

Exploración

In [1]:

```
!ls
```

```
clase_11b_bayes_word2vec.pdf  requirements.txt          tarea
3.ipynb
data                          Reuters21578-Apte-90Cat.tar.gz  tarea
3.pdf
```

In [2]:

```
!ls data
```

```
test  training
```

Training

In [3]:

```
!ls data/training/
```

acq	cpu	income	naphtha	rape-oil
sun-oil				
alum	crude	instal-debt	nat-gas	rapeseed
sunseed				
barley	dfl	interest	nickel	reserves
tea				
bop	dlr	ipi	nkr	retail
tin				
carcass	dmk	iron-steel	nzdlr	rice
trade				
castor-oil	earn	jet	oat	rubber
unknown				
cocoa	fuel	jobs	oilseed	rye
veg-oil				
coconut	gas	l-cattle	orange	ship
wheat				
coconut-oil	gnp	lead	palladium	silver
wpi				
coffee	gold	lei	palmkernel	sorghum
yen				
copper	grain	lin-oil	palm-oil	soybean
zinc				
copra-cake	groundnut	livestock	pet-chem	soy-meal
corn	groundnut-oil	lumber	platinum	soy-oil
cotton	heat	meal-feed	potato	strategic-meta
l				
cotton-oil	hog	money-fx	propane	sugar
cpi	housing	money-supply	rand	sun-meal

Dataset: Acquisition

In [4]:

```
!ls data/training/acq
```

0000005	0001151	0002088	0003205	0004272	0005419	0006378	00075
98	0008473						
0000007	0001152	0002090	0003210	0004284	0005428	0006384	00076
01	0008483						
0000027	0001160	0002093	0003211	0004300	0005436	0006388	00076
17	0008493						
0000028	0001166	0002095	0003237	0004306	0005441	0006393	00076
18	0008497						
0000044	0001169	0002102	0003241	0004311	0005442	0006394	00076
27	0008510						
0000065	0001171	0002107	0003260	0004319	0005444	0006396	00076
30	0008513						
0000078	0001183	0002112	0003284	0004323	0005446	0006425	00076
37	0008514						
0000087	0001204	0002117	0003294	0004329	0005448	0006428	00076
38	0008527						
0000090	0001256	0002120	0003299	0004330	0005449	0006431	00076
41	0008541						
0000095	0001257	0002122	0003301	0004345	0005457	0006446	00076
42	0008545						
0000096	0001260	0002123	0003310	0004363	0005460	0006474	00076
53	0008549						
0000112	0001262	0002124	0003312	0004364	0005461	0006489	00076
56	0008550						
0000114	0001271	0002125	0003332	0004365	0005462	0006492	00076
57	0008562						
0000117	0001273	0002131	0003338	0004368	0005463	0006496	00076
59	0008582						
0000132	0001278	0002132	0003339	0004389	0005465	0006499	00076
65	0008587						
0000133	0001281	0002139	0003344	0004398	0005466	0006502	00076
69	0008588						
0000143	0001287	0002140	0003348	0004401	0005469	0006505	00076
71	0008593						
0000185	0001296	0002141	0003366	0004404	0005470	0006509	00076
74	0008594						
0000209	0001303	0002142	0003372	0004408	0005476	0006510	00076
76	0008595						
0000212	0001308	0002143	0003385	0004422	0005483	0006514	00076
77	0008600						
0000214	0001316	0002149	0003387	0004428	0005489	0006525	00076
83	0008610						
0000224	0001320	0002150	0003388	0004432	0005495	0006531	00076
84	0008624						
0000234	0001321	0002155	0003389	0004440	0005498	0006533	00076
88	0008625						
0000236	0001326	0002157	0003391	0004452	0005501	0006535	00076
96	0008626						
0000254	0001329	0002161	0003392	0004458	0005503	0006541	00077
02	0008630						
0000255	0001333	0002163	0003398	0004467	0005505	0006544	00077
07	0008638						
0000256	0001334	0002165	0003401	0004471	0005508	0006560	00077
10	0008642						
0000258	0001343	0002168	0003402	0004472	0005524	0006566	00077
11	0008645						
0000259	0001349	0002170	0003409	0004474	0005528	0006572	00077
12	0008652						
0000260	0001373	0002174	0003418	0004476	0005529	0006574	00077
16	0008654						
0000262	0001375	0002180	0003420	0004481	0005530	0006582	00077

17	0008661						
0000265	0001380	0002190	0003430	0004482	0005531	0006587	00077
21	0008667						
0000268	0001384	0002193	0003433	0004483	0005535	0006589	00077
25	0008669						
0000269	0001392	0002202	0003436	0004487	0005537	0006600	00077
26	0008670						
0000271	0001393	0002210	0003439	0004500	0005548	0006601	00077
27	0008673						
0000275	0001396	0002215	0003440	0004507	0005549	0006608	00077
28	0008674						
0000279	0001397	0002218	0003445	0004511	0005550	0006611	00077
30	0008683						
0000281	0001407	0002220	0003449	0004519	0005552	0006619	00077
31	0008696						
0000282	0001410	0002222	0003451	0004527	0005556	0006620	00077
34	0008717						
0000289	0001412	0002225	0003462	0004528	0005571	0006626	00077
45	0008722						
0000298	0001415	0002226	0003474	0004529	0005574	0006634	00077
55	0008739						
0000304	0001417	0002228	0003491	0004531	0005575	0006635	00077
59	0008746						
0000309	0001420	0002231	0003517	0004538	0005577	0006637	00077
61	0008749						
0000313	0001428	0002234	0003519	0004549	0005578	0006638	00077
64	0008760						
0000314	0001431	0002238	0003520	0004557	0005579	0006644	00077
69	0008768						
0000315	0001441	0002239	0003521	0004558	0005580	0006653	00077
72	0008776						
0000318	0001446	0002240	0003528	0004573	0005587	0006662	00077
76	0008781						
0000319	0001450	0002242	0003530	0004577	0005592	0006665	00077
78	0008782						
0000333	0001452	0002243	0003532	0004604	0005593	0006678	00077
82	0008787						
0000336	0001462	0002246	0003535	0004612	0005598	0006684	00077
84	0008793						
0000337	0001463	0002247	0003539	0004624	0005599	0006686	00077
90	0008805						
0000339	0001464	0002249	0003540	0004631	0005602	0006690	00077
94	0008811						
0000352	0001478	0002250	0003543	0004632	0005606	0006691	00077
95	0008814						
0000353	0001492	0002251	0003544	0004643	0005613	0006699	00078
05	0008821						
0000354	0001499	0002252	0003548	0004644	0005620	0006709	00078
13	0008825						
0000357	0001500	0002253	0003562	0004648	0005621	0006710	00078
19	0008830						
0000358	0001501	0002257	0003575	0004653	0005623	0006713	00078
20	0008844						
0000378	0001526	0002258	0003586	0004654	0005636	0006722	00078
22	0008870						
0000388	0001532	0002283	0003604	0004661	0005660	0006726	00078
24	0008872						
0000390	0001542	0002299	0003612	0004668	0005665	0006730	00078
26	0008878						
0000392	0001549	0002312	0003622	0004675	0005670	0006733	00078
27	0008881						

0000393 37	0001558 0008886	0002361	0003627	0004677	0005682	0006742	00078
0000395 48	0001564 0008894	0002371	0003653	0004678	0005688	0006752	00078
0000399 57	0001566 0008906	0002375	0003664	0004680	0005690	0006761	00078
0000400 65	0001567 0008913	0002378	0003666	0004684	0005693	0006762	00078
0000406 66	0001568 0008921	0002388	0003667	0004686	0005720	0006772	00078
0000408 72	0001580 0008929	0002389	0003670	0004689	0005754	0006776	00078
0000428 75	0001585 0008941	0002392	0003672	0004695	0005755	0006782	00078
0000430 93	0001588 0008942	0002397	0003677	0004701	0005764	0006784	00078
0000433 94	0001592 0008944	0002400	0003679	0004705	0005767	0006785	00078
0000434 96	0001596 0008949	0002408	0003680	0004734	0005769	0006787	00078
0000436 99	0001602 0008950	0002422	0003683	0004751	0005771	0006789	00078
0000441 05	0001606 0008953	0002431	0003692	0004752	0005772	0006790	00079
0000442 06	0001608 0008961	0002443	0003698	0004766	0005788	0006801	00079
0000444 08	0001612 0009000	0002449	0003700	0004767	0005792	0006804	00079
0000446 10	0001613 0009007	0002468	0003701	0004770	0005795	0006810	00079
0000447 16	0001616 0009012	0002479	0003704	0004771	0005805	0006842	00079
0000449 22	0001618 0009033	0002501	0003709	0004774	0005808	0006843	00079
0000450 26	0001621 0009050	0002505	0003716	0004776	0005809	0006847	00079
0000453 31	0001624 0009053	0002533	0003717	0004777	0005813	0006851	00079
0000456 34	0001629 0009058	0002536	0003719	0004778	0005817	0006864	00079
0000457 36	0001630 0009062	0002539	0003723	0004788	0005842	0006883	00079
0000460 37	0001640 0009073	0002551	0003727	0004790	0005859	0006906	00079
0000463 40	0001646 0009074	0002558	0003735	0004794	0005860	0006909	00079
0000464 55	0001648 0009076	0002567	0003741	0004797	0005861	0006915	00079
0000466 66	0001650 0009080	0002571	0003753	0004798	0005863	0006923	00079
0000467 69	0001651 0009089	0002582	0003754	0004799	0005864	0006931	00079
0000476 70	0001654 0009094	0002587	0003760	0004811	0005868	0006935	00079
0000480 78	0001655 0009095	0002589	0003762	0004814	0005869	0006942	00079
0000488 82	0001665 0009096	0002609	0003778	0004815	0005870	0006950	00079
0000498 86	0001670 0009102	0002616	0003783	0004823	0005872	0006954	00079
0000501	0001673	0002620	0003784	0004824	0005878	0006967	00079

87	0009105						
0000505	0001693	0002625	0003794	0004825	0005883	0006969	00079
91	0009120						
0000508	0001695	0002640	0003801	0004826	0005889	0006983	00079
98	0009122						
0000509	0001696	0002647	0003803	0004840	0005902	0006986	00080
03	0009132						
0000510	0001700	0002651	0003811	0004844	0005903	0006988	00080
06	0009138						
0000512	0001703	0002655	0003829	0004851	0005904	0006990	00080
11	0009144						
0000531	0001704	0002660	0003836	0004852	0005906	0006991	00080
13	0009145						
0000548	0001708	0002684	0003837	0004853	0005907	0006993	00080
30	0009148						
0000573	0001715	0002688	0003839	0004862	0005909	0006994	00080
36	0009152						
0000578	0001737	0002690	0003849	0004867	0005911	0007004	00080
43	0009155						
0000580	0001746	0002694	0003867	0004873	0005925	0007007	00080
52	0009156						
0000595	0001754	0002696	0003892	0004923	0005927	0007011	00080
53	0009158						
0000597	0001757	0002698	0003893	0004926	0005946	0007024	00080
55	0009163						
0000606	0001778	0002704	0003900	0004933	0005947	0007029	00080
56	0009168						
0000608	0001782	0002710	0003901	0004934	0005955	0007030	00080
61	0009169						
0000611	0001783	0002711	0003905	0004936	0005957	0007031	00080
67	0009178						
0000620	0001793	0002716	0003917	0004937	0005959	0007034	00080
70	0009179						
0000635	0001794	0002725	0003923	0004944	0005961	0007048	00080
73	0009182						
0000640	0001800	0002737	0003924	0004951	0005974	0007055	00080
83	0009183						
0000643	0001803	0002738	0003929	0004954	0005978	0007057	00080
84	0009189						
0000648	0001805	0002740	0003934	0004972	0005979	0007065	00080
87	0009192						
0000669	0001806	0002746	0003936	0004974	0005981	0007069	00080
89	0009198						
0000670	0001809	0002758	0003939	0004986	0005988	0007072	00080
90	0009200						
0000686	0001814	0002759	0003940	0004987	0005990	0007077	00080
93	0009203						
0000690	0001815	0002764	0003941	0004988	0005996	0007082	00080
94	0009205						
0000691	0001817	0002765	0003944	0005002	0005998	0007083	00080
95	0009206						
0000700	0001819	0002766	0003945	0005007	0005999	0007086	00080
96	0009210						
0000704	0001821	0002771	0003946	0005016	0006035	0007088	00081
02	0009215						
0000722	0001831	0002796	0003947	0005028	0006037	0007092	00081
12	0009227						
0000727	0001839	0002797	0003955	0005045	0006038	0007098	00081
14	0009235						
0000730	0001854	0002804	0003959	0005059	0006047	0007099	00081
19	0009243						

0000757	0001864	0002811	0003961	0005060	0006075	0007108	00081
32	0009249						
0000761	0001870	0002813	0003969	0005063	0006093	0007111	00081
35	0009252						
0000765	0001871	0002816	0003970	0005067	0006097	0007115	00081
41	0009257						
0000768	0001872	0002821	0003973	0005070	0006098	0007130	00081
42	0009263						
0000770	0001874	0002826	0003982	0005076	0006099	0007131	00081
43	0009269						
0000774	0001878	0002860	0003988	0005078	0006107	0007137	00081
44	0009270						
0000778	0001880	0002875	0003993	0005083	0006112	0007141	00081
55	0009274						
0000780	0001885	0002881	0003994	0005099	0006114	0007150	00081
59	0009275						
0000806	0001891	0002890	0004002	0005103	0006118	0007154	00081
82	0009276						
0000808	0001893	0002891	0004003	0005111	0006125	0007155	00081
85	0009277						
0000814	0001894	0002893	0004005	0005115	0006129	0007169	00081
96	0009278						
0000815	0001900	0002897	0004006	0005126	0006140	0007181	00082
18	0009283						
0000822	0001904	0002901	0004007	0005130	0006145	0007186	00082
28	0009284						
0000832	0001905	0002902	0004008	0005131	0006147	0007187	00082
42	0009286						
0000844	0001908	0002904	0004013	0005135	0006156	0007189	00082
46	0009287						
0000857	0001909	0002906	0004024	0005136	0006159	0007198	00082
47	0009288						
0000896	0001911	0002908	0004031	0005137	0006160	0007244	00082
49	0009319						
0000897	0001912	0002920	0004046	0005143	0006161	0007245	00082
51	0009339						
0000904	0001913	0002929	0004084	0005147	0006170	0007247	00082
55	0009365						
0000907	0001915	0002931	0004095	0005159	0006188	0007264	00082
60	0009388						
0000936	0001918	0002934	0004096	0005164	0006189	0007275	00082
61	0009392						
0000938	0001954	0002936	0004097	0005173	0006195	0007286	00082
81	0009393						
0000947	0001963	0002946	0004104	0005191	0006198	0007293	00082
82	0009403						
0000950	0001965	0002948	0004109	0005204	0006205	0007314	00082
84	0009405						
0000953	0001972	0002951	0004110	0005211	0006208	0007318	00082
96	0009407						
0000956	0001974	0002952	0004112	0005215	0006209	0007327	00083
00	0009408						
0000990	0001979	0002953	0004117	0005216	0006210	0007331	00083
04	0009421						
0001002	0001982	0002961	0004122	0005221	0006224	0007334	00083
09	0009422						
0001005	0001983	0002979	0004124	0005225	0006227	0007338	00083
10	0009444						
0001007	0001985	0002983	0004134	0005227	0006242	0007343	00083
13	0009452						
0001016	0001986	0002984	0004137	0005232	0006244	0007344	00083

17	0009456						
0001020	0001995	0002986	0004143	0005235	0006249	0007349	00083
21	0009460						
0001028	0001997	0002999	0004144	0005238	0006250	0007352	00083
25	0009461						
0001031	0002000	0003001	0004145	0005244	0006251	0007358	00083
28	0009465						
0001032	0002002	0003007	0004149	0005265	0006257	0007359	00083
33	0009470						
0001035	0002008	0003012	0004163	0005273	0006266	0007361	00083
39	0009476						
0001044	0002009	0003016	0004164	0005280	0006267	0007364	00083
45	0009484						
0001053	0002010	0003025	0004166	0005289	0006269	0007368	00083
48	0009489						
0001056	0002012	0003028	0004173	0005309	0006283	0007374	00083
51	0009498						
0001057	0002019	0003041	0004182	0005311	0006293	0007391	00083
54	0009502						
0001058	0002021	0003047	0004193	0005317	0006294	0007395	00083
55	0009508						
0001060	0002022	0003048	0004196	0005318	0006296	0007403	00083
56	0009512						
0001063	0002025	0003050	0004197	0005319	0006306	0007414	00083
59	0009515						
0001065	0002029	0003067	0004199	0005331	0006313	0007422	00083
69	0009518						
0001066	0002030	0003071	0004201	0005338	0006315	0007425	00083
71	0009519						
0001070	0002031	0003074	0004215	0005340	0006322	0007430	00083
72	0009530						
0001075	0002032	0003086	0004217	0005344	0006327	0007438	00083
75	0009541						
0001076	0002035	0003094	0004222	0005347	0006331	0007442	00083
85	0009547						
0001088	0002039	0003108	0004225	0005352	0006332	0007444	00083
89	0009549						
0001095	0002043	0003119	0004226	0005357	0006337	0007446	00083
90	0009551						
0001099	0002049	0003128	0004227	0005362	0006339	0007452	00083
94	0009554						
0001101	0002052	0003130	0004234	0005364	0006342	0007506	00083
95	0009556						
0001102	0002054	0003140	0004237	0005368	0006344	0007508	00083
98	0009557						
0001103	0002063	0003147	0004240	0005369	0006345	0007521	00084
00	0009565						
0001112	0002065	0003158	0004242	0005377	0006354	0007523	00084
28	0009576						
0001120	0002072	0003159	0004243	0005381	0006358	0007538	00084
33	0009584						
0001127	0002073	0003181	0004245	0005384	0006361	0007545	00084
38							
0001128	0002074	0003186	0004247	0005391	0006362	0007562	00084
50							
0001129	0002076	0003188	0004249	0005401	0006367	0007584	00084
51							
0001134	0002083	0003189	0004262	0005405	0006372	0007595	00084
63							
0001137	0002084	0003197	0004263	0005409	0006376	0007596	00084
64							

0001141 0002087 0003198 0004265 0005417 0006377 0007597 0008465

Algún archivo

In [1]:

```
!cat data/training/acq/0000005
```

Dataset: CPU

In [6]:

```
!ls data/training/cpu
```

0003746 0003797 0003816

Buscar más Frecuentes

In [7]:

```
import os
trainingFiles = []
for root, dirs, files in os.walk("data/training"):
    for file in files:
        trainingFiles.append([root, file])

trainingFiles[0:10]
```

Out[7]:

```
[['data/training/napththa', '0004556'],
 ['data/training/napththa', '0005167'],
 ['data/training/soybean', '0004358'],
 ['data/training/soybean', '0008468'],
 ['data/training/soybean', '0000885'],
 ['data/training/soybean', '0002152'],
 ['data/training/soybean', '0001505'],
 ['data/training/soybean', '0001131'],
 ['data/training/soybean', '0005250'],
 ['data/training/soybean', '0006770']]
```

Agrupar

In [8]:

```
import pandas as pd
```

In [9]:

```
labels = ['root', 'file']
df = pd.DataFrame.from_records(trainingFiles, columns=labels)
```

In [10]:

```
df[0:10]
```

Out[10]:

	root	file
0	data/training/naphtha	0004556
1	data/training/naphtha	0005167
2	data/training/soybean	0004358
3	data/training/soybean	0008468
4	data/training/soybean	0000885
5	data/training/soybean	0002152
6	data/training/soybean	0001505
7	data/training/soybean	0001131
8	data/training/soybean	0005250
9	data/training/soybean	0006770

In [11]:

```
df.describe()
```

Out[11]:

	root	file
count	11413	11413
unique	91	9598
top	data/training/earn	0000167
freq	2877	15

GroupBy

In [12]:

```
df.groupby('root').describe()
```

Out[12]:

	file			
	count	unique	top	freq
root				
data/training/acq	1650	1650	0008228	1
data/training/alum	35	35	0006562	1
data/training/barley	37	37	0005296	1
data/training/bop	75	75	0007232	1
data/training/carcass	50	50	0003480	1
data/training/castor-oil	1	1	0007237	1
data/training/cocoa	55	55	0006724	1
data/training/coconut	4	4	0007537	1
data/training/coconut-oil	4	4	0006017	1
data/training/coffee	111	111	0007110	1
data/training/copper	47	47	0004030	1
data/training/copra-cake	2	2	0000167	1
data/training/corn	181	181	0002996	1
data/training/cotton	39	39	0001012	1
data/training/cotton-oil	1	1	0004459	1
data/training/cpi	69	69	0008538	1
data/training/cpu	3	3	0003797	1
data/training/crude	389	389	0008210	1
data/training/dfi	2	2	0007273	1
data/training/dlr	131	131	0007511	1
data/training/dmk	10	10	0002014	1
data/training/earn	2877	2877	0002387	1
data/training/fuel	13	13	0001919	1
data/training/gas	37	37	0008210	1
data/training/gnp	101	101	0007956	1
data/training/gold	94	94	0001527	1
data/training/grain	433	433	0003501	1
data/training/groundnut	5	5	0001490	1
data/training/groundnut-oil	1	1	0000222	1
data/training/heat	14	14	0008363	1
...

data/training/potato	3	3	0001242	1
	file			
	count	unique	top	freq
root				
data/training/propane	3	3	0000493	1
data/training/rand	2	2	0006570	1
data/training/rape-oil	5	5	0005713	1
data/training/rapeseed	18	18	0003141	1
data/training/reserves	55	55	0004016	1
data/training/retail	23	23	0003643	1
data/training/rice	35	35	0008181	1
data/training/rubber	37	37	0007882	1
data/training/rye	1	1	0000066	1
data/training/ship	197	197	0007474	1
data/training/silver	21	21	0008689	1
data/training/sorghum	24	24	0003919	1
data/training/soy-meal	13	13	0000045	1
data/training/soy-oil	14	14	0009255	1
data/training/soybean	78	78	0000885	1
data/training/strategic-metal	16	16	0001906	1
data/training/sugar	126	126	0008590	1
data/training/sun-meal	1	1	0008798	1
data/training/sun-oil	5	5	0002727	1
data/training/sunseed	11	11	0008613	1
data/training/tea	9	9	0007320	1
data/training/tin	18	18	0005895	1
data/training/trade	369	369	0004004	1
data/training/unknown	1830	1830	0000345	1
data/training/veg-oil	87	87	0000226	1
data/training/wheat	212	212	0002815	1
data/training/wpi	19	19	0003218	1
data/training/yen	45	45	0007586	1
data/training/zinc	21	21	0004895	1

91 rows × 4 columns

Ordernar y sacar los 10

Los 11 en realidad, porque hay una categoría unknown

In [13]:

```
largest = df.groupby('root').count().nlargest(11, columns="file")
largest
```

Out[13]:

	file
root	
data/training/learn	2877
data/training/unknown	1830
data/training/acq	1650
data/training/money-fx	538
data/training/grain	433
data/training/crude	389
data/training/trade	369
data/training/interest	347
data/training/wheat	212
data/training/ship	197
data/training/corn	181

In [14]:

```
largest.describe()
```

Out[14]:

	file
count	11.000000
mean	820.272727
std	891.546195
min	181.000000
25%	279.500000
50%	389.000000
75%	1094.000000
max	2877.000000

Iterate through values:

In [15]:

```
for row in df.iterrows():
    print(row[0], row[1]["root"], row[1]["file"])
    break
```

0 data/training/naphtha 0004556

In [16]:

```
for row in largest.iterrows():
    print(row[0], row[1]["file"])
```

```
data/training/earn 2877
data/training/unknown 1830
data/training/acq 1650
data/training/money-fx 538
data/training/grain 433
data/training/crude 389
data/training/trade 369
data/training/interest 347
data/training/wheat 212
data/training/ship 197
data/training/corn 181
```

Refactor

In [17]:

```
path = "data/training"
def getAllFiles(path):
    allFiles = []
    for root, dirs, files in os.walk(path):
        for file in files:
            allFiles.append([root, file])

    return allFiles
```

In [18]:

```
def getTop10Categories(files):
    labels = ['root', 'file']
    df = pd.DataFrame.from_records(files, columns=labels)
    largest = df.groupby('root').count().nlargest(11, columns="file")
    return [[row[0], row[1]["file"]] for row in largest.iterrows() if "unknown"
not in row[0]]
```


In [19]:

```
def getFilesInFolders(folderList):
    allFiles = []
    for folder in folderList:
        for root, dirs, files in os.walk(folder):
            for file in files:
                allFiles.append([root, file])

    return allFiles
```

In [20]:

```
def concatFiles(allFiles):
    return [str(r[0]) + "/" + str(r[1]) for r in allFiles]
```

Preview

All Files

In [21]:

```
trainingFiles = getAllFiles("data/training")
print(len(trainingFiles))
print(trainingFiles[0])
```

```
11413
['data/training/naphtha', '0004556']
```

Top 10 Categories

In [22]:

```
top10 = getTop10Categories(trainingFiles)
print(len(top10))
print(top10[0])
```

```
10
['data/training/earn', 2877]
```

Top 10 Folders

In [23]:

```
top10Folders = [f[0] for f in top10]
top10Folders
```

Out[23]:

```
['data/training/earn',
 'data/training/acq',
 'data/training/money-fx',
 'data/training/grain',
 'data/training/crude',
 'data/training/trade',
 'data/training/interest',
 'data/training/wheat',
 'data/training/ship',
 'data/training/corn']
```

Top 10 Files

In [24]:

```
top10Files = getFilesInFolders(top10Folders)
print(len(top10Files))
print("\n".join(concatFiles(top10Files)[:500]))
```

```
7193
data/training/earn/0006903
data/training/earn/0001638
data/training/earn/0000416
data/training/earn/0003049
data/training/earn/0001883
data/training/earn/0006551
data/training/acq/0004263
data/training/acq/0007759
data/training/acq/0005147
data/training/acq/0004862
data/training/money-fx/0007282
data/training/crude/0005910
data/training/trade/0003619
data/training/interest/0000568
data/training/ship/0009439
```

Check 'earn' Length

In [25]:

```
earnList = [r for r in top10Files if "earn" in r[0]]
len(earnList)
```

Out[25]:

```
2877
```

Eliminar signos de puntuación

In [26]:

```
testFile = os.path.join(trainingFiles[0][0], trainingFiles[0][1])
testFile
```

Out[26]:

```
'data/training/naphtha/0004556'
```

In [27]:

```
testFileOpen = open(testFile, "r")
testText = testFileOpen.read()
testFileOpen.close()
print(testText)
```

PAKISTAN-SWEDISH GOODS EXCHANGE AGREED

ISLAMABAD, March 18 - Pakistan and Sweden have signed a commodity exchange agreement for 88 mln dlrs each way, the Pakistan government announced.

