NATURAL LANGUAGE PROCESSING: AUTHOR STYLOMETRY 05

So what's the idea? Scholar T C Mendenhall in news paper said that an author stylic signature could be formed by counting how often he or she uses words of different lengths. For example if you pick a novel and count word lengths in a five thousand words segment and then plot a graph of the word length distribution. The curve would look pretty much the same no matter what part of the novel we had print, so isn't it cool? Will use this idea, will count the number of words of different lengths and then plot it, with this plot will try to understand the author's style of writing. This analysis is very basic one there are many advanced way of doing such classifications but since it is a very basic course will just motivate you with the idea so let's see this.