DS636 Lab 4

1. Please use the **train.csv** file
https://www.kaggle.com/c/home-depot-product-search-relevance/data

2. Please write a program to compute how many words (separated by space) in the "product_title" for each product and create a new column called "numofwords" to store the results.

   a. Please first try to implement this program through for loop and record the running time.

   b. Then re-implement the program with apply, lapply or other apply functions and compare the running time with the for-loop implementation.

3. The train.csv file contains a number of products and real customer search terms from Home Depot's website. The challenge is to predict a relevance score for the provided combinations of search terms and products.

One naïve method is to calculate the number of common words between the product and search terms, and then make relevance prediction accordingly. To make it more concrete, we will first investigate if there is any connection between the relevance score and the number of common words between product and search terms.
Please write a program to compute the average number of common words in each product and search term pair for each level of relevance score.
The results should contain the average number of common words for each relevance score level:

Relevance score level	Average number of common words in each pair
1	
…	
3	

For example, if there are only 3 pairs with relevance score 1. And there are 2 common words between product and search term pairs in total. Then we say the average number of common words for relevance score level 1 is 2/3.

(Please write an efficient program based on the experience in solving Q2.)