**山东大学 软件 学院**

**《自然语言处理》教学大纲**

编 写 人：孙宇清

编制时间：2024.02.25

**一、课程基本信息：**

|  |  |  |  |
| --- | --- | --- | --- |
| 课程名称 | **自然语言处理** | | |
| 英文名称 | **Natural Language Processing** | | |
| 课程编码 | **sd03032130** | | |
| 开课单位 | **软件学院** | | |
| 课程类别 | □通识教育必修课程 □通识教育核心课程  □通识教育选修课程 □学科基础平台课程  □专业基础课程 □专业必修课程。 √ □专业选修课程 | | |
| 课程性质 | □必修 √ □选修 | | |
| 学分 | **2.5** | 学时 | **32上课+16实验** |
| 适用专业 | **软件工程，人工智能，网络安全，大数据** | | |
| 先修课程 | 高级程序设计语言（C++,JAVA，python）、数据结构、机器学习、线性代数、概率与统计 | | |
| 课程网站 | **http://splab.sdu.edu.cn/zryycl1.htm** | | |

**二、课程描述**

自然语言处理（NLP）也称为计算语言学，研究如何采用计算手段处理人类语言的理论、技术和系统，包括语言表示、存储、处理和理解等，是人与计算机进行自然交互的重要手段，是信息时代最重要的智能技术之一。自然语言处理的相关应用无处不在，如网络搜索，医药问答，客户服务，语言翻译，学术报告等。

为了使得计算机能够理解自然语言，需要解决各种基础性的复杂任务，相关研究可以追溯到20世纪40年代末。经历了近80年的发展历史，自然语言处理技术取得了长足进步。近年来，以大数据和深度学习为代表的智能技术和高性能计算系统得到广泛应用，端到端的神经模型不需要传统的特征工程，大幅提升了多项自然语言处理任务的性能。本课程将从语言学知识出发，详细讲解处理语言任务的不同计算方法，以及深度学习在自然语言处理中的前沿进展。

本课程的特点是理论联系实际，教学方式包括授课、习题课、课堂讨论、作业和实验。通过上述教学方式培养学生的理论水平和实际动手能力，启发学生对理论知识和问题的思考，探索需求驱动的技术创新和应用创新。

**三、课程教学目标和教学要求**

【教学目标】

牢固掌握自然语言处理的基本概念、理论和经典算法，掌握解决自然语言典型任务的智能技术，了解端到端的预训练模型和深度学习等前沿技术和方法；具备应用智能技术解决基本问题的能力，包括模型设计、模型选择、模型实现、结果分析和全面评估等，学习必要的工程技巧，具备实施、训练、调试等工程能力；具备结合应用背景解决现实问题的分析能力和实践能力，培养发明新技术和应用的创新能力。

【教学要求】

通过本课程学习，学生将具备以下知识和能力：

1、自然语言处理的基本知识、理论和方法。

2、面向特定语言处理任务的现代工具学习和使用、改进技术和方法。

3、面向特定问题的解决方案设计能力，系统实现能力。

4、面向前沿问题的知识获取能力、自主学习能力、分析能力和探索能力。

**四、课程教学内容及学时分配**

|  |  |  |  |
| --- | --- | --- | --- |
| 章节 | 主 题 | 内 容 | 学时 |
| 绪论 | 自然语言处理介绍 | 什么是自然语言处理 | 1 |
| 第一章 | 词汇和词向量 | 词汇、词汇向量化，分布式词向量和评估，WordNet | 3 |
| 第二章 | 分词 | 中文分词 | 2 |
| 第三章 | 语言模型 | 语言模型，深度网络优化知识 | 2 |
| 第四章 | 句法分析 | 句法分析 | 2 |
| 第五章 | 文本序列标注 | 文本序列标注任务，词性标注 | 2 |
| 第六章 | 信息抽取 | 命名实体识别  关系抽取和知识图谱 | 2 |
|  | 困难问题解答 | 困难问题解答 | 2 |
| 第七章 | 文本分类 | 文本编码方法、文本分类 | 2 |
| 第八章 | 机器翻译 | 机器翻译 | 2 |
| 第九章 | 问答系统与自动摘要 | 问答系统、自动摘要和文本生成 | 2 |
| 第十章 | 大语言模型 | 注意力机制、模型结构、  大模型预训练和精调技术 | 4 |
| 第十一章 | 前沿问题讨论 | 长文本处理、数据标注、模型可解释性、伦理道德等社会问题 | 4 |
| 总结 | 复习总结 |  | 2 |

**五、评价考核方案**

课程成绩由**作业、**实验、大作业、课堂表现、线下交互、测试、期末考试组成，其中，作业、单元测试占比15-25%，采用线上提交和评审方式；实验和实践大作业占比20-30%，实验评定依据实验报告和代码运行结果判分，实践大作业通过答辩形式进行专家评审；课堂表现和线下问答占比5%，由平时的到课率以及课堂和课后问答和分享情况判定；期末考试占比50%。

**六、教材及参考资料**

【教材】

自然语言处理综论（第二版）,Daniel Jurafsky（D.朱夫斯凯），James H.Martin（J.H.马丁） 著，冯志伟 译。电子工业出版社.

【参考书】

1. 电子版英文教材Speech and Language Processing. Dan Jurafsky and James H. Martin. Free Pdf @ <https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf>
2. 统计自然语言处理（第二版）.宗成庆 著,清华大学出版社,出版时间：2013, ISBN：9787302319115
3. 统计学习方法. 李航 著, 清华大学出版社, 2013, ISBN：978-7-302-27595-4
4. 自然语言处理原理与技术实现,罗刚等。电子工业出版社。电子工业出版社, ISBN：978-7-121-28620-9
5. 课件参考: http://splab.sdu.edu.cn/zryycl1.htm
6. 课件参考: https://web.stanford.edu/class/cs224n/
7. 深度学习教材，电子版http://www.deeplearningbook.org/
8. Tools and corpora: https://nlp.stanford.edu/links/statnlp.html (Statistical natural language processing and corpus-based computational linguistics: An annotated list of resources)
9. 其他中文材料教师随堂提供或网络形式提供。