u'doomsday': 1, u'strong': 1, u'ten': 1, u'instrument': 1, u'construct': 1, u'next': 1, u'avail': 1, u'revenu': 1, u'185': 1, u'taken': 1, u'angri': 1, u'firm': 1, u'medium': 1, u'statement': 1, u'alastair': 1, u'dwindl': 1, u'39': 1, u'wrestl': 1, u'forward': 1, u'1040': 1, u'decemb': 1, u'hold': 1, u'160': 1, u'look': 1, u'1997': 1, u'16': 1, u'provid': 1, u'ferri': 1, u'cap': 1, u'hope': 1, u'becom': 1, u'posit': 1, u'arrang': 1, u'prospect': 1, u'150': 1, u'seven': 1, u'give': 1, u'flow': 1, u'cant': 1, u'exist': 1, u'seen': 1, u'work': 1, u'trade': 1, u'around': 1, u'16393': 1, u'subscrib': 1, u'develop': 1, u'dilut': 1, u'secur': 1, u'make': 1, u'offer': 1, u'afford': 1, u'complex': 1, u'reject': 1, u'suspend': 1, u'new': 1, u'negoti': 1, u'eight': 1, u'channel': 1, u'difficult': 1, u'week': 1, u'200': 1, u'crippl': 1, u'2004': 1, u'may': 1, u'finalis': 1, u'ingeni': 1, u'depreci': 1, u'stabil': 1, u'urg': 1, u'scenario': 1, u'choos': 1, u'grant': 1, u'well': 1, u'2003': 1, u'so': 1, u'allow': 1, u'time': 1, u'rout': 1}))

=====
file : 100593newsML.txt
=====

Anglo-French Channel Tunnel operator Eurotunnel Monday announced a deal giving its creditor banks 45.5 percent of the company in return for wiping out one billion pounds ($1.56 billion) of its debt. The long-awaited restructuring brings to an end months of wrangling between Eurotunnel and the 225 banks to which it owes nearly nine billion pounds ($14.1 billion). The deal, announced simultaneously in Paris and London, brings the company back from the brink of insolvency but leaves shareholders owning only 54.5 percent of the company. "The restructuring plan provides Eurotunnel with the medium-term financial stability to allow it to consolidate its substantial commercial achievements to date and to develop its operations," Eurotunnel co-chairman Alastair Morton said. The firm was now making a profit before interest, he added. Although shareholders will see their interests diluted, they were offered the prospect of a brighter future after months of uncertainty while Eurotunnel wrestled to reduce crippling interest payments negotiated during the tunnel's construction. Eurotunnel, which has taken around half the cross-Channel market from the European ferry companies, said a strong operating performance could allow it to pay its first dividend within the next 10 years. French co-chairman Patrick Ponsolle said shareholders would have to be patient before they could reap the benefits of the company's success. He called the debt restructuring plan "an acceptable compromise" for holders of Eurotunnel shares. The company said there was still considerable work to be done to finalise and agree on the details of the plan before it can be submitted to shareholders and the full 225 bank syndicate for approval, probably early in 1997. Monday's announcement followed two weeks of highly secretive negotiations between Eurotunnel and its six leading banks. This was extended to the 24 "instructing banks" at a meeting late last week in London. Eurotunnel said the debt-for-equity swap would be at 130 pence, or 10.40 francs, per share. That is considerably below the level of around 160 pence widely reported before announcement of the deal, and will reduce outstanding debt of 8.7 billion pounds ($13.6 billion) by 1.0 billion ($1.56 billion). The company said a further 3.7 billion pounds ($5.8 billion) of debt would be converted into new financial instruments and existing shareholders would be able to participate in this issue. If they choose not to take up free warrants entitling them to subscribe to this, Eurotunnel said shareholders' interests may be reduced further to just over 39 percent of the company by the end of December 2003. Eurotunnel's shares, which were suspended last week at 113.5 pence ahead of Monday's announcement, should resume trading on Tuesday, the company said.

[u'anglofrench', u'channel', u'tunnel', u'oper', u'eurotunnel', u'monday', u'announc', u'deal', u'give', u'creditor', u'bank', u'455', u'percent', u'compani', u'return', u'wipe', u'one', u'billion', u'pound', u'156', u'billion', u'debt', u'the', u'longawait', u'restructur', u'bring', u'end', u'month', u'wrangl', u'eurotunnel', u'225', u'bank', u'owe', u'nearli', u'nine', u'billion', u'pound', u'141', u'billion', u'the', u'deal', u'announc', u'simultan', u'pari', u'london', u'bring', u'compani', u'back', u'brink', u'insolv', u'leav', u'sharehold', u'own', u'545', u'percent', u'compani', u'the', u'restructur', u'plan', u'provid', u'eurotunnel', u'mediumterm', u'financi', u'stabil', u'allow', u'consolid', u'substanti', u'commerci', u'achiev', u'date', u'develop', u'oper', u'eurotunnel', u'cochairman', u'alastair', u'morton', u'said', u'the', u'firm', u'make', u'profit', u'interest', u'ad', u'although', u'sharehold', u'see', u'interest', u'dilut', u'offer', u'prospect', u'brighter', u'futur', u'month', u'uncertain', u'eurotunnel', u'wrestl', u'reduc', u'crippl', u'interest', u'payment', u'negoti', u'tunnel', u'construct', u'eurotunnel', u'taken', u'around', u'half', u'crosschannel', u'market', u'european', u'ferri', u'compani', u'said', u'strong', u'oper', u'perform', u'could', u'allow', u'pay', u'first', u'dividend', u'next', u'10', u'year', u'french', u'cochairman', u'patrick', u'ponsol', u'said', u'sharehold', u'would', u'patient', u'could', u'reap', u'benefit', u'compani', u'success', u'he', u'call', u'debt', u'restructur', u'plan', u'accept', u'compromis', u'holder', u'eurotunnel', u'share', u'the', u'compani', u'said', u'still', u'consider', u'work', u'done', u'finalis', u'agre', u'detail', u'plan', u'submit', u'sharehold', u'full', u'225', u'bank', u'syndic', u'approx', u'probabl', u'earli', u'1997', u'monday', u'announc', u'follow', u'two', u'week', u'highli', u'secret', u'negoti', u'eurotunnel', u'six', u'lead', u'bank', u'thi', u'extend', u'24', u'instruct', u'bank', u'meet', u'late', u'last', u'week', u'london', u'eurotunnel', u'said', u'debtforequ', u'swap', u'would', u'130', u'penc', u'1040', u'franc', u'per', u'share', u'that', u'consider', u'level', u'around', u'160', u'penc', u'wide', u'report', u'announc', u'deal', u'reduc', u'outstand', u'debt', u'87', u'billion', u'pound', u'136', u'billion', u'10', u'billion', u'156', u'billion', u'the', u'compani', u'said', u'37', u'billion', u'pound', u'58', u'billion', u'debt', u'would', u'convert', u'new', u'financi', u'instrument', u'exist', u'sharehold', u'would', u'abl', u'particip', u'issu', u'if', u'choos', u'take', u'free', u'warrant', u'entitl', u'subscrib', u'eurotunnel', u'said', u'sharehold', u'interest', u'may', u'reduc', u'39', u'percent', u'compani', u'end', u'decemb', u'2003', u'eurotunnel', u'share', u'suspend', u'last', u'week', u'1135', u'penc', u'ahead', u'monday', u'announc', u'resum', u'trade', u'tuesday', u'compani', u'said']

