**编译原理实践第6/7次课**

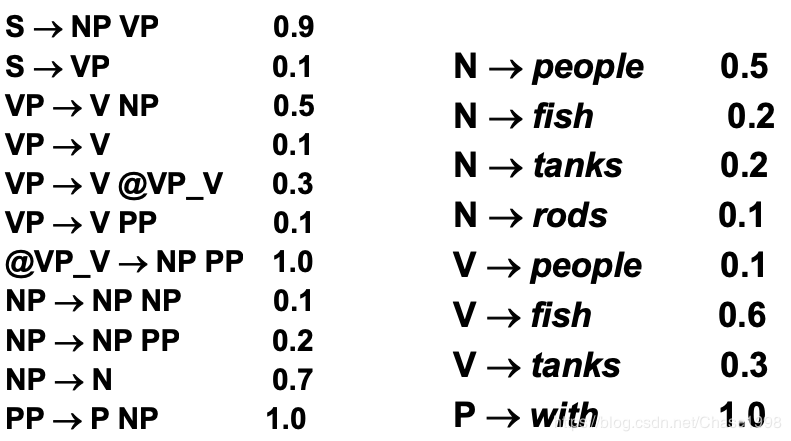
**（实现基于CKY算法的句法分析）**

说明:

(1) 本次(第六次和第七次)作业提交最晚时间为11月14日晚23:59.

(2) 需回答第2题、第3题和第5题.

1. 阅读 PCFG.ppt
2. 编写程序，给定如下图所示概率上下文无关文法（PCFG）,



注: 文法的终结符、非终结符、起始非终结符均可以从以个产生式中推出，这里不另外给出.

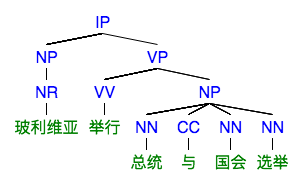
编程求句子fish people fish tanks的最优分析树 (即概率最大的分析树).

分析树的输出格式为第3题train.bracketed.txt文件所示格式. 如果当前句子不存在分析树（即解析不成功），输出None。

答:

2. 给定train.bracketed.txt文件, 其中第行表示一个句法树, 如第二行, 图形化该句法树如下图所示. (如果train.bracketed.txt中出现乱码, 请使用train.bracketed.gbk.txt)

( (IP (NP (NR 玻利维亚)) (VP (VV 举行) (NP (NN 总统) (CC 与) (NN 国会) (NN 选举)))) )



(1) 从该文件中抽取出所有的产生式, 并计算每条产生式的概率, 即P(RHS|LHS).

(2) 由于CYK算法要求文法符合CNF文法, 即每个产生式的右端至多只有两个二子节点, 需要对产生式进行二叉化.

二叉化例子:

原产生式:

A 🡪 B1 B2 … Bn

二叉化后:

@A 🡪 Bn-1 Bn

@A 🡪 Bn-2 @A

…

@A 🡪 B2 @A

A 🡪 B1 @A

从该文件中抽取出所有的二叉化后的产生式，并计算每条产生式的概率.

(3) 对test.txt文件中的每个句子, 求其最优分析树. 如果无分析树, 输出None. (如果test.txt中出现乱码, 请使用test.gbk.txt)

注: 如果求到的是二叉化分析树, 需要进一步还原为非二叉化分析树.

答: