**北京北大方正电子有限公司简介**

北京北大方正电子有限公司（以下简称“方正电子”）服务于印刷、传媒、出版、大数据、字库、数字教育等领域，致力于为客户提供领先的信息处理技术、产品、解决方案和增值服务，使最终用户可以随时随地、通过各种终端体会移动互联时代的信息化生活。

作为中国现代印刷与现代传媒技术革命的开创者与领导者，继激光照排、远程传版、彩色出版系统及新闻采编系统之后，方正电子持续推动行业进步与发展：

面向印刷企业数字化升级和面向印刷出版企业互联网转型提供先进的软件平台和方案，现已发展拥有了喷墨印刷、数码印刷、印前CTP三大硬件产品线，成为中国第一家具有自主知识产权的高端数字喷墨技术提供商;

在媒体深度融合的环境下，方正电子通过媒体内容生产、渠道传播、内容分发和运营等领域的实际应用，为传媒行业的深度融合转型之路提供持续动能。

方正字库已经融入现代人生活的方方面面，成为汉字书写文明在信息化时代的传承载体和全新表现形式……

致力信息技术，推动信息传播。方正电子将整合各方优势，立足中国，面向全球，为各行业用户提供先进的技术、实时的产品和专业的服务。

联系方式

地址：北京市海淀区上地信息产业基地五街9号方正大厦 100085

电话：010-82531670、15117993629

E-Mail：[luojuan@founder.com](mailto:luojuan@founder.com)

公司网址：<http://www.founder.com.cn/>

**方正电子与西电联培硕士科研项目介绍**

**一、互联网大数据挖掘-4人**

1、科研题目：互联网大数据挖掘

2、科研背景及重要性：

互联网的创新和应用日新月异，对社会生活的影响广泛深刻。特别是随着移动互联网的发展，信息发布和获取变得愈加便捷，发布方式也更加多样化，越来越多的网民参与网上内容的创造与传播，互联网已成为一个巨大的媒介平台和信息资源池。

如何对互联网大数据进行有效地分析挖掘，及时准确便捷地找到其中有价值的信息与知识是值得深入研究的问题，相关成果可应用于网络舆情监测、媒体内容挖掘、企业决策支持与竞争情报分析、新闻出版选题策划等领域。该项目研究是本企业重要技术战略之一，是本企业互联网大数据业务、媒体大数据业务的重要技术基础。

3、科研具体内容方向：

以互联网大数据为研究对象，结合互联网搜索技术、自然语言处理技术、多媒体信息处理技术、机器学习和深度学习技术，研究大规模互联网数据的分析挖掘方法，研发大规模数据环境下的网络内容挖掘、知识库构建、社会关系分析、信息传播分析、多媒体目标检测、跨媒体检索系统。

a、研究互联网大数据的分析挖掘方法，包括研究适合大规模微博、微信等短文本数据的有效分类、聚类、句法分析、特征提取算法以及噪音数据过滤方法，结合内容分析与人物关系分析，研究移动互联网环境下的话题检测与评价模型、传播过程与传播规律、虚拟人物圈子分析方法等。

b、研究事件检测与追踪、事件要素抽取及关联分析技术，构建大规模互联网事件库，研究内容包括大规模数据环境下的主题检测、要素抽取、演化分析、观点挖掘、关联网络以及事件趋势预测等，解决社会化网络中数据噪音数据大、网络用语表达不规范、话题漂移较快等问题，用于互联网中突发及热点事件的发现、专题分析。

c、研究面向互联网数据的领域知识库构建技术，包括适合开放域的领域本体识别、关系抽取、置信度评估等算法以及知识库标引关键技术，结合业务形成的结构化、半结构化知识数据，研究特定领域知识库的要素抽取、实体链接、语义消歧、知识融合等方法。

d、研究面向互联网图片视频数据的挖掘技术，包括特征提取、文字检测与识别、特定目标检测、实例搜索、大规模多媒体信息索引与检索、跨媒体检索等，并针对特定应用场景进行性能和效果的优化，实现针对互联网中海量图片及视频数据的高效处理。

