# 实验二

**要求：**利用给定的中英文文本序列（见 Chinese.txt 和 English.txt），分别利用以下给定的中英文分词工具进行分词并对不同分词工具产生的结果进行简要**对比分析**, 将实验过程与结果写成实验报告，实验课结束后提交。

## 中文分词工具

**Jieba**（重点），尝试**三种分词模式**与**自定义词典**功能：
https://github.com/fxsjy/jieba

SnowNLP：
https://github.com/isnowfy/snownlp

THULAC：
https://github.com/thunlp/THULAC-Python

NLPIR：
https://github.com/tsroten/pynlpir
https://blog.csdn.net/weixin_34613450/article/details/78695166

StanfordCoreNLP：
https://github.com/Lynten/stanford-corenlp

HanLP（选做，需要额外安装 Microsoft Visual C++ 14.0）
https://github.com/hankcs/pyhanlp

## 英文分词工具

NLTK：
http://www.nltk.org/index.html
https://github.com/nltk/nltk
https://www.jianshu.com/p/9d232e4a3c28

SpaCy：
https://blog.csdn.net/u012436149/article/details/79321112
https://spacy.io/usage/linguistic-features#section-tokenization

StanfordCoreNLP：
https://github.com/Lynten/stanford-corenlp