

批准立项年份	2006
通过验收年份	2008

教育部重点实验室年度报告

(2016 年 1 月—— 2016 年 12 月)

实验室名称：数据工程与知识工程教育部重点实验室

实验室主任：杜小勇

实验室联系人/联系电话：孙水/010-62515228

E-mail 地址：sunshui@ruc.edu.cn

依托单位名称：中国人民大学

依托单位联系人/联系电话：石源/010-62514955

2017 年 3 月 15 日填报

填写说明

一、年度报告中各项指标只统计当年产生的数据，起止时间为1月1日至12月31日。年度报告的表格行数可据实调整，不设附件，请做好相关成果支撑材料的存档工作。年度报告经依托高校考核通过后，于次年3月31日前在实验室网站公开。

二、“研究水平与贡献”栏中，各项统计数据均为本年度由实验室人员在本实验室完成的重大科研成果，以及通过国内外合作研究取得的重要成果。其中：

1. “论文与专著”栏中，成果署名须有实验室。专著指正式出版的学术著作，不包括译著、论文集等。未正式发表的论文、专著不得统计。

2. “奖励”栏中，取奖项排名最靠前的实验室人员，按照其排名计算系数。系数计算方式为： $1/\text{实验室最靠前人员排名}$ 。例如：在某奖项的获奖人员中，排名最靠前的实验室人员为第一完成人，则系数为1；若排名最靠前的为第二完成人，则系数为 $1/2=0.5$ 。实验室在年度内获某项奖励多次的，系数累加计算。部委（省）级奖指部委（省）级对应国家科学技术奖相应系列奖。一个成果若获两级奖励，填报最高级者。未正式批准的奖励不统计。

3. “承担任务研究经费”指本年度内实验室实际到账的研究经费、运行补助费和设备更新费。

4. “发明专利与成果转化”栏中，某些行业批准的具有知识产权意义的国家级证书（如：新医药、新农药、新软件证书等）视同发明专利填报。国内外同内容专利不得重复统计。

5. “标准与规范”指参与制定国家标准、行业/地方标准的数量。

三、“研究队伍建设”栏中：

1. 除特别说明统计年度数据外，均统计相关类型人员总数。固定人员指高等学校聘用的聘期2年以上的全职人员；流动人员指访问学者、博士后研究人员等。

2. “40岁以下”是指截至当年年底，不超过40周岁。

3. “科技人才”和“国际学术机构任职”栏，只统计固定人员。

4. “国际学术机构任职”指在国际学术组织和学术刊物任职情况。

四、“开放与运行管理”栏中：

1. “承办学术会议”包括国际学术会议和国内学术会议。其中，国内学术会议是指由主管部门或全国性一级学会批准的学术会议。

2. “国际合作项目”包括实验室承担的自然科学基金委、科技部、外专局等部门主管的国际科技合作项目，参与的国际重大科技合作计划/工程（如：ITER、CERN等）项目研究，以及双方单位之间正式签订协议书的国际合作项目。

一、简表

实验室名称		数据工程与知识工程教育部重点实验室				
研究方向 (据实增删)		研究方向 1	数据库			
		研究方向 2	互联网知识管理			
		研究方向 3	信息资源管理			
实验室主任	姓名	杜小勇	研究方向	数据管理技术、语义网技术、智能信息检索技术		
	出生日期	1963 年 8 月 22 日	职称	教授	任职时间	1999 年
实验室副主任 (据实增删)	姓名	张斌	研究方向	企业文件档案管理、知识管理、档案学基础理论、信息资源管理、CIO 制度设计		
	出生日期	1967 年 3 月 8 日	职称	教授	任职时间	2008 年
学术委员会主任	姓名	陆汝钤	研究方向	人工智能		
	出生日期	1935 年 2 月 15 日	职称	院士	任职时间	1999 年
研究水平 与贡献	论文与专著	发表论文	SCI	42 篇	EI	57 篇
		科技专著	国内出版	4 部	国外出版	0 部
	奖励	国家自然科学奖	一等奖	0 项	二等奖	0 项
		国家技术发明奖	一等奖	0 项	二等奖	0 项
		国家科学技术进步奖	一等奖	0 项	二等奖	0 项
		省、部级科技奖励	一等奖	1 项	二等奖	1 项
	项目到账总经费	1805.32 万元	纵向经费	1338.72 万元	横向经费	466.6 万元
	发明专利与成果转化	发明专利	申请数	0 项	授权数	0 项
		成果转化	转化数	0 项	转化总经费	0 万元
	标准与规范	国家标准	0 项		行业/地方标准	0 项
研究队伍建设	科技人才	实验室固定人员	67 人	实验室流动人员		1 人
		院士	0 人	千人计划		长期 0 人 短期 0 人

