Домашна работа по тема клосификакия по Бейс.

Разработете Бейсов класификатор, базиран на документи на български език, представляващи текстове от вашата имайл поща. Съберете корпус от по 10 мейла в категории: 1. Университет и 2. друго. Съставете по ваше усмотрение чанта с думи, като включите формите на думите, така, както вие прецените. Например – пиано, пиана, пианист, пинисти (това са проидводни и форми на думата).

Разработете Бейсов класификатор по следните стъпки:

1. Съставете Речника, който е използван във вашия корпус, като направите малък текст на Питон (може и друго), който прави списък с всички РАЗЛИЧНИ словоформи, които се срещат в корпуса. Преди да се започне този операция, в текста трябва да се заменят скбите, запетаите, точките и т.н. с шпации.Така разграничителите за всяка сровоформа са шпации отпред и отзад.
2. Като се основавате на материала в лекцията и за примерното броене и за формулите:

А) направете НА РЪКА р т.е. ползвайки напр. Уорд и Ексел небходимите преброявания и пресмятания.

Б) Може да направите, ако искате същото, на на Питон.

1. Класирайта две нови документа от ващия мейл-обмен, като ползвате обяснената методика. Възможно е да се наложи да логаритмувате произведението от вероятности, за да не губите точност. Не заблавяйте да антилогаритмувате резултата.