



大数据分析与应用课程说明

Intro to Big Data Analysis and Application

张华平 副教授 博士

Email: kevinzhang@bit.edu.cn



<http://www.nlpir.org/>

@ICTCLAS张华平博士

大数据搜索与挖掘实验室 (BDSM@BIT)

2016-9

大数据分析与应用/张华平



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

课程微信群（一周后过期）



大数据分析与应用硕士班2016



该二维码7天内(9月27日前)有效，重新进入将更新

大数据分析与应用/张华平



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



课程体系

➤ 理论篇

- 科学的大数据观 2h;
- 机器学习与数据挖掘方法 2h;

➤ 技术篇

- 大数据云平台架构（Hadoop, Spark架构、计算范式与应用实践） 2h
- JZSearch大数据语义精准搜索引擎（索引、增量、） 4h;
- NLP-IR大数据语义挖掘平台 8h: 语义基础（ICTCLAS汉语分词、新词发现与关键词识别、命名实体识别）、文档分析（文本分类、文本聚类与话题发现、自动摘要）、情感分析、过滤;
- 知识图谱自动生成与应用2h

大数据分析与应用/张华平



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



➤ 应用篇

- 新媒体：传播创新与国家电网头条应用 2h
- 国电：国家电网的大数据应用 2h
- 无人驾驶：吴甘沙（驭势科技CEO，原Intel中国研究院院长），无人驾驶算法与应用 2h
- 公安：公安大数据应用实践 1h；
- 安全：特定群体的自动发现与分析 1h；
- 金融：金融大数据风控 1h
- 文本：大数据平台下的文本算子场景研发 1h
- 日志：网络日志用户画像 1h
- 采集：大数据云采集与智能提取 1h



参考书

➤ 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)

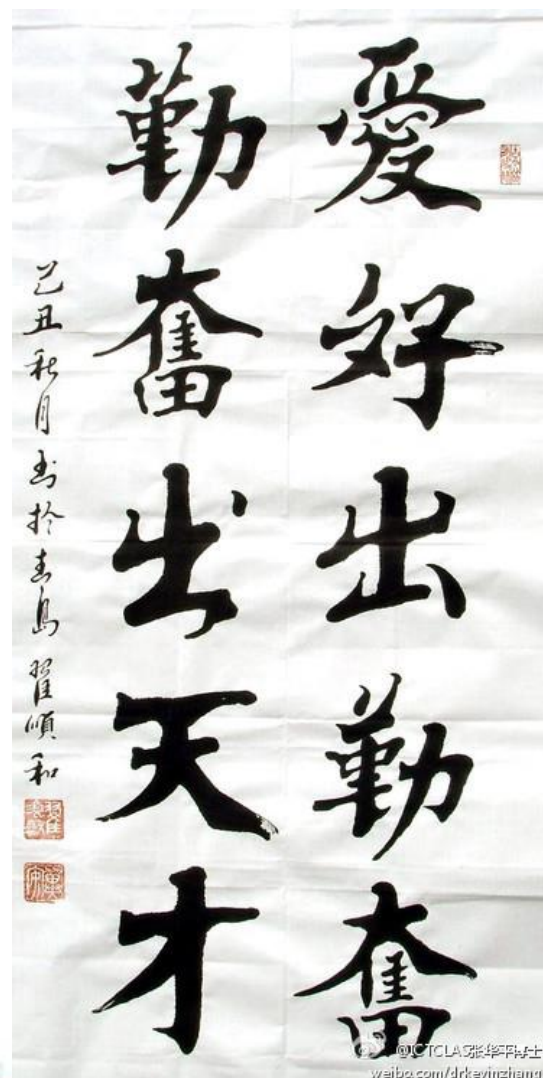


大数据分析与应用/张华平

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

➤ 兴趣第一

- 感兴趣找方法，不感兴趣找借口；
- 教育第一原则是培养对科学或者具体学科的兴趣，扼杀青年的兴趣，罪莫大焉；
- 再好的学问，以面目可憎的形象出现，年轻人也不可能接受。佛家无色无相，却幻化万象，以渡众生。



➤ 知行合一

- 明 王守仁 《传习录》 卷 教育家：陶行知
- 王守仁，号阳明先生，中国明代最著名的思想家、哲学家、文学家和军事家。陆王心学之集大成者，非但精通儒家、佛家、道家，而且能够统军征战，是中国历史上罕见的全能大儒。封“先儒”，奉祀孔庙东庑第58位。
- 计算机科学尤其强调知行合一。

知行合一



结课成绩构成

➤ 平时20分

- 课堂考勤10分
- 互动(平时+期终800字感想)：10分；

➤ 期末综述(提交论文与PPT)：40分

- 考核要点：新、权威、凝炼；

➤ Demo开发（最好与综述匹配，提交代码，说明文档，演示PPT）：40分

- 考核要点：工作质量（有用、有趣、落地）工作量；
- 可以是某项技术Demo，也可以是成熟技术的新应用；使用开源等一切资源，但不能是**简单照搬抄袭（杀无赦）**

选题包括但不限于：自然语言处理（分词、标注、短文本、分类、聚类、话题发现、情感计算、情绪分析）、信息检索、社交网络分析、数据挖掘、用户画像、推荐系统、精准广告、知识图谱





如何拿90+？

- 比例控制在10%;
- 结课前3周提交综述与Demo, 选取前十名进入终极PK;
- PK赛（最后一周）：每人十分钟演讲+演示答辩，老师的打分权重50%，同学微信群投票占50%（每人只能选6位）



感谢关注聆听！



张华平

Email: kevinzhang@bit.edu.cn

微博: @ICTCLAS张华平博士

实验室官网:

<http://www.nlpir.org>



大数据千人会