	长江学者	特聘讲座	1人 0人	国家杰出青年基金	1人
	青年长江		0人	国家优秀青年基金	0人
	青年千人计划		1人	其他国家、省部级人才计划	7人
	自然科学基金委创新群体		0个	科技部重点领域创新团队	0个
	国际学术机构任职 (据实增删)	姓名	任职机构或组织		职务
		杜小勇	IEEE 执行委员会		委员
		杜小勇	TCDE 执行委员会		委员
		杜小勇	DASFAA 指导委员会		委员
		文继荣	ACM Transactions on Information Systems		编委
		文继荣	IEEE Transactions on Knowledge and Data Engineering		编委
安小米		国际标准化组织信息与文献委员会档案/文件管理分技术委员会		委员	
安小米		国际标准化组织文档管理应用委员会应用问题分技术委员会		委员	
访问学者	国内	0人	国外	0人	
博士后	本年度进站博士后	0人	本年度出站博士后	0人	
学科发展与人才培养	依托学科 (据实增删)	学科 1	计算机科学与技术	学科 2	图书情报与档案管理
	研究生培养	在读博士生	94人	在读硕士生	210人
	承担本科课程	7217 学时		承担研究生课程	2465 学时
	大专院校教材	1 部			
开放与运行管理	承办学术会议	国际	1 次	国内 (含港澳台)	5 次
	年度新增国际合作项目			1 项	
	实验室面积	3321M ²	实验室网址	http://www.deke.ruc.edu.cn/	
	主管部门年度经费投入	(直属高校不填)万元	依托单位年度经费投入	260 万元	

二、研究水平与贡献

1、主要研究成果与贡献

结合研究方向，简要概述本年度实验室取得的重要研究成果与进展，包括论文和专著、标准和规范、发明专利、仪器研发方法创新、政策咨询、基础性工作等。总结实验室对国家战略需求、地方经济社会发展、行业产业科技创新的贡献，以及产生的社会影响和效益。