Counter({u'eurotunnel': 11, u'billion': 10, u'compani': 9, u'said': 8, u'sharehold': 6, u'the': 6, u'bank': 5, u'announc': 5, u'debt': 4, u'would': 4, u'interest': 4, u'pound': 4, u'penc': 3, u'deal': 3, u'restructur': 3, u'oper': 3, u'share': 3, u'percent': 3, u'week': 3, u'monday': 3, u'plan': 3, u'reduc': 3, u'consider': 2, u'month': 2, u'bring': 2, u'london': 2, u'around': 2, u'financi': 2, u'last': 2, u'could': 2, u'156': 2, u'225': 2, u'10': 2, u'tunnel': 2, u'end': 2, u'negoti': 2, u'cochairman': 2, u'allow': 2, u'compromis': 1, u'six': 1, u'particip': 1, u'insolv': 1, u'approx': 1, u'130': 1, u'136': 1, u'follow': 1, u'entitl': 1, u'simultan': 1, u'brink': 1, u'24': 1, u'consolid': 1, u'anglofrench': 1, u'160': 1, u'late': 1, u'brighter': 1, u'swap': 1, u'545': 1, u'outstand': 1, u'warrant': 1, u'take': 1, u'earli': 1, u'1135': 1, u'return': 1, u'wide': 1, u'half': 1, u'report': 1, u'trade': 1, u'reap': 1, u'wipe': 1, u'longawait': 1, u'58': 1, u'success': 1, u'level': 1, u'patrick': 1, u'provid': 1, u'achiev': 1, u'37': 1, u'crosschannel': 1, u'still': 1, u'resum': 1, u'wrestl': 1, u'see': 1, u'year': 1, u'morton': 1, u'87': 1, u'creditor': 1, u'lead': 1, u'profit': 1, u'abl': 1, u'benefit': 1, u'per': 1, u'leav': 1, u'new': 1, u'franc': 1, u'full': 1, u'patient': 1, u'extend': 1, u'debtforequ': 1, u'french': 1, u'if': 1, u'free': 1, u'although': 1, u'dividend': 1, u'strong': 1, u'convert': 1, u'ahead': 1, u'syndic': 1, u'ponsol': 1, u'uncertain': 1, u'first': 1, u'own': 1, u'owe': 1, u'tuesday': 1, u'within': 1, u'455': 1, u'one': 1, u'instruct': 1, u'done': 1, u'submit': 1, u'market': 1, u'highli': 1, u'instrument': 1, u'two': 1, u'construct': 1, u'next': 1, u'secret': 1, u'call': 1, u'subscrib': 1, u'taken': 1, u'firm': 1, u'alastair': 1, u'agre': 1, u'39': 1, u'mediumterm': 1, u'pari': 1, u'back': 1, u'decemb': 1, u'holder': 1, u'he': 1, u'1997': 1, u'ferri': 1, u'work': 1, u'commerci': 1, u'nine': 1, u'meet': 1, u'prospect': 1, u'ad': 1, u'give': 1, u'wrangl': 1, u'accept': 1, u'exist': 1, u'substanti': 1, u'nearli': 1, u'dilut': 1, u'perform': 1, u'pay': 1, u'make': 1, u'offer': 1, u'detail': 1, u'that': 1, u'futur': 1, u'suspend': 1, u'probabl': 1, u'channel': 1, u'european': 1, u'crippl': 1, u'thi': 1, u'141': 1, u'may': 1, u'choos': 1, u'date': 1, u'develop': 1, u'payment': 1, u'stabil': 1, u'1040': 1, u'issu': 1, u'2003': 1, u'finalis': 1}))

=====
file : 100618newsML.txt
=====

Anglo-French Channel Tunnel operator Eurotunnel on Monday announced a deal giving creditor banks 45.5 percent of the company in return for wiping out one billion pounds ($1.56 billion) of its debt mountain. The long-awaited restructuring brings to an end months of wrangling between Eurotunnel and the 225 banks to which it owes nearly nine billion pounds. The deal, announced simultaneously in Paris and London, brings the company back from the brink of insolvency but leaves shareholders owning only 54.5 percent of the company. "The restructuring plan provides Eurotunnel with the medium term financial stability to allow it to consolidate its substantial commercial achievements to date and to develop its operations," Eurotunnel co-chairman Alastair Morton said. The firm was now making a profit before interest, he added. Although shareholders will see their interests diluted, they were offered the prospect of a brighter future after months of uncertainty while Eurotunnel wrestled to reduce crippling interest payments negotiated during the tunnel's construction. Eurotunnel, which has taken around half the cross-Channel market from the European ferry companies, said a strong operating performance could allow it to pay its first dividend within the next 10 years. French co-chairman Patrick Ponsolle said shareholders would have to be patient before they could reap the benefits of the company's success. He called the debt restructuring plan "an acceptable compromise" for holders of Eurotunnel shares. The company said in a statement there was still considerable work to be done to finalise and agree the details of the plan before it can be submitted to shareholders and the full 225 bank syndicate for approval, probably early in 1997. Monday's announcement followed two weeks of highly secretive negotiations between Eurotunnel and its six leading banks. This was extended to the 24 "instructing banks" at a meeting late last week in London. Eurotunnel said the debt-for-equity swap would be at 130 pence, or 10.40 francs, per share. That is considerably below the level of around 160 pence widely reported in the run up to the deal, and will reduce outstanding debt of 8.7 billion pounds by 1.0 billion. The company said a further 3.7 billion pounds of debt would be converted into new financial instruments and existing shareholders would be able to participate in this issue. If they choose not to take up free warrants entitling them to subscribe to this, Eurotunnel said
```

shareholders' interests may be reduced further to just over 39 percent of the company by the end of December 2003. Eurotunnel's shares, which were suspended last week at 113.5 pence ahead of Monday's announcement, should resume trading on Tuesday, the company said. (\$1=6393 Pound)

[u'anglofrench', u'channel', u'tunnel', u'oper', u'eurotunnel', u'monday', u'announc', u'deal', u'give', u'creditor', u'bank', u'455', u'percent', u'compani', u'return', u'wipe', u'one', u'billion', u'pound', u'156', u'billion', u'debt', u'mountain', u'the', u'longawait', u'restructur', u'bring', u'end', u'month', u'wrangl', u'eurotunnel', u'225', u'bank', u'owe', u'nearli', u'nine', u'billion', u'pound', u'the', u'deal', u'announc', u'simultan', u'pari', u'london', u'bring', u'compani', u'back', u'brink', u'insolv', u'leav', u'sharehold', u'own', u'545', u'percent', u'compani', u'the', u'restructur', u'plan', u'provid', u'eurotunnel', u'medium', u'term', u'financi', u'stabil', u'allow', u'consolid', u'substanti', u'commerci', u'achiev', u'date', u'develop', u'oper', u'eurotunnel', u'cochairman', u'alastair', u'morton', u'said', u'the', u'firm', u'make', u'profit', u'interest', u'ad', u'although', u'sharehold', u'see', u'interest', u'dilut', u'offer', u'prospect', u'brighter', u'futur', u'month', u'uncertain', u'eurotunnel', u'westl', u'reduc', u'crippl', u'interest', u'payment', u'negoti', u'tunnel', u'construct', u'eurotunnel', u'taken', u'around', u'half', u'crosschannel', u'market', u'european', u'ferri', u'compani', u'said', u'strong', u'oper', u'perform', u'could', u'allow', u'pay', u'first', u'dividend', u'within', u'next', u'10', u'year', u'french', u'cochairman', u'patrick', u'ponsol', u'said', u'sharehold', u'would', u'patient', u'could', u'reap', u'benefit', u'compani', u'success', u'he', u'call', u'debt', u'restructur', u'plan', u'accept', u'compromis', u'holder', u'eurotunnel', u'share', u'the', u'compani', u'said', u'statement', u'still', u'consider', u'work', u'done', u'finalis', u'agre', u'detail', u'plan', u'submit', u'sharehold', u'full', u'225', u'bank', u'syndic', u'approv', u'proabl', u'earli', u'1997', u'monday', u'announc', u'follow', u'two', u'week', u'highli', u'secret', u'negoti', u'eurotunnel', u'six', u'lead', u'bank', u'thi', u'extend', u'24', u'instruct', u'bank', u'meet', u'late', u'last', u'week', u'london', u'eurotunnel', u'said', u'debtforequ', u'swap', u'would', u'130', u'penc', u'1040', u'franc', u'per', u'share', u'that', u'consider', u'level', u'around', u'160', u'penc', u'wide', u'report', u'run', u'deal', u'reduc', u'outstand', u'debt', u'87', u'billion', u'pound', u'10', u'billion', u'the', u'compani', u'said', u'37', u'billion', u'pound', u'debt', u'would', u'convert', u'new', u'financi', u'instrument', u'exist', u'sharehold', u'would', u'abl', u'particip', u'issu', u'if', u'choos', u'take', u'free', u'warrant', u'entitl', u'subscrib', u'eurotunnel', u'said', u'sharehold', u'interest', u'may', u'reduc', u'39', u'percent', u'compani', u'end', u'decemb', u'2003', u'eurotunnel', u'share', u'suspend', u'last', u'week', u'1135', u'penc', u'ahead', u'monday', u'announc', u'resum', u'trade', u'tuesday', u'compani', u'said', u'16393', u'pound']

Counter({u'eurotunnel': 11, u'compani': 9, u'said': 8, u'sharehold': 6, u'billion': 6, u'the': 6, u'bank': 5, u'pound': 5, u'debt': 4, u'announc': 4, u'would': 4, u'interest': 4, u'penc': 3, u'deal': 3, u'restructur': 3, u'oper': 3, u'share': 3, u'percent': 3, u'week': 3, u'monday': 3, u'plan': 3, u'reduc': 3, u'consider': 2, u'month': 2, u'bring': 2, u'london': 2, u'around': 2, u'financi': 2, u'last': 2, u'225': 2, u'10': 2, u'tunnel': 2, u'end': 2, u'negoti': 2, u'cochairman': 2, u'allow': 2, u'compromis': 1, u'six': 1, u'particip': 1, u'insolv': 1, u'approv': 1, u'130': 1, u'follow': 1, u'entitl': 1, u'simultan': 1, u'brink': 1, u'24': 1, u'consolid': 1, u'anglofrench': 1, u'160': 1, u'late': 1, u'brighter': 1, u'swap': 1, u'545': 1, u'outstand': 1, u'warrant': 1, u'take': 1, u'earli': 1, u'1135': 1, u'return': 1, u'wide': 1, u'half': 1, u'report': 1, u'trade': 1, u'reap': 1, u'wipe': 1, u'longawait': 1, u'term': 1, u'success': 1, u'level': 1, u'patrick': 1, u'provid': 1, u'achiev': 1, u'37': 1, u'crosschannel': 1, u'still': 1, u'resum': 1, u'westl': 1, u'see': 1, u'year': 1, u'run': 1, u'morton': 1, u'87': 1, u'creditor': 1, u'lead': 1, u'profit': 1, u'abl': 1, u'benefit': 1, u'per': 1, u'leav': 1, u'new': 1, u'franc': 1, u'full': 1, u'patient': 1, u'extend': 1, u'debtforequ': 1, u'french': 1, u'if': 1, u'free': 1, u'although': 1, u'dividend': 1, u'convert': 1, u'ahead': 1, u'syndic': 1, u'16393': 1, u'ponsol': 1, u'uncertain': 1, u'first': 1, u'own': 1, u'owe': 1, u'tuesday': 1, u'within': 1, u'455': 1, u'one': 1, u'instruct': 1, u'done': 1, u'submit': 1, u'market': 1, u'highli': 1, u'strong': 1, u'instrument': 1, u'156': 1, u'two': 1, u'construct': 1, u'next': 1, u'secret': 1, u'call': 1, u'subscrib': 1, u'taken': 1, u'firm': 1, u'medium': 1, u'statement': 1, u'alastair': 1, u'agre': 1, u'39': 1, u'pari': 1, u'back': 1, u'decemb': 1, u'holder': 1, u'he': 1, u'1997': 1, u'ferri': 1, u'work': 1, u'commerci': 1, u'nine': 1, u'meet': 1, u'prospect': 1, u'ad': 1, u'give': 1, u'wrangl': 1, u'accept': 1, u'exist': 1, u'substanti': 1, u'nearli': 1, u'mountain': 1, u'dilut': 1, u'perform': 1, u'pay': 1, u'make': 1, u'offer': 1, u'detail': 1, u'that': 1, u'futur': 1, u'suspend': 1, u'proabl': 1, u'channel': 1, u'european': 1, u'crippl': 1, u'thi': 1, u'may': 1, u'choos': 1, u'date': 1, u'develop': 1, u'payment': 1, u'stabil': 1, u'1040': 1, u'issu': 1, u'2003': 1, u'finalis': 1))

file : 130040newsML.txt

Britain's Barclays Plc said on Monday it was in talks with U.S. group Morgan Stanley about its global custody business and banking sources confirmed the unit is up for sale. Barclays and Morgan Stanley were coy in responding to press speculation, saying only that they are "in discussions to explore the potential for future co-operation". The move would be another step in the consolidation of the global custody business which is becoming dominated by those banks who are willing to make the large front-end investment in systems which are necessary to later rake in fees. Barclays' custody business has 350 to 400 staff worldwide, with some working in the Far East, but has for some time been treated as not in the bank's core business. Securities worth some 150 billion pounds (\$239.2 billion) have been placed with the unit. In its interim results, published in August, Barclays included custody in its "Businesses in Transition" category which includes restructuring businesses in France and the United States. The category including custody made a profit of 23 million pounds in the half-year to June 30. The report said, "Businesses in Transition primarily comprises lendings and other assets that are unlikely to be of long-term interest to the Group or that require significant restructuring." It was not clear when the deal will be finalised but analysts said it would mark a further step in the consolidation of the global custody market. "Consolidation is a natural if you look at the economic of the business," said John Leonard, banking analyst at Salomon Brothers. "It's a scale economy business." Leonard said that some banks had taken the decision not to invest the large sums needed to get computer systems up to scratch to be competitive. Custody businesses range from traditional safekeeping of securities, which in Britain is dominated by Lloyds TSB and Royal Bank of Scotland, to performance measurement and stock lending. As such, it has ceased to be the largely risk-free business it once was although risk levels are still relatively low. "It guarantees a steady income stream but it's not totally insensitive to market volumes," Leonard added. Analysts said Barclays has presumably decided that without further investment, the business will be too small to compete on the global stage. They expect Morgan Stanley to pay well under 100 million pounds for the business. "It could well be less than fifty million (pounds)," said one. Finance workers' union BIFU attacked the leak of the discussions, saying staff face an uncertain future if the business was sold. It called on the bank to "come clean" over its intentions. (\$1=6270 Pound)

[u'britain', u'barclay', u'plc', u'said', u'monday', u'talk', u'us', u'group', u'morgan', u'stanley', u'global', u'custodi', u'busi', u'bank', u'sourc', u'confirm', u'unit', u'sale', u'barclay', u'morgan', u'stanley', u'coy', u'respond', u'press', u'specul', u'say', u'discuss', u'explor', u'potenti', u'futur', u'cooper', u'the', u'move', u'would', u'anoth', u'step', u'consolid', u'global', u'custodi', u'busi', u'becom', u'domin', u'bank', u'will', u'make', u'larg', u'frontend', u'invest', u'system', u'necessari', u'later', u'rake', u'fee', u'barclay', u'custodi', u'busi', u'350', u'400', u'stuff', u'worldwid', u'work', u'far', u'east', u'time', u'treat', u'bank', u'core', u'busi', u'secur', u'worth', u'150', u'billion', u'pound', u'2392', u'billion', u'place', u'unit', u'in', u'interim', u'result', u'publish', u'august', u'barclay', u'includ', u'custodi', u'busi', u'transit', u'categori', u'includ', u'restructur', u'busi', u'franc', u'unit', u'state', u'the', u'categori', u'includ', u'custodi', u'made', u'profit', u'23', u'million', u'pound', u'halfyear', u'june', u'30', u'the', u'report', u'said', u'busi', u'transit', u'primarili', u'compris', u'lend', u'asset', u'unlik', u'longterm', u'interest', u'group', u'requir', u'signific', u'restructur', u'it', u'clear', u'deal', u'finalis', u'analyst', u'said', u'would', u'mark', u'step', u'consolid', u'global', u'custodi', u'market', u'consolid', u'natur', u'look', u'econom', u'busi', u'said', u'john', u'leonard', u'bank', u'analyst', u'salomon', u'brother', u'it', u'scale', u'economi', u'busi', u'leonard', u'said', u'bank', u'taken', u'decis', u'invest', u'larg', u'sum', u'need', u'get', u'comput', u'system', u'scratch', u'competit', u'custodi', u'busi', u'rang', u'tradit', u'safekeep', u'secur', u'britain', u'domin', u'lloyd', u'tsb', u'royal', u'bank', u'scotland', u'perform', u'measur', u'stock', u'lend', u'as', u'ceas', u'larg', u'riskfre', u'busi', u'although', u'risk', u'level', u'still', u'rel', u'low', u'it', u'guarante', u'steadi', u'incom', u'stream', u'total', u'insensit', u'market', u'volum', u'leonard', u'ad', u'analyst', u'said', u'barclay', u'presum', u'decid', u'without', u'invest', u'busi', u'small', u'compet', u'global', u'stage', u'they', u'expect', u'morgan', u'stanley', u'pay', u'well', u'100', u'million', u'pound', u'busi', u'it', u'could', u'well', u'less', u'fifti', u'million', u'pound', u'said', u'one', u'financ', u'worker', u'union', u'bifu', u'attack', u'leak', u'say', u'stuff', u'face', u'uncertain', u'futur', u'busi', u'sold', u'it', u'call', u'bank', u'come', u'clean', u'intent', u'16270', u'pound']

