

中国科学院大学

课程编号: B0912014Y

试题专用纸(B卷)

课程名称: 自然语言处理

任课教师: 宗成庆

注意事项:

1. 考核方式: 课堂开卷;
2. 请在规定时间内将本试卷和和答题纸一并交回。

1. 标注题 (10 分)

有如下一段文字:

6月25日是第33个全国土地日,今年的主题是“节约集约用地,严守耕地红线”。江苏在全国首创“责任+激励、行政+市场”耕地保护新机制,筑牢粮食“耕”基……守住用好每一寸土地的号角声愈发嘹亮,此次全国土地日主场活动前后,将组织策划“耕地保护进万村”“全国耕地保护全媒体创意作品征集”“在希望的田野上主题直播”等宣传活动,持续提升全社会保护耕地、节约集约和依法依规用地的意识。(文/冯唯歆 图/金裕)

- (1) 请对上面这段文字进行词语切分。要求:用明确的标记符号表示词语的边界,如两个或三个空格,或者用“/”。
- (2) 请列出上面这段文字中所有的命名实体,说明其类型。
- (3) 请说明下面句子的语义角色:
全国土地日主场将组织策划“耕地保护进万村”等宣传活动。

2. 计算题 (10 分)

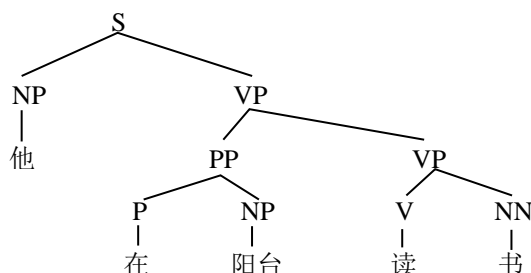
假设有如下三个句子:

- (a) 他喜欢大模型。
- (b) 大模型让他着迷。
- (c) 他没用过大模型。

- (1) 请以词为单位计算句子(a)和(b)之间的编辑距离。
- (2) 用上述三个句子作为训练语料,采用2元语法模型、“加1”数据平滑方法,请计算句子“没喜欢过大模型。”的概率。

3. 分析标注题 (10 分)

有如下短语结构树:



(1) 请根据该短语结构树画出其依存关系图。

(2) 请标出该句子的语义角色。

4. 设计题（10 分）

请设计一种方法，抽取出汉语文本中所有的语义相近的词对（近义词对）。写出基本思路 and 实现步骤。

5. 设计题（10 分）

假设一篇较长的文章被意外地打乱了段落和句子的顺序。请设计一种方法，将这些句子的顺序重新排列起来，恢复原来的语序。写出基本思路 and 实现步骤，需要时给出计算公式。

6. 设计题（10 分）

机器翻译系统输出的译文需要在流畅性（fluency）、正确性（correctness）和语义充分性（sufficiency）三个方面符合人工译员的要求。请设计一种方法，如何自动计算机器译文的充分性？写出基本思路 and 实现步骤，必要时给出计算公式。