

华东师范大学期末试卷（A）

2013—2014 学年第一学期

课程名称：自然语言处理导论

学生姓名：_____

学号：_____

专业：_____

年级/班级：_____

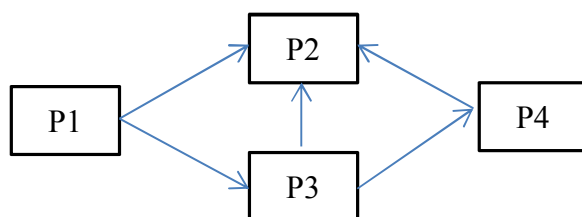
课程性质：专业选修

一	二	三	四	五	六	总分	阅卷人签名

一、简答题（共 30 分）

1. 人类关于语言的语义知识表现在什么能力上？（3 分）
2. 语料库与语言知识库的区别是什么？（3 分）
3. Web 信息检索中，PageRank 算法与向量空间模型之间的主要区别是什么？（3 分）
4. WordNet 如何表示概念？概念间的关系有哪几种？（3 分）
5. 简述格语法分析步骤。（4 分）
6. 语义知识跟句法知识的区别在哪里？（3 分）
7. 什么是篇章？（3 分）
8. Web 搜索引擎系统由哪几部分组成？（4 分）
9. 合一运算用于句法分析中有什么作用？（4 分）

二、下图表示了四个网页及他们之间的链接关系。请利用 PageRank 初始算法对各个网页的权威性进行初始化，并计算经过第一次循环重新赋值后各自的权值。（10 分）



三、分词：有如下词典（表 1），每个词及单字的费用 $C = -\log P$, P 为该词或单字在语料库中出现的概率。对“鞭炮声响彻夜空”进行分词。(共 15 分)

- (1) 分别给出正向和逆向最大匹配法的分词结果。(3 分)
- (2) 画出词图，给出最短路径法的分词结果。(4 分)
- (3) 用最大概率法分词，给出详细计算过程及结果。(8 分)

表 1

词	费用	单字	费用
鞭炮	1.5	鞭	4
炮声	2	炮	3.5
声响	2.5	声	3
响彻	3	响	2.5
彻夜	2	彻	5
夜空	1	夜	3
		空	4

四、句法分析

有上下文无关文法如下，

- 1. $S \rightarrow NP VP$
 - 2. $NP \rightarrow NP PP$
 - 3. $NP \rightarrow N$
 - 4. $PP \rightarrow P NP$
 - 5. $VP \rightarrow V NP$
 - 6. $VP \rightarrow VP PP$

用上述文法进行句法结构分析，需给出所有可能的结果。(共 15 分)

- (1) 用由底向上方法分析“**John/N ate/V fish/N with/P bone/N**”的句子结构，只需画出最后分析树。(5 分)
- (2) 用 Earley 算法分析上述句子的句法结构，给出详细过程。（10 分）

五、简述基于大规模文档集的问答系统的系统构成及工作流程。（10 分）

六、用 Viterbi 算法对下面句子进行词性标注。假设名词(n)和动词(v)的初始状态概率分布相同。（20 分）

Time/n-v flies/v-n like/p-v an/det arrow/n.

表 2、词性转移概率（例： $p(v \rightarrow n)=0.4$ ）

词性	n	v	p	det
n	0.2	0.2	0.1	0.1
v	0.4	0.2	0.2	0.1
p	0.3	0.1	0.05	0.3
det	0.9	0.05	0.01	0.00

表 3、词语输出概率

词语	词性	概率($\times 10^{-2}$)		词语	词性	概率($\times 10^{-2}$)
time	n	5		like	p	5
time	v	1		like	v	4
fly	v	6		an	det	20
fly	n	2		arrow	n	1