Counter({u'busi': 13, u'bank': 7, u'said': 7, u'custodi': 7, u'barclay': 5, u'pound': 5, u'it': 5, u'global': 4, u'consolid': 3, u'stanley': 3, u'larg': 3, u'million': 3, u'unit': 3, u'includ': 3, u'leonard': 3, u'analyst': 3, u'morgan': 3, u'invest': 3, u'the': 3, u'stuff': 2, u'group': 2, u'discuss': 2, u'categori': 2, u'restructur': 2, u'market': 2, u'step': 2, u'transit': 2, u'would': 2, u'system': 2, u'say': 2, u'britain': 2, u'secur': 2, u'futur': 2, u'lend': 2, u'billion': 2, u'domin': 2, u'well': 2, u'frontend': 1, u'less': 1, u'scratch': 1, u'press': 1, u'signific': 1, u'interim': 1, u'lloyd': 1, u'still': 1, u'decid': 1, u'23': 1, u'tsb': 1, u'requir': 1, u'400': 1, u'east': 1, u'worth': 1, u'pay': 1, u'move': 1, u'risk': 1, u'get': 1, u'far': 1, u'rake': 1, u'volum': 1, u'they': 1, u'report': 1, u'measur': 1, u'level': 1, u'brother': 1, u'2392': 1, u'compris': 1, u'ceas': 1, u'decis': 1, u'16270': 1, u'respond': 1, u'deal': 1, u'core': 1, u'natur': 1, u'result': 1, u'cooper': 1, u'confirm': 1, u'profit': 1, u'leak': 1, u'sum': 1, u'publish': 1, u'state': 1, u'rel': 1, u'necessari': 1, u'fifti': 1, u'franc': 1, u'august': 1, u'sold': 1, u'small': 1, u'plc': 1, u'although': 1, u'100': 1, u'becom': 1, u'come': 1, u'stage': 1, u'coy': 1, u'could': 1, u'30': 1, u'worldwid': 1, u'place': 1, u'riskfre': 1, u'salomon': 1, u'rang': 1, u'primarili': 1, u'finalis': 1, u'royal': 1, u'one': 1, u'expect': 1, u'unlik': 1, u'total': 1, u'fee': 1, u'busi': 1, u'union': 1, u'mark': 1, u'specul': 1, u'attack': 1, u'call': 1, u'treat': 1, u'interest': 1, u'taken': 1, u'john': 1, u'guarante': 1, u'tradit': 1, u'asset': 1, u'worker': 1, u'low': 1, u'made': 1, u'uncertain': 1, u'look': 1, u'insensit': 1, u'work': 1, u'us': 1, u'will': 1, u'150': 1, u'comput': 1, u'ad': 1, u'stream': 1, u'longterm': 1, u'halfyear': 1, u'sourc': 1, u'explor': 1, u'in': 1, u'need': 1, u'incom': 1, u'as': 1, u'perform': 1, u'anoth': 1, u'make': 1, u'scotland': 1, u'econom': 1, u'economi': 1, u'compet': 1, u'presum': 1, u'stock': 1, u'competit': 1, u'financ': 1, u'clean': 1, u'monday': 1, u'350': 1, u'june': 1, u'bifu': 1, u'scale': 1, u'intent': 1, u'clear': 1, u'later': 1, u'sale': 1, u'face': 1, u'without': 1, u'potenti': 1, u'time': 1, u'safekeep': 1, u'steadi': 1, u'talk': 1))

file : 137871newsML.txt

Italian business tycoon Silvio Berlusconi on Wednesday lost an appeal in the London High Court against the transfer of documents to Italian authorities investigating allegations of fraud and false accounting. Britain's Serious Fraud Office seized 15 bags of documents in April from the office of London lawyer David Mills who represents the interests of former Italian Prime Minister Berlusconi in Britain. The judges, Lord Justice Simon Brown and Mr Justice Gage, rejected all the grounds of the appeal, but Brown agreed to hear an application by the appellants on Friday morning, which could lead to an appeal to the House of Lords, although legal sources said this was unlikely to succeed. The hearing stemmed from an appeal by Berlusconi, who was Italian Prime Minister for seven months in 1994 and is on trial in Milan on corruption charges, his company Fininvest and company president Fedele Confalonieri against a court ruling earlier this year that the papers should be sent to Italy. Brown said the papers were linked to fraud and false accounting allegations amounting to 51 million pounds (\$82 million) which had been "surreptitiously removed from Fininvest and used for criminal purposes." He noted that prosecutions were already underway in Italy against Berlusconi for bribing Revenue inspectors and for making illicit donations to former Prime Minister Bettino Craxi. Italian

investigators suspect Berlusconi who, among other holdings, owns three TV channels and the football club AC Milan, may have used offshore companies to channel payments to politicians. At the end of his judgement Brown pointed out that, "It is imperative to recognise, however, that none of the applicants has yet been convicted of anything and that nothing I have said should be thought to raise the least presumption of guilt against them." He rejected all the grounds of the appeal on the basis that the Italian request was not too wide, that it was justified and that the offences in question were not "political". One of the grounds of appeal had been that the SFO had gone to Berlusconi's lawyer's office on a "fishing expedition" -- an unspecific search -- which is illegal under English law. Brown described the case as a "wide-ranging, multi-faceted, international fraud involving far-reaching allegations against a large number of individuals in connection with an even larger number of companies." The SFO said it was "happy" with the ruling. "The SFO's fight against fraud on an international level has been strengthened by this welcome judgement," it said in a statement. The SFO said the judges had rejected all of the main arguments put forward by the appellants. "We look forward to the speedy transfer of the documents to Milan so that the investigators can get on with their job," it said. A hearing on Friday will decide whether Berlusconi can appeal against today's ruling to the British House of Lords. SFO senior assistant director Chris Dickson said the judges in the current hearing were aware that, if any House of Lords hearing did not take place until next year, this would mean the papers would arrive too late to be used in Italian proceedings. He said that if the appellants were not given leave to appeal on Friday he expected the papers to be transferred immediately, although this was a decision for the Home Office. (\$1=.6260 Pound)