Pakistan's exports under the agreement will include raw cotton, cotton products, cotton textiles, steel products, molasses, naphtha and fresh and dried fruits.

Swedish exports to Pakistan will include medical and laboratory equipment, electrical telecommunication equipment, diesel engine spares, mining and security equipment, road-building and construction machinery, fertilisers and palm oil.

Fuerza bruta

In [28]:

```
puntuacion = [",", ".", ";", ":", "<", ">", "-"]
```

In [29]:

```
sinPuntos = testText

for punto in puntuacion:
    sinPuntos = sinPuntos.replace(punto, "")

print(sinPuntos)
```

PAKISTANSWEDISH GOODS EXCHANGE AGREED

ISLAMABAD March 18 Pakistan and Sweden have signed a commodity exchange agreement for 88 mln dlrs each way the Pakistan government announced

Pakistan's exports under the agreement will include raw cotton cotton products cotton textiles steel products molasses naphtha and fresh and dried fruits

Swedish exports to Pakistan will include medical and laboratory equipment electrical telecommunication equipment diesel engine spares mining and security equipment roadbuilding and construction machinery fertilisers and palm oil

Problemas de este método:

- Tener todos los signos de puntuación en el arreglo
- Eficiencia

Expresiones regulares

In [30]:

```
import re, string
```

In [31]:

```
sinPuntos = testText

regex = re.compile('[%s]' % re.escape(string.punctuation))
sinPuntos = regex.sub('', sinPuntos)

print(sinPuntos)
```

PAKISTANSWEDISH GOODS EXCHANGE AGREED

ISLAMABAD March 18 Pakistan and Sweden have signed a commodity exchange agreement for 88 mln dlrs each way the Pakistan government announced

Pakistans exports under the agreement will include raw cotton cotton products cotton textiles steel products molasses naphtha and fresh and dried fruits

Swedish exports to Pakistan will include medical and laboratory equipment electrical telecommunication equipment diesel engine spares mining and security equipment roadbuilding and construction machinery fertilisers and palm oil

String Translate

In [32]:

```
string.punctuation
```

Out[32]:

```
'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

In [33]:

```
sinPuntos = testText
sinPuntos = sinPuntos.translate(str.maketrans('', '', string.punctuation))
print(sinPuntos)
```

PAKISTANSWEDISH GOODS EXCHANGE AGREED

ISLAMABAD March 18 Pakistan and Sweden have signed a commodity exchange agreement for 88 mln dlrs each way the Pakistan government announced

Pakistans exports under the agreement will include raw cotton cotton products cotton textiles steel products molasses naphtha and fresh and dried fruits

Swedish exports to Pakistan will include medical and laboratory equipment electrical telecommunication equipment diesel engine spares mining and security equipment roadbuilding and construction machinery fertilisers and palm oil

Comparación

Source (<https://stackoverflow.com/questions/265960/best-way-to-strip-punctuation-from-a-string-in-python>)

```
replace    : 28.4436721802
regex      : 6.86155414581
translate  : 2.12455511093
```

Eliminar números

In [34]:

```
sinNumeros = re.sub('\d', '', sinPuntos)
print(sinNumeros)
```

PAKISTANSWEDISH GOODS EXCHANGE AGREED

ISLAMABAD March Pakistan and Sweden have signed a commodity exchange agreement for mln dlrs each way the Pakistan government announced

Pakistans exports under the agreement will include raw cotton cotton products cotton textiles steel products molasses naphtha and fresh and dried fruits

Swedish exports to Pakistan will include medical and laboratory equipment electrical telecommunication equipment diesel engine spares mining and security equipment roadbuilding and construction machinery fertilisers and palm oil

Eliminar Stopwords

Al igual que en el caso anterior, podríamos crear un arreglo y eliminar los stopwords. Una vez más, tendríamos que construir (o [descargar \(https://github.com/stopwords-iso/stopwords-en\)](https://github.com/stopwords-iso/stopwords-en)) un listado de estas.

Para este caso, ocuparemos el [Natural Language Toolkit \(http://www.nltk.org/\)](http://www.nltk.org/) para hacer esto.

In [35]:

```
from nltk.corpus import stopwords
```

In [36]:

```
stop = set(stopwords.words('english'))
list(stop)[0:10]
```

Out[36]:

```
['have',
 'into',
 'them',
 'which',
 'through',
 'those',
 'isn',
 'until',
 'that',
 'but']
```

In [37]:

```
sinStop = [i for i in sinNumeros.lower().split() if i not in stop]
print(" ".join(sinStop))
```

pakistanswedish goods exchange agreed islamabad march pakistan swede
n signed commodity exchange agreement mln dlrs way pakistan governme
nt announced pakistans exports agreement include raw cotton cotton p
roducts cotton textiles steel products molasses naphtha fresh dried
fruits swedish exports pakistan include medical laboratory equipment
electrical telecommunication equipment diesel engine spares mining s
ecurity equipment roadbuilding construction machinery fertilisers pa
lm oil

Tokenize

In [38]:

```
from nltk import word_tokenize
import nltk
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /home/clgadel/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

Out[38]:

True

In [39]:

```
tokenized = [i for i in word_tokenize(sinNumeros.lower()) if i not in stop]
```

In [40]:

```
tokenized[0:20]
```

Out[40]:

```
['pakistanswedish',  
'goods',  
'exchange',  
'agreed',  
'islamabad',  
'march',  
'pakistan',  
'sweden',  
'signed',  
'commodity',  
'exchange',  
'agreement',  
'mln',  
'dlrs',  
'way',  
'pakistan',  
'government',  
'announced',  
'pakistans',  
'exports']
```

Stemming

In [41]:

```
from nltk.stem.porter import PorterStemmer  
porter = PorterStemmer()
```

In [42]:

```
stemText = [porter.stem(i.lower()) for i in word_tokenize(sinNumeros) if i.lower()  
( ) not in stop]
```


In [43]:

```
stemText[0:20]
```

Out[43]:

```
['pakistanswedish',  
'good',  
'exchang',  
'agre',  
'islamabad',  
'march',  
'pakistan',  
'sweden',  
'sign',  
'commod',  
'exchang',  
'agreement',  
'mln',  
'dlr',  
'way',  
'pakistan',  
'govern',  
'announc',  
'pakistan',  
'export']
```

Refactor

In [44]:

```
def processDocument(document):  
    sinPuntos = document.translate(str.maketrans('', '', string.punctuation))  
    sinNumeros = re.sub('\d', '', sinPuntos)  
    stemText = [porter.stem(i.lower()) for i in word_tokenize(sinNumeros) if i.lower() not in stop]  
    return stemText
```

In [45]:

```
def readFile(filename):  
    file = open(filename)  
    try:  
        read = file.read()  
    except:  
        print(filename)  
    file.close()  
    return read
```

In [46]:

```
def readProcessDocument(documentFilename):  
    read = readFile(documentFilename)  
    parsed = processDocument(read)  
    return parsed
```

Test

In [47]:

```
processDocument(testText)[0:20]
```

Out[47]:

```
['pakistanswedish',  
'good',  
'exchang',  
'agre',  
'islamabad',  
'march',  
'pakistan',  
'sweden',  
'sign',  
'commod',  
'exchang',  
'agreement',  
'mln',  
'dlr',  
'way',  
'pakistan',  
'govern',  
'announc',  
'pakistan',  
'export']
```

Refactor Extract Files

Ahora con la carpeta y una lista de las categorías. Tuve que cambiarlos porque los test podrían no tener los mismos Top 10 de frecuencia.

In [48]:

```
# Usage: getFilesInFolders("data/training", ["acq", "earn", "money-fx"])  
def getFilesInFolders(path, folderList):  
    allFiles = []  
    for root, dirs, files in os.walk(path):  
        if any(word in root for word in folderList):  
            for file in files:  
                allFiles.append([root, file])  
  
    return allFiles
```

In [49]:

```
categories = ["acq", "earn", "money-fx", "grain", "crude", "trade", "interest",  
"ship", "wheat", "corn"]
```

Test

In [50]:

```
getFilesInFolders("data/training", categories)[:500]
```

Out[50]:

```
[['data/training/grain', '0000574'],  
 ['data/training/crude', '0001269'],  
 ['data/training/interest', '0001988'],  
 ['data/training/acq', '0009235'],  
 ['data/training/acq', '0004109'],  
 ['data/training/acq', '0001478'],  
 ['data/training/trade', '0007488'],  
 ['data/training/earn', '0006039'],  
 ['data/training/earn', '0005406'],  
 ['data/training/earn', '0009218'],  
 ['data/training/earn', '0004637'],  
 ['data/training/earn', '0001089'],  
 ['data/training/earn', '0006397'],  
 ['data/training/corn', '0002244'],  
 ['data/training/money-fx', '0007525']]
```

Extract Files

In [51]:

```
trainingFiles = getFilesInFolders("data/training", categories)  
print(len(trainingFiles))
```

7193

Training Set

In [52]:

```
trainDocuments = [[filename.split("/")[2], readFile(filename)] for filename in c  
oncatFiles(trainingFiles[:11])]  
print(len(trainDocuments))
```

654

Training Data and Target

In [53]:

```
trainData = [content for file, content in trainDocuments]  
trainTarget = [file for file, content in trainDocuments]
```

Preview

In [54]:

```
trainDocumentName = trainTarget[0]
trainDocumentText = trainData[0]
print(trainDocumentName, trainDocumentText)
```

grain

ZIMBABWE MAIZE HARVEST LOWER AFTER BUMPER CROPS HARARE, March 3

- Zimbabwe's maize crop in 1986/87 (April/ March) is likely to be slightly over 1.6 mln tonnes, against 1.83 mln in 1985/86, Grain Marketing Board figures show.

Maize exports for 1986/87 to January 31, 1987 totalled 315,000 tonnes, with about a further 40,000 tonnes expected to be exported in February and March, against 285,000 tonnes in 1985/86. Domestic usage is estimated at 650,000-900,000 tonnes, depending on how other crops are affected by current poor rains. Last year's consumption was around 700,000 tonnes.

Zimbabwe has around two mln stonnes of surplus maize in storage, accumulated after two years of bumper harvests.

Count Vectorizer

In [55]:

```
from sklearn.feature_extraction.text import CountVectorizer
```

In [56]:

```
count_vect = CountVectorizer(analyzer = processDocument)
```

Fit

In [57]:

```
count_vect.fit(trainData);
```

Fit Vocabulary

In [58]:

```
print(count_vect.vocabulary_)
```

{'zimbabw': 5921, 'maiz': 3100, 'harvest': 2314, 'lower': 3048, 'bumper': 707, 'crop': 1226, 'harar': 2295, 'march': 3127, 'april': 285, 'like': 2975, 'slightli': 4798, 'mln': 3323, 'tonn': 5330, 'grain': 2196, 'market': 3138, 'board': 591, 'figur': 1925, 'show': 4730, 'export': 1827, 'januari': 2712, 'total': 5341, 'expect': 1811, 'februari': 1892, 'domest': 1519, 'usag': 5564, 'estim': 1760, 'depend': 1378, 'affect': 97, 'current': 1255, 'poor': 3986, 'rain': 4190, 'last': 2878, 'year': 5886, 'consumpt': 1118, 'around': 325, 'two': 5442, 'stonn': 5003, 'surplu': 5108, 'storag': 5008, 'accumul': 40, 'ccc': 837, 'accept': 30, 'wheat': 5764, 'bid': 556, 'w': 5670, 'africa': 102, 'countri': 1189, 'washington': 5706, 'commod': 1031, 'credit': 1215, 'corport': 1162, 'bonu': 605, 'cover': 1197, 'sale': 4540, 'us': 5562, 'west': 5752, 'african': 103, 'agricultur': 120, 'depart': 1375, 'said': 4530, 'dark': 1292, 'northern': 3559, 'spring': 4917, 'shipment': 4711, 'may': 3168, 'june': 2770, 'dlr': 1491, 'per': 3870, 'made': 3090, 'peavey': 3852, 'compani': 1040, 'paid': 3767, 'form': 2010, 'inventori': 2653, 'addit': 66, 'still': 4990, 'avail': 402, 'enhanc': 1700, 'program': 4088, 'initi': 2589, 'announc': 247, 'octob': 3617, 'canada': 759, 'corn': 1155, 'decis': 1327, 'unjustifi': 5512, 'yeutter': 5899, 'trade': 5355, 'repres': 4350, 'clayton': 954, 'find': 1934, 'today': 5313, 'import': 2520, 'injur': 2591, 'canadian': 760, 'farmer': 1869, 'small': 4809, 'inconceiv': 2543, 'reason': 4242, 'measur': 3197, 'statement': 4958, 'follow': 1988, 'lead': 2904, 'could': 1178, 'result': 4386, 'rash': 4213, 'protectionist': 4110, 'action': 57, 'throughout': 5279, 'world': 5849, 'french': 2053, 'grower': 2232, 'recent': 4253, 'indic': 2560, 'challenge': 866, 'gluten': 2158, 'feed': 1896, 'europ': 1768, 'examin': 1787, 'close': 971, 'believ': 521, 'base': 476, 'fact': 1845, 'care': 789, 'evalu': 1771, 'appropri': 281, 'respons': 4377, 'say': 4575, 'step': 4975, 'take': 5156, 'pl': 3938, 'hondura': 2422, 'usda': 5569, 'unit': 5510, 'state': 4956, 'sign': 4749, 'public': 4135, 'law': 2890, 'agreement': 118, 'provid': 4115, 'worth': 5856, 'wheatwheat': 5767, 'flour': 1975, 'cornsorghum': 1157, 'tallow': 5162, 'suppli': 5102, 'fiscal': 1946, 'end': 1682, 'septemb': 4657, 'privat': 4067, 'trader': 5357, 'competit': 1049, 'basi': 478, 'purchas': 4143, 'author': 393, 'issu': 2681, 'acr': 52, 'idl': 2494, 'would': 5858, 'littl': 2998, 'impact': 2513, 'acreag': 53, 'prompt': 4099, 'cropland': 1227, 'everi': 1778, 'accord': 35, 'report': 4349, 'save': 4572, 'loan': 3010, 'transport': 5377, 'defici': 1344, 'payment': 3833, 'assess': 357, 'propos': 4106, 'cotton': 1177, 'sorghum': 4858, 'barley': 471, 'almost': 177, 'likelihood': 2976, 'provisiion': 4119, 'caus': 822, 'larg': 2871, 'remov': 4328, 'product': 4083, 'quit': 4178, 'return': 4397, 'typic': 5456, 'farm': 1868, 'favor': 1878, 'particip': 3804, 'usual': 5577, 'reduct': 4274, 'seed': 4630, 'permit': 3880, 'allow': 173, 'forego': 1998, 'plant': 3944, 'receiv': 4252, 'pct': 3844, 'use': 5571, 'produc': 4082, 'high': 2376, 'productionhigh': 4084, 'risk': 4437, 'area': 303, 'cost': 1171, 'keith': 2791, 'collin': 1002, 'director': 1439, 'econom': 1613, 'analysi': 226, 'staff': 4937, 'heart': 2338, 'belt': 527, 'get': 2136, 'much': 3409, 'one': 3659, 'ildl': 2500, 'equival': 1734, 'bushel': 719, 'bale': 439, 'determin': 1403, 'whether': 5770, 'need': 3469, 'weigh': 5735, 'cash': 809, 'rate': 4214, 'price': 4057, 'rise': 4435, 'level': 2944, 'reduc': 4273, 'also': 186, 'must': 3422, 'consid': 1106, 'your': 5904, 'give': 2148, 'differ': 1425, 'low': 3047, 'greater': 2207, 'appli': 275, 'pay': 3830, 'variabl': 5600, 'cicumst': 927, 'want': 5689, 'go': 2166, 'along': 180, 'attract': 378, 'incent': 2533, 'admit': 79, 'margin': 3131, 'first': 1941, 'toward': 5347, 'goal': 2167, 'separ': 4655, 'govern': 2186, 'speech': 4897, 'earlier': 1598, 'nation': 3443, 'associ': 356, 'secretari': 4625, 'richard': 4418, 'lyng': 3079, 'late': 2882, 'implement': 2518, 'sinc': 4763, 'signup': 4753, 'month': 3361, 'ec': 1611, 'set': 4665, 'licenc': 2959, 'pari':

3792, 'european': 1769, 'commun': 1035, 'commiss': 1024, 'award': 413, 'free': 2047, 'maximum': 3167, 'rebat': 4244, 'currenc': 1254, 'ecu': 1617, 'sourc': 4864, 'request': 4356, 'reject': 4313, 'sharp': 4691, 'drop': 1560, 'chicago': 897, 'analyst': 227, 'modest': 3342, 'declin': 1329, 'soybean': 4880, 'intent': 2620, 'due': 1577, 'est': 1757, 'gmt': 2163, 'averag': 404, 'guess': 2251, 'rang': 4205, 'well': 5743, 'improv': 2524, 'retir': 4394, 'pik': 3926, 'introduc': 2648, 'quesstim': 4169, 'compar': 1044, 'substanti': 5056, 'stock': 4995, 'less': 2937, 'interest': 2628, 'express': 1830, 'among': 214, 'schedul': 4584, 'eep': 1626, 'tactic': 5149, 'tool': 5333, 'gener': 2124, 'polici': 3975, 'comment': 1018, 'question': 4171, 'expand': 1808, 'soviet': 4875, 'union': 5508, 'directli': 1438, 'repli': 4347, 'highest': 2379, 'reagan': 4233, 'administr': 76, 'dont': 1528, 'preempt': 4029, 'told': 5318, 'continu': 1128, 'select': 4636, 'success': 5063, 'pressur': 4047, 'beghinsay': 510, 'increas': 2547, 'capital': 777, 'financ': 1931, 'expans': 1809, 'sugar': 5068, 'group': 2229, 'own': 3753, 'itali': 2685, 'gruppo': 2238, 'ferruzzi': 1910, 'rais': 4195, 'franc': 2037, 'threeforon': 5273, 'share': 4689, 'invest': 2654, 'certif': 859, 'presid': 4043, 'jeanmarc': 2724, 'vern': 5615, 'stage': 4939, 'new': 3485, 'bring': 666, 'billion': 567, 'fund': 2086, 'plan': 3942, 'includ': 2539, 'possibl': 4002, 'acquisit': 51, 'starch': 4953, 'haubourdin': 2318, 'sever': 4672, 'buy': 724, 'instal': 2607, 'apart': 268, 'three': 5270, 'factori': 1847, 'germani': 2134, 'britain': 668, 'spain': 4884, 'netherland': 3479, 'denmark': 1369, 'put': 4149, 'tag': 5151, 'fail': 1852, 'acquir': 50, 'develop': 1409, 'either': 1640, 'sector': 4627, 'kaysersberg': 2787, 'subsidiari': 5053, 'anoth': 251, 'major': 3101, 'transform': 5370, 'divis': 1489, 'fullyfledg': 2082, 'chemic': 889, 'hold': 2415, 'talk': 5161, 'ad': 61, 'onto': 3672, 'bours': 624, 'near': 3456, 'future': 2093, 'asc': 339, 'seek': 4631, 'offer': 3623, 'process': 4080, 'rough': 4481, 'rice': 4417, 'kansa': 2780, 'citi': 942, 'stabil': 4933, 'conserv': 1105, 'servic': 4662, 'corpor': 1161, 'deliv': 1357, 'pound': 4013, 'mill': 3283, 'spokesman': 4907, 'cdt': 840, 'offeror': 3624, 'notifi': 3570, 'mge': 3248, 'start': 4954, 'hfc': 2373, 'contract': 1129, 'minneapoli': 3303, 'exchang': 1793, 'fructos': 2067, 'syrup': 5142, 'cftc': 861, 'approv': 282, 'next': 3495, 'week': 5729, 'pat': 3819, 'henderson': 2361, 'spokesperson': 4909, 'tuesday': 5422, 'meet': 3209, 'review': 4409, 'submit': 5042, 'decemb': 1325, 'lot': 3038, 'tank': 5165, 'truck': 5409, 'sweeten': 5124, 'agent': 109, 'deriv': 1391, 'wet': 5758, 'commonli': 1033, 'call': 746, 'soft': 4832, 'drink': 1558, 'beverag': 549, 'deliveri': 1359, 'ship': 4710, 'facil': 1843, 'design': 1393, 'symbol': 5138, 'hf': 2372, 'cst': 1241, 'correspond': 1167, 'juli': 2768, 'quot': 4180, 'hundred': 2464, 'weight': 5736, 'minimum': 3298, 'fluctuat': 1978, 'cent': 847, 'cwt': 1265, 'hope': 2428, 'user': 5572, 'activ': 58, 'involv': 2660, 'processor': 4081, 'miller': 3284, 'bottler': 618, 'potenti': 4011, 'type': 5454, 'peopl': 3866, 'list': 2994, 'industri': 2568, 'util': 5581, 'chines': 905, 'hit': 2401, 'frost': 2065, 'snow': 4822, 'peke': 3854, 'rapese': 4208, 'east': 1605, 'china': 904, 'suffer': 5066, 'consider': 1107, 'damag': 1282, 'spell': 4899, 'unusu': 5541, 'cold': 994, 'weather': 5725, 'daili': 1276, 'temperatur': 5208, 'day': 1301, 'five': 1951, 'degre': 1350, 'centigrad': 850, 'fellow': 1902, 'jiangsu': 2737, 'anhui': 244, 'hubei': 2453, 'henan': 2360, 'make': 3102, 'earli': 1597, 'sow': 4876, 'difficult': 1427, 'heavy': 2341, 'blanket': 582, 'central': 852, 'south': 4866, 'jilin': 2739, 'north': 3556, 'liaon': 2952, 'leav': 2916, 'farmland': 1871, 'muddi': 3410, 'plough': 3959, 'paper': 3785, 'rainfal': 4193, 'yangtz': 5883, 'higher': 2377, 'normal': 3552, 'sunday': 5086, 'part': 3800, 'guangdong': 2245, 'particularli': 3806, 'arid': 311, 'dri': 1556, 'season': 4615, 'mark': 3134, 'flood': 1970, 'gave': 2112, 'detail': 1400, 'news': 3490, 'agenc': 108, 'prospect': 4107, 'sown': 487