4、学生毕业时需掌握程度：

熟练掌握互联网大数据挖掘常用算法，如NLP领域的分词、自动分类、自动聚类、话题检测、命名实体识别等，并深入掌握某种算法的具体原理。至少掌握一种编程语言，如Python、C++、Java等。

**二、XML排版系统书籍模板制作方法-2人**

1、科研题目：XML排版系统书籍模板制作方法

2、科研背景及重要性：

XML排版系统是国家数字复合出版工程的重要课题，XML排版系统通过模板与XML的合成完成数字出版物的自动化排版。模板制作工具是该系统的重要组成部分，目前该系统已经在科技期刊方面提供快了速模板制作工具。书籍的版式比科技期刊更加多样化，其模板制作方法与科技期刊也会存在较大的差异。

3、科研具体内容方向：

分析书籍的版式特点，设计书籍模板的快速制作方法，提供书籍快速模板制作工具。

4、学生毕业时需掌握程度：

了解XML排版系统原理，能够利用当前工具制作书籍模板，了解书籍的版式特点并形成总结报告，形成书籍快速模板制作的完整方案，书籍快速模板制作工具达到可使用程度。

**三、机器学习/人工智能在文档结构分析领域的应用-2人**

1、科研题目：机器学习/人工智能在文档结构分析领域的应用

2、科研背景及重要性：

对各教辅、科技等各类型图书的结构化内容进行识别分析，形成标准的xml格式，以便进行多元化的输出应用 ，对图书内容的结构化是数字化应用及发布的基础。

3、科研具体内容方向：

由于图书品类众多、版式丰富，随着图书结构化处理的深入，版式及内容特征会越来越复杂，甚至会出现一些冲突的情况，给图书结构化分析带来一定的困扰。通过机器学习/人工智能等相关技术与图书结构化工作相结合，降低结构化工作的复杂度，提升结构化工作的准确性。

4、学生毕业时需掌握程度：

熟练应用机器学习各类算法和开发语言，并且对图书的结构和docbook标准熟练应用。

**四、工具书数字化在线编纂系统-2人**

1、科研题目：工具书数字化在线编纂系统

2、科研背景及重要性：

辞海、大百科等工具书数字化编纂应用系统，是国家复合出版工程的课题。

3、科研具体内容方向：

针对工具书类的图书在编纂过程的特点，对工具书在线编纂提供在线编辑器、内容审核、条目查询、审核权限流程等一系列功能，其中内容审核通过一系列的算法设计，来对工具书的书证落空、历史纪念等一系列内容进行辅助检查，辅助人工快速发现内容中的知识性错误。

4、学生毕业时需掌握程度：

熟练使用相关开发语言，熟悉工具书内容检查相关算法设计，数据数据库设计和优化

**五、图数据库技术-2人**

1、科研题目：图数据库技术

2、科研背景及重要性：

基于三元组的知识图谱来构建语义网络，从而实现机器对于领域知识的理解和推理，是当下人工智能的重要方向之一。为了支撑海量三元组的存储与查询，图数据库技术至关重要，是知识图谱的基础技术。

3、科研具体内容方向：

研究国内外图数据库的现状及对比，基于gStore图数据库实现海量数据的管理及快速图检索，参与gStore工程化研发，满足企业实际集成应用需求

4、学生毕业时需掌握程度：

了解国内外现有图数据库的技术路线及优缺点；掌握gStore、Neo4j两种图数据库的开发应用能力；能够参与gStore的研发团队，并掌握gStore图数据库技术性能优化策略，掌握搭建高可用分布式图数据库集群的技术；

**六、内容管理和发布-2人**

1、科研题目：内容管理和发布

2、科研背景及重要性：

充分利用移动互联网、云计算、大数据、人工智能技术，为媒体实现业务融合、管理融合、用户融合、数据融合; 全面支撑新闻业务移动化、信息传播移动化，通过打造大数据支撑下的集融媒体报道指挥、新闻策采编发、智能决策管理、经营管理为一体的超融合平台，实现融媒体新闻生产、决策、经营的一体化，并能根据各报社融合发展的不同进程和特点，通过业务流程动态组合、数据用户资源的互通共享，灵活支持报社的管理变革、组织变革和流程再造，让超融合平台成为报社融合转型的安全、高效、可控的技术引擎。