[u'italian', u'busi', u'ttycoon', u'silvio', u'berlusconi', u'wednesday', u'lost', u'appeal', u'london', u'high', u'court', u'transfer', u'document', u'italian', u'author', u'investig', u'alleg', u'fraud', u'fals', u'account', u'britain', u'seriou', u'fraud', u'offic', u'seiz', u'15', u'bag', u'document', u'april', u'offic', u'london', u'lawyer', u'david', u'mill', u'repres', u'interest', u'former', u'italian', u'prime', u'minist', u'berlusconi', u'britain', u'the', u'judg', u'lord', u'justic', u'simon', u'brown', u'mr', u'justic', u'gage', u'reject', u'ground', u'appeal', u'brown', u'agre', u'hear', u'applic', u'appel', u'friday', u'morn', u'could', u'lead', u'appeal', u'hous', u'lord', u'although', u'legal', u'sourc', u'said', u'unlik', u'succeed', u'the', u'hear', u'stem', u'appeal', u'berlusconi', u'italian', u'prime', u'minist', u'seven', u'month', u'1994', u'trial', u'milan', u'corrupt', u'charg', u'compani', u'fininvest', u'compani', u'presid', u'fedel', u'confalonieri', u'court', u'rule', u'earlier', u'year', u'paper', u'sent', u'itali', u'brown', u'said', u'paper', u'link', u'fraud', u'fals', u'account', u'alleg', u'amount', u'51', u'million', u'pound', u'82', u'million', u'surreptiti', u'remov', u'fininvest', u'use', u'crimin', u'purpos', u'he', u'note', u'prosecut', u'alreadi', u'underway', u'itali', u'berlusconi', u'bribe', u'revenu', u'inspector', u'make', u'illicit', u'donat', u'former', u'prime', u'minist', u'bettino', u'craxi', u'italian', u'investig', u'suspect', u'berlusconi', u'among', u'hold', u'own', u'three', u'tv', u'channel', u'football', u'club', u'ac', u'milan', u'may', u'se', u'offshor', u'compani', u'channel', u'payment', u'politician', u'at', u'end', u'judgement', u'brown', u'point', u'it', u'imper', u'recognis', u'howev', u'none', u'applic', u'yet', u'convict', u'anyth', u'noth', u'i', u'said', u'thought', u'rais', u'least', u'presumpt', u'guilt', u'he', u'reject', u'ground', u'appeal', u'basil', u'italian', u'request', u'wide', u'justifi', u'offenc', u'question', u'polit', u'one', u'ground', u'appeal', u'sfo', u'gone', u'berlusconi', u'lawyer', u'offic', u'fish', u'expedit', u'unspecif', u'search', u'illeg', u'english', u'law', u'brown', u'describ', u'case', u'widerang', u'multifacet', u'intern', u'fraud', u'involv', u'farreach', u'alleg', u'larg', u'number', u'individu', u'connect', u'even', u'larger', u'number', u'compani', u'the', u'sfo', u'said', u'happi', u'rule', u'the', u'sfo', u'fight', u'fraud', u'intern', u'level', u'strengthen', u'welcom', u'judgement', u'said', u'statement', u'the', u'sfo', u'said', u'judg', u'reject', u'main', u'argument', u'put', u'forward', u'appeal', u'we', u'look', u'forward', u'speedi', u'transfer', u'document', u'milan', u'investig', u'get', u'job', u'said', u'a', u'hear', u'friday', u'decid', u'whether', u'berlusconi', u'appeal', u'today', u'rule', u'british', u'hous', u'lord', u'sfo', u'senior', u'assist', u'director', u'chri', u'dickson', u'said', u'judg', u'current', u'hear', u'awar', u'hous', u'lord', u'hear', u'take', u'place', u'next', u'year', u'would', u'mean', u'paper', u'would', u'arriv', u'late', u'use', u'italian', u'proceed', u'he', u'said', u'appeal', u'given', u'leav', u'appeal', u'friday', u'expect', u'paper', u'transfer', u'immedi', u'although', u'decis', u'home', u'offic', u'16260', u'pound']

Counter({u'said': 9, u'appeal': 8, u'italian': 7, u'berlusconi': 7, u'hear': 5, u'fraud': 5, u'brown': 5, u'sfo': 5, u'the': 5, u'paper': 4, u'offic': 4, u'lord': 4, u'compani': 4, u'ground': 3, u'minist': 3, u'prime': 3, u'milan': 3, u'investig': 3, u'friday': 3, u'transfer': 3, u'use': 3, u'appel': 3, u'he': 3, u'hous': 3, u'judg': 3, u'alleg': 3, u'reject': 3, u'document': 3, u'rule': 3, u'london': 2, u'applic': 2, u'fals': 2, u'number': 2, u'judgement': 2, u'lawyer': 2, u'year': 2, u'million': 2, u'although': 2, u'justic': 2, u'would': 2, u'fininvest': 2, u'forward': 2, u'former': 2, u'account': 2, u'intern': 2, u'pound': 2, u'britain': 2, u'court': 2, u'itali': 2, u'channel': 2, u'16260': 1, u'none': 1, u'month': 1, u'chri': 1, u'mill': 1, u'illicit': 1, u'yet': 1, u'involv': 1, u'presumpt': 1, u'football': 1, u'simon': 1, u'tv': 1, u'april': 1, u'late': 1, u'busi': 1, u'decis': 1, u'main': 1, u'sent': 1, u'sourc': 1, u'wide': 1, u'donat': 1, u'next': 1, u'dickson': 1, u'politician': 1, u'look': 1, u'lost': 1, u'anyth': 1, u'level': 1, u'51': 1, u'wednesday': 1, u'underway': 1, u'bag': 1, u'larg': 1, u'leav': 1, u'fish': 1, u'surreptiti': 1, u'82': 1, u'home': 1, u'even': 1, u'awar': 1, u'given': 1, u'lead': 1, u'legal': 1, u'gage': 1, u'senior': 1, u'suspect': 1, u'recognis': 1, u'convict': 1, u'current': 1, u'fedel': 1, u'fight': 1, u'we': 1, u'prosecut': 1, u'stem': 1, u'job': 1, u'succeed': 1, u'put': 1, u'basil': 1, u'search': 1, u'speedi': 1, u'english': 1, u'could': 1, u'larger': 1, u'british': 1, u'david': 1, u'trial': 1, u'amount': 1, u'statement': 1, u'place': 1, u'decid': 1, u'presid': 1, u'mean': 1, u'among': 1, u'bettino': 1, u'own': 1, u'point': 1, u'ac': 1, u'multifacet': 1, u'one': 1, u'expect': 1, u'unlik': 1, u'alreadi': 1, u'craxi': 1, u'agre': 1, u'silvio': 1, u'describ': 1, u'question': 1, u'three': 1, u'least': 1, u'happi': 1, u'revenu': 1, u'morn': 1, u'interest': 1, u'illeg': 1, u'immedi': 1, u'individu': 1, u'today': 1, u'ttycoon': 1, u'offshor': 1, u'club': 1, u'widerang': 1, u'argument': 1, u'expedit': 1, u'hold': 1, u'case': 1, u'high': 1, u'15': 1, u'1994': 1, u'whether': 1, u'seiz': 1, u'remov': 1, u'imper': 1, u'arriv': 1, u'guilt': 1, u'mr': 1, u'purpos': 1, u'howev': 1, u'connect': 1, u'gone': 1, u'strengthen': 1, u'it': 1, u'repres': 1, u'corrupt': 1, u'farreach': 1, u'at': 1, u'unspecif': 1, u'crimin': 1, u'inspector': 1, u'welcom': 1, u'author': 1, u'make': 1, u'get': 1, u'note': 1, u'take': 1, u'noth': 1, u'polit': 1, u'may': 1, u'assist': 1, u'earlier': 1, u'director': 1, u'rais': 1, u'end': 1, u'law': 1, u'payment': 1, u'charg': 1, u'a': 1, u'proceed': 1, u'i': 1, u'seriou': 1, u'confalonieri': 1, u'request': 1, u'justifi': 1, u'thought': 1, u'bribe': 1, u'link': 1, u'offenc': 1}))

tf-idf for each word

	100554newsML.txt	100593newsML.txt	100618newsML.txt	130040newsML.txt	137871newsML.txt
10	0.064077978184	0.0790335897544	0.0825906290022	0.0	0.0
100	0.0	0.0	0.0	0.0470401010308	0.0
1040	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
1135	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
130	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
136	0.0	0.0590056849221	0.0	0.0	0.0
141	0.0	0.0590056849221	0.0	0.0	0.0
15	0.0	0.0	0.0	0.0	0.0394572624027
150	0.0385969909435	0.0	0.0	0.0379516573231	0.0
156	0.0	0.0952108301304	0.0497479766054	0.0	0.0
16	0.0478399754196	0.0	0.0	0.0	0.0
160	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
16260	0.0	0.0	0.0	0.0	0.0394572624027
16270	0.0	0.0	0.0	0.0470401010308	0.0
16393	0.0385969909435	0.0	0.0497479766054	0.0	0.0
185	0.0478399754196	0.0	0.0	0.0	0.0
1994	0.0	0.0	0.0	0.0	0.0394572624027
1997	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
200	0.0478399754196	0.0	0.0	0.0	0.0
2003	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
2004	0.0478399754196	0.0	0.0	0.0	0.0
225	0.032038989092	0.0790335897544	0.0825906290022	0.0	0.0
23	0.0	0.0	0.0	0.0470401010308	0.0
2392	0.0	0.0	0.0	0.0470401010308	0.0
24	0.0	0.0476054150652	0.0497479766054	0.0	0.0
30	0.0	0.0	0.0	0.0470401010308	0.0
350	0.0	0.0	0.0	0.0470401010308	0.0
37	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
39	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
400	0.0771939818871	0.0	0.0	0.0379516573231	0.0
455	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
51	0.0	0.0	0.0	0.0	0.0394572624027
545	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
58	0.0	0.0590056849221	0.0	0.0	0.0
82	0.0	0.0	0.0	0.0	0.0394572624027
87	0.0	0.0476054150652	0.0497479766054	0.0	0.0
abil	0.0478399754196	0.0	0.0	0.0	0.0
abl	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
ac	0.0	0.0	0.0	0.0	0.0394572624027
accept	0.0	0.0476054150652	0.0497479766054	0.0	0.0
account	0.0	0.0	0.0	0.0	0.0789145248053
achiev	0.0	0.0476054150652	0.0497479766054	0.0	0.0
ad	0.0539044153324	0.0332427736336	0.0347389204097	0.026501572388	0.0
admit	0.0478399754196	0.0	0.0	0.0	0.0
afford	0.0478399754196	0.0	0.0	0.0	0.0
agre	0.0539044153324	0.0332427736336	0.0347389204097	0.0	0.0222295333743
ahead	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
alastair	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
alleg	0.0	0.0	0.0	0.0	0.118371787208

allow	0.032038989092	0.0790335897544	0.0825906290022	0.0	0.0
alreadi	0.0385969909435	0.0	0.0	0.0	0.0318338708633
analyst	0.0771939818871	0.0	0.0	0.113854971969	0.0
anglofrench	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
angri	0.0478399754196	0.0	0.0	0.0	0.0
announc	0.096116967276	0.197583974386	0.165181258004	0.0	0.0
annual	0.0478399754196	0.0	0.0	0.0	0.0
anoth	0.0	0.0	0.0	0.0470401010308	0.0
anyth	0.0	0.0	0.0	0.0	0.0394572624027
appeal	0.0	0.0	0.0	0.0	0.315658099221
appel	0.0	0.0	0.0	0.0	0.118371787208
applic	0.0	0.0	0.0	0.0	0.0789145248053
approv	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
april	0.0	0.0	0.0	0.0	0.0394572624027
argument	0.0	0.0	0.0	0.0	0.0394572624027
arrang	0.0478399754196	0.0	0.0	0.0	0.0
arriv	0.0	0.0	0.0	0.0	0.0394572624027
ask	0.0478399754196	0.0	0.0	0.0	0.0
asset	0.0	0.0	0.0	0.0470401010308	0.0
assist	0.0	0.0	0.0	0.0	0.0394572624027
attack	0.0	0.0	0.0	0.0470401010308	0.0
august	0.0	0.0	0.0	0.0470401010308	0.0
author	0.0	0.0	0.0	0.0	0.0394572624027
avail	0.0478399754196	0.0	0.0	0.0	0.0
avoid	0.0478399754196	0.0	0.0	0.0	0.0
awar	0.0	0.0	0.0	0.0	0.0394572624027
bag	0.0	0.0	0.0	0.0	0.0394572624027
bank	0.188665453663	0.166213868168	0.173694602049	0.185511006716	0.0
bankruptci	0.0478399754196	0.0	0.0	0.0	0.0
barclay	0.0	0.0	0.0	0.235200505154	0.0
basi	0.0	0.0	0.0	0.0	0.0394572624027
becom	0.0385969909435	0.0	0.0	0.0379516573231	0.0
benefit	0.0	0.0476054150652	0.0497479766054	0.0	0.0
berlusconi	0.0	0.0	0.0	0.0	0.276200836819
bettino	0.0	0.0	0.0	0.0	0.0394572624027
bifu	0.0	0.0	0.0	0.0470401010308	0.0
billion	0.134761038331	0.332427736336	0.208433522458	0.053003144776	0.0
bond	0.0478399754196	0.0	0.0	0.0	0.0
bribe	0.0	0.0	0.0	0.0	0.0394572624027
brighter	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
bring	0.032038989092	0.0790335897544	0.0825906290022	0.0	0.0
brink	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
britain	0.0	0.0	0.0	0.0759033146463	0.0636677417267
british	0.0	0.0	0.0	0.0	0.0394572624027
brother	0.0	0.0	0.0	0.0470401010308	0.0
brown	0.0	0.0	0.0	0.0	0.197286312013
buisi	0.0	0.0	0.0	0.0470401010308	0.0
busi	0.0	0.0	0.0	0.493371545201	0.0318338708633
busiest	0.0478399754196	0.0	0.0	0.0	0.0
cap	0.0478399754196	0.0	0.0	0.0	0.0
case	0.0	0.0	0.0	0.0	0.0394572624027
cash	0.0956799508391	0.0	0.0	0.0	0.0
categori	0.0	0.0	0.0	0.0940802020617	0.0
ceas	0.0	0.0	0.0	0.0470401010308	0.0
channel	0.0269522076662	0.0332427736336	0.0347389204097	0.0	0.0444590667487
charg	0.0	0.0	0.0	0.0	0.0394572624027
choos	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
chri	0.0	0.0	0.0	0.0	0.0394572624027
clean	0.0	0.0	0.0	0.0470401010308	0.0
clear	0.0	0.0	0.0	0.0470401010308	0.0
club	0.0	0.0	0.0	0.0	0.0394572624027
cochairman	0.064077978184	0.0790335897544	0.0825906290022	0.0	0.0
collaps	0.0478399754196	0.0	0.0	0.0	0.0
come	0.0771939818871	0.0	0.0	0.0379516573231	0.0
commerci	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
compani	0.323426491995	0.299184962702	0.312650283687	0.0	0.0889181334974
compet	0.0	0.0	0.0	0.0470401010308	0.0
competit	0.0	0.0	0.0	0.0470401010308	0.0
complex	0.0478399754196	0.0	0.0	0.0	0.0
compris	0.0	0.0	0.0	0.0470401010308	0.0
compromis	0.0	0.0476054150652	0.0497479766054	0.0	0.0
comput	0.0	0.0	0.0	0.0470401010308	0.0
confalonieri	0.0	0.0	0.0	0.0	0.0394572624027
confer	0.0956799508391	0.0	0.0	0.0	0.0
confirm	0.0	0.0	0.0	0.0470401010308	0.0
connect	0.0	0.0	0.0	0.0	0.0394572624027
consider	0.064077978184	0.0790335897544	0.0825906290022	0.0	0.0
consolid	0.0269522076662	0.0332427736336	0.0347389204097	0.0795047171641	0.0
constitut	0.0478399754196	0.0	0.0	0.0	0.0
construct	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
convert	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
convict	0.0	0.0	0.0	0.0	0.0394572624027
cooper	0.0	0.0	0.0	0.0470401010308	0.0
core	0.0	0.0	0.0	0.0470401010308	0.0
corrupt	0.0	0.0	0.0	0.0	0.0394572624027
cost	0.0478399754196	0.0	0.0	0.0	0.0
court	0.0	0.0	0.0	0.0	0.0789145248053
coy	0.0	0.0	0.0	0.0470401010308	0.0
craxi	0.0	0.0	0.0	0.0	0.0394572624027
creditor	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
crimin	0.0	0.0	0.0	0.0	0.0394572624027
cripl	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
crosschannel	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
current	0.0771939818871	0.0	0.0	0.0	0.0318338708633
custodi	0.0	0.0	0.0	0.329280707216	0.0
date	0.0	0.0476054150652	0.0497479766054	0.0	0.0
david	0.0	0.0	0.0	0.0	0.0394572624027
deal	0.188665453663	0.0997283209007	0.104216761229	0.026501572388	0.0
debt	0.192233934552	0.158067179509	0.165181258004	0.0	0.0
debtforequ	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
decemb	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
decid	0.0	0.0	0.0	0.0379516573231	0.0318338708633
decis	0.0	0.0	0.0	0.0379516573231	0.0318338708633
depreci	0.0478399754196	0.0	0.0	0.0	0.0
describ	0.0	0.0	0.0	0.0	0.0394572624027
despit	0.0478399754196	0.0	0.0	0.0	0.0
develop	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
dickson	0.0	0.0	0.0	0.0	0.0394572624027
difficult	0.0478399754196	0.0	0.0	0.0	0.0
dilut	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0

director	0.0	0.0	0.0	0.0	0.0394572624027
discuss	0.0	0.0	0.0	0.0940802020617	0.0
dividend	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
document	0.0	0.0	0.0	0.0	0.118371787208
domin	0.0	0.0	0.0	0.0940802020617	0.0
donat	0.0	0.0	0.0	0.0	0.0394572624027
doomsday	0.0478399754196	0.0	0.0	0.0	0.0
dwindl	0.0478399754196	0.0	0.0	0.0	0.0
earli	0.096116967276	0.0395167948772	0.0412953145011	0.0	0.0
earlier	0.0	0.0	0.0	0.0	0.0394572624027
east	0.0	0.0	0.0	0.0470401010308	0.0
econom	0.0	0.0	0.0	0.0470401010308	0.0
economi	0.0	0.0	0.0	0.0470401010308	0.0
end	0.0269522076662	0.0664855472672	0.0694778408194	0.0	0.0222295333743
english	0.0	0.0	0.0	0.0	0.0394572624027
entitl	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
equiti	0.0478399754196	0.0	0.0	0.0	0.0
european	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
eurotunnel	0.32038989092	0.434684743649	0.454248459512	0.0	0.0
exist	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
expect	0.0	0.0	0.0	0.0379516573231	0.0318338708633
expedit	0.0	0.0	0.0	0.0	0.0394572624027
explor	0.0	0.0	0.0	0.0470401010308	0.0
extend	0.0	0.0476054150652	0.0497479766054	0.0	0.0
face	0.0	0.0	0.0	0.0470401010308	0.0
fals	0.0	0.0	0.0	0.0	0.0789145248053
far	0.0	0.0	0.0	0.0470401010308	0.0
farreach	0.0	0.0	0.0	0.0	0.0394572624027
fedel	0.0	0.0	0.0	0.0	0.0394572624027
fee	0.0	0.0	0.0	0.0470401010308	0.0
ferri	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
fifti	0.0	0.0	0.0	0.0470401010308	0.0
fight	0.0	0.0	0.0	0.0	0.0394572624027
finalis	0.0269522076662	0.0332427736336	0.0347389204097	0.026501572388	0.0