7, 'hectar': 2348, 'rural': 4511, 'fair': 1854, 'comparison': 1045, 'sudan': 5065, 'grade': 2192, 'number': 3586, 'better': 547, 'except': 1791, 'durum': 1584, 'shall': 4684, 'period': 3878, 'august': 385, 'load': 3009, 'portland': 3995, 'feb': 1891, 'seven': 4669, 'six': 4774, 'wait': 5673, 'merchant': 3227, 'grant': 2200, 'zero': 5916, 'blah': 580, 'spotlight': 4914, 'asian': 343, 'drought': 1561, 'gavin': 2113, 'greenwood': 2214, 'reutersauthor': 4402, 'hong': 2423, 'kong': 2837, 'geograph': 2129, 'divers': 1482, 'asia': 342, 'link': 2985, 'scientist': 4601, 'reintensif': 4309, 'complex': 1055, 'littleunderstood': 2999, 'el': 1642, 'nino': 3510, 'pattern': 3826, 'accuweather': 44, 'inc': 2531, 'commerci': 1022, 'forecast': 1995, 'sugarcane': 5069, 'australia': 389, 'tea': 5179, 'planter': 3946, 'sri': 4927, 'lanka': 2870, 'face': 1842, 'serious': 4659, 'loss': 3036, 'respect': 4375, 'unless': 5515, 'arriv': 328, 'time': 5296, 'break': 653, 'offic': 3625, 'offici': 3626, 'meteorologist': 3238, 'wen': 5746, 'wei': 5734, 'po': 3967, 'worst': 5855, 'provinc': 4117, 'without': 5828, 'adequ': 68, 'threaten': 5269, 'eight': 1635, 'key': 2803, 'sichuan': 4740, 'lowest': 3051, 'cut': 1261, 'store': 5009, 'water': 5709, 'volum': 5651, 'mani': 3112, 'river': 4440, 'hydroelectr': 2477, 'power': 4016, 'shortag': 4722, 'household': 2440, 'upper': 5553, 'reach': 4227, 'centuri': 853, 'run': 4507, 'aground': 122, 'across': 54, 'media': 3202, 'record': 4263, 'celsiu': 844, 'harbin': 2296, 'northeast': 3557, 'onethird': 3666, 'ministri': 3301, 'plantat': 3945, 'reuter': 4400, 'concern': 1070, 'grow': 2231, 'belowavera': 526, 'queensland': 4167, 'coast': 987, 'western': 5753, 'local': 3013, 'meteorolog': 3237, 'bureau': 713, 'fear': 1888, 'present': 4041, 'yet': 5898, 'pose': 4000, 'threat': 5268, 'autumnwint': 401, 'certainli': 858, 'heighten': 2351, 'memori': 3220, 'devast': 1408, 'coincid': 993, 'occurr': 3612, 'bare': 468, 'understood': 5490, 'phenomenon': 3905, 'known': 2831, 'although': 188, 'cautiou': 824, 'asiapacif': 345, 'region': 4293, 'disrupt': 1467, 'singl': 4766, 'role': 4460, 'studi': 5034, 'specialis': 4891, 'data': 1295, 'event': 1775, 'uniqu': 5509, 'alway': 192, 'effect': 1630, 'pronounc': 4101, 'spanish': 4885, 'christ': 918, 'child': 902, 'appear': 273, 'christma': 921, 'warm': 5695, 'air': 127, 'bear': 500, 'cloud': 974, 'shift': 4708, 'indonesian': 2566, 'archipelago': 299, 'peru': 3888, 'mingl': 3296, 'pacif': 3761, 'wind': 5808, 'strengthen': 5018, 'pump': 4141, 'sea': 4609, 'indonesia': 2565, 'ocean': 3613, 'equilibrium': 1729, 'send': 4648, 'surg': 5107, 'warmer': 5696, 'back': 424, 'collid': 1001, 'heat': 2339, 'deflect': 1348, 'rainbear': 4191, 'away': 414, 'vari': 5599, 'mayseptemb': 3171, 'droughtstricken': 1563, 'annual': 249, 'coastal': 988, 'strip': 5025, 'half': 2270, 'diminish': 1435, 'raini': 4194, 'draw': 1549, 'come': 1013, 'secondari': 4621, 'occur': 3611, 'winter': 5814, 'hurt': 2468, 'german': 2133, 'mild': 3277, 'berlin': 537, 'field': 1923, 'kill': 2810, 'might': 3273, 'probabl': 4074, 'limit': 2979, 'southern': 4869, 'widespread': 5790, 'ground': 2227, 'night': 3503, 'sunshin': 5090, 'led': 2919, 'heav': 2340, 'furthermor': 2091, 'surfac': 5106, 'thaw': 5243, 'stand': 4945, 'repair': 4339, 'special': 4890, 'thin': 5251, 'harrow': 2311, 'welltim': 5745, 'applic': 277, 'nitrogen': 3513, 'necessari': 3464, 'wheatsugar': 5766, 'ago': 114, 'busi': 720, 'weekli': 5731, 'custom': 1258, 'rose': 4473, 'growth': 2234, 'slow': 4802, 'jakarta': 2707, 'calendar': 739, 'stagnat': 4940, 'embassi': 1665, 'stapl': 4951, 'food': 1991, 'fall': 1859, 'perform': 3876, 'output': 3729, 'drawdown': 1550, 'look': 3030, 'promis': 4096, 'rubber': 4500, 'copra': 1152, 'gain': 2100, 'overall': 3733, 'signific': 4751, 'palm': 3773, 'oil': 3636, 'kernel': 2801, 'constraint': 1113, 'diseas': 1459, 'resist': 4368, 'fertil': 1911, 'land': 2865, 'insect': 2600, 'pest': 3894, 'reluct': 4323, 'underlin': 5486, 'outbreak': 3716, 'wereng': 5750, 'brown': 683, 'hopper': 2430, 'offset': 3627, 'yield': 5900, 'forc': 1994, 'ban': 446, 'insecticid': 2601, 'varieti': 5602, 'loweryield':

3050, 'coffe': 992, 'plywood': 3964, 'revenu': 4406, 'brought': 682, 'crude': 1234, 'octsept': 3618, 'crush': 1236, 'meal': 3192, 'deval u': 1407, 'rupiah': 4510, 'demand': 1363, 'even': 1774, 'slight': 47 97, 'neg': 3470, 'mid': 3262, 'dramat': 1547, 'slump': 4808, 'see': 4629, 'harmoni': 2306, 'prime': 4060, 'minist': 3299, 'jacqu': 2703, 'chirac': 909, 'visit': 5640, 'frequent': 2054, 'rancor': 4202, 'dis put': 1466, 'begin': 511, 'fade': 1850, 'libyan': 2958, 'bomb': 601, 'thing': 5252, 'past': 3816, 'war': 5690, 'didnt': 1421, 'happen': 2 294, 'good': 2180, 'cooper': 1146, 'terror': 5225, 'oblig': 3602, 'd eal': 1310, 'seri': 4658, 'alleg': 162, 'terrorist': 5226, 'target': 5168, 'libya': 2957, 'refus': 4288, 'britishbas': 670, 'plane': 394 3, 'overfli': 3736, 'territori': 5224, 'circuit': 935, 'rout': 4485, 'anger': 239, 'ask': 347, 'identifi': 2493, 'felt': 1903, 'strong': 5027, 'enough': 1705, 'pleas': 3955, 'access': 31, 'portug': 3996, 'join': 2750, 'becom': 504, 'retali': 4391, 'saw': 4574, 'deliber': 1356, 'move': 3394, 'exclud': 1798, 'pois': 3972, 'impos': 2522, 'sw ing': 5125, 'tariff': 5169, 'avert': 405, 'minut': 3308, 'hissen': 2 398, 'habr': 2265, 'chad': 863, 'support': 5104, 'train': 5364, 'ar m': 318, 'score': 4604, 'push': 4148, 'troop': 5404, 'common': 1032, 'japan': 2716, 'rest': 4378, 'open': 3677, 'relat': 4315, 'markedl i': 3137, 'irrit': 2677, 'remain': 4324, 'top': 5334, 'cap': 770, 'r oot': 4472, 'evil': 1782, 'intern': 2634, 'subsidis': 5054, 'sell': 4638, 'vast': 5604, 'amount': 216, 'excess': 1792, 'therebi': 5248, 'eat': 1609, 'summari': 5080, 'usdanoaa': 5570, 'moder': 3340, 'doub lecrop': 1535, 'eastern': 1606, 'guangxi': 2247, 'guanghdong': 2246, 'hunan': 2463, 'jiangxi': 2738, 'fujian': 2074, 'zhejiang': 5918, 'j oint': 2751, 'commerc': 1020, 'moistur': 3347, 'lessen': 2938, 'irri g': 2676, 'abund': 27, 'inund': 2650, 'wash': 5705, 'requir': 4357, 'replant': 4346, 'light': 2973, 'shower': 4731, 'topsoil': 5338, 'dr yness': 1567, 'plain': 3941, 'veget': 5607, 'note': 3565, 'deng': 13 67, 'leader': 2905, 'xiaop': 5876, 'ming': 3295, 'pao': 3782, 'newsp ap': 3493, 'influenc': 2581, 'whole': 5777, 'economi': 1615, 'contex t': 1125, 'remark': 4326, 'situat': 4773, 'point': 3971, 'pig': 392 5, 'fed': 1893, 'proper': 4103, 'place': 3939, 'avoid': 409, 'counci l': 1179, 'decid': 1326, 'unchang': 5479, 'despit': 1396, 'summer': 5081, 'cabinet': 731, 'spent': 4901, 'yuan': 5909, 'antidrought': 25 6, 'work': 5845, 'hui': 2458, 'bao': 465, 'shanxi': 4688, 'hebei': 2 344, 'shandong': 4686, 'eas': 1603, 'snowfal': 4823, 'project': 409 2, 'worsen': 5854, 'expert': 1820, 'electr': 1645, 'congress': 1091, 'soil': 4836, 'exempt': 1802, 'senat': 4647, 'committe': 1028, 'bil l': 565, 'tomorrow': 5325, 'regul': 4298, 'rotat': 4478, 'alfalfa': 152, 'multiyear': 3417, 'grass': 2201, 'legum': 2925, 'row': 4487, 'socal': 4826, 'sodbust': 4831, 'highli': 2381, 'erod': 1742, 'los e': 3035, 'feder': 1894, 'benefit': 532, 'later': 2883, 'sen': 4646, 'edward': 1625, 'zorinski': 5927, 'dneb': 1504, 'sponsor': 4911, 'pr actic': 4018, 'eros': 1743, 'strictli': 5022, 'argentin': 306, 'buen o': 694, 'previou': 4052, 'bread': 651, 'nil': 3505, 'sunflowerse': 5088, 'cumul': 1252, 'bracket': 636, 'thousand': 5265, 'soymeal': 48 81, 'seen': 4633, 'hamburg': 2278, 'exceed': 1790, 'newslett': 3492, 'significantli': 4752, 'least': 2915, 'america': 202, 'januarmarc h': 2715, 'quarter': 4164, 'quantiti': 4162, 'argentina': 307, 'braz il': 648, 'given': 2149, 'chang': 872, 'sens': 4651, 'view': 5630, 'polish': 3977, 'scale': 4577, 'butter': 722, 'onward': 3675, 'behin d': 514, 'pickup': 3918, 'bought': 621, 'midnovemb': 3271, 'ussr': 5 576, 'domin': 1520, 'system': 5143, 'increa': 2546, 'post': 4003, 'u nkknown': 5513, 'destin': 1397, 'taiwan': 5155, 'preced': 4021, 'four week': 2030, 'switch': 5129, 'prior': 4065, 'twothird': 5447, 'mexic o': 3246, 'main': 3096, 'buyer': 725, 'cake': 736, 'largest': 2874, 'venezuela': 5609, 'decreas': 1331, 'dominican': 1521, 'republ': 435 2, 'panama': 3778, 'turkey': 5433, 'thailand': 5240, 'korea': 2839, 'upcom': 5545, 'israel': 2680, 'cypru': 1270, 'saudi': 4571, 'arabi

a': 291, 'danforth': 1288, 'bond': 603, 'missouri': 3317, 'republica
n': 4353, 'john': 2748, 'christoph': 922, 'feedgrain': 1898, 'missis
sippi': 3316, 'collect': 999, 'aid': 125, 'austin': 387, 'schlick':
4591, 'similar': 4759, 'provis': 4118, 'hous': 2439, 'committeepas
s': 1029, 'emerg': 1667, 'disast': 1445, 'assist': 355, 'taken': 515
7, 'destroy': 1399, 'leve': 2943, 'vulner': 5668, 'floor': 1971, 'in
tens': 2618, 'technolog': 5187, 'concept': 1069, 'moscow': 3383, 'br
oadcast': 673, 'radio': 4183, 'bad': 428, 'realiz': 4238, 'third': 5
254, 'cereal': 856, 'rouen': 4480, 'port': 3990, 'algeria': 155, 'co
mpris': 1061, 'friday': 2056, 'poland': 3974, 'grainoilse': 2197, 'a
djust': 74, 'oilse': 3641, 'fob': 1984, 'roast': 4450, 'groundnuts
e': 2228, 'size': 4780, 'pellet': 3855, 'preserv': 4042, 'oneil': 36
63, 'fengbang': 1905, 'counti': 1188, 'done': 1526, 'none': 3533, 'l
eft': 2921, 'xu': 5881, 'jinfeng': 2740, 'middleag': 3266, 'villag':
5634, 'edg': 1619, 'shanghai': 4687, 'sum': 5077, 'dilemma': 1430,
'tri': 5391, 'let': 2940, 'richer': 4420, 'build': 699, 'home': 242
1, 'popul': 3988, 'seventh': 4670, 'arabl': 293, 'turn': 5436, 'ne
t': 3477, 'rapid': 4209, 'industrialis': 2569, 'countrysid': 1190,
'gobbl': 2169, 'peasant': 3851, 'afford': 100, 'lost': 3037, 'smalle
r': 4810, 'mouth': 3393, 'began': 509, 'nearli': 3459, 'famili': 186
3, 'built': 702, 'gone': 2179, 'railway': 4189, 'line': 2984, 'alrea
di': 184, 'matter': 3163, 'leadership': 2906, 'nationwid': 3445, 'qu
ota': 4181, 'convers': 1143, 'abus': 28, 'occupi': 3610, 'unlaw': 55
14, 'wast': 5707, 'resourc': 4372, 'articl': 331, 'press': 4046, 'ex
plain': 1822, 'great': 2206, 'cultiv': 1250, 'badli': 430, 'anyth':
267, 'obstacl': 3604, 'effort': 1632, 'ensur': 1709, 'profit': 4086,
'cheap': 884, 'fertilis': 1912, 'diesel': 1424, 'advanc': 82, 'resi
d': 4363, 'bonus': 606, 'argu': 309, 'simplest': 4762, 'solut': 484
4, 'problem': 4076, 'chen': 891, 'zuyuan': 5931, 'communist': 1037,
'parti': 3801, 'listen': 2995, 'selfish': 4637, 'rule': 4504, 'confl
ict': 1088, 'social': 4827, 'editori': 1622, 'reform': 4286, 'long':
3022, 'term': 5222, 'object': 3601, 'consum': 1116, 'inadequ': 2529,
'particular': 3805, 'wealth': 5723, 'grave': 2202, 'hall': 2275, 'an
cestor': 230, 'certain': 857, 'freedom': 2048, 'expir': 1821, 'dat
e': 1297, 'halfway': 2273, 'worri': 5852, 'anxieti': 265, 'stem': 49
73, 'organis': 3701, 'room': 4470, 'individu': 2564, 'reassur': 424
3, 'consist': 1108, 'longterm': 3029, 'agre': 116, 'greec': 2209, 'm
adrid': 3092, 'cargil': 790, 'compania': 1041, 'de': 1307, 'abasteci
miento': 5, 'sa': 4519, 'cindasa': 932, 'peseta': 3891, 'kilo': 281
2, 'final': 1929, 'valencia': 5590, 'second': 4620, 'bound': 623, 'r
otterdam': 4479, 'ghent': 2138, 'unspecifi': 5538, 'specif': 4894,
'graviti': 2203, 'ireland': 2669, 'belgium': 519, 'absenc': 22, 'del
ay': 1353, 'yesterday': 5896, 'posit': 4001, 'advoc': 90, 'congressi
on': 1092, 'rep': 4338, 'glenn': 2150, 'english': 1698, 'dokla': 151
2, 'spark': 4887, 'controversi': 1139, 'trim': 5396, 'scaleddown': 4
578, 'version': 5617, 'equit': 1732, 'pertain': 3886, 'simpl': 4761,
'treat': 5385, 'equal': 1728, 'narrow': 3437, 'amend': 201, 'charl':
878, 'stenholm': 4974, 'rtex': 4498, 'unclear': 5480, 'congressman':
1093, 'choic': 914, 'option': 3690, 'howev': 2444, 'polit': 3978, 'r
ealiti': 4237, 'unavail': 5474, 'cutback': 1262, 'arlan': 317, 'stan
geland': 4949, 'rminn': 4446, 'harold': 2307, 'volkmer': 5649, 'dm
o': 1499, 'reopen': 4334, 'percentag': 3872, 'economist': 1616, 'opp
on': 3684, 'prematur': 4033, 'uk': 5465, 'merg': 3229, 'london': 302
0, 'potato': 4010, 'meat': 3198, 'legal': 2922, 'advis': 87, 'instru
ct': 2610, 'amalgam': 194, 'lpfa': 3053, 'lmfe': 3005, 'gafta': 209
8, 'soya': 4878, 'bean': 499, 'somfa': 4851, 'timet': 5298, 'membe
r': 3218, 'consult': 1115, 'englebright': 1696, 'merger': 3230, 'ult
im': 5467, 'combin': 1012, 'discuss': 1458, 'baltic': 443, 'freigh
t': 2052, 'biffex': 558, 'intend': 2616, 'soon': 4854, 'jame': 2709,
'macki': 3088, 'aim': 126, 'regulatori': 4301, 'achiev': 45, 'recogn
it': 4258, 'financi': 1932, 'act': 56, 'legisl': 2923, 'recognis': 4

257, 'rie': 4427, 'structur': 5032, 'secur': 4628, 'commit': 1026, 'enabl': 1679, 'progress': 4090, 'stronger': 5029, 'chairman': 865, 'format': 2012, 'elmer': 1657, 'exist': 1806, 'valu': 5593, 'andor': 233, 'transship': 5378, 'entir': 1715, 'help': 2357, 'park': 3796, 'ridg': 4425, 'ill': 2501, 'american': 203, 'afbf': 95, 'far': 1865, 'foreign': 1999, 'dean': 1313, 'kleckner': 2822, 'tribun': 5392, 'su bsid': 5051, 'injuri': 2592, 'upheld': 5548, 'countervail': 1187, 'd uti': 1587, 'specialist': 4892, 'went': 5748, 'ottawa': 3711, 'testi fi': 5229, 'actual': 59, 'grown': 2233, 'exert': 1804, 'upward': 555 6, 'worldwid': 5850, 'hangzhou': 2289, 'japanes': 2718, 'largescal': 2873, 'peak': 3848, 'short': 4721, 'reli': 4319, 'human': 2461, 'ani m': 245, 'though': 5263, 'earn': 1599, 'saturday': 4570, 'attent': 3 75, 'especi': 1753, 'spare': 4886, 'autumn': 400, 'sweet': 5123, 'pa ddi': 3764, 'highyield': 2384, 'fight': 1924, 'devot': 1411, 'stres s': 5020, 'vital': 5642, 'repeatedli': 4343, 'stabl': 4935, 'outli n': 3724, 'encourag': 1681, 'suburb': 5059, 'pretax': 4048, 'asid': 346, 'money': 3353, 'salari': 4539, 'worker': 5846, 'chu': 924, 'fen gb': 1904, 'outsid': 3731, 'unpaid': 5520, 'keep': 2789, 'pan': 377 7, 'huashan': 2452, 'scienc': 4599, 'raw': 4220, 'materi': 3156, 'ro ad': 4448, 'infrastructur': 2584, 'case': 808, 'leas': 2912, 'advant ag': 83, 'tangibl': 5164, 'siphon': 4768, 'intermedi': 2633, 'bureau craci': 714, 'corrupt': 1169, 'boost': 611, 'enthusiasm': 1713, 'sha rpli': 4693, 'hgca': 2374, 'provision': 4120, 'excis': 1797, 'attac k': 370, 'surpris': 5110, 'disappoint': 1443, 'attempt': 373, 'libe r': 2954, 'rounnd': 4483, 'multilater': 3414, 'negoti': 3472, 'incom prehens': 2542, 'barrier': 475, 'cave': 826, 'protection': 4109, 'da rwin': 1294, 'e': 1594, 'stolt': 5001, 'strain': 5012, 'uscanadian': 5568, 'relationship': 4316, 'picken': 3917, 'rig': 4428, 'dalla': 12 80, 'boon': 610, 'texa': 5233, 'oilman': 3639, 'headi': 2329, 'ope r': 3678, 'lifetim': 2968, 'count': 1183, 'recov': 4264, 'eventu': 1 776, 'slip': 4800, 'wont': 5837, 'barrel': 474, 'tour': 5345, 'promo t': 4098, 'autobiographi': 397, 'fee': 1895, 'order': 3695, 'stimu l': 4991, 'mee': 3208, 'opec': 3676, 'bpd': 632, 'nicosia': 3498, 'c eil': 841, 'pipelin': 3930, 'ecuador': 1618, 'middl': 3265, 'surve y': 5112, 'landslid': 2866, 'breach': 650, 'iraq': 2667, 'earthqua k': 1602, 'shut': 4738, 'four': 2027, 'arabian': 292, 'neutral': 348 3, 'zone': 5926, 'kuwait': 2848, 'float': 1969, 'hisham': 2397, 'naz er': 3451, 'televis': 5204, 'visnew': 5641, 'cyprusbas': 1271, 'auth orit': 395, 'lift': 2971, 'equiti': 1733, 'partial': 3802, 'stop': 5 005, 'entitl': 1718, 'insist': 2604, 'debt': 1319, 'tripoli': 5400, 'sent': 4653, 'telex': 5205, 'inequ': 2573, 'variou': 5603, 'iraqi': 2668, 'letter': 2942, 'baghdad': 431, 'difficulti': 1428, 'subject': 5040, 'refin': 4282, 'covert': 1199, 'discount': 1454, 'hydrocarbo n': 2476, 'packag': 3762, 'mexican': 3244, 'reserv': 4362, 'proven': 4114, 'liquid': 2989, 'petroleo': 3899, 'mexicano': 3245, 'pemex': 3 856, 'flatten': 1960, 'previouli': 4053, 'natur': 3447, 'ga': 2097, 'cubic': 1249, 'feet': 1901, 'mainli': 3097, 'exploratori': 1825, 's uspend': 5115, 'offshor': 3629, 'campech': 757, 'account': 37, 'occi dent': 3608, 'oxi': 3756, 'big': 560, 'lo': 3008, 'angel': 238, 'pet roleum': 3900, 'corp': 1160, 'shiviyacu': 4714, 'block': 586, 'ab': 3, 'peruvian': 3889, 'amazon': 197, 'jungl': 2772, 'depth': 1388, 'd rill': 1557, 'explor': 1824, 'southeast': 4867, 'mile': 3278, 'bp': 631, 'ltd': 3060, 'british': 669, 'co': 984, 'plc': 3954, 'bpl': 63 4, 'replac': 4345, 'stg': 4985, 'histor': 2399, 'depress': 1386, 'it em': 2688, 'charg': 876, 'chief': 899, 'execut': 1801, 'david': 130 0, 'kendal': 2797, 'mirror': 3310, 'restructur': 4385, 'undertaken': 5492, 'warn': 5697, 'larger': 2872, 'lend': 2928, 'caraca': 785, 'un determin': 5497, 'serious': 4660, 'energi': 1688, 'mine': 3293, 'art uro': 335, 'hernandez': 2367, 'grisanti': 2223, 'pdvsa': 3846, 'cep e': 854, 'incom': 2540, 'speak': 4888, 'miraflor': 3309, 'palac': 37 72, 'deputi': 1389, 'fernando': 1907, 'santo': 4555, 'alvit': 191,

