Google 1gram questions (hdfs://data/1gram  [link](http://storage.googleapis.com/books/ngrams/books/datasetsv2.html)):

1. For each year available, plot the size of the set of words used.  Year on the x-axis, number of words on y-axis.

**Ans:**

This question requires number of words used in each year from 1505 to 2008.

**Approach:**

1.       Map all the words in the data in the format (year, 1)

2.      Apply Reduce operation using year as key to get the number of words in year.

3. Plot the graph with year on x-axis and number of words on y-axis

Example: 1505 11055

1506 18569

**Year – Word Graph:**

