语言处理基础——NTLK 入门及英文语料库处理

了解基础的 Python NTLK 操作，给定文本 text_en.txt,完成以下练习：

1. 文本预处理
   (1) 分词、提取词干
   (2) 去停用词
   (3) 标点符号过滤
   (4) 低频词过滤（n <= threshold）
   (5) 绘制离散图，查看指定单词（Elizabeth, Darcy, Wickham, Bingley, Jane）在文中的分布位置
   (6) 对前 20 个有意义的高频词，绘制频率分布图
2. 了解 n-gram 在词性标注及训练中的使用：
   http://www.nltk.org/book/ch05.html#n_gram_tagger_index_term
3. 思考文本处理需要考虑哪些问题？以及中文在文本处理时和英文可能有哪些不同？