

中文分词、词性标注工具包

实验要求

1. 实现一个中文分词、词性标注工具包，工具包中**必须**包含以下分词算法：FMM、BMM、最短路径分词法、最大概率法分词法和基于HMM的分词法，同时也需要包含基于HMM的词性标注法。
2. 实验报告中需要简要介绍每个算法以及它们的伪代码。同时也需要测试不同算法的性能效果。

注意事项

1. 工具包中所要求实现的算法在附件中的课件中有介绍，大家也可以在网上自学。
2. 工具包的代码中要有必要的说明、注释，上传的压缩包中**必须**包含各个算法的使用样例。
3. 附件中 词性标注@人民日报199801.txt 为我们提供的文本数据，大家需要自行切分出训练验证数据以及测试数据，以测试不同算法的效果。大家也可以在网络上搜集其他数据来提高工具包的性能。
4. 大家也可以自学并实现基于CRF（条件随机场）算法的分词和切词方法。