'lent': 2932, 'repay': 4340, 'elabor': 1643, 'arrang': 326, 'millio
n': 3286, 'flag': 1957, 'tanker': 5166, 'temporari': 5209, 'protec
t': 4108, 'kuwaiti': 2849, 'troubl': 5406, 'persian': 3882, 'gulf':
2257, 'confirm': 1087, 'york': 5901, 'transfer': 5369, 'registr': 42
95, 'iran': 2664, 'launch': 2888, 'silkworm': 4755, 'missil': 3314,
'superpow': 5097, 'vow': 5661, 'traffic': 5360, 'tehran': 5191, 'chi
neseomad': 906, 'carri': 797, 'temporarili': 5210, 'creat': 1214, 'in
sur': 2612, 'defens': 1341, 'warship': 5703, 'escort': 1747, 'mideas
t': 3267, 'inevit': 2574, 'ted': 5189, 'dafflisio': 1275, 'noth': 35
68, 'donald': 1523, 'k': 2778, 'mcivor': 3185, 'exxon': 1838, 'xon':
5877, 'senior': 4650, 'vice': 5621, 'interview': 2643, 'nonopec': 35
42, 'gap': 2102, 'widen': 5787, 'trend': 5389, 'firm': 1939, 'capa
c': 772, 'weve': 5759, 'discoveri': 1457, 'bilion': 564, 'bulk': 70
4, 'prove': 4113, 'lie': 2964, 'discov': 1456, 'oneon': 3664, 'thous
andth': 5266, 'contain': 1120, 'ever': 1777, 'super': 5092, 'giant':
2139, 'found': 2023, 'conclus': 1075, 'cantarel': 768, 'alaska': 13
9, 'slope': 4801, 'prudho': 4127, 'bay': 486, 'plenti': 3957, 'oppor
tun': 3685, 'particulartli': 3807, 'deped': 1377, 'upon': 5552, 'reg
im': 4292, 'ideal': 2491, 'constant': 1111, 'oppos': 3686, 'subsid
i': 5052, 'expens': 1816, 'herrington': 2369, 'deplet': 1379, 'thrus
t': 5282, 'valid': 5591, 'highlight': 2382, 'play': 3951, 'increasin
gli': 2548, 'worlkd': 5851, 'fuel': 2072, 'outlook': 3725, 'temper':
5207, 'brad': 637, 'schade': 4583, 'employ': 1674, 'chart': 879, 'no
nfarm': 3536, 'payrol': 3836, 'job': 2745, 'treasuri': 5383, 'drov
e': 1564, 'technic': 5183, 'prudenti': 4125, 'bach': 422, 'fred': 20
43, 'leiner': 2927, 'threemonth': 5275, 'penetr': 3860, 'ralli': 419
8, 'wednesday': 5728, 'bullish': 706, 'chartist': 881, 'fals': 1861,
'breakout': 655, 'signal': 4750, 'bearish': 502, 'doubl': 1534, 'por
tend': 3991, 'steep': 4972, 'tend': 5213, 'scenario': 4581, 'eilee
n': 1638, 'rico': 4423, 'rel': 4314, 'cautious': 825, 'optimist': 36
89, 'curv': 1257, 'inflat': 2578, 'dollar': 1514, 'monetari': 3352,
'eurodollar': 1767, 'maintain': 3098, 'got': 2183, 'extra': 1835,
'suspens': 5116, 'thursday': 5286, 'runup': 4508, 'shortliv': 4727,
'carrol': 799, 'mcente': 3183, 'mcginley': 3184, 'brian': 659, 'sin
ger': 4765, 'induc': 2567, 'within': 5827, 'tremend': 5388, 'diamon
d': 1416, 'shamrock': 4685, 'wti': 5872, 'dlrsbbl': 1494, 'refiner
i': 4284, 'publish': 4137, 'lee': 2920, 'r': 4182, 'raymond': 4223,
'closur': 973, 'spokeswoman': 4910, 'mention': 3223, 'name': 3433,
'opert': 3679, 'fo': 1982, 'mediterranean': 3204, 'capciti': 773,
'jerom': 2729, 'intellig': 2615, 'influenti': 2582, 'journal': 276
0, 'antwerp': 263, 'paul': 3828, 'mlotok': 3326, 'salomon': 4543, 'b
rother': 681, 'bunch': 708, 'upgrad': 5547, 'shaken': 4683, 'indepe
d': 2553, 'effici': 1631, 'superior': 5094, 'royal': 4490, 'dutch':
1585, 'rd': 4225, 'feel': 1900, 'downstream': 1540, 'piw': 3936, 'e
lsewher': 1660, 'reclaim': 4255, 'italian': 2686, 'netback': 3478,
'bank': 457, 'recoveri': 4266, 'arab': 290, 'abu': 26, 'dhabi': 141
3, 'emir': 1668, 'predict': 4028, 'wam': 5687, 'gcc': 2114, 'bahrai
n': 432, 'oman': 3654, 'qatar': 4152, 'uae': 5459, 'restrain': 4382,
'curb': 1253, 'boom': 609, 'experi': 1818, 'budget': 692, 'defici
t': 1345, 'aris': 313, 'recess': 4254, 'fresh': 2055, 'spend': 4900,
'bugdet': 697, 'gross': 2226, 'dirham': 1441, 'warplan': 5698, 'ves
sel': 5619, 'iranian': 2666, 'hour': 2438, 'militari': 3281, 'ina':
2527, 'latest': 2885, 'strike': 5024, 'dealt': 1312, 'accur': 41,
'blow': 587, 'safe': 4524, 'immedi': 2510, 'kharg': 2804, 'island':
2679, 'termin': 5223, 'hamilton': 2279, 'haml': 2280, 'denver': 137
3, 'round': 4482, 'right': 4429, 'cnooc': 981, 'pearl': 3850, 'yello
w': 5892, 'sq': 4921, 'km': 2825, 'lufeng': 3065, 'shenzhen': 4704,
'readi': 4231, 'enter': 1710, 'restrict': 4384, 'eleven': 1653, 'in
terven': 2641, 'speaker': 4889, 'parliament': 3798, 'hojatoleslam':
2413, 'akbar': 133, 'hashemi': 2316, 'rafsanjani': 4186, 'waterwa
y': 5711, 'intervent': 2642, 'insecur': 2602, 'lebanon': 2917, 'repe

at': 4342, 'everywher': 1780, 'defenc': 1339, 'disclos': 1451, 'nav
i': 3449, 'prepar': 4036, 'regular': 4299, 'offshoot': 3628, 'irna':
2670, 'disclosur': 1452, 'erect': 1737, 'site': 4772, 'antiship': 2
59, 'strait': 5013, 'hormuz': 2432, 'wide': 5786, 'narrowest': 3438,
'artilleri': 333, 'longest': 3026, 'lesser': 2939, 'therefor': 524
9, 'other': 3710, 'pancanadian': 3780, 'calgari': 740, 'alberta': 14
0, 'properti': 4104, 'whitehal': 5773, 'pctown': 3845, 'cp': 1200,
'partner': 3809, 'retain': 4390, 'royalti': 4491, 'api': 269, 'dist
il': 1472, 'gasolin': 2105, 'held': 2353, 'primari': 4058, 'revis':
4410, 'institut': 2609, 'statist': 4961, 'bulletin': 705, 'residu':
4365, 'reiter': 4312, 'pact': 3763, 'kingdom': 2816, 'stabilis': 49
34, 'spa': 4882, 'fulli': 2079, 'adher': 70, 'never': 3484, 'circums
t': 937, 'hemispher': 2359, 'glut': 2157, 'architect': 300, 'fix': 1
955, 'turmoil': 5435, 'briefli': 663, 'shown': 4732, 'clearli': 959,
'contact': 1119, 'stick': 4987, 'jamaica': 2708, 'rilwanu': 4431,
'lukman': 3067, 'nigerian': 3502, 'awar': 412, 'manipul': 3116, 'sa
tisfi': 4569, 'fundament': 2087, 'condit': 1076, 'sheikh': 4701, 'al
i': 157, 'alkhalifa': 159, 'alsabah': 185, 'alqaba': 183, 'deni': 13
68, 'overproduc': 3741, 'energydril': 1690, 'vorman': 5657, 'housto
n': 2442, 'drastic': 1548, 'rapidli': 4210, 'deflat': 1347, 'bubbl':
687, 'spot': 4913, 'oilpatch': 3640, 'falter': 1862, 'deliver': 135
8, 'costli': 1175, 'lock': 3015, 'resurg': 4388, 'locat': 3014, 'thi
nk': 5253, 'upturn': 5555, 'immin': 2512, 'crandal': 1209, 'theyr':
5250, 'bet': 544, 'balanc': 436, 'welcom': 5742, 'oilfield': 3638,
'surviv': 5113, 'plung': 3963, 'spur': 4919, 'soar': 4825, 'roustab
out': 4484, 'command': 1016, 'premium': 4035, 'wage': 5671, 'hugh':
2457, 'ht': 2451, 'baromet': 473, 'health': 2332, 'tradit': 5359,
'slowest': 4805, 'dive': 1480, 'postworld': 4009, 'ii': 2497, 'expe
rienc': 1819, 'sharpest': 4692, 'ike': 2499, 'kerridg': 2802, 'lique
fi': 2988, 'feasibl': 1889, 'trillion': 5395, 'approach': 280, 'geor
g': 2130, 'gaspar': 2106, 'robert': 4452, 'baird': 433, 'cycl': 126
7, 'gradual': 2194, 'anticip': 255, 'dedic': 1332, 'mean': 3193, 'ga
sper': 2107, 'search': 4614, 'conduct': 1078, 'unexplor': 5500, 'rev
ert': 4408, 'biggest': 562, 'mobil': 3337, 'mob': 3336, 'exempl': 17
88, 'conoco': 1097, 'dupont': 1581, 'dd': 1306, 'deepest': 1335, 'pl
atform': 3949, 'huge': 2456, 'mesa': 3232, 'partnership': 3811, 'ml
p': 3327, 'convinc': 1145, 'verg': 5613, 'tenneco': 5217, 'tgt': 523
8, 'skeptic': 4784, 'joe': 2746, 'foster': 2021, 'life': 2966, 'toky
o': 5316, 'dlrunti': 1496, 'exportimport': 1828, 'tito': 5305, 'aya
l': 417, 'ayala': 418, 'seminar': 4644, 'complet': 1054, 'distribu
t': 1474, 'nueva': 3584, 'teapa': 5181, 'salina': 4541, 'cruz': 123
8, 'construct': 1114, 'phase': 3904, 'liquifi': 2990, 'lpg': 3054,
'santa': 4553, 'ammonia': 212, 'lazaro': 2897, 'cardena': 788, 'equ
ip': 1731, 'maya': 3169, 'rather': 4215, 'diversifi': 1484, 'np': 35
78, 'npeeq': 3579, 'truste': 5413, 'appoint': 278, 'bankruptci': 46
3, 'court': 1195, 'overse': 3742, 'chapter': 875, 'proceed': 4079,
'thwart': 5287, 'hostil': 2435, 'pursu': 4146, 'unsecur': 5535, 'cr
editor': 1219, 'whose': 5782, 'claim': 949, 'nielsen': 3499, 'duches
n': 1575, 'utah': 5580, 'test': 5228, 'reaffirm': 4232, 'jukwey': 27
67, 'lago': 2859, 'defend': 1340, 'confer': 1082, 'colleagu': 998,
'herebi': 2365, 'wish': 5821, 'emphas': 1670, 'nigeria': 3501, 'uph
old': 5549, 'extraordinari': 1836, 'reconfirm': 4260, 'mislead': 331
1, 'contrari': 1133, 'shortfal': 4725, 'geneva': 2127, 'restor': 438
1, 'refer': 4280, 'idea': 2490, 'suggest': 5071, 'resort': 4371, 'de
stock': 1398, 'verifi': 5614, 'alon': 179, 'disassoci': 1444, 'succe
ed': 5062, 'differenti': 1426, 'formerli': 2014, 'postpon': 4005, 'i
ndefinit': 2552, 'moment': 3350, 'contempl': 1122, 'benchmark': 529,
'crucial': 1233, 'heavili': 2343, 'drawn': 1551, 'prais': 4020, 'co
ntribut': 1136, 'criticis': 1225, 'longheld': 3027, 'confid': 1083,
'hand': 2284, 'describ': 1392, 'tragic': 5361, 'walker': 5680, 'sim
on': 4760, 'quito': 4179, 'resum': 4387, 'section': 4626, 'agrio': 1

21, 'balao': 437, 'supplement': 5100, 'regist': 4294, 'mercal': 322
5, 'assign': 354, 'cartel': 805, 'hook': 2427, 'colombian': 1006, 's
queez': 4924, 'slide': 4795, 'owe': 3749, 'eighth': 1636, 'latin': 2
886, 'jaim': 2706, 'lusinchi': 3075, 'guarante': 2248, 'head': 2327,
'venezuelan': 5610, 'hector': 2349, 'hurtado': 2469, 'interior': 26
32, 'travel': 5381, 'coordin': 1147, 'relief': 4321, 'renew': 4330,
'jenni': 2726, 'kantyka': 2782, 'cargo': 791, 'dubai': 1571, 'inspi
r': 2606, 'goe': 2170, 'charter': 880, 'recur': 4268, 'dissoci': 146
8, 'firsthalf': 1944, 'brent': 658, 'blend': 583, 'there': 5247, 'i
m': 2504, 'trust': 5412, 'abil': 13, 'lucki': 3063, 'subroto': 5046,
'origin': 3703, 'vienna': 5629, 'philip': 3908, 'shehadi': 4700, 'r
iyadh': 4443, 'fallen': 1860, 'evid': 1781, 'king': 2815, 'fahd': 18
51, 'nomin': 3525, 'artifici': 332, 'stake': 4941, 'aramco': 294, 't
exaco': 5234, 'chevron': 895, 'enjoy': 1701, 'flexibl': 1967, 'pic
k': 3916, 'sought': 4860, 'fulfil': 2077, 'contractu': 1131, 'chv':
926, 'nippon': 3511, 'ventur': 5612, 'san': 4549, 'francisco': 204
0, 'usa': 5563, 'tokyobas': 5317, 'jointventur': 2752, 'leasehold':
2914, 'onshor': 3671, 'california': 743, 'colorado': 1007, 'oklahom
a': 3646, 'montana': 3358, 'dakota': 1279, 'wyom': 5874, 'outer': 37
18, 'continent': 1126, 'shelf': 4702, 'qtr': 4156, 'preliminari': 40
32, 'releas': 4317, 'direct': 1437, 'yearago': 5887, 'favour': 1879,
'baker': 434, 'tax': 5173, 'person': 3884, 'divid': 1487, 'bureacra
ci': 712, 'administ': 75, 'harm': 2304, 'tension': 5220, 'abat': 6,
'ankara': 246, 'pull': 4140, 'sismik': 4770, 'aegean': 92, 'semioff
ici': 4645, 'anatolian': 229, 'naval': 3448, 'turkish': 5434, 'dista
nc': 1470, 'popular': 3989, 'headlin': 2330, 'resolv': 4370, 'nato':
3446, 'bore': 614, 'fruit': 2069, 'stay': 4964, 'gune': 2259, 'tops
el': 5337, 'hurriyet': 2467, 'front': 2064, 'page': 3765, 'resolut':
4369, 'cumhuriyet': 1251, 'milliyet': 3288, 'ident': 2492, 'crisi':
1223, 'frozen': 2066, 'basic': 479, 'disagr': 1442, 'unsolv': 5537,
'confront': 1089, 'assur': 360, 'athen': 364, 'calcul': 738, 'ton
e': 5327, 'reflect': 4285, 'ct': 1243, 'sour': 4863, 'ural': 5557,
'bonni': 604, 'singapor': 4764, 'heavier': 2342, 'weakest': 5722,
'an': 224, 'cif': 929, 'la': 2852, 'weaker': 5721, 'weak': 5719, 'c
rack': 1207, 'es': 1744, 'sider': 4743, 'grademar': 2193, 'mar': 312
4, 'lt': 3059, 'attaka': 372, 'supertank': 5098, 'raid': 4187, 'esca
l': 1745, 'broke': 676, 'monthlong': 3363, 'lull': 3068, 'celebr': 8
42, 'hail': 2267, 'failur': 1853, 'victori': 5626, 'communiqu': 103
6, 'jetti': 2734, 'afternoon': 104, 'struck': 5031, 'nearbi': 3457,
'column': 1010, 'smoke': 4814, 'billo': 568, 'lloyd': 3002, 'ton':
5326, 'avaj': 403, 'independ': 2554, 'took': 5332, 'khark': 2805,
'nowruz': 3577, 'cyru': 1272, 'ardeshir': 302, 'subcommitte': 5039,
'hear': 2336, 'repeal': 4341, 'windfal': 5809, 'taxpay': 5175, 'sho
ot': 4719, 'doe': 1509, 'opposit': 3687, 'cours': 1194, 'forward': 2
020, 'kind': 2814, 'settl': 4667, 'debat': 1315, 'written': 5864, 'u
rg': 5558, 'unp': 5519, 'corpu': 1163, 'christi': 919, 'champlin': 8
68, 'transact': 5366, 'asset': 353, 'aggress': 112, 'excpand': 1800,
'presenc': 4040, 'fit': 1950, 'profil': 4085, 'previous': 4054, 'ne
twork': 3481, 'southland': 4871, 'slc': 4792, 'citgo': 941, 'potetni
t': 4012, 'outlet': 3723, 'agredd': 117, 'feedstock': 1899, 'rome':
4466, 'coupon': 1192, 'payabl': 3831, 'cct': 839, 'matur': 3165, 's
ixmonthli': 4777, 'repurchas': 4355, 'indirect': 2562, 'inject': 259
0, 'add': 65, 'indirectli': 2563, 'via': 5620, 'counteract': 1185,
'elev': 1652, 'indian': 2557, 'delhi': 1355, 'banker': 460, 'broke
r': 678, 'deposit': 1381, 'abolish': 17, 'modif': 3344, 'borrow': 61
6, 'shortterm': 4728, 'depositor': 1383, 'reinvest': 4311, 'manufact
ur': 3122, 'tata': 5170, 'steel': 4971, 'setter': 4666, 'bombay': 60
2, 'rupe': 4509, 'stockbrok': 4996, 'investor': 2658, 'enthusiast':
1714, 'convert': 1144, 'nonconvert': 3531, 'debentur': 1316, 'brigh
t': 665, 'westminst': 5755, 'respond': 4376, 'england': 1695, 'endor
s': 1686, 'strongli': 5030, 'match': 3154, 'lloyl': 3003, 'clear': 9

57, 'nwbl': 3591, 'way': 5716, 'morn': 3374, 'toler': 5320, 'strenu
ou': 5019, 'prevent': 4051, 'freddi': 2044, 'mac': 3082, 'mortgag':
3381, 'old': 3647, 'dealer': 1311, 'fhlbb': 1917, 'guilder': 2256,
'amsterdam': 221, 'duth': 1586, 'tender': 5214, 'twelveday': 5440,
'met': 3234, 'full': 2078, 'nineday': 3507, 'alloc': 171, 'suffic
i': 5067, 'durat': 1582, 'abbey': 7, 'martin': 3147, 'cherrin': 893,
'window': 5810, 'dead': 1308, 'center': 849, 'someth': 4848, 'fairl
i': 1857, 'william': 5798, 'hint': 2395, 'grigg': 2221, 'santow': 45
56, 'versu': 5618, 'broader': 674, 'grew': 2218, 'minu': 3307, 'aggr
eg': 111, 'absolut': 23, 'slowdown': 4803, 'stephen': 4976, 'slife
r': 4796, 'lehman': 2926, 'anem': 237, 'risen': 4436, 'persist': 388
3, 'comfort': 1015, 'stanc': 4944, 'perhap': 3877, 'grip': 2222, 'fr
agil': 2034, 'chanc': 869, 'choos': 915, 'monday': 3351, 'today': 5
443, 'marketdriven': 3139, 'focus': 1986, 'instead': 2608, 'buoy': 7
10, 'rowena': 4488, 'whelan': 5768, 'wors': 5853, 'outcom': 3717, 'v
isibl': 5639, 'unreservedli': 5530, 'chase': 882, 'manhattan': 3111,
'andrew': 235, 'wroblewski': 5866, 'sterl': 4978, 'rebound': 4245,
'revers': 4407, 'index': 2555, 'basket': 480, 'midday': 3264, 'dete
rior': 1402, 'nonoil': 3541, 'undermin': 5487, 'motor': 3388, 'begu
n': 512, 'doubt': 1536, 'realli': 4239, 'card': 787, 'sentiment': 46
54, 'danger': 1289, 'chancellor': 870, 'exchequ': 1796, 'nigel': 350
0, 'lawson': 2894, 'chri': 917, 'tinker': 5299, 'phillip': 3910, 'dr
ew': 1554, 'faster': 1875, 'partli': 3808, 'aberr': 9, 'expeditur':
1814, 'ian': 2478, 'harwood': 2315, 'warburg': 5691, 'confeder': 10
81, 'oversea': 3743, 'importintens': 2521, 'manufacturer': 3123, 'gi
lt': 2143, 'betterthanexpect': 548, 'evapor': 1773, 'stead': 4968,
'turnabout': 5437, 'attribut': 379, 'opinion': 3680, 'poll': 3980,
'marplan': 3143, 'allianc': 169, 'democrat': 1364, 'neck': 3467, 'l
abour': 2855, 'greet': 2215, 'seem': 4632, 'voter': 5659, 'indiffe
r': 2561, 'observ': 3603, 'economist': 1727, 'dent': 1372, 'margare
t': 3129, 'thatcher': 5242, 'upset': 5554, 'percept': 3874, 'elect':
1644, 'sensit': 4652, 'caution': 823, 'wari': 5694, 'inflationredu
c': 2580, 'curtail': 1256, 'bundesbank': 709, 'frankfurt': 2042, 'pl
edg': 3956, 'drain': 1545, 'ahead': 124, 'compens': 1047, 'governmen
town': 2188, 'behalf': 513, 'specul': 4896, 'disburs': 1447, 'facto
r': 1846, 'takeup': 5159, 'balnc': 442, 'outflow': 3719, 'circul': 9
36, 'barclay': 467, 'slowli': 4806, 'longer': 3024, 'interim': 2631,
'deutsch': 1406, 'ag': 106, 'dbkgf': 1304, 'dresdner': 1552, 'drsd
f': 1565, 'commerzbank': 1023, 'cbkgf': 831, 'fuer': 2073, 'gemeinwi
rtschaft': 2121, 'bkkgf': 577, 'desir': 1394, 'gaug': 2111, 'reactio
n': 4229, 'refinanc': 4283, 'undercut': 5484, 'longerd': 3025, 'four
th': 2029, 'repo': 4348, 'suntrust': 5091, 'sti': 4986, 'sun': 5083,
'florida': 1973, 'georgia': 2131, 'ionterest': 2661, 'mediums': 320
6, 'outright': 3730, 'pass': 3813, 'allot': 172, 'volcker': 5647, 'c
onnect': 1096, 'testimoni': 5230, 'abl': 14, 'detect': 1401, 'johnso
n': 2749, 'gro': 2224, 'brasilia': 647, 'rumour': 4506, 'file': 192
7, 'extends': 1832, 'lawsuit': 2895, 'dilson': 1432, 'funaro': 2084,
'attend': 374, 'imf': 2508, 'space': 4883, 'session': 4664, 'band':
454, 'midfebruari': 3268, 'brussel': 686, 'belgian': 518, 'team': 5
180, 'scope': 4603, 'copi': 1150, 'journalist': 2761, 'adopt': 81,
'emphasi': 1671, 'ratio': 4216, 'gnp': 2164, 'real': 4234, 'noninte
rest': 3539, 'expeditur': 1815, 'iron': 2671, 'distort': 1473, 'bur
den': 711, 'sustain': 5117, 'impat': 2514, 'firmer': 1940, 'overhau
l': 3737, 'facilit': 1844, 'subscript': 5049, 'nonresid': 3547, 'vol
atil': 5646, 'remaind': 4325, 'staley': 4942, 'lacey': 2856, 'unexpe
ct': 5499, 'sluggish': 4807, 'demonstr': 1365, 'karnoski': 2784, 'do
ldrum': 1513, 'perceiv': 3871, 'side': 4741, 'eager': 1596, 'selle
r': 4639, 'golden': 2172, 'gate': 2109, 'norman': 3553, 'quinn': 417
6, 'marketplac': 3141, 'echo': 1612, 'flow': 1976, 'stiff': 4988, 'm
eantim': 3195, 'refund': 4287, 'id': 2486, 'imperi': 2516, 'texon':
5235, 'txon': 5450, 'privatelyheld': 4068, 'specialti': 4893, 'plas

