

课程名称：语言信息处理

一、课程编码：0700002

课内学时：__48__ 学分：__3__

二、适用专业：计算机应用技术，计算机软件与理论

三、先修课程：计算理论，算法分析与设计，数理统计

四、教学目的：

通过本课程的学习，使研究生：

- 1、了解语言信息处理的主要研究内容与发展历程；
- 2、了解文法基础理论，掌握典型的自然语言形式模型；
- 3、掌握语言信息处理，特别是词法、句法、语义、语用等多层次语言分析的基本原理与主要技术方法；
- 4、了解主要的自然语言生成技术；
- 5、掌握语言信息处理，特别是机器翻译、信息检索、信息抽取等工程应用的核心问题与主要技术方法；
- 6、了解语言信息处理技术的前沿及发展趋势。

五、教学方式：

课堂讲授，材料自学与课堂讨论，专题讲座相结合。

六、教学主要内容及对学生的要求：

第1章 引论

3 学时

- 1.1 问题提出
- 1.2 基本概念
- 1.3 学科定位
- 1.4 处理层次及研究内容
- 1.5 发展历程
- 1.6 系统构成

第一部分 知识表示及文法

第2章 语言学基础知识及知识表示

4 学时

- 2.1 基本概念
- 2.2 词法基础知识
 - 2.2.1 词的构成
 - 2.2.2 词法规则
 - 2.2.3 词法分析
 - 2.2.4 汉语分词
- 2.3 句法基础知识
 - 2.3.1 词及词类
 - 2.3.2 句法成分及句法结构
 - 2.3.3 句法规则

- 2.4 语义和逻辑表达
 - 2.4.1 语义和逻辑形式
 - 2.4.2 词义和歧义
 - 2.4.3 基本逻辑形式语言
 - 2.4.4 一阶谓词逻辑
 - 2.4.6 论旨角色
- 2.5 语用和常识
 - 2.5.1 知识的表示和推理
 - 2.5.2 局部篇章上下文和指代
 - 2.5.3 常识的使用
- 2.6 篇章结构
- 2.7 小结

第3章 文法基础理论

5 学时

- 3.1 引论
- 3.2 模式匹配
- 3.3 基于语法驱动的文法
 - 3.3.1 上下文无关文法
 - 3.3.2 转换生成文法
 - 3.3.3 扩充转移网络
- 3.4 基于语义的文法
 - 3.4.1 语义文法
 - 3.4.2 格文法
 - 3.4.3 语义网络
- 3.5 基于知识的文法
 - 3.5.1 概念依存文法
 - 3.5.2 优选语义学
- 3.6 基于逻辑推理和人工智能的文法
 - 3.6.1 蒙塔格文法
 - 3.6.2 系统文法
 - 3.6.3 故事文法
 - 3.6.4 SC 文法
- 3.7 基于合一运算的文法
 - 3.7.1 功能合一文法
 - 3.7.2 词汇功能文法
- 3.8 小结

***专题 谓词逻辑与基于逻辑的 NLP**

第二部分 语言分析与生成

第4章 语言分析技术

15 学时

- 4.1 词法分析技术
 - 4.1.1 概述
 - 4.1.2 知识回顾
 - 4.1.3 屈折语形态分析

- 4.1.4 汉语自动分词
- 4.1.5 词性标注
- 4.1.6 词汇语义计算
- 4.2 句法分析技术**
 - 4.2.1 概述
 - 4.2.2 知识回顾
 - 4.2.3 基于上下文无关文法的句法分析
 - 4.2.4 基于概率上下文无关文法的句法分析
 - 4.2.5 依存句法分析
 - 4.2.6 浅层句法分析
- 4.3 语义分析技术**
 - 4.3.1 概述
 - 4.3.2 知识回顾
 - 4.3.3 语法驱动的分析技术
 - 4.3.4 语义驱动的分析技术
 - 4.3.5 语义角色标注
- 4.4 小结**

***专题：图模型在 NLP 中的应用**

***专题：NLP 中的机器学习**

第5章 篇章分析与会话处理技术

3 学时

- 5.1 篇章问题概述
- 5.2 篇章标注模型
- 5.3 衔接
- 5.4 连贯
- 5.5 言语行为模型
- 5.5 会话管理
- 5.6 小结

第6章 语言生成技术

2 学时

- 6.1 概述
- 6.2 自然语言生成发展与现状
- 6.3 主要的自然语言文本生成技术
- 6.4 语言生成技术应用
- 6.5 小结

第三部分 应用技术

第7章 机器翻译

4 学时

- 7.1 概述
- 7.2 基于规则的机器翻译
 - 7.2.1 一个基于规则的机器翻译系统实例
- 7.3 基于实例的机器翻译
 - 7.3.1 双语实例库的构建
 - 7.3.2 相似实例的检索

7.3.3 最终译文的生成

7.4 统计机器翻译

7.4.1 噪声信道模型

7.4.2 基于短语的翻译系统

7.4.3 基于句法的翻译系统

7.5 译文质量评估

*专题 NLP 和机器翻译中的深度学习

第8章 信息检索与问答系统

3 学时

8.1 信息检索技术

8.1.1 概述

8.1.2 信息检索模型

8.1.3 文档排序方法

8.1.4 评价标准

8.1.5 发展趋势

8.2 问答系统

8.2.1 概述

8.2.2 系统构成

8.2.3 关键问题及主要方法

8.2.4 问答系统评测

8.2.5 发展趋势

第9章 信息抽取与自动摘要

3 学时

9.1 信息抽取

9.1.1 概述

9.1.2 系统构成

9.1.3 主要任务和关键技术

9.1.4 信息抽取系统评测

9.1.5 发展趋势

9.2 自动摘要

9.2.1 概述

9.2.2 关键问题及主要方法

9.2.3 自动摘要系统评测

9.2.4 发展趋势

第四部分 主题讨论与专题讲座

6 学时

语言信息处理相关问题主题讨论、邀请知名专家进行专题讲座

七、考核与成绩评定

成绩以百分制衡量。

成绩评定依据: 开放式问题讨论成绩占 40%，平时作业成绩占 30%，期末项目及报告成绩占 30%。

八、参考书及学生必读参考资料:

- 1、 Allen, J.F. Natural Language Understanding. The Benjamin/Cummings Publishing Company, Inc., Redwood City, CA. 1995. (自然语言理解 (第二版), James Allen 著, 刘群等译, 电子工业出版社. 2005)
- 2、 Jurafsky, D. and J. H. Martin. Speech and Language Processing-An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice Hall. 2000. (自然语言处理综论(第二版), Daniel Jurafsky, James H. Martin 著, 冯志伟、孙乐译, 电子工业出版社. 2005)
- 3、 Christopher D. Manning, HinrichSchütze. Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press. 1999. (苑春法等译. 统计自然语言处理基础. 北京: 电子工业出版社. 2005.)
- 4、 Philipp Koehn. Statistical Machine Translation. New York, NY: Cambridge University Press. 2009. 宗成庆, 张霄军译. 统计机器翻译. 北京: 电子工业出版社. 2012.
- 5、 冯志伟. 自然语言处理的形式模型. 安徽: 中国科学技术大学出版社. 2010.
- 6、 宗成庆. 统计自然语言处理(第二版). 北京: 清华大学出版社.2013.
- 7、 李航. 统计学习方法. 北京: 清华大学出版社. 2012.
- 8、 胡壮麟. 语言学教程(第四版). 北京: 北京大学出版社. 2013.
- 9、 吴军. 数学之美. 北京: 人民邮电出版社. 2012.

九、大纲撰写人：黄河燕、史树敏、鉴萍