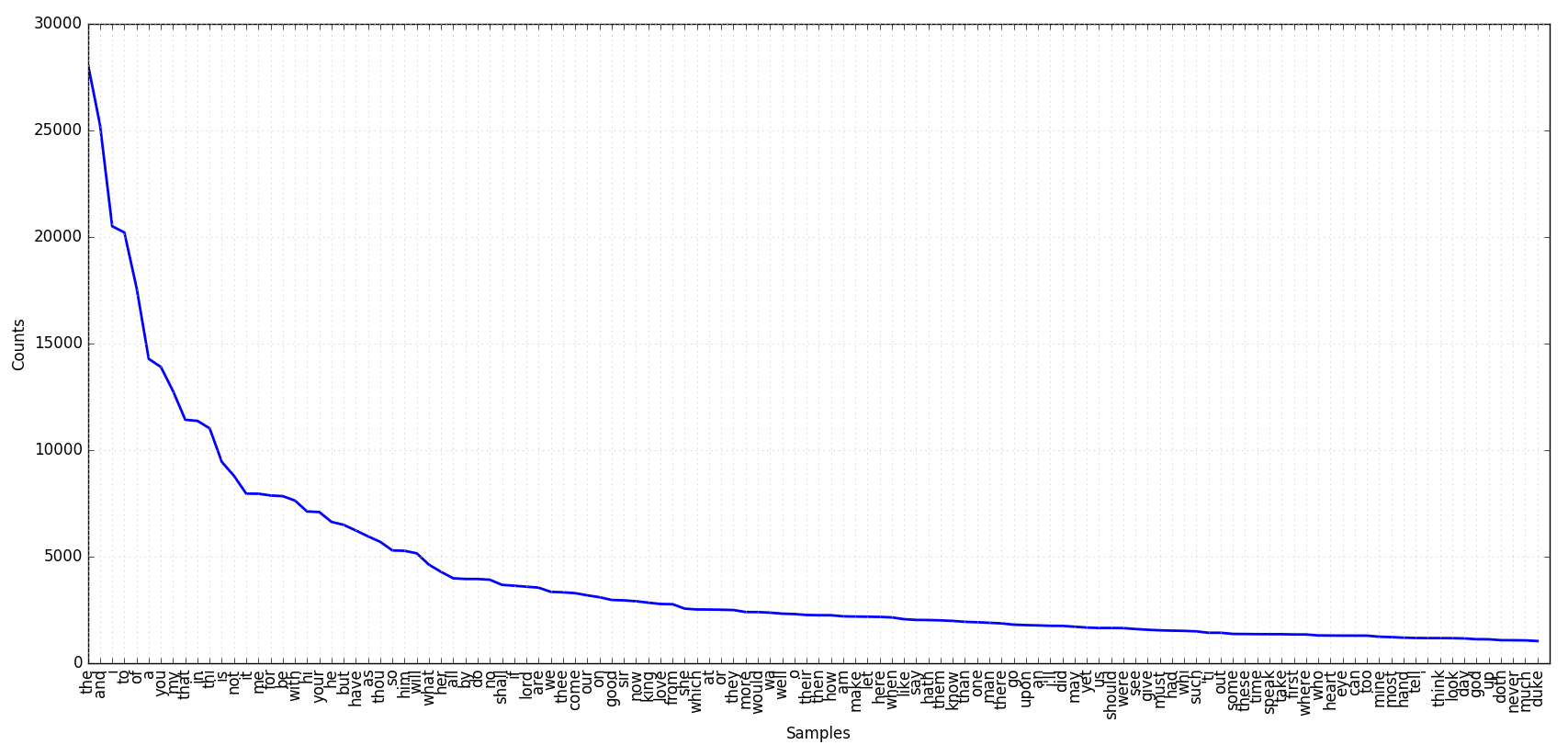
爬虫

网站分析

得到index.txt 枚举所有的标题 如101001

以及所有文件 如101001.txt

Word counting & stopping word



遍历后求出了所有的单词list.txt

取3000为界线 在list中得出stopword