

课程名称：大数据分析与应用

一、课程编码：

课内学时：32 学分：2

二、适用学科专业： 计算机专业硕士

三、先修课程： 无

四、教学目标

通过本课程的课堂学习与应用案例，建立科学的大数据观，掌握大数据架构、大数据精准语义搜索、大数据语义分析挖掘、知识图谱等关键技术，熟练使用常用的大数据搜索挖掘与可视化工具，提升大数据的综合应用能力。

五、教学方式

课堂学习、研讨班与应用实践

六、主要内容及学时分配

1. 科学的大数据观 2 学时
 - 1.1. 大数据的定义，科学发展渊源；
 - 1.2. 如何科学看待大数据？
 - 1.3. 如何把握大数据，分别从“知著”、“显微”、“晓义”三个层面阐述科学的大数据观。
2. 大数据技术平台与架构 4 学时
 - 2.1 云计算技术与开源平台搭建
 - 2.2 Hadoop、Spark 等数据架构、计算范式与应用实践
3. 机器学习与常用数据挖掘 4 学时
 - 3.1 常用机器学习算法：Bayes, SVM, 最大熵、深度神经网络等；
 - 3.2 常用数据挖掘技术：关联规则挖掘、分类、聚类、奇异点分析。
4. 大数据语义精准搜索 4 学时
 - 4.1. 通用搜索引擎与大数据垂直业务的矛盾；
 - 4.2. 大数据精准搜索的基本技术：快速增量在线倒排索引、结构化与非机构化数据融合、大数据排序算法、语义关联、自动缓存与优化机制；
 - 4.3. 大数据精准搜索语法：邻近搜索、复合搜索、情感搜索、精准搜索；
 - 4.4. JZSearch 大数据精准搜索应用案例：国家电网、中国邮政搜索、国家标准搜索、维吾尔语搜索、内网文档搜索、舆情搜索；
5. 非结构化大数据语义挖掘 10 学时
 - 5.1. 语义理解基础：ICTCLAS 与汉语分词
 - 5.2. 内容关键语义自动标引与词云自动生成；
 - 5.3. 大数据聚类；
 - 5.4. 大数据分类与信息过滤；
 - 5.5. 大数据去重、自动摘要；
 - 5.6. 情感分析与情绪计算；

- | | | |
|------|----------------------------|------|
| 6. | 知识图谱的大数据自动构建与应用 | 4 学时 |
| 6.1. | 知识图谱概念 | |
| 6.2. | 知识点的自动发现; | |
| 6.3. | 基于 bootstrapping 的知识大数据生成; | |
| 7. | 大数据应用案例剖析与综述答辩 | 8 学时 |
| 7.1. | 国家电网大数据应用案例 | |
| 7.2. | 新媒体传播创新与头条应用; | |

七、考核与成绩评定

平时研讨，期终大作业加综述答辩考核

八、参考书及学生必读参考资料

1. 张华平, 高凯, 黄河燕, 赵燕平. 大数据搜索与挖掘 [M]. 北京: 科学出版社, 2014. 5 (ISBN:978-7-03-040318-6)

九、大纲撰写人: 张华平