**Write UP**

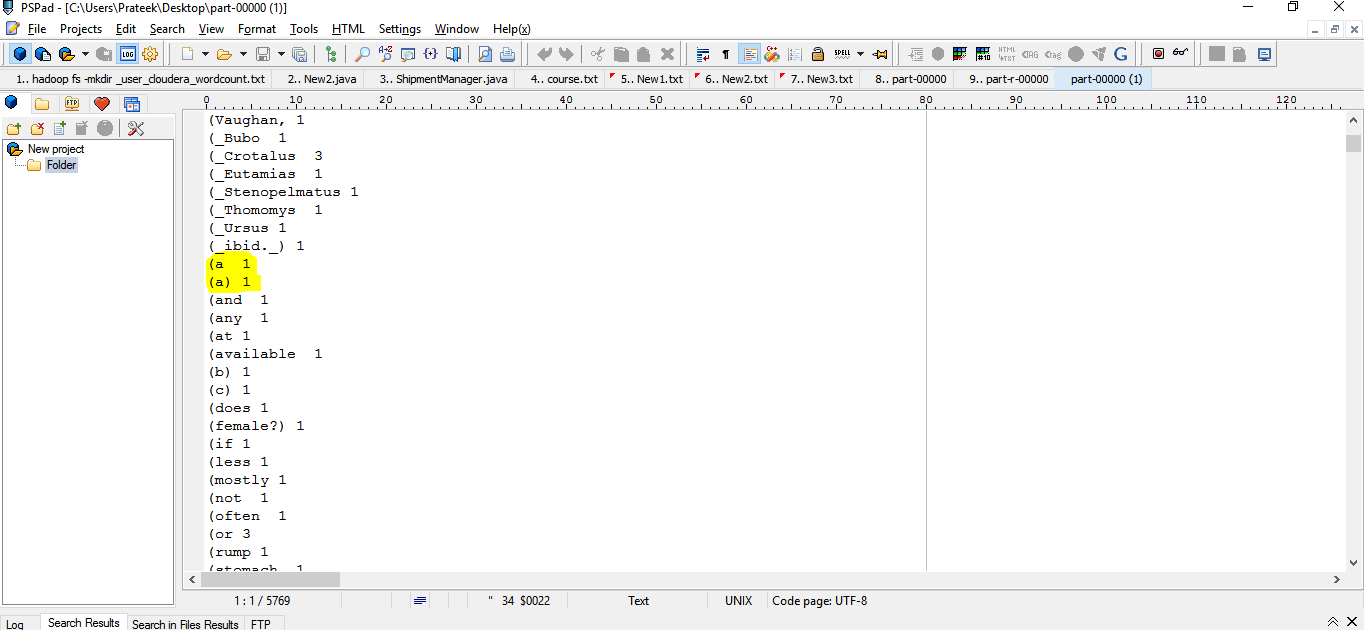
Hadoop-MapReduce-with-Modified-wordCount The original wordcount example given by Hadoop libraries counts the words which are basically separated by space. Although, it can not manipulate/identify special symbols and therefore it counts ( e.g. You and You! and You? ) all 3 as different words. Logically thinking, all these words should be counted as one (i. e. You).

Secondly, the basic program treats capital letters separately which results in different wordcounts ( e.g. You and you). This program modifies the mapper function such that it removes special symbols as well as converts everything to lowercase for eliminate difference made capitalization

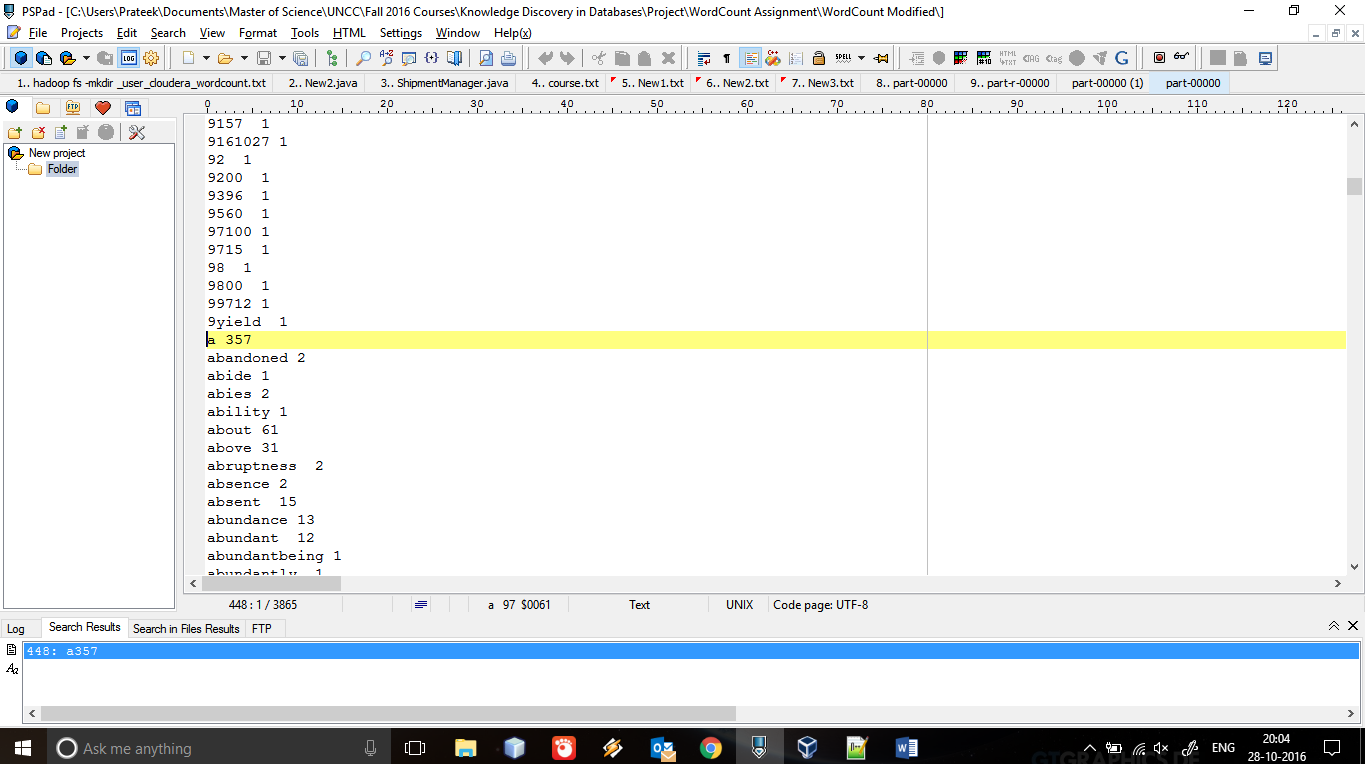
Similarly in our example of MammalBook **(a** and **(a)** etc are read as same words . And in output the mapping and reduding count differs as below..

WordCountV2 –

(a and (a) are considered different words



**Wprdcount modified output**



All such variations are disgarded and a occurs 357 times