1. 2016年，重点实验室在大数据与数据库系统研发方面，在华东师大-中国人民大学-InfoSys数据科学联合实验室的支持下，以广东省省级科技计划重大项目为工作聚焦点，围绕高通量大数据实时日志分析系统研发开展了深入的研究。研究团队针对日志型大数据的高通量装载，实时、多并发的个性化数据分析需求所带来的挑战性问题，设计并研发了ParaFlow日志大数据分析系统。该系统侧重以实体为中心的高并发数据分析处理能力，使用流水线式查询引擎，具备高通量的数据装载能力。初步的实验结果显示，ParaFlow比SparkSQL快至少一个数量级，并且具备非常高的并发查询分析处理能力，其中在日志数据列存储优化方面的研究成果被SIGMOD '17国际会议录用。此外，重点实验室在大数据管理系统基准研发方面逐步形成研发团队，围绕国家重点研发计划高端制造业大数据管理系统基准研发课题，逐步开展了一系列较为深入的调研和研究工作。
2. 互联网知识管理领域。主要面向大数据算法、大数据系统和大数据应用三个方向开展工作。包括(1)提出了一种新的语义稀疏表示方法对图片进行特征表示，利用了图片的标注信息来增强图片特征的代表能力，这些标注信息可以通过算法预测或者用户分享获得。(2)提出了一种从搜索结果中查询分面挖掘的方法。首先从搜索结果中挖掘出大量的以列表形式组织的平行信息，然后对这些信息进行去燥、赋权、聚类和排序，生成最终的查询分面。实验表明该方法可以有效的挖掘出高质量的查询分面。(3)建立了一个更加能够反应用户行为模式的主题模型。其创新点就在于设置不同的话题角色，进而每个用户可以同时拥有多种话题角色，对应着不同的话题兴趣。
3. 信息资源管理领域。获得国家社科重大课题“历史文化村镇数字化保护的理论和应用研究”资助，国家自然科学基金重点课题“我国信息资源产业发展政策及管理研究”225(71133006/G0314)结题并获得优秀，《深圳市政府开放数据行动计划研究报告》和《深圳市政府大数据应用研究报告》被深圳市经济贸易和信息化委员会采纳，所起草的《深圳市促进大数据行动计划(2016-2018)》得到深圳市政府批准发布，《信息资源管理术语及概念体系》提出的领域术语及核心概念构建原则被ISO/TC46/SC11“Information and documentation-Records Management-Vocabulary”草案、ISO/TC268/SC1 Smart Community Infrastructures “Guidelines on Data exchange and sharing for smart community infrastructures”、ISO/TC46 “ISOFDIS 5127:2016. Information and Documentation — Vocabulary”等3个ISO国际标准术语起草工作采用，被IEC智慧城市系统专业委员会术语工作组(IEC SyC on Smart Cities WG1)采用。

2、承担科研任务

2016年，实验室各团队成员新获各类项目39项，其中，国家重点研发计划项目1项，项目经费607万元；国家自然科学基金项目8项，经费541.72万元；国家社会科学基金项目4项，经费195万元；其他合作项目26项，经费490.1万元。

其中，李翠平教授承担国家重大科技专项课题一项（《高端制造大数据管理系统标准规范、评测基准和测试工具研制》），项目经费607万元。陈红教授承担国家自然科学基金重大项目一项（《面向物联网搜索的隐私保护理论和关键技术》），项目经费334.62万元。

截至2016年底（2006年至2016年），实验室先后承担各类科研项目674项。包括国家重大科研项目（核高基项目、863项目、973项目、重点研发计划）20项，国家自然科学基金项目105项，国家社会科学基金项目22项，其他纵向项目135项，横向项目285项，校内项目107项，与众多高校、企业开展了广泛的联合研究工作。实验室累计到账科研经费18734.71万元。其中，国家重大科研项目（核高基项目、863项目、973项目）科研经费4808.75万元，国家自然科学基金项目科研经费4044.18万元，国家社会科学基金项目科研经费669万元，其他纵向项目科研经费2731.84万元，横向项目科研经费4672.14万元，校内项目科研经费1808.8万元。

请选择本年度内主要重点任务填写以下信息：

序号	项目/课题名称	编号	负责人	起止时间	经费(万元)	类别
1	高端制造大数据管理系统标准规范、评测基准和测试工具研制	2016YFB1000702	李翠平	2016.6-2019.6	607	国家重大科技专项
2	面向物联网搜索的隐私保护理论和关键技术	61532021	陈红	2016.1-2020.12	334.62	国家自然科学基金(重大)
3	面向科技型中小企业的数字化服务创新生态系统价值共创机制研究	71571184	余艳	2016.1-2019.12	48.7	国家自然科学基金(面上)
4	噪声环境下的弱监督图像语义分割研究	61573363	卢志武	2016.1-2019.12	66	国家自然科学基金(面上)
5	Web信息检索中搜索结果个性化和多样化算法的融合技术研究	61502501	窦志成	2016.1-2018.12	20	国家自然科学基金(青年)
6	融合多网络社区身份的用户话题兴趣建模研究	61502502	赵鑫	2016.1-2018.12	24	国家自然科学基金(青年)
7	云计算环境下基于图模型的海量RDF数据管理关键技术研究	61502504	卢卫	2016.1-2018.12	26.2	国家自然科学基金(青年)
8	支持摘要搜索的数据库多维动态索引技术	61502503	魏哲巍	2016.1-2018.12	21	国家自然科学基金