financ	0.0385969909435	0.0	0.0	0.0379516573231	0.0
financi	0.064077978184	0.0790335897544	0.0825906290022	0.0	0.0
fininvest	0.0	0.0	0.0	0.0	0.0789145248053
firm	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
fish	0.0	0.0	0.0	0.0	0.0394572624027
fix	0.0478399754196	0.0	0.0	0.0	0.0
flow	0.0478399754196	0.0	0.0	0.0	0.0
follow	0.0	0.0476054150652	0.0497479766054	0.0	0.0
footbal	0.0	0.0	0.0	0.0	0.0394572624027
formula	0.0478399754196	0.0	0.0	0.0	0.0
forward	0.0385969909435	0.0	0.0	0.0	0.0636677417267
franc	0.0539044153324	0.0332427736336	0.0347389204097	0.026501572388	0.0
fraud	0.0	0.0	0.0	0.0	0.197286312013
free	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
french	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
friday	0.0	0.0	0.0	0.0	0.118371787208
frontend	0.0	0.0	0.0	0.0470401010308	0.0
futur	0.0539044153324	0.0332427736336	0.0347389204097	0.053003144776	0.0
gage	0.0	0.0	0.0	0.0	0.0394572624027
game	0.0478399754196	0.0	0.0	0.0	0.0
given	0.0	0.0	0.0	0.0	0.0394572624027
global	0.0	0.0	0.0	0.188160404123	0.0
gone	0.0	0.0	0.0	0.0	0.0394572624027
grant	0.0478399754196	0.0	0.0	0.0	0.0
ground	0.0	0.0	0.0	0.0	0.118371787208
group	0.0385969909435	0.0	0.0	0.0759033146463	0.0
guarante	0.0	0.0	0.0	0.0470401010308	0.0
guilt	0.0	0.0	0.0	0.0	0.0394572624027
half	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
halfyear	0.0	0.0	0.0	0.0470401010308	0.0
happen	0.0478399754196	0.0	0.0	0.0	0.0
happi	0.0	0.0	0.0	0.0	0.0394572624027
hear	0.0	0.0	0.0	0.0	0.197286312013
help	0.0478399754196	0.0	0.0	0.0	0.0
high	0.0	0.0	0.0	0.0	0.0394572624027
highli	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
hold	0.0385969909435	0.0	0.0	0.0	0.0318338708633
holder	0.0	0.0476054150652	0.0497479766054	0.0	0.0
home	0.0	0.0	0.0	0.0	0.0394572624027
hope	0.0478399754196	0.0	0.0	0.0	0.0
hous	0.0	0.0	0.0	0.0	0.118371787208
howev	0.0	0.0	0.0	0.0	0.0394572624027
illeg	0.0	0.0	0.0	0.0	0.0394572624027
illicit	0.0	0.0	0.0	0.0	0.0394572624027
im	0.0956799508391	0.0	0.0	0.0	0.0
immedi	0.0	0.0	0.0	0.0	0.0394572624027
imper	0.0	0.0	0.0	0.0	0.0394572624027
includ	0.0	0.0	0.0	0.141120303093	0.0
incom	0.0	0.0	0.0	0.0470401010308	0.0
individu	0.0	0.0	0.0	0.0	0.0394572624027
ingeni	0.0478399754196	0.0	0.0	0.0	0.0
insensit	0.0	0.0	0.0	0.0470401010308	0.0
insolv	0.0	0.0476054150652	0.0497479766054	0.0	0.0
inspector	0.0	0.0	0.0	0.0	0.0394572624027
instruct	0.0	0.0476054150652	0.0497479766054	0.0	0.0
instrument	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
intent	0.0	0.0	0.0	0.0470401010308	0.0
interim	0.0	0.0	0.0	0.0470401010308	0.0
intern	0.0	0.0	0.0	0.0	0.0789145248053
invest	0.0771939818871	0.0	0.0	0.113854971969	0.0
investig	0.0	0.0	0.0	0.0	0.118371787208
involv	0.0	0.0	0.0	0.0	0.0394572624027
issu	0.096116967276	0.0395167948772	0.0412953145011	0.0	0.0
itali	0.0	0.0	0.0	0.0	0.0789145248053
italian	0.0	0.0	0.0	0.0	0.276200836819
job	0.0	0.0	0.0	0.0	0.0394572624027
john	0.0	0.0	0.0	0.0470401010308	0.0
judg	0.0	0.0	0.0	0.0	0.118371787208
judgement	0.0	0.0	0.0	0.0	0.0789145248053
june	0.0	0.0	0.0	0.0470401010308	0.0
justic	0.0	0.0	0.0	0.0	0.0789145248053
justifi	0.0	0.0	0.0	0.0	0.0394572624027
larg	0.0	0.0	0.0	0.113854971969	0.0318338708633
larger	0.0	0.0	0.0	0.0	0.0394572624027
late	0.0	0.0395167948772	0.0412953145011	0.0	0.0264249884878
later	0.0	0.0	0.0	0.0470401010308	0.0

law	0.0	0.0	0.0	0.0	0.0394572624027
lawyer	0.0	0.0	0.0	0.0	0.0789145248053
lead	0.0	0.0395167948772	0.0412953145011	0.0	0.0264249884878
leak	0.0	0.0	0.0	0.0470401010308	0.0
leav	0.0269522076662	0.0332427736336	0.0347389204097	0.0	0.0222295333743
legal	0.0	0.0	0.0	0.0	0.0394572624027
lend	0.0	0.0	0.0	0.0940802020617	0.0
leonard	0.0	0.0	0.0	0.141120303093	0.0
level	0.022796004616	0.0281165250203	0.0293819563807	0.022414860184	0.0188015969485
lifelin	0.0478399754196	0.0	0.0	0.0	0.0
like	0.0478399754196	0.0	0.0	0.0	0.0
link	0.0	0.0	0.0	0.0	0.0394572624027
lloyd	0.0	0.0	0.0	0.0470401010308	0.0
london	0.0269522076662	0.0664855472672	0.0694778408194	0.0	0.0444590667487
longawait	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
longterm	0.0	0.0	0.0	0.0470401010308	0.0
look	0.032038989092	0.0	0.0	0.0315033038917	0.0264249884878
lord	0.0	0.0	0.0	0.0	0.157829049611
lost	0.0	0.0	0.0	0.0	0.0394572624027
low	0.0	0.0	0.0	0.0470401010308	0.0
main	0.0	0.0	0.0	0.0	0.0394572624027
major	0.0478399754196	0.0	0.0	0.0	0.0
make	0.022796004616	0.0281165250203	0.0293819563807	0.022414860184	0.0188015969485
mark	0.0	0.0	0.0	0.0470401010308	0.0
market	0.0269522076662	0.0332427736336	0.0347389204097	0.053003144776	0.0
massiv	0.0478399754196	0.0	0.0	0.0	0.0
maximum	0.0478399754196	0.0	0.0	0.0	0.0
mean	0.0	0.0	0.0	0.0	0.0394572624027
measur	0.0	0.0	0.0	0.0470401010308	0.0
medium	0.0385969909435	0.0	0.0497479766054	0.0	0.0
mediumterm	0.0	0.0590056849221	0.0	0.0	0.0
meet	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
milan	0.0	0.0	0.0	0.0	0.118371787208
million	0.096116967276	0.0	0.0	0.0945099116751	0.0528499769756
minist	0.0	0.0	0.0	0.0	0.118371787208
monday	0.0539044153324	0.0997283209007	0.104216761229	0.026501572388	0.0
month	0.0269522076662	0.0664855472672	0.0694778408194	0.0	0.0222295333743
morgan	0.0	0.0	0.0	0.141120303093	0.0
morn	0.0	0.0	0.0	0.0	0.0394572624027
morton	0.128155956368	0.0395167948772	0.0412953145011	0.0	0.0
mountain	0.0	0.0	0.0616613347257	0.0	0.0
mr	0.0	0.0	0.0	0.0	0.0394572624027
multifacet	0.0	0.0	0.0	0.0	0.0394572624027
natur	0.0	0.0	0.0	0.0470401010308	0.0
nearli	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
necessari	0.0	0.0	0.0	0.0470401010308	0.0
need	0.0	0.0	0.0	0.0470401010308	0.0
negoti	0.032038989092	0.0790335897544	0.0825906290022	0.0	0.0
new	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
news	0.0956799508391	0.0	0.0	0.0	0.0
nobodi	0.0956799508391	0.0	0.0	0.0	0.0
note	0.115790972831	0.0	0.0	0.0	0.0318338708633
noth	0.0	0.0	0.0	0.0	0.0394572624027
number	0.0	0.0	0.0	0.0	0.0789145248053
offenc	0.0	0.0	0.0	0.0	0.0394572624027
offer	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
offic	0.0	0.0	0.0	0.0	0.157829049611
offshor	0.0	0.0	0.0	0.0	0.0394572624027
oper	0.096116967276	0.118550384632	0.123885943503	0.0	0.0
outstand	0.0	0.0476054150652	0.0497479766054	0.0	0.0
owe	0.0	0.0476054150652	0.0497479766054	0.0	0.0
paper	0.0	0.0	0.0	0.0	0.157829049611
pari	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
particip	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
patient	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
patrick	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
pay	0.0808566229986	0.0332427736336	0.0347389204097	0.026501572388	0.0
payment	0.0539044153324	0.0332427736336	0.0347389204097	0.0	0.0222295333743
penc	0.096116967276	0.118550384632	0.123885943503	0.0	0.0
percent	0.096116967276	0.118550384632	0.123885943503	0.0	0.0
perform	0.0539044153324	0.0332427736336	0.0347389204097	0.026501572388	0.0
place	0.0	0.0	0.0	0.0379516573231	0.0318338708633
plan	0.064077978184	0.118550384632	0.123885943503	0.0	0.0
plc	0.0	0.0	0.0	0.0470401010308	0.0
point	0.0	0.0	0.0	0.0	0.0394572624027
polit	0.0	0.0	0.0	0.0	0.0394572624027
politician	0.0	0.0	0.0	0.0	0.0394572624027
ponsol	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
posit	0.0478399754196	0.0	0.0	0.0	0.0
potenti	0.0	0.0	0.0	0.0470401010308	0.0
