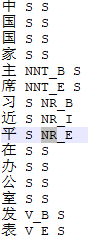
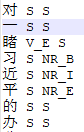
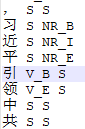
# 概率图模型--条件随机场

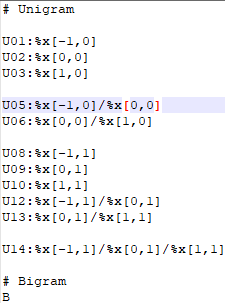
## 语料预处理：

为了获得更多的语义信息，从第二题的原始语料中可以提取更多的词性信息。所以先写一个脚本从第二题的原始语料中保留词性信息中的动词、称谓词、副词等5种词性信息，然后处理成单字形式的部分语料如下图：



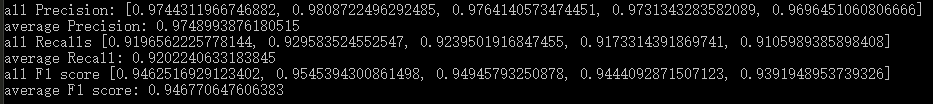
## 训练和测试：

1. 编写模板文件：文件内容如下



1. 开始训练： crf\_learn.exe -f 5 template.txt train\_data.txt model
2. 利用训练得到的模板文件进行测试 crf\_test.exe -m model test\_data.txt -o result
3. 编写脚本，统计测试结果文件中的TP , TN, FP, FN 数量，计算 precision Recall 和F1值。

统计结果如下图所示：



5次交叉验证得到的平均的准确率为97.49% , 召回率92.02%, F1 Score为94.68%