tic': 3947, 'consent': 1103, 'promissori': 4097, 'deak': 1309, 'matt
hey': 3164, 'preciou': 4023, 'metal': 3235, 'minori': 3306, 'arkad
i': 315, 'kuhlmann': 2847, 'wholesal': 5779, 'nonferr': 3537, 'absor
b': 24, 'cpc': 1202, 'milan': 3276, 'control': 1138, 'eridania': 174
0, 'zuccherifici': 5928, 'nazionali': 3452, 'erdimi': 1736, 'princip
l': 4062, 'millng': 3289, 'gwf': 2260, 'beverli': 550, 'hill': 2387,
'calif': 742, 'alden': 143, 'affili': 98, 'newlyform': 3488, 'onw':
3674, 'manag': 3110, 'gecc': 2117, 'approxim': 283, 'orangeco': 369
3, 'oj': 3642, 'holder': 2416, 'lake': 2861, 'fla': 1956, 'sharehol
d': 4690, 'summit': 5082, 'resourec': 4373, 'sanwa': 4557, 'portugue
s': 3998, 'anwat': 264, 'oportobas': 3682, 'banco': 450, 'portugu':
3997, 'investmento': 2657, 'sarl': 4561, 'bpi': 633, 'sister': 477
1, 'yen': 5894, 'portugus': 3999, 'permiss': 3879, 'atlantico': 368,
'stateown': 4959, 'oporto': 3681, 'inform': 2583, 'acceler': 29, 'o
livetti': 3650, 'sgsthomson': 4680, 'ing': 2586, 'c': 729, 'olivm':
3651, 'semiconductor': 4642, 'stet': 4979, 'societa': 4830, 'finanz
iaria': 1933, 'telefonica': 5197, 'pa': 3758, 'thomsoncsf': 5259, 't
csfpa': 5177, 'dispos': 1464, 'thomson': 5258, 'sg': 4677, 'microele
ttronica': 3261, 'semiconducteur': 4641, 'civil': 946, 'conclud': 10
74, 'queri': 4168, 'kohlberg': 2834, 'kravi': 2842, 'owensillinoi':
3752, 'oi': 3634, 'oii': 3635, 'prefer': 4030, 'owen': 3750, 'twa':
5439, 'settlement': 4668, 'usair': 5565, 'u': 5457, 'tran': 5365,
'airlin': 129, 'notic': 3569, 'knowingli': 2829, 'willingli': 5801,
'interfer': 2629, 'consumm': 1117, 'piedmont': 3922, 'aviat': 407,
'pie': 3921, 'litig': 2997, 'pittsburgh': 3934, 'delawar': 1352, 'c
arl': 792, 'icahn': 2484, 'vote': 5658, 'proport': 4105, 'mcandrew':
3172, 'forb': 1993, 'revlon': 4411, 'pictur': 3919, 'nwp': 3592, 't
aft': 5150, 'tfb': 5237, 'sold': 4839, 'brief': 662, 'dudley': 1576,
'narragansett': 3436, 'narr': 3435, 'bass': 481, 'fort': 2016, 'lin
dner': 2983, 'linder': 2982, 'jone': 2754, 'vine': 5635, 'jnsv': 274
4, 'vulcan': 5667, 'vul': 5666, 'braintre': 640, 'mass': 3152, 'comp
etitor': 1050, 'shoe': 4715, 'buyout': 726, 'nunber': 3588, 'recomme
nd': 4259, 'aprov': 287, 'dismiss': 1461, 'frivol': 2062, 'substan
c': 5055, 'compli': 1056, 'hudson': 2454, 'writeoff': 5863, 'dome':
1518, 'toronto': 5340, 'dmp': 1501, 'subordin': 5043, 'dividend': 1
488, 'retail': 4389, 'markborough': 3135, 'nonmarkborough': 3540, 'e
stat': 1759, 'pension': 3865, 'pclb': 3841, 'tsseedman': 5418, 'die
go': 1423, 'envis': 1724, 'attphilip': 377, 'cgct': 862, 'telephon':
5200, 'telegraph': 5198, 'telecommun': 5195, 'bv': 727, 'apt': 288,
'cie': 928, 'general': 2125, 'telephoniqu': 5201, 'soontobeprivati
s': 4856, 'succe': 5061, 'marc': 3125, 'mathieu': 3158, 'societ': 48
29, 'anonym': 250, 'sat': 4565, 'foresaw': 2003, 'uss': 5575, 'att':
369, 'tn': 5309, 'nv': 3590, 'gloeilampenfabrieken': 2154, 'pgloa':
3901, 'du': 1569, 'midi': 3269, 'mcdppa': 3182, 'mutual': 3424, 'pr
ivatis': 4071, 'makeup': 3104, 'citizen': 945, 'morgan': 3372, 'guar
anti': 2249, 'banqu': 464, 'neufeliz': 3482, 'schlumberg': 4592, 'mal
let': 3107, 'algemen': 154, 'nederland': 3468, 'abnna': 15, 'owner':
3754, 'vicepresid': 5624, 'wim': 5806, 'huisman': 2459, 'envisag':
1725, 'satapt': 4566, 'win': 5807, 'research': 4361, 'aptsat': 289,
'essprx': 1756, 'adapt': 63, 'standard': 4946, 'draft': 1544, 'aban
don': 4, 'introduc': 2649, 'wider': 5788, 'candidaci': 765, 'sieme
n': 4748, 'siegf': 4746, 'alli': 168, 'schneider': 4595, 'schnpa': 4
596, 'jeumontschneid': 2735, 'sweden': 5121, 'telefon': 5196, 'lm':
3004, 'ericsson': 1738, 'ericst': 1739, 'matra': 3159, 'matrpa': 31
60, 'bouygu': 625, 'bouypa': 626, 'italtel': 2687, 'telecom': 5194,
'ntlto': 3583, 'clabir': 948, 'clg': 964, 'agr': 115, 'popsicl': 39
87, 'greenwich': 2213, 'conn': 1095, 'ambrit': 199, 'abi': 10, 'sar
a': 4560, 'sle': 4793, 'worht': 5844, 'license': 2961, 'maker': 310
3, 'distributor': 1475, 'novelti': 3575, 'courier': 1193, 'montrea
l': 3365, 'em': 1662, 'calgarybas': 741, 'messeng': 3233, 'gelco': 2
119, 'gel': 2118, 'complement': 1053, 'heller': 2355, 'broad': 672,

'governor': 2189, 'diversif': 1483, 'thrift': 5277, 'handl': 2287, 'function': 2085, 'armslength': 321, 'serv': 4661, 'strength': 5017, 'word': 5843, 'enterpris': 1711, 'loot': 3032, 'leagu': 2908, 'tabl': 5146, 'arrest': 327, 'srd': 4925, 'cleveland': 963, 'compos': 1059, 'nonexect': 3535, 'purpos': 4145, 'monitor': 3354, 'boston': 617, 'cravath': 1211, 'swain': 5119, 'moor': 3369, 'macandrew': 3083, 'rev': 4403, 'outstand': 3732, 'formal': 2011, 'tesco': 5227, 'extend': 1831, 'hillard': 2388, 'tscol': 5414, 'contest': 1124, 'p': 3757, 'penc': 3858, 'dc': 1305, 'dilut': 1433, 'norfolk': 3550, 'nsc': 3582, 'issuabl': 2682, 'oustand': 3715, 'safeguard': 4525, 'scientific': 4600, 'sfe': 4675, 'prussia': 4128, 'sanchez': 4550, 'comput': 1063, 'softwar': 4835, 'malvern': 3109, 'ownership': 3755, 'cyacq': 1266, 'cyclop': 1268, 'citicorp': 944, 'reidashman': 4305, 'rea': 4226, 'divest': 1485, 'clara': 950, 'entertain': 1712, 'steven': 4981, 'reid': 4304, 'founder': 2025, 'ten': 5211, 'sept': 4656, 'osr': 3708, 'osrc': 3709, 'bethpag': 546, 'ny': 3593, 'telcom': 5193, 'former': 2013, 'film': 1928, 'mostli': 3385, 'knickerbock': 2826, 'jean': 2723, 'nobel': 3520, 'bofor': 598, 'stockholm': 4998, 'noblst': 3522, 'ordnanc': 3697, 'weapon': 5724, 'contraven': 1135, 'swedish': 5122, 'uncov': 5482, 'unaccept': 5471, 'regrett': 4297, 'incid': 2535, 'histori': 2400, 'larserik': 2876, 'thunholm': 5284, 'ander': 231, 'carlberg': 793, 'inquiri': 2597, 'reveal': 4405, 'smuggl': 4816, 'tylan': 5452, 'tyln': 5453, 'furnac': 2089, 'carson': 804, 'kahn': 2779, 'harri': 2309, 'justic': 2774, 'dept': 1387, 'dart': 1293, 'supermarket': 5096, 'sgl': 4678, 'unsolicit': 5536, 'confidenti': 1085, 'bidder': 557, 'brochur': 675, 'book': 607, 'employe': 1675, 'suspect': 5114, 'leverag': 2945, 'friendli': 2059, 'willing': 5799, 'understand': 5489, 'prohibit': 4091, 'unreason': 5528, 'togeth': 5315, 'leonard': 2934, 'lieberman': 2965, 'dougherti': 1537, 'murray': 3420, 'levin': 2947, 'implic': 2519, 'volit': 5648, 'waiv': 5674, 'aspect': 349, 'path': 3823, 'chain': 864, 'rickel': 4422, 'centr': 851, 'greenmail': 2212, 'enact': 1680, 'poison': 3973, 'pill': 3927, 'suit': 5072, 'fiduciari': 1922, 'unwant': 5543, 'suitor': 5074, 'takeov': 5158, 'hecla': 2347, 'hl': 2406, 'green': 2211, 'creek': 1221, 'amselco': 219, 'miner': 3294, 'engag': 1692, 'engin': 1694, 'goldsilverleadzinc': 2176, 'ore': 3698, 'bodi': 596, 'admiralti': 78, 'southwest': 4872, 'juneau': 2771, 'munford': 3418, 'mfd': 3247, 'bradbury': 638, 'dyer': 1590, 'paragon': 3788, 'solar': 4838, 'ssdn': 4930, 'inkey': 2593, 'miami': 3253, 'sundanc': 5085, 'condom': 1077, 'spermicid': 4902, 'ingredi': 2588, 'sexual': 4674, 'transmiss': 5373, 'cite': 939, 'forschung': 2015, 'eurocapit': 1766, 'shortli': 4726, 'usbas': 5566, 'europharmaceut': 1770, 'vagin': 5588, 'distributorship': 1476, 'drug': 1566, 'fda': 1886, 'lengthi': 2931, 'antisept': 258, 'unrel': 5529, 'know': 2828, 'brand': 643, 'carlton': 794, 'exclus': 1799, 'oneyear': 3668, 'obtain': 3605, 'publicli': 4136, 'definit': 1346, 'divestitur': 1486, 'ameritrust': 206, 'amtr': 223, 'cleveland': 962, 'alfr': 153, 'lerner': 2935, 'prog': 4087, 'gat': 2108, 'asu': 363, 'detroit': 1405, 'von': 5656, 'princip': 4061, 'drexell': 1555, 'burnham': 718, 'lambert': 2862, 'donaldson': 1524, 'lufkin': 3066, 'jenrett': 2727, 'counderwrit': 1180, 'owenscorn': 3751, 'ocf': 3614, 'toledo': 5319, 'ohio': 3632, 'fibergla': 1920, 'previouslyannounc': 4055, 'foam': 1983, 'atla': 365, 'roof': 4469, 'undisclosed': 5498, 'xtra': 5880, 'xtr': 5879, 'fruehauf': 2068, 'ftr': 2071, 'rentco': 4333, 'trailer': 5363, 'wholli': 5780, 'up': 5544, 'calmat': 748, 'czm': 1274, 'brierley': 664, 'zealand': 5914, 'chrysler': 923, 'benefici': 530, 'bnl': 590, 'allentown': 166, 'portion': 3994, 'portfolio': 3993, 'ctl': 1245, 'monro': 3356, 'cabl': 733, 'michigan': 3257, 'tennesse': 5218, 'arkansa': 316, 'louisiana': 3044, 'ksz': 2845, 'balli': 441, 'bli': 584, 'nugget': 3585, 'casino': 810, 'hotel': 2437, 'atlant': 366, 'jersey': 2732, 'parcel': 3789, 'assumpt': 359, 'branch': 642, 'deerfield': 1336, 'beach': 498, 'hollyw

ood': 2419, 'guardian': 2250, 'goldom': 2174, 'micro': 3258, 'smsi': 4815, 'acuir': 60, 'supermac': 5095, 'supplier': 5103, 'disc': 1448, 'drive': 1559, 'subsystem': 5058, 'appl': 274, 'mountain': 3391, 'scheringplough': 4587, 'sgp': 4679, 'mull': 3413, 'dr': 1543, 'scholl': 4598, 'madison': 3091, 'nj': 3515, 'merril': 3231, 'lynch': 3078, 'mer': 3224, 'schere': 4586, 'puerto': 4138, 'calni': 749, 'clni': 970, 'pepsico': 3869, 'pep': 3868, 'mateo': 3155, 'intendsd': 2617, 'oppenheim': 3683, 'altern': 187, 'franchise': 2038, 'taco': 5148, 'bell': 522, 'restaur': 4380, 'oregon': 3699, 'petit': 3897, 'boulangeri': 622, 'bakeri': 435, 'seattl': 4617, 'nine': 3506, 'novemb': 3576, 'lariv': 2875, 'inadeuq': 2530, 'conting': 1127, 'iv': 2691, 'correct': 1164, 'batm': 484, 'sec': 4618, 'antitakeov': 260, 'mart': 3145, 'kresg': 2844, 'underli': 5485, 'fw': 2094, 'woolworth': 5840, 'z': 5912, 'stevenson': 4982, 'forerunn': 2002, 'casual': 814, 'scatter': 4580, 'downtown': 1541, 'suburban': 5060, 'gencorp': 2123, 'gy': 2261, 'wagner': 5672, 'afg': 101, 'akron': 134, 'ohioba': 3633, 'tire': 5302, 'plasticmak': 3948, 'aerospac': 94, 'receipt': 4251, 'reynold': 4412, 'will': 5794, 'fargo': 1867, 'wfc': 5760, 'shearson': 4696, 'axp': 416, 'sure': 5105, 'headquart': 2331, 'segment': 4634, 'rko': 4445, 'licens': 2960, 'station': 4960, 'wor': 5842, 'khj': 2806, 'imbal': 2507, 'nyse': 3596, 'bridg': 661, 'bind': 569, 'syndic': 5140, 'promptli': 4100, 'lear': 2911, 'sieglers': 4747, 'withdrew': 5824, 'usx': 5579, 'x': 5875, 'roderick': 4456, 'speed': 4898, 'club': 975, 'pain': 3768, 'uncorrect': 5481, 'abroad': 20, 'wrong': 5867, 'massiv': 3153, 'fundment': 2088, 'qualiti': 4160, 'compet': 1048, 'global': 2151, 'raider': 4188, 'stockhold': 4997, 'allcash': 161, 'preferenti': 4031, 'discourag': 1455, 'bko': 578, 'oildril': 3637, 'cancel': 763, 'californiabas': 744, 'utilicorp': 5582, 'ucu': 5463, 'virginia': 5637, 'dominion': 1522, 'becam': 503, 'chfd': 896, 'jefferson': 2725, 'bristol': 667, 'va': 5585, 'warronton': 5701, 'iroquoi': 2673, 'ibl': 2480, 'rebuff': 4247, 'lyon': 3080, 'su': 5037, 'assert': 351, 'healthvest': 2335, 'hvt': 2473, 'medic': 3203, 'eastwood': 1608, 'hospit': 2434, 'memphi': 3221, 'healthcar': 2333, 'reph': 4344, 'pesch': 3890, 'alpha': 182, 'bei': 515, 'beih': 516, 'iveyrowton': 2693, 'atlanta': 367, 'nashvil': 3441, 'tennbas': 5216, 'moodi': 3368, 'downgrad': 1539, 'statu': 4962, 'uncertain': 5477, 'baa': 420, 'debtsevic': 1322, 'love': 3046, 'tenci': 5307, 'staf': 4936, 'buildier': 701, 'norrel': 3554, 'builder': 700, 'standstil': 4948, 'conven': 1140, 'adjourn': 73, 'reconven': 4262, 'proxi': 4121, 'clearanc': 958, 'solicit': 4842, 'accordingli': 36, 'lima': 2978, 'entireti': 1716, 'jose': 2758, 'palomino': 3776, 'conad': 1066, 'slim': 4799, 'inti': 2644, 'goven': 2185, 'aeroperu': 93, 'flagship': 1958, 'jet': 2733, 'fleet': 1964, 'nonfinanci': 3538, 'applianc': 276, 'copper': 1151, 'empresa': 1676, 'nacion': 3431, 'tintaya': 5300, 'arequipa': 305, 'neither': 3473, 'enriqu': 1707, 'estremadoyro': 1761, 'antonio': 262, 'almenara': 175, 'remit': 4327, 'amfac': 209, 'ama': 193, 'liberti': 2956, 'h': 2264, 'hooogoven': 2426, 'cirex': 938, 'ijmuiden': 2498, 'groep': 2225, 'gloeielampenfabrieken': 2153, 'precis': 4025, 'wax': 5715, 'mould': 3390, 'turnov': 5438, 'car': 784, 'highvalu': 2383, 'stipul': 4993, 'toyota': 5353, 'pnb': 3966, 'manila': 3115, 'philippin': 3909, 'toyot': 5352, 'bankrupt': 462, 'peso': 3892, 'tie': 5290, 'delta': 1362, 'foreclos': 1996, 'wireless': 5818, 'cawll': 827, 'japanbas': 2717, 'exactli': 1786, 'margaretthatch': 3130, 'regard': 4289, 'wireless': 5819, 'unawar': 5475, 'foundat': 2024, 'novamin': 3572, 'breakwat': 656, 'antitrust': 261, 'enforc': 1691, 'disastr': 1446, 'rampan': 4200, 'mania': 3113, 'howard': 2443, 'metzenbaum': 3243, 'that': 5241, 'americantur': 205, 'aaix': 0, 'adventur': 84, 'goaqc': 2168, 'liabil': 2951, 'issuanc': 2683, 'pursuant': 4147, 'reorgan': 4336, 'hanson': 2293, 'han': 2283, 'pcr': 3843, 'scm': 4602, 'unbc': 5476, 'masontown': 3151, 'wichita': 5784, 'wro': 5865, 'fountain': 202

6, 'fga': 1916, 'colo': 1003, 'newli': 3486, 'plaza': 3953, 'flyfai
r': 1980, 'shefra': 4699, 'vacat': 5587, 'nobuo': 3524, 'ito': 2690,
'itj': 2689, 'rival': 4439, 'specifi': 4895, 'digit': 1429, 'idc':
2489, 'telesi': 5202, 'entiti': 1717, 'kokusai': 2836, 'denshin': 1
371, 'denwa': 1374, 'myer': 3426, 'myr': 3427, 'chavin': 883, 'le':
2903, 'rg': 4413, 'dickinson': 1420, 'unabl': 5470, 'warrant': 570
0, 'pend': 3859, 'decre': 1330, 'bit': 575, 'submers': 5041, 'tric
o': 5393, 'tro': 5403, 'disciplin': 1450, 'alleggheni': 164, 'intl':
2645, 'pittsburghbas': 3935, 'document': 1508, 'investig': 2656, 'o
ngo': 3669, 'probe': 4075, 'companyown': 1043, 'aircraft': 128, 'rou
tin': 4486, 'violat': 5636, 'consolid': 1109, 'classact': 953, 'comp
laint': 1052, 'pennsylvania': 3863, 'numer': 3587, 'lawyer': 2896,
'illeg': 2502, 'freez': 2051, 'unfair': 5501, 'vigor': 5631, 'resig
n': 4366, 'buckley': 689, 'gtx': 2243, 'revalu': 4404, 'electron': 1
649, 'phoenix': 3912, 'computer': 1064, 'read': 4230, 'carteret': 80
7, 'savingscbc': 4573, 'morristown': 3379, 'baltimor': 444, 'md': 31
90, 'admir': 77, 'plu': 3961, 'fairfax': 1856, 'maryland': 3149, 'wa
llenber': 5684, 'fought': 2022, 'londonbas': 3021, 'erik': 1741, 'p
enser': 3864, 'smb': 4811, 'st': 4932, 'core': 1153, 'empir': 1673,
'forvaltn': 2019, 'providentia': 4116, 'thre': 5267, 'nobl': 3521,
'b': 419, 'crown': 1232, 'expensis': 1817, 'farflung': 1866, 'preda
t': 4026, 'volvo': 5654, 'volv': 5653, 'copco': 1148, 'ast': 361, 's
tora': 5007, 'koppaberg': 2838, 'skp': 4789, 'oust': 3714, 'skf': 47
85, 'skfr': 4786, 'skanska': 4782, 'skb': 4783, 'frederik': 2046, 'l
undberg': 3073, 'wrest': 5861, 'zurichbas': 5930, 'tycoon': 5451, 'd
iari': 1418, 'alfa': 151, 'laval': 2889, 'alf': 150, 'concentr': 106
8, 'asea': 341, 'fring': 2060, 'ranco': 4201, 'rni': 4447, 'dublin':
1574, 'sieb': 4744, 'nordbanken': 3549, 'auction': 380, 'fermenta':
1906, 'collater': 997, 'frmsst': 2063, 'refaat': 4279, 'elsay': 165
9, 'debtor': 1321, 'inabl': 2528, 'overdu': 3735, 'unoffici': 5518,
'suitabl': 5073, 'withdraw': 5823, 'undertak': 5491, 'sanction': 45
51, 'solv': 4845, 'gotabanken': 2184, 'industrivarden': 2570, 'kjel
l': 2820, 'brandstrom': 644, 'thought': 5264, 'trump': 5411, 'crosb
i': 1228, 'harper': 2308, 'hpr': 2445, 'strateg': 5014, 'yearold': 5
890, 'kidder': 2808, 'peabodi': 3847, 'theodor': 5244, 'cross': 122
9, 'harcourt': 2299, 'brace': 635, 'jovanovich': 2762, 'hbj': 2325,
'unanim': 5473, 'winthrop': 5816, 'knowlton': 2830, 'establish': 17
58, 'pertin': 3887, 'expediti': 1813, 'recapit': 4249, 'class': 952,
'votinmg': 5660, 'bwrlf': 728, 'vancouv': 5597, 'columbia': 1009,
'shrae': 4734, 'exercis': 1803, 'mcdermott': 3179, 'mdr': 3191, 'or
lean': 3705, 'appar': 271, 'incorpor': 2545, 'panamanian': 3779, 'in
valid': 2652, 'armstrong': 322, 'ack': 47, 'thomasvil': 5256, 'nc':
3454, 'gilliam': 2142, 'furnitur': 2090, 'suar': 4923, 'foot': 199
2, 'troutman': 5407, 'yearend': 5889, 'upholst': 5550, 'fb': 1880,
'luvern': 3076, 'pipeston': 3931, 'minnesota': 3304, 'millipor': 32
87, 'mili': 3280, 'protein': 4111, 'huntington': 2466, 'databas': 12
96, 'millicorp': 3285, 'interact': 2621, 'varian': 5601, 'var': 559
8, 'analytichem': 228, 'palo': 3774, 'alto': 189, 'closelyheld': 97
2, 'harbor': 2297, 'sampl': 4547, 'bancorp': 451, 'usbc': 5567, 'lin
coln': 2980, 'midyear': 3272, 'converg': 1142, 'cvgt': 1264, 'oaklea
f': 3598, 'auto': 396, 'neoax': 3475, 'noax': 3519, 'novatron': 357
3, 'lawrencevil': 2893, 'veeco': 5606, 'instrument': 2611, 'vee': 56
05, 'becomlet': 505, 'shelter': 4703, 'carryforward': 802, 'military
specif': 3282, 'avion': 408, 'compon': 1058, 'contractor': 1130, 'co
rroon': 1168, 'black': 579, 'cbl': 832, 'carroon': 800, 'poggiharris
on': 3970, 'rosskopf': 4476, 'rapp': 4211, 'schmidt': 4593, 'mont':
3357, 'fwf': 2095, 'alambr': 156, 'prsl': 4122, 'orang': 3692, 'c
onf': 1079, 'vernon': 5616, 'overthecount': 3746, 'bac': 421, 'unind
entifi': 5507, 'reeserv': 4276, 'downward': 1542, 'linotyp': 2986,
'gmbh': 2162, 'typeset': 5455, 'print': 4063, 'aldn': 144, 'reil':
4306, 'sydney': 5137, 'sunbeam': 5084, 'victa': 5625, 'lawn': 2892,

