NLP第一次大作业描述

1. 基本思路

今年才学的java所以就采用java来写这个程序，首先根据所给的词库路径建立一个词表，词表我用hashtable来写的。然后再对需要分词的文件一行一行的分词，输出写到result.txt里。

1. 分词算法

分词算法采用的是最大匹配算法，按从左至右正向最大匹配和从右到左反向最大匹配，当分词结果两种不一致时，按最少切分原则，取切分词数最少的一种，如果两种分词结果切分的词数一样，取反向最大匹配作为分词的结果

1. 遇到的问题
2. 是词表的问题，一开始我采用list来建立词表，但后来发现在分词匹配的过程中有问题，且不易完成匹配，需要复杂的过程才能完成搜索，上网上看了看别人的分享和建议，发现采用hashtable或者hashmap比较好，于是采用hashtable来建立词表，搜索匹配的过程直接用hashtable的成员函数containskey()即可完成搜索和匹配。

2，是前向和后向匹配结果不同是的处理，采用最好切分原则来处理。