

Assignment 5

Task 1

```
>>> w_dict.cache().sortByKey()
PythonRDD[15] at RDD at PythonRDD.scala:53
>>> w_dict.lookup("applicant")[0]
347
>>> w_dict.lookup("and")[0]
2
>>> w_dict.lookup("attack")[0]
504
>>> w_dict.lookup("protein")[0]
3014
>>> w_dict.lookup("car")[0]
612
```

Task 2

Top 50 words and their slopes:

```
['prisons'] 's parameter is: 0.42827000188750064
['fascism'] 's parameter is: 0.36049003225738074
['mc'] 's parameter is: 0.3501389826182507
['mitsubishi'] 's parameter is: 0.3430627546506681
['management'] 's parameter is: 0.342440124763022
['roster'] 's parameter is: 0.3378696272861401
['semifinal'] 's parameter is: 0.33322296523332134
['accordance'] 's parameter is: 0.32599875893335933
['failures'] 's parameter is: 0.32446783991994543
['orders'] 's parameter is: 0.3179789365848749
['claim'] 's parameter is: 0.3118270005005011
['maize'] 's parameter is: 0.31172460446341455
['lola'] 's parameter is: 0.30732030526556714
['erred'] 's parameter is: 0.30657503064176583
['trinidad'] 's parameter is: 0.30609543325474103
['construed'] 's parameter is: 0.3056458429421953
['registers'] 's parameter is: 0.30396053977798615
['commissioners'] 's parameter is: 0.30290035574417673
['turf'] 's parameter is: 0.3024751794981476
['ligament'] 's parameter is: 0.30017195933209795
['nos'] 's parameter is: 0.2974205684417545
['steamtown'] 's parameter is: 0.29665523381591447
['reunification'] 's parameter is: 0.29472022023366823
['death'] 's parameter is: 0.2942333175411066
['catalina'] 's parameter is: 0.2922269541216992
['ways'] 's parameter is: 0.29185319236508384
['yuri'] 's parameter is: 0.29009536304448835
['wildcard'] 's parameter is: 0.28763057851786905
['heroic'] 's parameter is: 0.2875143144096061
['undoubtedly'] 's parameter is: 0.2866334154287944
['apps'] 's parameter is: 0.286383446006574
```

```
['hamer'] 's parameter is: 0.28462614713579426
['neglect'] 's parameter is: 0.2842478944967335
['kitchen'] 's parameter is: 0.2840654499400074
['efficiently'] 's parameter is: 0.2821340587293734
['whatever'] 's parameter is: 0.2808455316334704
['couples'] 's parameter is: 0.2794845969529191
['nara'] 's parameter is: 0.27874463762671436
['outright'] 's parameter is: 0.2787038386226588
['medal'] 's parameter is: 0.2786930207840932
['nfc'] 's parameter is: 0.27831314650947864
['stint'] 's parameter is: 0.2782802061470775
['jwp'] 's parameter is: 0.2771024781683967
['marjoribanks'] 's parameter is: 0.2765811798603899
['jackie'] 's parameter is: 0.2762589428190093
['distracted'] 's parameter is: 0.27609644514412585
['tunisia'] 's parameter is: 0.27497250270860935
['kidnapped'] 's parameter is: 0.27473070601436556
['angered'] 's parameter is: 0.2745816165165282
['shoppers'] 's parameter is: 0.27412621824015787
```

Task 3

```
test accuracy: 0.98655
recall: 0.71618
precision 0.99209
F1 score 0.83185
There are 1 false positive cases.
14683768
```

This false positive case 14683768 is about an Australian restaurant . It may contains similar words due to this regional coincidence. Meanwhile, in its Controversies section some legal issues are described. Words like “management”, “claimed”, “engage” appear so increase its possibility to be classified as a positive case.