'mower': 3396, 'sheep': 4697, 'shear': 4695, 'gardena': 2104, 'pacc
ar': 3759, 'pcar': 3840, 'univers': 5511, 'uvrhold': 5583, 'questa
r': 4170, 'str': 5010, 'redeem': 4270, 'accru': 38, 'amex': 208, 'lp
x': 3056, 'wallboard': 5683, 'nonbind': 3529, 'gypsum': 2262, 'norwe
st': 3563, 'telemedia': 5199, 'harrowsmith': 2312, 'equinox': 1730,
'magazin': 3093, 'privatelyown': 4069, 'lifestyl': 2967, 'seventhla
rgest': 4671, 'englishlanguag': 1699, 'edit': 1620, 'editiion': 162
1, 'midapril': 3263, 'clearwat': 960, 'fine': 1935, 'channel': 873,
'minor': 3305, 'hillsdown': 2390, 'cornwal': 1158, 'chill': 903, 'f
ish': 1947, 'pate': 3821, 'painewebberpwj': 3770, 'shaer': 4682, 's
h': 4681, 'mitchel': 3318, 'hutchin': 2471, 'painewebb': 3769, 'ja
n': 2710, 'client': 965, 'topic': 5336, 'klm': 2823, 'klma': 2824,
'commonwealth': 1034, 'bcoml': 495, 'flight': 1968, 'schipol': 458
9, 'airport': 130, 'feeder': 1897, 'jacob': 2701, 'cote': 1176, 'do
r': 1530, 'zurich': 5929, 'suchard': 5064, 'chocol': 913, 'ordinar
i': 3696, 'eurasbank': 1765, 'creditanstaltbankverein': 1217, 'cabvv
i': 735, 'herrhausen': 2368, 'eura': 1764, 'creditanstalt': 1216, 'g
renfel': 2217, 'banca': 447, 'damerica': 1283, 'ditalia': 1478, 'ban
kamerica': 458, 'bacn': 427, 'f': 1840, 'wilhelm': 5793, 'christia
n': 920, 'nonbank': 3528, 'sheet': 4698, 'parent': 3790, 'ffmc': 191
4, 'confidata': 1084, 'wall': 5682, 'street': 5016, 'stockslor': 499
9, 'lor': 3033, 'loral': 3034, 'climb': 968, 'amid': 211, 'rumor': 4
505, 'ge': 2115, 'jack': 2698, 'welch': 5741, 'doesnt': 1510, 'expos
ur': 1829, 'heard': 2337, 'accuraci': 42, 'wouldnt': 5859, 'biogen':
572, 'bgenf': 552, 'cambridg': 751, 'switzerland': 5130, 'waterma
n': 5710, 'gillett': 2141, 'gs': 2239, 'pen': 3857, 'authoris': 394,
'csac': 1240, 'francin': 2039, 'grace': 2191, 'foyergomez': 2032,
'alic': 158, 'lundgren': 3074, 'laser': 2877, 'photon': 3914, 'laz
r': 2898, 'orlando': 3704, 'radix': 4184, 'organ': 3700, 'gluch': 21
56, 'jr': 2763, 'lichter': 2962, 'pierr': 3924, 'schoenheim': 4597,
'roger': 4457, 'kirk': 2818, 'friedkin': 2057, 'fukuhara': 2076, 'j
ay': 2719, 'watnick': 5712, 'ira': 2663, 'goldstein': 2177, 'thurma
n': 5285, 'sasser': 4564, 'michael': 3255, 'clinger': 969, 'uic': 54
64, 'teck': 5188, 'hock': 2412, 'pte': 4133, 'finalis': 1930, 'overs
eachines': 3744, 'pariba': 3793, 'bangkok': 455, 'citibank': 943, 'n
a': 3430, 'national': 3444, 'mccormack': 3174, 'mmc': 3329, 'globe':
2152, 'stamford': 4943, 'metallurg': 3236, 'cypm': 1269, 'silicon':
4754, 'magnesium': 3095, 'ferrosilicon': 1909, 'aluminum': 190, 'mc
lean': 3186, 'mii': 3274, 'cranford': 1210, 'code': 991, 'csx': 124
2, 'sealand': 4612, 'transpacifichawaiigum': 5376, 'tent': 5221, 't
ranspacif': 5375, 'lancer': 2864, 'crowley': 1231, 'maritim': 3133,
'hire': 2396, 'lender': 2929, 'motion': 3386, 'hawaii': 2321, 'gua
m': 2244, 'humana': 2462, 'hum': 2460, 'declar': 1328, 'louisvil': 3
045, 'ky': 2850, 'treatment': 5386, 'twice': 5441, 'ffbn': 1913, 'ba
nkeast': 459, 'benh': 533, 'nashua': 3440, 'nh': 3496, 'fsb': 2070,
'cablesystem': 734, 'acn': 49, 'heritag': 2366, 'hci': 2326, 'subsc
rib': 5048, 'town': 5349, 'compton': 1062, 'hawaiian': 2322, 'garde
n': 2103, 'willowbrook': 5802, 'fireston': 1938, 'florenc': 1972, 'm
tsr': 3408, 'salt': 4544, 'titanium': 5304, 'zirconium': 5924, 'rar
e': 4212, 'earth': 1601, 'unpat': 5521, 'tech': 5182, 'whollyown': 5
781, 'ron': 4467, 'mcarthur': 3173, 'tobacco': 5312, 'confectionar
i': 1080, 'vend': 5608, 'wholes': 5778, 'ctl': 1244, 'marketech': 3
140, 'englewood': 1697, 'cliff': 966, 'diatron': 1419, 'nutrit': 358
9, 'diantron': 1417, 'physician': 3915, 'patientreadi': 3824, 'infu
s': 2585, 'therapi': 5246, 'outpati': 3728, 'jersemi': 2731, 'inpat
i': 2596, 'bed': 506, 'borgwarn': 615, 'bor': 613, 'ok': 3643, 'york
bas': 5902, 'dubili': 1572, 'safeti': 4526, 'mne': 3333, 'auergesell
schaft': 384, 'msa': 3403, 'pti': 4134, 'johannesburg': 2747, 'boar
t': 592, 'anglo': 242, 'angl': 240, 'boartmsa': 593, 'ln': 3006, 'ak
zo': 135, 'arnhem': 323, 'akzoa': 136, 'aarnoud': 1, 'loudon': 3042,
'fibr': 1921, 'enka': 1702, 'debtequ': 1320, 'lotu': 3040, 'assces

s': 350, 'bluefish': 589, 'fulltext': 2080, 'retriev': 4396, 'machin': 3086, 'ibm': 2481, 'compat': 1046, 'goodyear': 2182, 'gt': 2241, 'valuat': 5595, 'chemlawn': 890, 'chem': 887, 'cal': 737, 'mankowski': 3117, 'wnx': 5833, 'arbitrageur': 298, 'shot': 4729, 'mayb': 3170, 'debra': 1318, 'strohmaier': 5026, 'regularli': 4300, 'elliott': 1656, 'schlang': 4590, 'prescott': 4039, 'ball': 440, 'turben': 5429, 'white': 5772, 'knight': 2827, 'rosemary': 4474, 'morbelli': 3371, 'ingal': 2587, 'snyder': 4824, 'servicemast': 4663, 'lp': 3052, 'svm': 5118, 'rollin': 4463, 'rol': 4458, 'pesticid': 3895, 'herbicid': 2364, 'valuabl': 5594, 'residenti': 4364, 'catapult': 816, 'language': 2869, 'entrepren': 1720, 'sevic': 4673, 'refil': 4281, 'transworld': 5379, 'dot': 1533, 'perfect': 3875, 'counsel': 1181, 'bucks tein': 690, 'incomplet': 2541, 'moorehandley': 3370, 'mhco': 3251, 'barker': 470, 'jmr': 2743, 'quaker': 4158, 'upland': 5551, 'sole y': 4841, 'fbc': 1881, 'advertis': 86, 'rioght': 4434, 'affirm': 99, 'edmin': 1623, 'colodni': 1004, 'expedit': 1812, 'satisfactori': 4568, 'dilig': 1431, 'women': 5836, 'txa': 5449, 'bancshar': 453, 'bankgalleria': 461, 'fonderen': 1989, 'gulfway': 2258, 'tabgalleria': 5145, 'rubbermaid': 4501, 'rbd': 4224, 'wooster': 5841, 'vike': 5632, 'brush': 685, 'broom': 680, 'clean': 956, 'tike': 5294, 'toy': 5351, 'dna': 1502, 'dnam': 1503, 'walter': 5686, 'g': 2096, 'mize': 3322, 'represent': 4351, 'lite': 2996, 'lean': 2910, 'beef': 507, 'catt l': 820, 'holbrook': 2414, 'deem': 1334, 'mhc': 3250, 'foreclosur': 1997, 'hanov': 2292, 'ofp': 3630, 'tower': 5348, 'mlnsquarefeet': 3325, 'videotron': 5628, 'exhibit': 1805, 'promexpo': 4094, 'paolo': 3783, 'di': 1415, 'torino': 5339, 'californian': 745, 'turin': 5431, 'istituto': 2684, 'bancario': 449, 'fullyown': 2083, 'wilmington': 5803, 'valley': 5592, 'relev': 4318, 'holiday': 2417, 'hia': 2375, 'comsat': 1065, 'cq': 1204, 'satellit': 4567, 'hinet': 2394, 'inroom': 2599, 'video': 5627, 'assum': 358, 'rossignol': 4475, 'ski': 4787, 'boot': 612, 'tenni': 5219, 'dynastar': 1593, 'cfa': 860, 'warrington': 5702, 'van': 5596, 'ommeren': 3656, 'ph': 3902, 'ommna': 3658, 'lgt': 2950, 'goldsil': 2175, 'onehalf': 3662, 'conrac': 1099, 'cax': 828, 'conract': 1100, 'advisor': 88, 'goldman': 2173, 'sach': 4521, 'maxim': 3166, 'avon': 410, 'giorgio': 2147, 'fragranc': 2035, 'wiesbaden': 5791, 'unveil': 5542, 'directorgener': 1440, 'inf': 2576, 'apolonio': 270, 'ruiz': 4503, 'ligero': 2972, 'prevail': 4050, 'lag': 2858, 'fashion': 1873, 'wine': 5811, 'fouryear': 2031, 'fad': 1849, 'lack': 2857, 'oecd': 3620, 'businessmen': 721, 'realis': 4235, 'shrink': 4735, 'deregul': 1390, 'pin': 3928, 'boca': 595, 'raton': 4219, 'pryor': 4129, 'panel': 3781, 'argentinabrazil': 308, 'jump': 2769, 'dant': 1291, 'caputo': 783, 'integr': 2614, 'gatt': 2110, 'taupo': 5172, 'halt': 2276, 'bilater': 563, 'ministeri': 3300, 'ineffici': 2572, 'rational': 4218, 'emphasis': 1672, 'punta': 4142, 'del': 1351, 'uruguay': 5561, 'impetu': 2517, 'endeavour': 1683, 'ridden': 4424, 'worthwhil': 5857, 'standpoint': 4947, 'ough t': 3713, 'mechan': 3199, 'reliev': 4322, 'amortis': 215, 'tunisia': 5428, 'egypt': 1633, 'morocco': 3375, 'jordan': 2756, 'framework': 2036, 'chip': 907, 'react': 4228, 'angrili': 243, 'retaliatori': 4392, 'allegedli': 163, 'live': 3000, 'microchip': 3259, 'eiaj': 1634, 'shoichi': 4716, 'saba': 4520, 'irrat': 2674, 'assess': 352, 'government': 2187, 'reconsid': 4261, 'emot': 1669, 'bias': 553, 'slap': 4790, 'cutpric': 1263, 'americanmad': 204, 'salvag': 4546, 'slash': 4791, 'regret': 4296, 'viceforeign': 5623, 'zhou': 5919, 'nan': 3434, 'extern': 1834, 'commission': 1025, 'willi': 5797, 'clercq': 961, 'affair': 96, 'whichev': 5771, 'dane': 1287, 'baldrig': 438, 'koreantaiwan': 2841, 'malcolm': 3106, 'advisori': 89, 'widerang': 5789, 'fujitsu': 2075, 'fairchild': 1855, 'lastli': 2880, 'acid': 46, 'unresolv': 5531, 'consensu': 1102, 'graham': 2195, 'hillier': 2389, 'deleg': 1354, 'roadbut': 4449, 'translat': 5372, 'quickli': 4173, 'hard': 2300, 'coupl': 1191, 'tough': 5343, 'ludicr': 3064, 'immens': 2

511, 'nobodi': 3523, 'answer': 252, 'nonproduct': 3545, 'ration': 42
17, 'ineffect': 2571, 'prioriti': 4066, 'realist': 4236, 'mission':
3315, 'taipei': 5154, 'shop': 4720, 'elimin': 1655, 'undersecretar
i': 5488, 'allen': 165, 'walli': 5685, 'eishiro': 1639, 'saito': 453
5, 'keidanren': 2790, 'stave': 4963, 'subcabinetlevel': 5038, 'antag
onis': 253, 'halv': 2277, 'cope': 1149, 'abolit': 18, 'trigger': 539
4, 'accus': 43, 'arnott': 324, 'appeal': 272, 'waiver': 5675, 'contr
adict': 1132, 'australian': 390, 'flaw': 1961, 'coverag': 1198, 'jus
tif': 2775, 'dubiou': 1573, 'dairi': 1278, 'peanut': 3849, 'samuel':
4548, 'ambassador': 198, 'invit': 2659, 'ninetytwo': 3509, 'bargai
n': 469, 'tourism': 5346, 'vs': 5663, 'finish': 1936, 'toughen': 534
4, 'dan': 1285, 'rostenkowski': 4477, 'defer': 1342, 'gephardt': 213
2, 'exact': 1785, 'committess': 1030, 'wave': 5714, 'wrap': 5860, 's
prinkel': 4918, 'beryl': 541, 'futher': 2092, 'engend': 1693, 'depre
ci': 1385, 'postwar': 4008, 'reneg': 4329, 'restraint': 4383, 'sol
e': 4840, 'relianc': 4320, 'fault': 1877, 'undesir': 5496, 'tigrel':
5293, 'consequ': 1104, 'acknowledg': 48, 'attain': 371, 'programm':
4089, 'creditworthi': 1220, 'allevi': 167, 'mediumterm': 3207, 'bla
me': 581, 'municip': 3419, 'gear': 2116, 'liberalis': 2955, 'freer':
2050, 'arifin': 312, 'siregar': 4769, 'antara': 254, 'expansionar
i': 1810, 'budgetari': 693, 'sumatran': 5079, 'medan': 3201, 'sumatr
a': 5078, 'urgent': 5560, 'delor': 1361, 'featur': 1890, 'promin': 4
095, 'hk': 2404, 'slid': 4794, 'censu': 846, 'reexport': 4278, 'entr
epot': 1719, 'outpac': 3727, 'india': 2556, 'countertrad': 1186, 'no
ncommunist': 3530, 'jute': 2777, 'machineri': 3087, 'stc': 4966, 'mm
tc': 3331, 'bloc': 585, 'swap': 5120, 'textil': 5236, 'envoy': 1726,
'nakason': 3432, 'yasuhiro': 5885, 'matsunaga': 3161, 'kyodo': 285
1, 'usjapan': 5574, 'kansai': 2781, 'tell': 5206, 'intensifi': 2619,
'ronald': 4468, 'meanwhil': 3196, 'ldp': 2902, 'supercomput': 5093,
'surplus': 5109, 'rude': 4502, 'miniist': 3297, 'onno': 3670, 'foru
m': 2018, 'aggrav': 110, 'indebt': 2551, 'enourm': 1706, 'weaken': 5
720, 'customsclear': 1260, 'dollaryen': 1516, 'chipmak': 908, 'lind
a': 2981, 'siegl': 4745, 'strongarm': 5028, 'complianc': 1057, 'preda
tori': 4027, 'miti': 3319, 'battl': 485, 'breath': 657, 'scrap': 460
7, 'harder': 2301, 'recalcitr': 4248, 'abid': 11, 'hot': 2436, 'pete
r': 3896, 'wolff': 5834, 'inexpens': 2575, 'escap': 1746, 'grey': 22
19, 'loophol': 3031, 'sneak': 4820, 'jeopardi': 2728, 'dump': 1580,
'nonamerican': 3527, 'nec': 3462, 'keller': 2792, 'capel': 774, 'be
st': 543, 'carol': 795, 'ryavec': 4515, 'matt': 3162, 'aizawa': 132,
'outfox': 3720, 'scheme': 4585, 'steve': 4980, 'donovan': 1527, 'mm
i': 3330, 'kk': 2821, 'annoy': 248, 'shrug': 4736, 'guidanc': 2253,
'easier': 1604, 'campaign': 755, 'foreignmad': 2000, 'mix': 3320,
'urgenc': 5559, 'kilobit': 2813, 'dram': 1546, 'dynam': 1592, 'rand
om': 4204, 'mere': 3228, 'transit': 5371, 'megabit': 3210, 'comebac
k': 1014, 'watch': 5708, 'fate': 1876, 'consortium': 1110, 'thorni':
5261, 'sia': 4739, 'winner': 5812, 'yendollar': 5895, 'killen': 281
1, 'jitter': 2741, 'flex': 1966, 'muscl': 3421, 'prolif': 4093, 'sou
nd': 4861, 'strategi': 5015, 'aldo': 145, 'roldan': 4459, 'economet
r': 1614, 'tackl': 5147, 'movement': 3395, 'pure': 4144, 'pretti': 4
049, 'brazilian': 649, 'background': 425, 'vengeanc': 5611, 'ico': 2
485, 'reestablish': 4277, 'intransig': 2647, 'anyon': 266, 'els': 16
58, 'nineti': 3508, 'eighti': 1637, 'lb': 2899, 'distant': 1471, 'si
xcent': 4775, 'jorio': 2757, 'dauster': 1298, 'rio': 4433, 'janeir
o': 2711, 'breakdown': 654, 'focu': 1985, 'steppedup': 4977, 'loew
i': 3016, 'prudentialbach': 4126, 'motiv': 3387, 'cant': 767, 'nervo
u': 3476, 'tommi': 5324, 'eshleman': 1749, 'asa': 337, 'uncertain
i': 5478, 'forfeit': 2008, 'melicia': 3215, 'inflationari': 2579, 'e
nviron': 1722, 'inclin': 2537, 'mute': 3423, 'oversold': 3745, 'hono
r': 2425, 'ethanol': 1763, 'stretch': 5021, 'cocoa': 990, 'flat': 19
59, 'throw': 5280, 'collaps': 996, 'hagu': 2266, 'modestli': 3343,
'patrick': 3825, 'leahi': 2909, 'dvt': 1588, 'melcher': 3214, 'dmon

t': 1500, 'voluntari': 5652, 'donat': 1525, 'concession': 1072, 'ban
gladesh': 456, 'seneg': 4649, 'passag': 3814, 'kenya': 2799, 'cost
a': 1172, 'rica': 4415, 'malaysia': 3105, 'yearonyear': 5891, 'suppl
ementari': 5101, 'januaryfebruari': 2714, 'mask': 3150, 'bfg': 551,
'twomonth': 5446, 'extrem': 1837, 'hinder': 2392, 'ho': 2411, 'high
level': 2380, 'juic': 2766, 'hajim': 2268, 'tamura': 5163, 'ward': 5
692, 'spokesmen': 4908, 'prepay': 4037, 'everyth': 1779, 'tune': 542
6, 'maamoun': 3081, 'youssef': 5905, 'warpln': 5699, 'flew': 1965,
'sorti': 4859, 'iranan': 2665, 'clash': 951, 'boat': 594, 'men': 32
22, 'northen': 3558, 'tip': 5301, 'wer': 5749, 'sunk': 5089, 'occu
p': 3609, 'fled': 1963, 'deploy': 1380, 'hy': 2474, 'ssn': 4931, 'st
yx': 5036, 'sink': 4767, 'fraction': 2033, 'explos': 1826, 'fire': 1
937, 'cb': 830, 'civilian': 947, 'regenc': 4290, 'cruis': 1235, 'm
v': 3425, 'regent': 4291, 'monmouth': 3355, 'renov': 4331, 'star': 4
952, 'fiveyear': 1954, 'irv': 2678, 'v': 5584, 'counter': 1184, 'cam
pina': 758, 'compound': 1060, 'unrest': 5532, 'antiinfl': 257, 'nort
hwest': 3561, 'sao': 4558, 'paulo': 3829, 'wilson': 5804, 'gome': 21
78, 'moura': 3392, 'stoppag': 5006, 'came': 753, 'seamen': 4613, 'pe
trobra': 3898, 'forestal': 2006, 'calm': 747, 'willingess': 5800, 'm
arin': 3132, 'shipown': 4713, 'protest': 4112, 'unfairli': 5502, 'di
e': 1422, 'accid': 33, 'electrocut': 1647, 'weekend': 5730, 'steam':
4970, 'dw': 1589, 'handler': 2288, 'graan': 2190, 'mij': 3275, 'ge
m': 2120, 'discharg': 1449, 'grainsoilse': 2198, 'melvin': 3217, 'si
m': 4757, 'memorandum': 3219, 'tom': 5322, 'kay': 2786, 'tonnag': 53
31, 'congressmen': 1094, 'rkan': 4444, 'defeat': 1338, 'crusher': 12
37, 'overnight': 3740, 'shipmentgreec': 4712, 'lasthalf': 2879, 'sai
dtaiwan': 4534, 'olein': 3649, 'yesterdaypakistan': 5897, 'retend':
4393, 'weekth': 5732, 'red': 4269, 'wehat': 5733, 'aprilnov': 286,
'elig': 1654, 'medium': 3205, 'colombia': 1005, 'malt': 3108, 'eep
h': 1627, 'flaxse': 1962, 'tonen': 5328, 'pakistan': 3771, 'saidac
t': 4531, 'timechart': 5297, 'imprt': 2525, 'saiddri': 4533, 'gree
k': 2210, 'redman': 4272, 'exacerb': 1784, 'ascertain': 340, 'andre
a': 234, 'papander': 3784, 'aggressor': 113, 'friend': 2058, 'powe
l': 4015, 'dayold': 1302, 'docemart': 1506, 'harbour': 2298, 'dures
s': 1583, 'sarney': 4562, 'despatch': 1395, 'defiant': 1343, 'mood':
3367, 'do': 1505, 'quick': 4172, 'denounc': 1370, 'unpatriot': 552
2, 'cruzado': 1239, 'captain': 782, 'ghana': 2137, 'almir': 176, 'pa
zzionotto': 3837, 'baseless': 477, 'bbc': 490, 'moslem': 3384, 'casp
ar': 811, 'weinberg': 5737, 'whatev': 5763, 'lane': 2867, 'propagand
a': 4102, 'saddam': 4522, 'hussein': 2470, 'legitim': 2924, 'what':
5762, 'navig': 3450, 'arent': 304, 'sympathet': 5139, 'carrier': 79
8, 'kitti': 2819, 'hawk': 2323, 'rescu': 4360, 'ferri': 1908, 'zeebr
ugg': 5915, 'herald': 2363, 'townsend': 5350, 'thoresen': 5260, 'div
er': 1481, 'passeng': 3815, 'trap': 5380, 'dover': 1538, 'capsiz': 7
81, 'pier': 3923, 'squabbl': 4922, 'thoma': 5255, 'hammer': 2281, 'a
bidjan': 12, 'ivori': 2694, 'ivorian': 2695, 'chamber': 867, 'monthl
i': 3362, 'seafar': 4610, 'rollback': 4462, 'roman': 4464, 'gralewic
z': 2199, 'percenti': 3873, 'concili': 1073, 'crew': 1222, 'fbt': 18
84, 'payout': 3835, 'montval': 3366, 'qtli': 4155, 'div': 1479, 'flt
i': 1977, 'shr': 4733, 'mth': 3405, 'triton': 5401, 'trro': 5408, 't
h': 5239, 'jolla': 2753, 'avg': 406, 'discontinuu': 1453, 'realiti': 4
240, 'rsi': 4494, 'split': 4905, 'dec': 1323, 'deduct': 1333, 'metpr
o': 3240, 'mpr': 3398, 'harleysvil': 2303, 'qtrli': 4157, 'parker':
3797, 'pkd': 3937, 'tulsa': 5423, 'okla': 3645, 'quarterli': 4166,
'biflyx': 559, 'nd': 3455, 'grove': 2230, 'lpp': 3055, 'cbtf': 835,
'fairmont': 1858, 'wva': 5873, 'microdym': 3260, 'mcdi': 3180, 'oca
la': 3607, 'idb': 2487, 'idbx': 2488, 'cclr': 838, 'hike': 2385, 'ti
verton': 5306, 'danah': 1286, 'dhr': 1414, 'rale': 4196, 'threefortw
o': 5274, 'fe': 1887, 'sfp': 4676, 'cathay': 819, 'airway': 131, 'ca
phhk': 775, 'swire': 5126, 'swpchk': 5133, 'hongkong': 2424, 'hkbhh
k': 2405, 'wausau': 5713, 'wsau': 5868, 'wi': 5783, 'pershar': 3881,

