

复旦大学研究生课程教学大纲

1. 课程概要/Course Summary			
课程名称（中文） Course Title (Chinese)	高级人工智能技术（互联网大数据处理技术）		
课程名称（英文） Course Title (English)	Advanced Artificial Intelligence Technology （Processing Technologies for Internet Big Data）		
开课院系 School/Department	软件学院	开课学期 semester	2021-2022 学年第一学期 Autumn 2021
课程代码 Course Code	ME630048	任课教师 Instructor(s)	曾剑平
授课语言 Teaching Language	中文	适用学科专业 Discipline/Specialization	软件工程、电子信息
学分数 Course Credit(s)	3	教学周数 Weeks	共 16 周
总学时 Teaching Hours in Total	共 54 学时	实验/实践学时 Hours for Experiments /Practice	共 0 学时
预修课程要求 Pre-requisite Course(s)	计算机网络、概率论		
课程简介 Course Introduction (150-300 字)	课程内容以互联网大数据处理技术体系为线索和重点，从互联网信息内容获取、特征提取、大数据语义、表示模型、存储入手，围绕新闻分类、社交媒体挖掘、评论信息挖掘等典型应用等关键点展开。同时，结合技术发展前沿，介绍相关的理论与技术，讲解这些技术用于解决互联网信息内容处理问题的应用方法，课程的特点是理论与实用相结合。		
2. 教学目标/Course Objective (100-200 字)			
本课程主要针对计算机、软件和大数据等专业在职研究生，主要讲解互联网大数据技术体系、原理、模型、算法和关键技术，并对相关领域的技术发展前沿进行梳理和讲解，让学生既能熟悉技术的相关基础，又能了解发展趋势。介绍互联网大数据技术的爬虫、Web 信息提取、文本预处理、大数据语义、分类器、主题建模以及相关的应用技术。通过本课程教学，使学生对互联网大数据技术及应用有一个全面的了解，掌握基本的信息内容采集与分析方法，并且具备一定的实际运用和研发能力。			
3. 教学内容及进度安排/Course Content & Schedule			
课次 No.	教学周 Week	教学内容及预期效果 Content & Expected Achievement	作业/实验/实践 Assignment
1	第 1 周	互联网大数据技术概述：特点、概念、技术体系与研究进展	

2	第 2 周	网络爬虫相关基础，Web 服务器、HTML 语言、HTTP 协议、以及爬虫的共性技术	
3	第 3 周	动态页面技术：原理、技术实现方法、以及 DeepWeb 技术	爬虫技术相关文献阅读
4	第 4 周	主题爬虫技术：架构、主题表示、相关度计算、爬行策略与实现技术	
5	第 5 周	微博信息采集、爬虫与反爬虫、爬虫技术讨论	
6	第 6 周	爬虫技术文献讨论	
7	第 7 周	Web 信息提取技术原理、Web 信息提取的技术实现、研究进展	
8	第 8 周	分词技术：基于词典、基于统计、词性识别方法、歧义处理、新词、相关分词工具和语言实现	
9	第 9 周	大数据语义：词汇级别语义、句子级别语义命名实体识别的技术原理与实现技术。	
10	第 10 周	信息抽取、大数据语义文献讨论	
11	第 11 周	文本表示：特征选择、特征提取 LSI、以及向量空间模型	
12	第 12 周	基于向量空间的分类器技术原理、实现方法、相关工具、性能评估	互联网大数据分析应用作业
13	第 13 周	文本的概率模型与分类器集成：n-gram、PLSI	
14	第 14 周	主题建模技术：原理、LDA	
15	第 15 周	社交媒体内容分析技术	
16	第 16 周	课程论文交流	
17	考试周		
18			

4. 课程考核及成绩评定/Course Assessment & Grading

考核形式* Assessment Criteria	权重 Percentage	评定标准 Assessment Standard
出勤 Attendance	10%	根据考勤记录
课堂表现 Participation	30%	根据课堂报告、课堂互动、回答、参与度
作业/实验/实践 Assignment(s)		
课程论文 Course Paper	60%	根据课程论文的工作量、主要贡献、规范性等
开卷考试 Open-book exam		

闭卷考试 Close-book exam					
其他 Other(s)					
* 各项考核指标可自由设置，无相关要求的可填 0%，总权重为 100%。					
5. 教材/Textbook(s) （如使用自编讲义，请在“名称”列中备注说明）					
序号 No.	名称 Title	主编/作者 Author(s)	标准书号 ISBN	出版机构 Publisher	出版日期 Publication Date
1	互联网大数据处理技术与应用	曾剑平	9787302463719	清华大学出版社	2017.4
2					
3					
4					
5					
6					
6. 任课教师简介/Profile of Instructor(s) （教学科研经历简介，300 字左右）					
<p>曾剑平，博士，副教授，在复旦大学从事教学科研工作。教学方面，讲授过的课程包括《信息内容安全》（本科）、《计算机取证技术》（本科）、《Python 程序设计》（本科）、《互联网大数据技术》（研究生）、《大数据安全》（研究生）等。科研方面，主要从事大数据安全、人工智能安全、社交媒体挖掘应用、主动防御等方面的研究，在 IEEE TIFS、Computers & Security、Knowledge-based Systems 等国内外学术期刊和会议上发表论文 80 多篇，获得大数据技术、网络安全方面相关授权专利 7 项，担任《计算机工程与应用》、《计算机工程》、《International Journal of Network Security》期刊的编委，担任 IEEE ASID 2020 会议主席。出版《互联网大数据处理技术与应用》、《Python 爬虫大数据采集与挖掘》两本书。</p>					
办公地址 Office Add	江湾校区交叉二号楼 A6027		办公时间 Office Hour		
联系邮箱 Email Add	zjp@fudan.edu.cn		联系电话 Contact phone	13564317273	
教师签名 Signature	曾剑平		日期 Date	2021/6/3	