

Advanced Programming HW 3

By Divam Gupta 2014038

When we enter the third option we get a list words which are mostly stop words. The proportion , relative frequency and rankings of the stop words are approximately same as the top words used in english language.

Similarly when we enter a name of any author the top words are generally stop words like [the , of , a , to and the] and the rankings of these stop words are approximately same of all the authors.

After removing all the stop words we can get a different frequency distribution of every authors. And that relate to the areas of research that author is doing. Another interesting observation is that the top words of each author are related.

For example for Dr. Aman Gupta top keywords are:

base, core, data, detect, detector, elements, in, ip, mobile, network, probability, service , signaling , station

And we can see these are related to Signals and systems.

Also by using the frequency distributions of words of all the authors we can predict the author of a paper by matching its keywords.