3、科研具体内容方向：

针对媒体融合特点，结合国内外技术和应用场景，研究内容编辑、审核、发布，反馈等方面的技术，主要技术方向包括以下可供选择：1）机器写稿技术，2）智能评论技术，3）可视化模版设计工具研发，4）视频处理和视频存储发布方向。

4、学生毕业时需掌握程度：

熟练使用相关开发语言，熟悉一定的智能算法，熟练掌握前端开发技术，熟悉视频的处理、存储和发布。

**七、自动化测试-3人**

1、科研题目：敏捷开发模式下的自动化测试研究

2、科研背景及重要性：

目前信息时代，需求变化快，交付周期成为企业核心竞争力，轻量级的更能适应变化的敏捷开发模式盛行，在敏捷开发模式下，自动化测试是必不可少的，也是实行敏捷开发的关键环节。

3、科研具体内容方向：

自动化测试主要包括基于UI的自动化测试、接口自动化测试、客户端工具的自动化测试。

a、研究APP产品的UI及接口自动化方案，包括Android及IOS自动化测试；

b、研究出飞翔排版工具的图片效果比对的自动化测试方法，能通过自动化进行图片比对测试，不仅能判断其是否与预期相同，还能根据需求对于一些合理的不同进行过滤。

4、学生毕业时需掌握程度：

至少掌握一种编程语言，如Python、C++、Java等；

在RobotFramwork框架下进行UI自动化测试、接口自动化测试；

实现并掌握图片比对的自动化测试。

**八、数字图像对比-2人**

1、科研题目：数字图像对比

2、科研背景及重要性：

出版印刷领域是方正电子的核心业务之一，自从748工程开始，方正电子一直在此领域占据优势地位。此领域的核心问题是把集图形、图像、文字为一身的页面描述文件光栅化成设备可接受的光栅点阵，然后送到输出设备成像。随着技术的发展，出版社在教材印刷上已经摒弃了通过胶片传版的方式，逐渐改造成CTP出版。此时就带来了一个问题：如何保证不同印点的光栅化结果是“正确的”。

目前的正确性检查是基于人工判断的，费时费力。这一点严重阻碍了“电子胶片”在出版物印刷上的普及。

近年来，模式识别、神经网络、机器学习等新方法在计算机视觉领域取得了令人瞩目的进步，这些基础技术的进步为解决上述问题提供了新的思路和方法。

3、科研具体内容方向：

本课题以不同分辨率下成像的不同光栅图像为研究对象，研制一种高效并且识别率高的方法来替代人工检查，从而提高出版物的印制效率。

4、学生毕业时需掌握程度：

a、具有独立检索科研文献，针对特定技术难点提出符合课题限定条件的方法，并通过程序和实验数据验证方法有效；

b、熟悉和理解图像/图形在该领域的技术应用进展和基础算法，并能够结合已有算法，通过改进以满足本领域应用要求。

|  |  |
| --- | --- |
| **北京北大方正电子有限公司**  **联培项目介绍** | |
| 招生专业 | 计算机、软件、数学、电子信息、通信工程、自动化、模式识别、图形图像处理 |
| 招生人数 | 合计19人 |
| 学制 | 3年 |
| 奖学金 | 学费及奖学金政策按学校政策执行，企业每月提供培养费、实习报酬、安排宿舍 |
| 录取方式 | 初试成绩达到各学院复试线，由学院与企业联合复试确定人选 |
| 学习方式 | 学校课程+企业课程两个阶段，第一学期学生在学校完成校内课程学习，其余时间在联培企业完成企业课程学习、科研实践、开题答辩、中期考核、论文答辩等培养环节 |
| 学籍说明 | 录取学生均为西电正式注册学籍研究生，毕业后颁发西电毕业证与学位证 |
| 就业方式 | 联合培养硕士生接受企业的培养费资助并通过甲方的考核认定条件下，毕业后承诺为甲方服务两年，未能完成服务视为联培学生违约，并将承担违约责任，退还企业已经支付的全部培养费。 |
| 报考要求 | 全国高校优秀计算机相关专业学子 |
| 备注 | 学习及工作地点：北京 |