	研究					(青年)
--	----	--	--	--	--	------

注：请依次以国家重大科技专项、“973”计划（973）、“863”计划（863）、国家自然科学基金（面上、重点和重大、创新研究群体计划、杰出青年基金、重大科研计划）、国家科技（攻关）、国防重大、国际合作、省部重大科技计划、重大横向合作等为序填写，并在类别栏中注明。只统计项目/课题负责人是实验室人员的任务信息。只填写所牵头负责的项目或课题。若该项目或课题为某项目的子课题或子任务，请在名称后加*号标注。

三、研究队伍建设

1、各研究方向及研究队伍

研究方向	学术带头人	主要骨干
1 数据库	杜小勇	王珊、陈红、李翠平、陆嘉恒
2 互联网知识管理	文继荣	毛基业、左美云、卢志武、窦志成
3 信息资源管理	张斌	冯惠玲、赵国俊、安小米

2.本年度固定人员情况

序号	姓名	类型	性别	最后学位	聘任职务	年龄	在实验室工作年限
1	安小米	研究人员	女	博士	教授	51	10
2	柴云鹏	研究人员	男	博士	副教授	33	7
3	朝乐门	研究人员	男	博士	副教授	37	4
4	陈红	研究人员	女	博士	教授	51	10
5	陈晋川	研究人员	男	博士	副教授	38	7
6	陈文萍	研究人员	女	博士	讲师	40	8
7	陈跃国	研究人员	男	博士	副教授	38	7
8	崔鹏	研究人员	男	博士	副教授	43	3
9	窦志成	研究人员	男	博士	副教授	36	2
10	杜小勇	研究人员	男	博士	教授	53	10
11	范举	研究人员	男	博士	副教授	32	1
12	冯惠玲	研究人员	女	博士	教授	63	10
13	何军	研究人员	男	博士	教授	54	10
14	胡鹤	研究人员	男	博士	副教授	40	8
15	胡婷	管理人员	女	硕士	助理研究员	31	5
16	金琴	研究人员	女	博士	副教授	44	3
17	李翠平	研究人员	女	博士	教授	45	10
18	李德英	研究人员	女	博士	教授	51	10
19	李锡荣	研究人员	男	博士	副教授	33	5
20	梁彬	研究人员	男	博士	副教授	43	10
21	刘家俊	研究人员	男	博士	副教授	32	1

22	刘桃	研究人员	女	博士	副教授	35	8
23	刘耀鸿	技术人员	男	硕士	工程师	30	4
24	刘越男	研究人员	女	博士	教授	42	8
25	卢卫	研究人员	男	博士	副教授	35	2
26	卢志武	研究人员	男	博士	副教授	38	3
27	陆嘉恒	研究人员	男	博士	教授	41	4
28	马林青	研究人员	女	博士	讲师	34	7
29	毛基业	研究人员	男	博士	教授	53	3
30	孟小峰	研究人员	男	博士	教授	52	10
31	牛力	研究人员	男	博士	副教授	34	10
32	钱毅	研究人员	男	博士	副教授	43	10
33	秦波	研究人员	女	博士	讲师	39	7
34	任明	研究人员	女	博士	副教授	36	5
35	石文昌	研究人员	男	博士	教授	52	7
36	孙辉	研究人员	女	博士	讲师	39	6
37	孙水	管理人员	男	硕士	助理研究员	37	3
38	覃飙	研究人员	男	博士	副教授	44	10
39	覃雄派	研究人员	男	博士	讲师	45	10
40	唐跃进	研究人员	男	博士	副教授	58	10
41	王健	研究人员	女	硕士	教授	55	10
42	王璞巍	研究人员	男	博士	讲师	37	7
43	王秋月	研究人员	女	博士	讲师	42	6
44	王珊	研究人员	女	硕士	教授	73	10
45	王永才	研究人员	男	博士	副教授	36	2
46	魏哲巍	研究人员	男	博士	副教授	30	2
47	文继荣	研究人员	男	博士	教授	44	3
48	夏天	研究人员	男	博士	副教授	38	6
49	谢丽	研究人员	女	博士	副教授	50	1
50	徐拥军	研究人员	男	博士	副教授	41	8
51	许洁萍	研究人员	女	博士	副教授	50	10
52	许伟	研究人员	男	博士	副教授	35	5
53	杨刚	研究人员	男	博士	讲师	37	6
54	杨孟辉	研究人员	男	博士	副教授	47	8
55	余艳	研究人员	女	博士	讲师	36	4
56	张斌	研究人员	男	博士	教授	49	10
57	张美芳	研究人员	女	博士	教授	52	10

