

```
1 =begin
2 class Word
3   def initialize (total, count)
4     @total = total
5     @count = count
6   end
7   def total
8     @total
9   end
10  def count
11    @count
12  end
13
14  #average
15  def average
16    return @total/@count
17  end
18
19  def increment_total(increment)
20    @total += increment
21  end
22  def increment_count
23    @count += 1
24  end
25 end
26
27 =end
28
29 class String
30   def alpha?
31     !!match(/^[:alpha:]]+$/ )
32   end
33 end
34 class Object
35   def in(ary)
36     ary.include?(self)
37   end
38 end
39 def spaced_out(string)
40   out = ''
41   string.each_char do |letter|
42     if letter.alpha? or letter == " "
43       out += letter
44     else
45       out += ' '+letter+' '
46     end
47   end
48   return out
49 end
50 end
```

```

51
52 $anti_words = []
53 anti = open('100words.txt', 'r')
54 anti.each do |word|
55   $anti_words+=word
56 end
57
58
59
60 $all_words = {}
61
62 #file_names = %w'amazon_cells_labelled.txt','imdb_labelled
   .txt','yelp_labelled.txt'
63
64 def analyze(source)
65   filename = source.to_s
66   data = open(filename,'r')
67   text = data.readlines
68   text.each do |line|
69     score = line[-2].to_i
70     text = spaced_out(line.chomp('\t').downcase).split(' '
   )
71     text.each do |item|
72       if item.in $all_words
73         $all_words[item][0] += (score)
74         $all_words[item][1]+=1
75       elsif not item.in $all_words
76         $all_words[item] = [score, 1]
77       end
78     end
79   end
80 end
81
82 def output(dict, file)
83   dict.each do |word|
84     file.puts word
85   end
86 end
87
88 =begin
89 file_names.each do |file|
90   analyze(file)
91 end
92 =end
93 analyze('amazon_cells_labelled.txt')
94 analyze('imdb_labelled.txt')
95 analyze('yelp_labelled.txt')
96
97 puts $all_words.keys.length
98 $anti_words.each do |word|

```

```
99     $all_words.delete(word.chomp)
100 end
101
102 $all_words.each_key do |key|
103     if not key.alpha?
104         $all_words.delete(key)
105     end
106 end
107
108 puts $all_words.keys.length
109
110 out = open('sentiment2.txt','w')
111 output($all_words, out)
112
113 out.close
114
115
```