湖南工业大学研究生课程教学大纲

课程编号：00812030

编写人：吴岳忠

编写日期：2021-12-06

课程中文名称	自然语言处理				
课程英文名称	Natural Language Processing				
开课学期	春季	开课单位	计算机学院		
课程类别	计算机科学与技术专业、电子信息专业 专业课				
主讲教师	吴岳忠	职称	副教授	联系电话	13973363474
教学团队成员	朱艳辉 万烂军				
学时	32				
学分	32	教学及考核方式	理论教学与实践教学相结合 课程论文		
面向学科（专业学位领域）	计算机科学与技术、电子信息			考核方式	☐ 考试 ☒ 考查
预修课程	线性代数 概率论与数理统计 微积分 程序设计语言 人工智能导论 机器学习				

课程内容：

《自然语言处理》是计算机专业的一门专业课，是人工智能领域的主要内容，研究用电子计算机模拟人的语言交际过程，使计算机能理解和运用人类社会的自然语言，实现人机之间的自然语言通信，以代替人的部分脑力劳动，包括查询资料、解答问题、摘录文献、汇编资料以及一切有关自然语言信息的加工处理。

第一部分 自然语言领域知识概述

2 学时

包括人工智能发展历史综述、不同领域自然语言处理应用问题及方法、自然语言处理基本技术方法概述等

第二部分 语言模型

6 学时

包括自然语言处理相关语言模型基础理论以及相关平滑处理技术等（如 N-Gram; Katz 算法等）

第三部分 序列标注问题

6 学时

包括序列标注学习及其相关应用等（如离散/连续马尔科夫模型、中文分词、词性标注、命名实体识别等）

第四部分 句法分析

6 学时

包括自然语言中语法分析等基础理论知识等（如上下文无关文法、自上而下句法分析、概率上下文无关文法、最大似然训练、依存语法树等）

第五部分 语义分析

6 学时

包括自然语言中语义分析等基础理论知识等（如语义角色、语义角色标注、基于句法树方法等）

第六部分 情感分析技术

4 学时

包括情感分析技术理论与方法等（如感情倾向性分析等）

第七部分 问答系统

2 学时

包括聊天机器人实现的原理、方法等

课程内容英文简介

This course provides an introduction to the field of natural language processing (or computational linguistics), including both analysis and generation. Speech processing, machine translation, and computational approaches to language acquisition and language evolution are also given some attention. A wide range of linguistic phenomena, including phonology, morphology, syntax and semantics will be treated, and examples will come from various languages. We will be concerned both with how well particular approaches solve practical problems and with how well they model human data. On the basis of the above, students are required to finish a course project so that they can perform practice through releasing their potential, searching related literature, and summarizing latest theoretical and technique innovations.

课程教学目标及重点、难点：

本课程的任务是要使学生掌握自然语言处理的基本概念、基本原理、主要方法、开源工具，以及发展现状，紧跟自然语言处理某一研究领域的国内外最新理论和技术，综合运用自然语言处理知识，结合真实应用场景分析设计更加具体、科学、规范的解决方案。本课程是一门实践性很强的课程，它是各种开发经验的总结与提炼，学生不但应注意概念、原理、方法、技术的掌握，也应该注重方法、技术的实际应用。本课程教学难点为语言模型、语法分析、语义分析、NLP 常用深度学习算法等。

教学要求：

1. 了解人工智能发展历程和自然语言处理研究内容、基本方法及其发展。

2. 掌握自然语言处理相关语言模型基础理论以及相关平滑处理技术。
3. 掌握马尔科夫模型、中文分词、词性标注、命名实体识别等技术及其应用。
4. 了解数据清洗，掌握缺失值、异常值和重复值的处理方法。
5. 了解特征预处理，掌握规范化和标准化的处理方法，掌握独热编码、CountVectorizer 和 TF-IDF 基本原理和使用方法。
6. 掌握上下文无关语法、自上而下句法分析、概率上下文无关语法、最大似然训练、依存语法树等自然语言中语法分析等基础理论知识，掌握语义角色、语义角色标注、基于句法树方法等自然语言中语义分析等基础理论知识。
7. 了解情感分析相关概念、场景，掌握情感分析的一般流程，能够设计实现一个行业应用。
8. 掌握 NLP 常用的分类算法、聚类算法以及深度学习算法，结合具体场景进行工程实践。
9. 了解聊天机器人的实现原理和过程，掌握从句子理解、句法分析、同义词提取等方面实现聊天机器人的技术。
10. 了解自然语言处理中的高级课题，把握自然语言处理发展动态和方向。

教材及主要参考书：

1. 陈鄞. 自然语言处理基本理论和方法（第二版）. 哈尔滨工业大学出版社. 2017 年
2. 宗成庆. 统计自然语言处理. 清华大学出版社, 2013 年
3. Daniel Jurafsky (D. 朱夫斯凯), James H. Martin (J. H. 马丁). 自然语言处理综论（第二版）. 电子工业出版社, 2018 年
4. 涂铭, 刘祥, 刘树春著. Python 自然语言处理实战：核心技术与算法. 机械工业出版社, 2018 年
5. 周元哲. Python 自然语言处理（微课版）. 清华大学出版社, 2021 年

大作业：

基于深度学习的实际生产场景情感分析系统设计。