58	张宁	研究人员	女	博士	副教授	39	2
59	张孝	研究人员	男	博士	副教授	44	10
60	张延松	研究人员	男	博士	副教授	43	8
61	赵国俊	研究人员	男	学士	教授	62	10
62	赵素云	研究人员	女	博士	副教授	37	5
63	赵鑫	研究人员	男	博士	讲师	31	1
64	周春来	研究人员	男	博士	副教授	40	6
65	周烜	研究人员	男	博士	副教授	39	6
66	朱青	研究人员	女	博士	副教授	53	10
67	左美云	研究人员	男	博士	教授	45	5

注：（1）固定人员包括研究人员、技术人员、管理人员三种类型，应为所在高等学校聘用的聘期2年以上的全职人员。（2）“在实验室工作年限”栏中填写实验室工作的聘期。

3、本年度流动人员情况

序号	姓名	类型	性别	年龄	职称	国别	工作单位	在实验室工作期限
1	周晓方	其他	男	54	教授	澳	昆士兰大学	5
2	陈熙霖	其他	男	52	研究员	中	中科院	2
3	商烁	其他	男	31	教授	中	中国石油大学	2

注：（1）流动人员包括“博士后研究人员、访问学者、其他”三种类型，请按照以上三种类型进行人员排序。（2）在“实验室工作期限”在实验室工作的协议起止时间。

四、学科发展与人才培养

1、学科发展

简述实验室所依托学科的年度发展情况，包括科学研究对学科建设的支撑作用，以及推动学科交叉与新兴学科建设的情况。

1. 信息学院组织完成了计算机科学与技术、软件工程、数学学科的教育部第四次学科评估工作以及计算机科学与技术专业工程教育认证预评估工作；组织完成了《信息学院“十三五”教材建设发展规划》。共引进4位海内外人才，包括首次引进1位非华裔外籍全职教授，任中国人民大学“新奥国际杰出教席”教授，同时还引进1位海外人才D岗讲师、1位普通岗副教授和1位普通岗讲师。4位引进人才中3位在海外获得博士学位，1位有海外名校访问经历。
2. 组织召开了“数据科学专业建设研讨会”。参加人员包括重点实验室学位委员会主任陆汝钐院士在内部分成员、111引智基地的海外专家以及中国人民大学统计与大数据研究院的艾春荣教授等。会议重点围绕数据科学这一新专业的设置及人才培养模式等问题进行了研讨。
3. 组织“首届商务数据科学研讨会”。分别就“商务数据在创新创业中

的应用”以及“商务数据科学课程和教学体系”为主题进行了研讨。YeeCall 联合创始人、副总裁阮龙，云众数据创始人总经理曾毅，拉拉公园创始人 CEO 廖卓营，农总行科技局银行技术专家杨唯实处长以及来自人民大学、北京大学、对外经贸大学等高校的计算机、管理学等领域的专家参加了会议研讨。

