

s1280120 Taito Waragaya 15%
s1280116 Tatsumi Matsumoto 35%
s1280084 Makishi Ono 25%
s1280106 Keisuke Kano 25%

Authorship profiling for education

#pesudo code

Read 5 dataset of words divided by educational level (1-5)

Read text file to be analysed

Store the text to be analysed, word by word, in array 1

Create array 2 by eliminating duplicate words from array 1

Compare all elements of array 2 with the contents of the word list by education level 1 to level 5 and store the level (1 to 5) in the total value variable when a hit is found

Divide the total value by the total number of letters (total number of array 1) (this value is the vocabulary level)

Output an estimate of the person's education level from the value of the vocabulary level.