pound	0.182368036928	0.112466100081	0.146909781904	0.11207430092	0.037603193897
presid	0.0	0.0	0.0	0.0	0.0394572624027
press	0.0	0.0	0.0	0.0470401010308	0.0
presum	0.0	0.0	0.0	0.0470401010308	0.0
presumpt	0.0	0.0	0.0	0.0	0.0394572624027
pretti	0.0478399754196	0.0	0.0	0.0	0.0
primarili	0.0	0.0	0.0	0.0470401010308	0.0
prime	0.0	0.0	0.0	0.0	0.118371787208
probabl	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
proceed	0.0	0.0	0.0	0.0	0.0394572624027
profit	0.0269522076662	0.0332427736336	0.0347389204097	0.026501572388	0.0
prosecut	0.0	0.0	0.0	0.0	0.0394572624027
prospect	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
provid	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
publish	0.0	0.0	0.0	0.0470401010308	0.0
purpos	0.0	0.0	0.0	0.0	0.0394572624027
question	0.0	0.0	0.0	0.0	0.0394572624027
rais	0.0	0.0	0.0	0.0	0.0394572624027
rake	0.0	0.0	0.0	0.0470401010308	0.0
rang	0.0	0.0	0.0	0.0470401010308	0.0
reap	0.0	0.0476054150652	0.0497479766054	0.0	0.0
recognis	0.0	0.0	0.0	0.0	0.0394572624027
reduc	0.064077978184	0.118550384632	0.123885943503	0.0	0.0
reject	0.0385969909435	0.0	0.0	0.0	0.09550161259
rel	0.0	0.0	0.0	0.0470401010308	0.0
remov	0.0	0.0	0.0	0.0	0.0394572624027
report	0.0808566229986	0.0332427736336	0.0347389204097	0.026501572388	0.0
repres	0.0	0.0	0.0	0.0	0.0394572624027
request	0.0	0.0	0.0	0.0	0.0394572624027
requir	0.0	0.0	0.0	0.0470401010308	0.0

respond	0.0	0.0	0.0	0.0470401010308	0.0
restructur	0.0269522076662	0.0997283209007	0.104216761229	0.053003144776	0.0
result	0.0	0.0	0.0	0.0470401010308	0.0
resum	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
return	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
revenu	0.0385969909435	0.0	0.0	0.0	0.0318338708633
risk	0.0	0.0	0.0	0.0470401010308	0.0
riskfre	0.0	0.0	0.0	0.0470401010308	0.0
roll	0.0478399754196	0.0	0.0	0.0	0.0
rout	0.0478399754196	0.0	0.0	0.0	0.0
royal	0.0	0.0	0.0	0.0470401010308	0.0
rule	0.0	0.0	0.0	0.0	0.118371787208
run	0.0385969909435	0.0	0.0497479766054	0.0	0.0
safekeep	0.0	0.0	0.0	0.0470401010308	0.0
said	0.205164041544	0.224932200162	0.235055651046	0.156904021288	0.169214372536
sale	0.0	0.0	0.0	0.0470401010308	0.0
salomon	0.0	0.0	0.0	0.0470401010308	0.0
say	0.0	0.0	0.0	0.0940802020617	0.0
scale	0.0	0.0	0.0	0.0470401010308	0.0
scenario	0.0478399754196	0.0	0.0	0.0	0.0
scotland	0.0	0.0	0.0	0.0470401010308	0.0
scratch	0.0	0.0	0.0	0.0470401010308	0.0
search	0.0	0.0	0.0	0.0	0.0394572624027
secret	0.0	0.0476054150652	0.0497479766054	0.0	0.0
secur	0.0385969909435	0.0	0.0	0.0759033146463	0.0
seen	0.0478399754196	0.0	0.0	0.0	0.0
seiz	0.0	0.0	0.0	0.0	0.0394572624027
senior	0.0	0.0	0.0	0.0	0.0394572624027
sent	0.0	0.0	0.0	0.0	0.0394572624027
seriou	0.0	0.0	0.0	0.0	0.0394572624027
seven	0.0385969909435	0.0	0.0	0.0	0.0318338708633
sfo	0.0	0.0	0.0	0.0	0.197286312013
share	0.064077978184	0.118550384632	0.123885943503	0.0	0.0
sharehold	0.224272923644	0.237100769263	0.247771887006	0.0	0.0
signific	0.0	0.0	0.0	0.0470401010308	0.0
silvio	0.0	0.0	0.0	0.0	0.0394572624027
simon	0.0	0.0	0.0	0.0	0.0394572624027
simultan	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
small	0.0	0.0	0.0	0.0470401010308	0.0
sold	0.0	0.0	0.0	0.0470401010308	0.0
sourc	0.0	0.0	0.0	0.0379516573231	0.0318338708633
specul	0.0	0.0	0.0	0.0470401010308	0.0
speedi	0.0	0.0	0.0	0.0	0.0394572624027
spring	0.0478399754196	0.0	0.0	0.0	0.0
stabil	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
stabilis	0.0956799508391	0.0	0.0	0.0	0.0
staff	0.0	0.0	0.0	0.0940802020617	0.0
stage	0.0	0.0	0.0	0.0470401010308	0.0
stanley	0.0	0.0	0.0	0.141120303093	0.0
state	0.0	0.0	0.0	0.0470401010308	0.0
statement	0.032038989092	0.0	0.0412953145011	0.0	0.0264249884878
stadi	0.0	0.0	0.0	0.0470401010308	0.0
stem	0.0	0.0	0.0	0.0	0.0394572624027
step	0.0	0.0	0.0	0.0940802020617	0.0
stock	0.0	0.0	0.0	0.0470401010308	0.0
stream	0.0	0.0	0.0	0.0470401010308	0.0
strengthen	0.0	0.0	0.0	0.0	0.0394572624027
strong	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
submit	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
subscrib	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
substanti	0.0	0.0476054150652	0.0497479766054	0.0	0.0
succeed	0.0	0.0	0.0	0.0	0.0394572624027
success	0.0	0.0476054150652	0.0497479766054	0.0	0.0
sum	0.0	0.0	0.0	0.0470401010308	0.0
surreptiti	0.0	0.0	0.0	0.0	0.0394572624027
suspect	0.0	0.0	0.0	0.0	0.0394572624027
suspend	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
sustain	0.0478399754196	0.0	0.0	0.0	0.0
swap	0.064077978184	0.0395167948772	0.0412953145011	0.0	0.0
syndic	0.0	0.0476054150652	0.0497479766054	0.0	0.0
taken	0.0269522076662	0.0332427736336	0.0347389204097	0.026501572388	0.0
talk	0.0	0.0	0.0	0.0470401010308	0.0
term	0.0385969909435	0.0	0.0497479766054	0.0	0.0
thought	0.0	0.0	0.0	0.0	0.0394572624027
throw	0.0478399754196	0.0	0.0	0.0	0.0
time	0.0385969909435	0.0	0.0	0.0379516573231	0.0
today	0.0	0.0	0.0	0.0	0.0394572624027
told	0.0956799508391	0.0	0.0	0.0	0.0
total	0.0	0.0	0.0	0.0470401010308	0.0
trade	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
tradit	0.0	0.0	0.0	0.0470401010308	0.0
transfer	0.0	0.0	0.0	0.0	0.118371787208
transit	0.0	0.0	0.0	0.0940802020617	0.0
treat	0.0	0.0	0.0	0.0470401010308	0.0
trial	0.0	0.0	0.0	0.0	0.0394572624027
tsb	0.0	0.0	0.0	0.0470401010308	0.0
tuesday	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
tunnel	0.064077978184	0.0790335897544	0.0825906290022	0.0	0.0
tv	0.0	0.0	0.0	0.0	0.0394572624027
tycoon	0.0	0.0	0.0	0.0	0.0394572624027
uncertain	0.0	0.0	0.0	0.0470401010308	0.0
uncertainty	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
underway	0.0	0.0	0.0	0.0	0.0394572624027
union	0.0	0.0	0.0	0.0470401010308	0.0
unit	0.0	0.0	0.0	0.141120303093	0.0
unlik	0.0	0.0	0.0	0.0379516573231	0.0318338708633
unpaid	0.0478399754196	0.0	0.0	0.0	0.0
unspecif	0.0	0.0	0.0	0.0	0.0394572624027
urg	0.0478399754196	0.0	0.0	0.0	0.0
use	0.0	0.0	0.0	0.0	0.118371787208
volum	0.0	0.0	0.0	0.0470401010308	0.0
want	0.0956799508391	0.0	0.0	0.0	0.0
warrant	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
wednesday	0.0	0.0	0.0	0.0	0.0394572624027
week	0.032038989092	0.118550384632	0.123885943503	0.0	0.0
welcom	0.0	0.0	0.0	0.0	0.0394572624027
wide	0.0269522076662	0.0332427736336	0.0347389204097	0.0	0.0222295333743
widerang	0.0	0.0	0.0	0.0	0.0394572624027
wipe	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
work	0.0269522076662	0.0332427736336	0.0347389204097	0.026501572388	0.0

worker	0.0	0.0	0.0	0.0470401010308	0.0
worldwid	0.0	0.0	0.0	0.0470401010308	0.0
worth	0.0	0.0	0.0	0.0470401010308	0.0
wrangl	0.0	0.0476054150652	0.0497479766054	0.0	0.0
wrestl	0.032038989092	0.0395167948772	0.0412953145011	0.0	0.0
year	0.161713245997	0.0332427736336	0.0347389204097	0.0	0.0444590667487

cosine similarity for documents

	100554newsML.txt	100593newsML.txt	100618newsML.txt	130040newsML.txt	137871newsML.txt
100554newsML.txt	1.0	0.747076672441	0.772366997261	0.166756995412	0.109543969878
100593newsML.txt	0.747076672441	1.0	0.9746691127	0.122279601282	0.0838630885123
100618newsML.txt	0.772366997261	0.9746691127	1.0	0.123710873913	0.0898335710639
130040newsML.txt	0.166756995412	0.122279601282	0.123710873913	1.0	0.0688468107264
137871newsML.txt	0.109543969878	0.0838630885123	0.0898335710639	0.0688468107264	1.0

Process finished with exit code 0