# 人工智能-第四次课程作业说明

1. **问题描述**

给定语言的词典和语言的语法，现输入字符串语句S = “the wumpus is dead”。 请采用短语结构句法解析语句S，并输出句法解析树的线性表示 “[*S*:[*NP*:[*Article*: **the**] [*Noun*: **wumpus**]] [*VP*:[*Verb*: **is**][*Adjective*: **dead**]]]”。

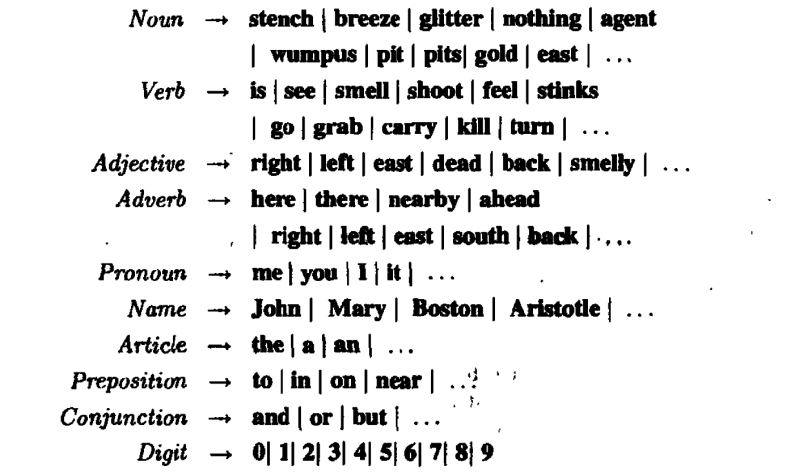


图1

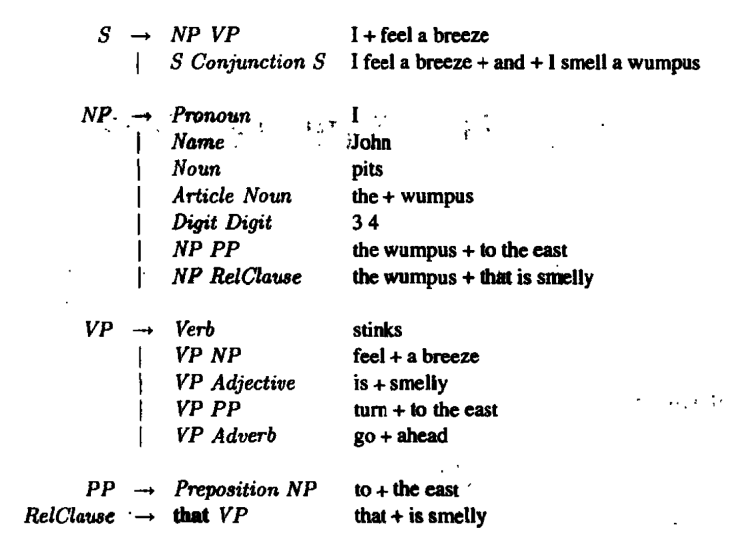


图2

1. **项目说明（Nlp类介绍）**

Nlp类中保存两个重要的成员变量，分别为字典和语法。字典的数据结构为：map< string, vector<LexRule> >，字典的键为词性，值为vector<LexRule>。形式如下：

1. {
2. "Noun": [
3. {"pos": "Noun", "lexicon": "stench", "prob": 0.05},
4. {"pos": "Noun", "lexicon": "breeze", "prob": 0.1}
5. ]
6. }

最外层的“Noun”代表该键储存所有名词的词汇规则，LexRule有三个成员变量，分别为词性（pos），词汇（lexicon），此种规则出现的概率（prob）。

语法的数据结构为：map< string, vector<CFGRule> >，字典的键为语法规则的左边，值为vector<CFGRule>。形式如下：

1. {
2. "S": [
3. {"lhs": "S", "rhs": ["NP", "VP"], "prob": 0.9},
4. {"lhs": "S", "rhs": ["S", "Conj", "S"], "prob": 0.1}
5. ]
6. }

最外层的“S”代表该键储存所有句子语法规则，CFGRule有三个成员变量，分别为规则左边（lhs），规则右边（rhs），此种规则出现的概率（prob）。上面的例子表示，一个句子有90%的概率会被拆解为名词性词组+动词性词组（NP+VP）的形式，有10%的概率会被拆解为句子+连接词+句子（S+Conj+S）的形式。

**3、项目运行**

使用打开Assignment4.sln项目，实现Nlp.cpp中的三个函数：1）get\_possible\_chunkings；

2）parse；3）parse\_rec。

完成函数后运行项目，针对“The Wumpus is dead”应得到图3中的结果。

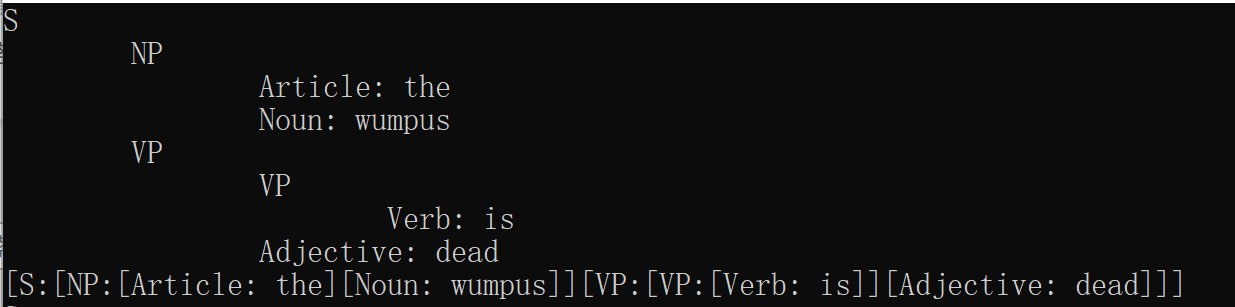


图3 句法解析树的线表示

**4、作业要求**

**4.1文档部分（2分）**

* 针对本题目撰写文档，保证文档结构完整性，文档可包括 题目、问题描述、实验方案、实验结果、分析、结论，也可参考其他标准文档或者学术论文的格式。
* 文档内容及结构均纳入评分范围。
* **文档：1分。源代码风格：1分**

**4.2源代码部分（8分）**

* 1个函数，8分，分数由解析树的正确率决定。由于该句子在给定句法下没有歧义，所以只有正确和错误两种结果。助教将比较标准答案的解析树和同学们提交的解析树是否完全相同。
* （评价：希望同学们独立完成。如果有困难，则可以参考其他同学的答案，但是不要抄袭）

**5、作业提交**

提交时间：第14周周四（12月08日）上课前

提交方式：将电子版源代码及文档打包，命名为“学号\_姓名\_Assignment4”，发送至邮箱tengh@seu.edu.cn。

注意：打包方式使用zip；打包时不要把项目中的.vs文件打包进去。