'restat': 4379, 'emco': 1666, 'longmont': 3028, 'pulitz': 4139, 'pl
tzc': 3960, 'loui': 3043, 'armor': 319, 'armr': 320, 'lana': 2863,
'webcor': 5727, 'awk': 415, 'teeco': 5190, 'partnerhsip': 3810, 'pr
oce': 4077, 'mccormick': 3175, 'mccrk': 3176, 'hunt': 2465, 'aati':
2, 'stonington': 5002, 'uam': 5460, 'lser': 3058, 'wipe': 5817, 'le
ngth': 2930, 'reincorpor': 4307, 'deb': 1314, 'philadelphia': 3907,
'bdm': 497, 'geoffrion': 2128, 'leclerc': 2918, 'dltll': 1497, 'deb
it': 1317, 'warehous': 5693, 'wclb': 5718, 'onetim': 3667, 'unprofi
t': 5523, 'illinois': 2503, 'indiana': 2558, 'invacar': 2651, 'ivcr':
2692, 'elyria': 1661, 'roy': 4489, 'weston': 5756, 'wstna': 5869,
'chester': 894, 'effecit': 1629, 'firstg': 1943, 'southeastern': 48
68, 'northwestern': 3562, 'fullyear': 2081, 'jmed': 2742, 'mo': 333
5, 'gruntal': 2236, 'gru': 2235, 'vtx': 5665, 'farmingdal': 1870, 'f
iveforfour': 1952, 'newscop': 3491, 'writedown': 5862, 'swissair': 5
128, 'swszz': 5135, 'swiss': 5127, 'pc': 3839, 'pcqt': 3842, 'bindle
y': 570, 'incbindth': 2532, 'indianapoli': 2559, 'ind': 2550, 'unrea
l': 5525, 'photograph': 3913, 'pscx': 4130, 'rochest': 4455, 'notet
o': 3566, 'jhsn': 2736, 'hershey': 2370, 'hso': 2450, 'pasadena': 38
12, 'russ': 4513, 'tog': 5314, 'rt': 4496, 'canbra': 762, 'dlrcommo
n': 1492, 'art': 329, 'faaa': 1841, 'westworld': 5757, 'wchi': 5717,
'forest': 2005, 'interco': 2625, 'gemini': 2122, 'vm': 5643, 'vmlp
z': 5644, 'depositari': 1382, 'rtc': 4497, 'stewartwarn': 4983, 'st
x': 5035, 'fasb': 1872, 'nonoper': 3543, 'nonrecur': 3546, 'plexu':
3958, 'zambia': 5913, 'halfyear': 2274, 'rite': 4438, 'srfi': 4926,
'harrisburg': 2310, 'foremost': 2001, 'whittak': 5775, 'wkr': 5829,
'disposit': 1465, 'hsa': 2449, 'mainten': 3099, 'tic': 5288, 'whitt
ar': 5776, 'dofascoxinc': 1511, 'bertramtrojan': 540, 'investcorp':
2655, 'benn': 534, 'marrel': 3144, 'juster': 2773, 'incur': 2549,
'yearearli': 5888, 'westinghous': 5754, 'earningsshr': 1600, 'twofo
ron': 5445, 'cinema': 933, 'kren': 2843, 'scott': 4605, 'gold': 217
1, 'ctsshr': 1246, 'newpark': 3489, 'premdor': 4034, 'wood': 5838,
'door': 1529, 'quantech': 4161, 'qant': 4151, 'alevittown': 148, 'm
idland': 3270, 'mm': 3328, 'coven': 1196, 'henri': 2362, 'ginsburg':
2145, 'bernard': 539, 'weinblatt': 5738, 'rsign': 4495, 'ginsberg':
2144, 'n': 3429, 'hecht': 2345, 'pse': 4131, 'pow': 4014, 'teletra
k': 5203, 'helm': 2356, 'salant': 4538, 'qslt': 4154, 'obion': 3600,
'adequaci': 69, 'ray': 4221, 'creditgor': 1218, 'dlrsd': 1495, 'pos
ttax': 4007, 'prechapt': 4022, 'logic': 3018, 'dlog': 1490, 'anahei
m': 225, 'gre': 2205, 'zinc': 5923, 'silver': 4756, 'smelt': 4812,
'snbl': 4818, 'salisbury': 4542, 'wlwa': 5831, 'endfebruari': 1684,
'wale': 5678, 'markdown': 3136, 'reorganis': 4337, 'zn': 5925, 'rai
nbow': 4192, 'safeway': 4527, 'rth': 4499, 'dist': 1469, 'allison':
170, 'all': 160, 'quinta': 4177, 'inn': 2595, 'inclqmr': 2538, 'tu
e': 5421, 'nynex': 3595, 'nyn': 3594, 'consecut': 1101, 'electrome
d': 1648, 'aeco': 91, 'technodyn': 5186, 'tnd': 5310, 'electro': 164
6, 'audio': 381, 'ead': 1595, 'tomaso': 5323, 'dtom': 1568, 'lire':
2993, 'nbct': 3453, 'loma': 3019, 'nettleton': 3480, 'lnf': 3007,
'transamerica': 5367, 'ta': 5144, 'age': 107, 'reut': 4399, 'hopki
n': 2429, 'minn': 3302, 'carryback': 801, 'spindl': 4903, 'yr': 590
6, 'aftertax': 105, 'twoforf': 5444, 'catena': 818, 'volvst': 5655,
'safvean': 4529, 'jacobson': 2702, 'och': 3615, 'ponsbach': 3984,
'fondkommiss': 1990, 'foil': 1987, 'jude': 2764, 'stjm': 4994, 'one
tenth': 3665, 'commenc': 1017, 'frank': 2041, 'fbh': 1882, 'briarcli
ff': 660, 'manor': 3120, 'macgregor': 3085, 'sport': 4912, 'mgsnd':
3249, 'rutherford': 4514, 'cpl': 1203, 'cntr': 983, 'davenport': 12
99, 'iowa': 2662, 'jcb': 2721, 'jackson': 2699, 'mich': 3254, 'payb
l': 3832, 'presplit': 4045, 'ah': 123, 'robin': 4453, 'qrah': 4153,
'richmond': 4421, 'robbin': 4451, 'stanley': 4950, 'swk': 5131, 'es
so': 1755, 'saf': 4523, 'essfpa': 1754, 'xonn': 5878, 'reintegr': 43
08, 'tai': 5152, 'vw': 5669, 'seat': 4616, 'wolfsburg': 5835, 'volks
wagen': 5650, 'vowgf': 5662, 'sociedad': 4828, 'espanola': 1752, 'au

tomovil': 399, 'turismo': 5432, 'werner': 5751, 'supervisor': 5099, 'boersenzeitung': 597, 'cc': 836, 'lilli': 2977, 'coat': 989, 'licia': 2963, 'welbilt': 5740, 'welb': 5739, 'hyde': 2475, 'poclain': 3968, 'par': 3786, 'milford': 3279, 'nmsb': 3518, 'reord': 4335, 'fb': 1883, 'citg': 940, 'omit': 3655, 'miss': 3313, 'ommit': 3657, 'quartelri': 4163, 'default': 1337, 'laargest': 2853, 'intertel': 2640, 'intla': 2646, 'chandler': 871, 'ariz': 314, 'ceradyn': 855, 'crdn': 1213, 'srv': 4928, 'scr': 4606, 'bermuda': 538, 'sherwood': 4706, 'sherwood': 4707, 'lease': 2913, 'sulpetro': 5076, 'oct': 3616, 'sulbath': 5075, 'robn': 4454, 'dayton': 1303, 'backlog': 426, 'asamera': 338, 'asm': 348, 'yrago': 5907, 'shoney': 4718, 'shon': 4717, 'tenn': 5215, 'envirosur': 1723, 'env': 1721, 'buffalo': 695, 'nov': 3571, 'joseph': 2759, 'saj': 4536, 'sand': 4552, 'sndcf': 4819, 'solang': 4837, 'santisauthor': 4554, 'onefor': 3660, 'secondquart': 4623, 'jerri': 2730, 'shattner': 4694, 'placement': 3940, 'aura': 386, 'penni': 3862, 'appreci': 279, 'hitachi': 2402, 'trademark': 5356, 'accessori': 32, 'disk': 1460, 'solidst': 4843, 'devic': 1410, 'printer': 4064, 'manifram': 3114, 'sapien': 4559, 'languag': 2868, 'straight': 5011, 'tei': 5192, 'unqualifi': 5524, 'audit': 382, 'auditor': 383, 'arthur': 330, 'andersen': 232, 'qualifi': 4159, 'subsequ': 5050, 'carter': 806, 'hawley': 2324, 'hale': 2269, 'montli': 3364, 'bachelor': 423, 'workforc': 5847, 'syosset': 5141, 'sheppard': 4705, 'morri': 3377, 'moliv': 3349, 'campbel': 756, 'soup': 4862, 'cpb': 1201, 'camden': 752, 'chubb': 925, 'paloma': 3775, 'biscuit': 574, 'ubisl': 5462, 'ub': 5461, 'unalloc': 5472, 'nlcsth': 3516, 'nixdorf': 3514, 'swpchk': 5134, 'caphhkg': 776, 'primarili': 4059, 'enlarg': 1703, 'obvious': 3606, 'taikoo': 5153, 'shing': 4709, 'dockyard': 1507, 'depot': 1384, 'thompson': 5257, 'tm': 5308, 'soni': 4853, 'forese': 2004, 'snet': 4821, 'tsunao': 5419, 'hashimoto': 2317, 'irregular': 2675, 'fivemonth': 1953, 'gloomi': 2155, 'camerarecord': 754, 'betaformat': 545, 'tape': 5167, 'vtr': 5664, 'compact': 1039, 'player': 3952, 'walkman': 5681, 'colour': 1008, 'esystem': 1762, 'esi': 1750, 'bayer': 488, 'leverkusen': 2946, 'baygf': 489, 'extent': 1833, 'materialis': 3157, 'polyurethan': 3983, 'freepor': 2049, 'fmr': 1981, 'mthli': 3406, 'paradis': 3787, 'parf': 3791, 'firstcorp': 1942, 'fcr': 1885, 'raleigh': 4197, 'centel': 848, 'cnt': 982, 'reuss': 4398, 'cellular': 843, 'tripl': 5398, 'electrosensor': 1650, 'mcdonald': 3181, 'mcd': 3178, 'daniel': 1290, 'fast': 1874, 'multipl': 3415, 'salad': 4537, 'bath': 483, 'burger': 716, 'wendy': 5747, 'justifi': 2776, 'mcm': 3187, 'casualti': 815, 'omaha': 3653, 'commut': 1038, 'bsci': 496, 'zfour': 5917, 'technitrol': 5185, 'tnl': 5311, 'modern': 3341, 'mocon': 3338, 'tcb': 5176, 'chl': 911, 'ben': 528, 'commercechem': 1021, 'outofst': 3726, 'lawmak': 2891, 'interst': 2639, 'hardhit': 2302, 'struggl': 5033, 'republicbank': 4354, 'albn': 141, 'mcorp': 3189, 'interfirst': 2630, 'bancorpor': 452, 'corner': 1156, 'nonperform': 3544, 'tidal': 5289, 'absorpt': 25, 'empti': 1677, 'vacanc': 5586, 'offens': 3622, 'montgomeri': 3360, 'adjac': 71, 'metropolitan': 3242, 'dallasfort': 1281, 'ifc': 2496, 'rpt': 4492, 'archriv': 301, 'virtual': 5638, 'secondlargest': 4622, 'roost': 4471, 'glynw': 2160, 'glynl': 2159, 'lethbridg': 2941, 'stafford': 4938, 'boy': 629, 'pbi': 3838, 'manni': 3118, 'moe': 3345, 'dylex': 1591, 'mellon': 3216, 'mel': 3211, 'firstquart': 1945, 'chargeoff': 877, 'loanloss': 3011, 'barn': 472, 'address': 67, 'guidelin': 2254, 'quarterend': 4165, 'bigger': 561, 'energybas': 1689, 'prudent': 4124, 'southwestern': 4873, 'ldc': 2900, 'privatesector': 4070, 'helpt': 2358, 'sovereign': 4874, 'reschedul': 4359, 'contributor': 1137, 'off': 3621, 'costcut': 1173, 'costeel': 1174, 'mn': 3332, 'usf': 5573, 'fg': 1915, 'gtsc': 2242, 'seismic': 4635, 'laboratori': 2854, 'boysmanni': 630, 'metromail': 3241, 'mtml': 3407, 'neb': 3461, 'capabl': 771, 'startup': 4955, 'fison': 1949, 'fisnl': 1948, 'pharmaceut': 3903, 'horticultur': 2433, 'oneforo

n': 3661, 'capitalis': 779, 'cra': 1205, 'melbourn': 3213, 'craa': 1206, 'forth': 2017, 'imput': 2526, 'taxfre': 5174, 'pid': 3920, 'substitut': 5057, 'notabl': 3564, 'steadili': 4969, 'overcapac': 3734, 'bougainvil': 620, 'buva': 723, 'comalco': 1011, 'cmac': 976, 'ironor': 2672, 'coal': 985, 'argyl': 310, 'unrealis': 5526, 'hedg': 2350, 'pir': 3932, 'tex': 5232, 'threefor': 5272, 'shortest': 4724, 'booker': 608, 'bokll': 599, 'agribusi': 119, 'rican': 4416, 'cement': 845, 'prn': 4072, 'withheld': 5825, 'adob': 80, 'adb': 64, 'walbro': 5677, 'walb': 5676, 'cass': 812, 'alltim': 174, 'automot': 398, 'carburetor': 786, 'unlik': 5516, 'pace': 3760, 'throttl': 5278, 'gm': 2161, 'interrupt': 2638, 'snat': 4817, 'lumberton': 3069, 'brunswick': 684, 'bc': 493, 'reichert': 4303, 'reichart': 4302, 'newlyacquir': 3487, 'recreat': 4267, 'bowl': 628, 'olymp': 3652, 'medal': 3200, 'game': 2101, 'adjoin': 72, 'stood': 5004, 'frederick': 2045, 'florjanc': 1974, 'presidentfin': 4044, 'esio': 1751, 'webber': 5726, 'pwm': 4150, 'parit': 3795, 'amph': 217, 'laidlaw': 2860, 'ldmfa': 2901, 'horizon': 2431, 'hrzb': 2448, 'bellingham': 524, 'binghamton': 571, 'trail': 5362, 'trl': 5402, 'bellevu': 523, 'seagram': 4611, 'vo': 5645, 'pont': 3985, 'spirit': 4904, 'ameron': 207, 'amn': 213, 'monterey': 3359, 'zim': 5920, 'zimr': 5922, 'norsk': 3555, 'vikingolj': 5633, 'oslo': 3706, 'mi': 3252, 'polo': 3981, 'jaguar': 2705, 'richardson': 4419, 'duin': 1578, 'chemclear': 888, 'cmcl': 978, 'unilater': 5506, 'unsuccess': 5539, 'workov': 5848, 'overhead': 3738, 'michel': 3256, 'billard': 566, 'berckman': 535, 'buccan': 688, 'blue': 588, 'dolphin': 1517, 'marcu': 3128, 'mrc': 3401, 'milwaukee': 3291, 'stb': 4965, 'baybank': 487, 'bbnk': 492, 'southernnet': 4870, 'sout': 4865, 'bralorn': 641, 'goodwil': 2181, 'patent': 3822, 'hug': 2455, 'mott': 3389, 'msm': 3404, 'hartford': 2313, 'mnl': 3334, 'wilton': 5805, 'wltn': 5830, 'woodridg': 5839, 'kentucki': 2798, 'kenca': 2796, 'lexington': 2949, 'semiannu': 4640, 'tssi': 5417, 'electrosound': 1651, 'esgrd': 1748, 'hauppaug': 2319, 'mors': 3380, 'mr': 3400, 'canton': 769, 'craftmaticcontour': 1208, 'crcc': 1212, 'trevos': 5390, 'aloett': 178, 'cosmet': 1170, 'alet': 147, 'macnealschwendl': 3089, 'heligmey': 2354, 'hmi': 2407, 'tirework': 5303, 'aleg': 146, 'talc': 5160, 'factual': 1848, 'advers': 85, 'unicorp': 5504, 'uac': 5458, 'mlnm': 3324, 'penney': 3861, 'jcp': 2722, 'jc': 2720, 'correctedmicropro': 1166, 'mpro': 3399, 'rafael': 4185, 'rykoffsexton': 4518, 'ryk': 4517, 'mrl': 3402, 'wick': 5785, 'pagurian': 3766, 'transco': 5368, 'chock': 912, 'onut': 3673, 'wiley': 5792, 'son': 4852, 'willb': 5795, 'handi': 2286, 'harman': 2305, 'hnh': 2410, 'riversid': 4441, 'rsgi': 4493, 'jacksonvil': 2700, 'gruop': 2237, 'nroc': 3581, 'wmx': 5832, 'oak': 3597, 'brook': 679, 'winstonaleam': 5813, 'oracl': 3691, 'orcl': 3694, 'belmont': 525, 'underwrit': 5495, 'alex': 149, 'absb': 21, 'tultex': 5424, 'ttt': 5420, 'martinsvil': 3148, 'morrison': 3378, 'morr': 3376, 'ala': 137, 'td': 5178, 'northview': 3560, 'novc': 3574, 'nord': 3548, 'nrd': 3580, 'kodak': 2832, 'ek': 1641, 'polyest': 3982, 'fiber': 1919, 'eastman': 1607, 'carryov': 803, 'partiallyori': 3803, 'filament': 1926, 'yarn': 5884, 'poy': 4017, 'carolina': 796, 'sc': 4576, 'older': 3648, 'kingsport': 2817, 'holsten': 2420, 'exit': 1807, 'kodel': 2833, 'insignific': 2603, 'scepter': 4582, 'interdyn': 2626, 'idyn': 2495, 'milpita': 3290, 'ctyn': 1247, 'hmo': 2408, 'hmoa': 2409, 'bug': 696, 'previouslyreport': 4056, 'method': 3239, 'companycontrol': 1042, 'biomed': 573, 'technicon': 5184, 'intang': 2613, 'astrom': 362, 'alot': 181, 'warwick': 5704, 'ri': 4414, 'district': 1477, 'conolog': 1098, 'cnlg': 980, 'somervil': 4847, 'tseng': 5415, 'tsng': 5416, 'newtown': 3494, 'coleco': 995, 'clc': 955, 'cabbag': 730, 'patch': 3820, 'kid': 2807, 'morton': 3382, 'handel': 2285, 'boveri': 627, 'und': 5483, 'bbczz': 491, 'baden': 429, 'sas': 4563, 'jaczz': 2704, 'bearer': 501, 'heck': 2346, 'hex': 2371, 'nitro': 3512, 'lifo': 2970, 'amsh': 220, 'tree': 5387, 'gnt': 2165, 'keltic': 2793, 'pr

o': 4073, 'willcox': 5796, 'gibb': 2140, 'wg': 5761, 'pn': 3965, 'kemper': 2795, 'kemc': 2794, 'cmc': 977, 'ibca': 2479, 'cineplex': 934, 'odeon': 3619, 'yrend': 5908, 'std': 4967, 'ami': 210, 'nln': 3517, 'artsway': 334, 'artw': 336, 'imag': 2506, 'himg': 2391, 'bicc': 554, 'biccl': 555, 'bradley': 639, 'brli': 671, 'bcd': 494, 'postfebruari': 4004, 'postsplit': 4006, 'rental': 4332, 'tenant': 5212, 'amcast': 200, 'acst': 55, 'ultramarc': 5468, 'hj': 2403, 'heinz': 2352, 'wtd': 5870, 'wtid': 5871, 'correctedinsituform': 1165, 'incinsua': 2536, 'arb': 295, 'arbbbr': 296, 'luxembourg': 3077, 'lesco': 2936, 'lsco': 3057, 'fiama': 1918, 'outgo': 3721, 'raym': 4222, 'generaldirector': 2126, 'guido': 2255, 'schmidtchiari': 4594, 'schill': 4588, 'hann': 2291, 'androsch': 236, 'unsatisfactori': 5533, 'pragu': 4019, 'statecontrol': 4957, 'austria': 391, 'austrian': 392, 'steyr': 4984, 'daimlerpuch': 1277, 'sdpvvi': 4608, 'cbt': 834, 'inspect': 2605, 'yugoslav': 5910, 'unfavor': 5503, 'yugoslavia': 5911, 'dim': 1434, 'belgrad': 520, 'subotica': 5044, 'unseason': 5534, 'dorman': 1531, 'yellowbrown': 5893, 'cast': 813, 'leaf': 2907, 'latesee': 2884, 'topdress': 5335, 'pasta': 3817, 'interemdi': 2627, 'gsm': 2240, 'iwc': 2696, 'coars': 986, 'thu': 5283, 'oat': 3599, 'resown': 4374, 'bolster': 600, 'endof': 1685, 'shultz': 4737, 'maggi': 3094, 'mcneil': 3188, 'recogn': 4256, 'explicitli': 1823, 'noteworthy': 3567, 'complain': 1051, 'lowerpr': 3049, 'rebuf': 4246, 'adamantli': 62, 'overtur': 3747, 'aboutfac': 19, 'lobbyst': 3012, 'ridiculous': 4426, 'somebodi': 4846, 'somewher': 4850, 'attitud': 376, 'judgment': 2765, 'havent': 2320, 'commentari': 1019, 'amstutz': 222, 'taught': 5171, 'necessarili': 3465, 'subsaharan': 5047, 'droughtreduc': 1562, 'strife': 5023, 'tropic': 5405, 'climat': 967, 'capita': 778, 'constitut': 1112, 'denatur': 1366, 'anglia': 241, 'abnorm': 16, 'wheatproduc': 5765, 'manoeuvr': 3119, 'markup': 3142, 'till': 5295, 'winterkil': 5815, 'accu': 39, 'colleg': 1000, 'czechoslovakia': 1273, 'romania': 4465, 'spotti': 4915, 'tonight': 5329, 'scant': 4579, 'inch': 2534, 'flurri': 1979, 'breadmak': 652, 'greg': 2216, 'mccune': 3177, 'parisbas': 3794, 'persuad': 3885, 'underway': 5494, 'jop': 2755, 'familiar': 1864, 'updat': 5546, 'marcel': 3126, 'cazal': 829, 'subproduct': 5045, 'apr': 284, 'januarydecemb': 2713, 'sunflower': 5087, 'linse': 2987, 'branpollard': 646, 'rape': 4207, 'branpollard': 645, 'holland': 2418, 'cuba': 1248, 'bulgaria': 703, 'canari': 761, 'corea': 1154, 'thou': 5262, 'cabinetlevel': 732, 'dormant': 1532, 'capitol': 780, 'guid': 2252, 'gail': 2099, 'martel': 3146, 'eff': 1628, 'hutton': 2472, 'pessimist': 3893, 'optim': 3688, 'germin': 2135, 'ukrain': 5466, 'caucasu': 821, 'precipit': 4024, 'bitte': 576, 'decad': 1324, 'toll': 5321, 'ampl': 218, 'wellinsul': 5744, 'besid': 542, 'livestock': 3001, 'graze': 2204, 'moisturedeplet': 3348, 'pastur': 3818, 'necessit': 3466, 'healthi': 2334, 'chernobyl': 892, 'contamin': 1121, 'embargo': 1663, 'trip': 5397, 'nelson': 3474, 'impend': 2515, 'customari': 1259, 'reenter': 4275, 'interpret': 2637, 'tightlip': 5292, 'content': 1123, 'spoken': 4906, 'fritz': 2061, 'fourmlntonn': 2028, 'embarrass': 1664, 'spurn': 4920, 'dlrper-tonn': 1493, 'dampen': 1284, 'critic': 1224, 'nonetheless': 3534, 'procedur': 4078, 'chortl': 916, 'housew': 2441, 'cheaper': 885, 'convent': 1141, 'wisdom': 5820, 'forgotten': 2009, 'surround': 5111, 'lastminut': 2881, 'unsurpass': 5540, 'rank': 4206, 'sixti': 4779, 'enrol': 1708, 'plummet': 3962, 'slower': 4804, 'paymentinkind': 3834, 'tumbl': 5425, 'greatest': 2208, 'contrast': 1134, 'budg': 691, 'higherproduc': 2378, 'kept': 2800, 'spread': 4916, 'counselor': 1182, 'lof': 3017, 'displac': 1462, 'usussr': 5578, 'redempt': 4271, 'hrw': 2447, 'hr': 2446, 'srw': 4929, 'sww': 5136, 'track': 5354, 'peoria': 3867, 'evansvil': 1772, 'cincinnati': 931, 'norfphil': 3551, 'amarillo': 195, 'lubbock': 3062, 'lou': 3041, 'stockton': 5000, 'duluth': 1579, 'rye': 4516, 'soyb': 4879, 'sorg': 4857, 'evnsvll': 1783, 'cinci': 930, 'mpl': 3397, 'baltnor': 445, 'phil': 3906, 'k

c': 2788, 'amarlo': 196, 'lubbck': 3061, 'outlay': 3722, 'alabama':
 138, 'burgess': 717, 'reced': 4250, 'lightli': 2974, 'often': 3631,
 'transmit': 5374, 'longawait': 3023, 'accompani': 34, 'beet': 508,
 'porter': 3992, 'rill': 4430, 'politicallyinfluenti': 3979, 'cane':
 766, 'recover': 4265, 'inroad': 2598, 'sugarcontain': 5070, 'cornwh
 eat': 1159, 'lunch': 3071, 'halfhour': 2271, 'lunchtim': 3072, 'conc
 ert': 1071, 'touch': 5342, 'phlx': 3911, 'australasian': 388, 'nicho
 la': 3497, 'giordano': 2146, 'educ': 1624, 'asianpacif': 344, 'excha
 ngetrad': 1795, 'pioneer': 3929, 'impress': 2523, 'truer': 5410, 'ar
 bitrag': 297, 'noncurr': 3532, 'mulford': 3412, 'shortdat': 4723, 's
 ixday': 4776, 'twoway': 5448, 'departur': 1376, 'dress': 1553, 'defu
 s': 1349, 'misread': 3312, 'threeday': 5271, 'categori': 817, 'unif
 i': 5505, 'hikmat': 2386, 'nashashibi': 3439, 'prerequisit': 4038,
 'plausibl': 3950, 'peg': 3853, 'model': 3339, 'dinar': 1436, 'trade
 weight': 5358, 'emsstyl': 1678, 'grid': 2220, 'theori': 5245, 'inter
 arab': 2623, 'ebb': 1610, 'infanc': 2577, 'pauciti': 3827, 'hamper':
 2282, 'miyazawa': 3321, 'manuel': 3121, 'committ': 1027, 'kiichi':
 2809, 'imfworld': 2509, 'simex': 4758, 'mercantil': 3226, 'treasury
 bond': 5384, 'secertari': 4619, 'interamerican': 2622, 'excchang': 1
 789, 'roll': 4461, 'meant': 3194, 'tighten': 5291, 'kuala': 2846, 'l
 umpur': 3070, 'ringgit': 4432, 'negara': 3471, 'prudenc': 4123, 'jaa
 far': 2697, 'bergsten': 536, 'enorm': 1704, 'sizeabl': 4781, 'reuter
 author': 4401, 'leo': 2933, 'melam': 3212, 'cme': 979, 'smithsonia
 n': 4813, 'era': 1735, 'overwhelm': 3748, 'unrealist': 5527, 'unnece
 ssari': 5517, 'policymak': 3976, 'gyrat': 2263, 'triplewitch': 5399,
 'salutari': 4545, 'witch': 5822, 'bar': 466, 'pit': 3933, 'dual': 1
 570, 'nearest': 3458, 'dellsher': 1360, 'conced': 1067, 'bottom': 61
 9, 'forex': 2007, 'cambist': 750, 'ica': 2483, 'abdullah': 8, 'aldak
 hil': 142, 'burgan': 715, 'hani': 2290, 'ramadan': 4199, 'beirut': 5
 17, 'riyad': 4442, 'threeyear': 5276, 'vicechairmen': 5622, 'ezzedi
 n': 1839, 'saidan': 4532, 'international': 2635, 'tuni': 5427, 'moha
 m': 3346, 'osman': 3707, 'bancair': 448, 'liban': 2953, 'walid': 567
 9, 'nasouli': 3442, 'ibrahim': 2482, 'buhindi': 698, 'imad': 2505,
 'bata': 482, 'treasur': 5382, 'safrica': 4528, 'rand': 4203, 'candi
 d': 764, 'whitesonli': 5774, 'psycholog': 4132, 'broken': 677, 'semi
 gilt': 4643, 'reintroduc': 4310, 'quiet': 4174, 'vagu': 5589, 'secre
 t': 4624, 'necess': 3463, 'soften': 4833, 'underton': 5493, 'overloo
 k': 3739, 'dollarmark': 1515, 'softer': 4834, 'karl': 2783, 'otto':
 3712, 'poehl': 3969, 'lothar': 3039, 'mueller': 3411, 'confus': 109
 0, 'exchanger': 1794, 'orient': 3702, 'unab': 5469, 'skorea': 4788,
 'korean': 2840, 'myra': 3428, 'macdonald': 3084, 'thrown': 5281, 'n
 earterm': 3460, 'beneficiari': 531, 'sixnat': 4778, 'stifl': 4989,
 'rush': 4512, 'sidelin': 4742, 'sooner': 4855, 'lewi': 2948, 'where
 a': 5769, 'retrac': 4395, 'halfpoint': 2272, 'resili': 4367, 'crossr
 at': 1230, 'turbul': 5430, 'quietli': 4175, 'await': 411, 'chienku
 o': 900, 'swollen': 5132, 'chicheng': 898, 'interbank': 2624, 'wan
 g': 5688, 'chaom': 874, 'headach': 2328, 'chienshien': 901, 'inlan
 d': 2594, 'endyear': 1687, 'resal': 4358, 'liradollar': 2992, 'dm':
 1498, 'liradmark': 2991, 'reap': 4241, 'parliamentari': 3799, 'hind
 sight': 2393, 'somewhat': 4849, 'inconclus': 2544, 'confin': 1086,
 'yoshiko': 5903, 'mori': 3373, 'liff': 2969, 'cbot': 833, 'internat
 ionalis': 2636, 'katsuyuki': 2785, 'okayasu': 3644, 'yamaichi': 588
 2, 'stimulu': 4992, 'tetsuya': 5231, 'dezuka': 1412, 'mind': 3292,
 'display': 1463, 'multipli': 3416, 'nomura': 3526, 'koki': 2835, 'c
 hiyojima': 910, 'nikko': 3504, 'withhold': 5826, 'deter': 1404, 'ch
 eapest': 886, 'latter': 2887}

Transform

In [59]:

```
X_train_counts = count_vect.transform(trainData)
```

In [132]:

```
print(X_train_counts.shape)
print(X_train_counts.toarray())
```

```
(654, 5932)
[[0 0 0 ..., 0 0 0]
 [0 0 0 ..., 0 0 0]
 [0 0 0 ..., 0 0 0]
 ...,
 [0 0 0 ..., 0 0 0]
 [0 0 0 ..., 0 0 0]
 [0 0 0 ..., 0 0 0]]
```

Visualizar

In [134]:

```
countMatrix = pd.DataFrame.from_records(X_train_counts.toarray())  
countMatrix
```


626	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
	0	1	2	3	4	5	6	7	8	9	...	5922	5923	5924	5925	5926	5927	5928	592
627	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
628	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
629	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
630	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
631	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
632	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
633	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0	0	0
634	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
635	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	1	0	0	0
636	0	0	0	0	0	0	0	0	1	0	...	0	0	0	0	0	0	0	0
637	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
638	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
639	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
640	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
641	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
642	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
643	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
644	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
645	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
646	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
647	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
648	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
649	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
650	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
651	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
652	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0
653	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0

654 rows × 5932 columns



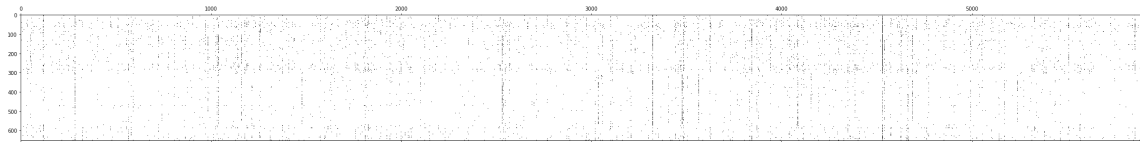
Visualize (2)

In [138]:

```
import matplotlib.pyplot as plt
from matplotlib.pyplot import figure, show
```

In [142]:

```
plt.figure(figsize=(40,40))
plt.spy(X_train_counts.toarray())
plt.show()
```



De la imagen se puede apreciar que hay palabras que al parecer están en muchos de los documentos.

Es posible que no se haya aplicado el analiyzer de forma correcta.

Se procederá a buscar cuales son las palabras que más incidencia tienen.

Más frecuentes

In [195]:

```
countMatrix.sum().nlargest(30)
```

Out[195]:

4530	1637
3323	1262
5663	902
1491	820
3127	670
1243	539
5886	492
3844	474
5562	423
3477	396
3036	383
1040	364
457	325
4689	325
4733	318
567	317
5355	284
3138	271
2531	263
4086	260
5858	255
3636	224
3485	214
2878	211
285	206
4057	204
1160	200
3659	197
3678	196
4540	196

dtype: int64

Más repetido

In [202]:

```
[k for k, v in count_vect.vocabulary_.items() if v == 4530]
```

Out[202]:

```
['said']
```

Todos los valores:

In [201]:

```
for index, data in countMatrix.sum().nlargest(30).iteritems():  
    print([k for k, v in count_vect.vocabulary_.items() if v == index])
```

```
['said']  
['mln']  
['vs']  
['dlr']  
['march']  
['ct']  
['year']  
['pct']  
['us']  
['net']  
['loss']  
['compani']  
['bank']  
['share']  
['shr']  
['billion']  
['trade']  
['market']  
['inc']  
['profit']  
['would']  
['oil']  
['new']  
['last']  
['april']  
['price']  
['corp']  
['one']  
['oper']  
['sale']
```

Algunas de las palabras encontradas son relevantes. Otras no tanto. Es difícil determinar de forma automática cuáles palabras son importantes y cuáles no. Se dejarán por el momento, sin embargo, se considera que su presencia debería afectar los resultados.

TF-idf Transform

In [61]:

```
from sklearn.feature_extraction.text import TfidfTransformer
```

Fit and Transform

In [62]:

```
tfidf_transformer = TfidfTransformer()
X_train_tfidf = tfidf_transformer.fit_transform(X_train_counts)
X_train_tfidf.shape
```

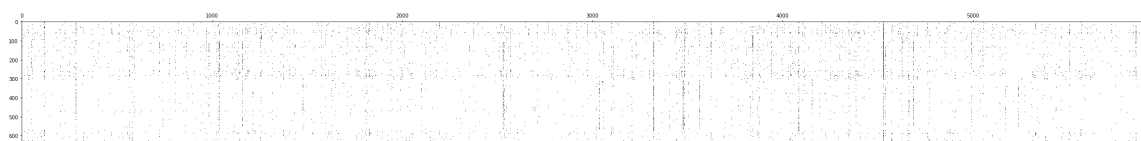
Out[62]:

(654, 5932)

Visualize

In [203]:

```
plt.figure(figsize=(40,40))
plt.spy(X_train_tfidf.toarray())
plt.show()
```



Naive Bayes

In [63]:

```
from sklearn.naive_bayes import MultinomialNB
```

In [64]:

```
clf = MultinomialNB().fit(X_train_tfidf, trainTarget)
```

Test Predict

In [65]:

```
testFile = "data/test/money-fx/0009664"
```

In [66]:

```
testTarget = [testFile.split("/")[2]]
testData = [readFile(testFile)]
print(testTarget)
print(testData)
```

```
['money-fx']
["\n\nU.K. MONEY MARKET GIVEN 53 MLN STG ASSISTANCE\n\n    LONDON, A
pril 8 - The Bank of England said it provided the\nmoney market with
53 mln stg assistance in the morning session.\n    This compares wit
h the bank's estimate of a shortage in the\nsystem of around 300 mln
stg which it earlier revised up from\n250 mln.\n    The central bank
made outright purchases of bank bills\ncomprising 46 mln stg in band
three at 9-3/4 pct and seven mln\nstg in band four at 9-11/16 pct.\n\n"]
```

In [67]:

```
X_new_counts = count_vect.transform(testData)
X_new_tfidf = tfidf_transformer.transform(X_new_counts)
```

In [68]:

```
predicted = clf.predict(X_new_tfidf)
```

In [69]:

```
for doc, category in zip(testTarget, predicted):
    print('%s => %s \n' % (doc, category))
```

```
money-fx => money-fx
```

Pipeline

In [70]:

```
from sklearn.pipeline import Pipeline
text_clf = Pipeline([('vect', CountVectorizer(analyzer = processDocument)),
                     ('tfidf', TfidfTransformer()),
                     ('clf', MultinomialNB()),
                     ])
```

In [71]:

```
categories = ["acq", "earn", "money-fx", "grain", "crude", "trade", "interest",
             "ship", "wheat", "corn"]
```

In [72]:

```
trainingFiles = getFilesInFolders("data/training", categories)
trainDocuments = [[filename.split("/")[2], readFile(filename)] for filename in c
concatFiles(trainingFiles)]
trainData = [content for file, content in trainDocuments]
trainTarget = [file for file, content in trainDocuments]
```

In [73]:

```
testFiles = getFilesInFolders("data/test", categories)
testDocuments = [[filename.split("/")[2], readFile(filename)] for filename in co
concatFiles(testFiles)]
testData = [content for file, content in testDocuments]
testTarget = [file for file, content in testDocuments]
```

In [74]:

```
text_clf.fit(trainData, trainTarget)
```

Out[74]:

```
Pipeline(memory=None,
          steps=[('vect', CountVectorizer(analyzer=<function processDocum
ent at 0x7fee3111cd90>,
          binary=False, decode_error='strict', dtype=<class 'numpy.int
64'>,
          encoding='utf-8', input='content', lowercase=True, max_df=1.
0,
          max_features=None, min_df=1, ngram_range=(1, 1), preprocess
o...inear_tf=False, use_idf=True)), ('clf', MultinomialNB(alpha=1.0,
class_prior=None, fit_prior=True))])
```

In [76]:

```
predicted = text_clf.predict(testData)
```

Comparar resultados

In [75]:

```
import numpy as np
```

In [95]:

```
np.mean(predicted == testTarget)
```

Out[95]:

```
0.79332615715823462
```

In [90]:

```
same = np.where(predicted == testTarget)
different = np.where(predicted != testTarget)
print("same:", len(same[0]))
print("diff:", len(different[0]))
print(1-int(len(different[0]))/int(len(same[0])))
```

```
same: 2211
```

```
diff: 576
```

```
0.7394843962008142
```

In [96]:

```
print(different)
```

```
(array([ 7, 10, 12, 21, 27, 34, 35, 40, 46, 56,
70,
75, 80, 84, 87, 103, 105, 109, 113, 119, 123,
124,
125, 127, 131, 132, 134, 136, 149, 151, 153, 155,
160,
161, 162, 163, 165, 167, 168, 169, 173, 174, 175,
176,
177, 182, 191, 192, 193, 201, 202, 203, 204, 205,
207,
208, 210, 215, 216, 218, 220, 221, 223, 224, 225,
227,
230, 233, 234, 236, 239, 242, 243, 245, 246, 250,
265,
278, 281, 285, 287, 293, 295, 297, 298, 299, 300,
301,
302, 303, 304, 305, 306, 308, 309, 315, 321, 325,
331,
338, 342, 344, 346, 347, 348, 349, 350, 351, 353,
355,
356, 357, 358, 360, 361, 363, 364, 365, 366, 367,
368,
369, 370, 371, 372, 373, 374, 375, 376, 378, 382,
383,
384, 385, 386, 387, 388, 389, 390, 391, 392, 393,
394,
395, 396, 397, 399, 400, 401, 402, 403, 404, 405,
406,
407, 408, 409, 410, 411, 412, 413, 414, 415, 416,
417,
418, 419, 420, 421, 422, 424, 425, 426, 427, 428,
429,
430, 431, 432, 433, 434, 435, 436, 437, 438, 439,
440,
441, 442, 443, 444, 445, 446, 447, 448, 449, 450,
451,
452, 453, 454, 455, 456, 457, 458, 459, 460, 461,
462,
463, 464, 467, 468, 470, 492, 503, 504, 509, 516,
519,
527, 530, 539, 546, 551, 584, 591, 593, 597, 634,
639,
647, 658, 667, 668, 679, 691, 705, 731, 755, 757,
766,
769, 775, 797, 803, 813, 822, 840, 842, 846, 850,
851,
857, 863, 865, 900, 917, 920, 929, 939, 945, 951,
973,
974, 986, 1014, 1040, 1063, 1074, 1086, 1095, 1109, 1119, 1
143,
1155, 1160, 1174, 1188, 1189, 1190, 1191, 1192, 1193, 1195, 1
198,
1201, 1204, 1205, 1206, 1207, 1209, 1214, 1216, 1217, 1218, 1
225,
1226, 1227, 1229, 1230, 1234, 1236, 1237, 1239, 1242, 1244, 1
245,
1248, 1249, 1250, 1251, 1252, 1253, 1254, 1255, 1260, 1261, 1
266,
1267, 1272, 1274, 1275, 1278, 1279, 1281, 1282, 1283, 1284, 1
287,
1288, 1298, 1300, 1301, 1304, 1305, 1306, 1307, 1308, 1309, 1
```



```
310, 1311, 1312, 1314, 1315, 1316, 1317, 1318, 1319, 1320, 1321, 1
322, 1323, 1324, 1325, 1326, 1327, 1328, 1329, 1330, 1331, 1332, 1
333, 1334, 1335, 1336, 1337, 1338, 1339, 1340, 1341, 1342, 1343, 1
344, 1345, 1346, 1347, 1348, 1349, 1351, 1352, 1353, 1354, 1355, 1
356, 1357, 1358, 1359, 1360, 1361, 1362, 1363, 1364, 1365, 1366, 1
368, 1369, 1370, 1371, 1372, 1373, 1374, 1376, 1377, 1378, 1379, 1
380, 1381, 1382, 1383, 1384, 1385, 1386, 1387, 1388, 1389, 1390, 1
391, 1392, 1393, 1707, 1759, 1855, 1972, 2124, 2129, 2369, 2405, 2
407, 2445, 2446, 2481, 2482, 2483, 2484, 2485, 2486, 2487, 2488, 2
489, 2490, 2491, 2492, 2493, 2494, 2495, 2496, 2497, 2498, 2499, 2
500, 2501, 2502, 2503, 2504, 2505, 2506, 2507, 2508, 2509, 2510, 2
511, 2512, 2513, 2514, 2515, 2516, 2517, 2518, 2519, 2520, 2521, 2
522, 2523, 2524, 2525, 2526, 2527, 2528, 2529, 2530, 2531, 2532, 2
533, 2534, 2535, 2536, 2537, 2538, 2539, 2540, 2541, 2542, 2543, 2
544, 2545, 2546, 2547, 2548, 2549, 2550, 2551, 2552, 2553, 2554, 2
555, 2556, 2557, 2558, 2559, 2560, 2561, 2562, 2563, 2564, 2565, 2
566, 2567, 2568, 2569, 2570, 2571, 2572, 2573, 2574, 2575, 2576, 2
577, 2578, 2579, 2580, 2581, 2582, 2583, 2584, 2585, 2586, 2587, 2
588, 2589, 2590, 2591, 2592, 2593, 2594, 2595, 2596, 2597, 2598, 2
599, 2600, 2601, 2602, 2603, 2604, 2605, 2606, 2607, 2609, 2619, 2
620, 2631, 2638, 2659, 2660, 2681, 2716, 2726, 2744, 2745, 2749, 2
753, 2760, 2762, 2767, 2768]))))
```

Busqueda de diferencias

In [106]:

```
testSample = testFiles[7]
print(testSample)
```

```
['data/test/grain', '0012662']
```

In [107]:

```
testSampleTarget = [testSample[0].split("/")[2]]
testSampleData = [readFile("/".join(testSample))]
print(testSampleTarget)
print(testSampleData)
```

```
['grain']
['\n\nU.S. SENATE PANEL VOTES TO LIMIT COUNTY LOAN RATE  CHANGES STA
RTING WITH 1988 CROPS\n\nBlah blah blah.\n&#3;\n\n']
```

In [108]:

```
predicted = text_clf.predict(testSampleData)
```

In [109]:

```
for doc, category in zip(testTarget, predicted):
    print('%s => %s \n' % (doc, category))
```

```
grain => earn
```

Refactor

In [126]:

```
def evaluateTestSample(testFiles, index):
    testSample = testFiles[index]
    print("Sample")
    print(testSample)
    print()
    testSampleTarget = [testSample[0].split("/")[2]]
    testSampleData = [readFile("/".join(testSample))]
    print(testSampleData)
    print()
    print("Predicted:")
    predicted = text_clf.predict(testSampleData)
    for doc, category in zip(testSampleTarget, predicted):
        print('%s => %s \n' % (doc, category))
```

In [127]:

```
evaluateTestSample(testFiles, 10)
```

Sample

```
['data/test/grain', '0011705']
```

```
['\n\nBANGLADESH NAVY UNLOADS GRAINS DURING PORT STRIKE\n\n    CHITT  
AGONG, Bangladesh, June 19 - The navy is unloading\nfoodgrains at Ch  
ittagong port following a strike by nearly\n1,000 dockworkers, Bangl  
adesh Shipping Corp officials said.\n    The navy was unloading 74,0  
00 tonnes of foodgrains from two\nships today, and four vessels lade  
n with some 90,000 tonnes of\ngains were waiting at the outer ancho  
rage, port officials\nsaid.\n    A spokesman for the workers said th  
ey would continue their\nstrike until authorities agree to their dem  
ands for higher pay\nand other benefits. The strike began on June 1  
1.\n    "Unloading of grains and other essential commodities started  
\nnormally after we called in navy personnel to help," a port\noffic  
ial told Reuters.\n    "The strike has caused no serious dislocatio  
n," he said.\n    The government meanwhile declared the port jobs an  
\n"essential service" and said the strikers could be dismissed\nunle  
ss they end the strike soon.\n\n']
```

Predicted:

grain => ship