

## 《自然语言理解》课程作业

课程编号：71253Z                      课程属性：专业基础课                      学时/学分：40/2

预修课程：概率论与数理统计、算法分析与程序设计

主讲人：宗成庆 研究员

联系方式：E-mail: cqzong@nlpr.ia.ac.cn      Tel. 6255 4263

### 一、作业目的：

通过本课程作业加深对自然语言理解基础理论的认识和了解，锻炼和提高分析问题、解决问题的能力。通过对具体项目的任务分析、技术调研、数据准备、算法设计和编码实现以及系统调试等几个环节的练习，基本掌握实现一个自然语言处理系统的基本过程。

### 二、作业题目：

#### 1. 实现一个汉语命名实体自动识别系统 (Named entity identification)

命名实体一般指如下几类专有名词：中国人名、外国人译名、地名、组织机构名、数字、日期和货币数量。可实现其中一种类型命名实体的识别。

#### 2. 实现一个汉语自动分词与词性自动标注系统 (Chinese word segmentation and POS tagging)

如果在本项目中不考虑命名实体识别问题，歧义消解是汉语自动分词和词性标注中的关键问题。

#### 3. 实现一个 (汉语/英语) 词义自动消歧系统 (Word sense disambiguation)

很多词汇具有一词多义的特点，但一个词在特定的上下文语境中其含义却是确定的。本项目要求实现系统能够自动根据不同上下文判断某一词的特定含义。

#### 4. 实现一个 (汉语/英语) 名词短语自动识别系统 (Noun phrase recognition)

请设计并实现汉语句子的名词短语自动识别系统。

#### 5. 实现一个汉语句子谓语自动识别系统 (Chinese predicate recognition)

谓语是句子的中心。请设计并实现一个汉语句子谓语自动识别系统。

#### 6. 实现一个汉语句子自动改写系统 (Chinese sentence paraphrasing)

请设计并实现一个汉语句子自动改写系统。

#### 7. 实现一个汉英词语自动对齐系统 (Chinese-English word alignment)

利用汉英双语句子对齐语料，设计并实现汉英词语自动对齐算法。

#### 8. 实现一个汉语-英语人名自动翻译系统 (Chinese-English person name translation)

该系统可以将汉语人名 (包括中国人名和外国人译名) 翻译成英语。

#### 9. 实现一个英文单词拼写错误自动检查系统 (English text proofreading)

请设计并实现英文文本中单词拼写错误自动检查系统。可以限定词汇量。

#### 10. 实现一个文本自动分类系统 (Text categorization)

依据某种文本分类标准，设计并实现一个分类自动分类系统。汉语文本或英语文本任选一种。

### 三、基本要求：

1) 每人可以选择其中的一个题目，也可以几个人合作完成一个题目，但彼此之间必须有明确的分工和要求。

2) 任何一个题目，都不限定采用的方法，可以采用基于规则的分析方法，也可以采用基于语料库的统计方法，还可以是几种方法的结合，鼓励方法创新，但必须有理论根据或实验数据依据。

3) 上述有的题目比较困难，如果不能找到合作的同学共同完成，可以选做上述某一题目中的部分工作，但请说明所做的部分与整个项目其它部分的关系。

4) 完成一份技术报告，报告内容包括：项目目标、国内外相关工作、自己在本项目中承担工作的不同点、实现系统 (或模块) 的核心思想和算法描述、系统主要模块流程、实验结果及分析。

5) 提交系统源代码和可执行程序，以保证实验系统可以正常编译和运行。如果是多人合作完成的，应提交最终集成的系统。

### 四、声明：

1. 希望 6 月 1 日前完成，报告中请留下作者的姓名、单位、联系电话和邮件地址。可以通过电子邮件提交或直接提交光盘。

2. 鼓励充分使用网络资源和其它一切可以利用的资源 (包括数据、语料、软件工具和论文资料等)，但严禁侵害他人知识产权，技术报告中必须明确说明所用资源的真实来源。

3. 鼓励相互交流、相互合作，但严禁抄袭他人工作，严禁伪造结果。