2、科教融合推动教学发展

简要介绍实验室人员承担依托单位教学任务情况，主要包括开设主讲课程、编写教材、教改项目、教学成果等，以及将本领域前沿研究情况、实验室科研成果转化为教学资源的情况。

1. 王珊教授领衔建设的《数据库系统概论》课程入选教育部第一批“国家级精品资源共享课”。《数据库系统概论》教材历经 30 多年实践，成为我国影响最大的数据库教材，经 5 次改版，累计发行 260 多万册。获得多项奖励，包括国家级优秀教材等。
2. 2016 年，信息学院共有十部教材申请了中国人民大学“十三五”本科规划教材第一批立项。
3. 2016 年，信息学院启动了 MOOC 课程建设工作，其中，1 门课程获得学校立项。今年已有两门 MOOC 课程在中国大学 MOOC 上上线，累计选课人数超过 3 万人。
4. 冯惠玲教授获得中国人民大学“数字记忆”厚重人才成长支持计划资助。2016 年信息资源管理学院共出版教程和专著 8 部，研究成果在信息资源管理术语及概念体系构建、数字信息保存、电子文件管理前沿研究等方面产生了一定国际影响和社会影响，对创新教学内容，促进教学改革，教学研用协同发展方面积累了一定经验。出版教材包括：《数字信息保存》、《Web 程序设计》、《外国档案管理学（第二版）》等。
5. 重点实验室承建的基于大数据的文科专业虚拟仿真国家级实验教学示范中心以大数据实验教学环境的建设为突破口，努力拓展数据资源，不断加强课程建设，充分整合优质资源，初步构建了综合性的虚拟实验教学平台。目前，中心已建立了包括金融网络数据、新闻热点数据等在内的 8 个大型数据及，并组织了专门团队负责数据的持续性更新与维护。

3、人才培养

(1) 人才培养总体情况

简述实验室人才培养的代表性举措和效果，包括跨学科、跨院系的人才交流和培养，与国内、国际科研机构或企业联合培养创新人才等。

1. 2016 年，信息学院“图灵班”及“明德图灵”计划顺利开展。9 月，2016 级“图灵班”顺利选拔 22 名学员并开展培养计划。10 月，“明德图灵”厚重人才成长支持计划顺利启动，并陆续开展参访百度、搜狗等活动。“明德图灵”厚重人才成长支持计划是在学校

支持下，信息学院结合学科特点和专业特色，优化整合现有的教育资源，着力培养学生的跨专业学习能力，为提升人才培养水平进行了积极尝试。该项目以“家国情怀，文理交融”为理念，把国情教育、专业知识教育融入人才培养的全过程，打破单一学科视角的培养模式，综合利用学院学科优势和学校社会科学的特长，全面提升学生的综合素质。

2. 信息资源管理学院“数字记忆”厚重人才成长支持计划（以下简称“数字记忆”计划）招生工作正式启动，并选定 25 名同学作为本次“数字记忆”计划的学员参与未来两年的学习研究活动。“数字记忆”计划以“兼纳文理，传承记忆”为理念，融合人文理念和科学精神，把学生的传统文化、信息资源和信息技术的教育，科研能力、创新意识和实践能力的培养，以及国情教育、道德养成等融入项目全过程。本项目充分利用全国第一个大型区域性数字记忆平台——“北京城市记忆”数字资源库，以及相关的专业教育、学术研究、国际交流及第二课堂成果，通过课程学习、专题讲坛、读书沙龙（兰台读书会）、参观学习、研究实践、课题研究、国际交流等方式，为优秀学生在人文素养、信息技术方面搭建成长成才的有效平台。
3. 通过与 InfoSys 公司开展实质性合作，重点实验室本年度先后派出 6 名研究生到 Infosys 在印度班加罗尔的总部进行实习，实习期每人 2-3 个月。同时，以华东师大-人民大学-InfoSys 联合实验室为平台，大量吸收在校研究生进入科研团队工作。
4. 互联网知识管理团队利用自建的时事探针系统，为会计、经济、新闻、法律、公共管理等专业的科研与教学提供支持，尤其是重点支持了一些跨学科、跨院系的交叉研究项目。商学院有多位博士研究生在平台的支持下，完成了毕业论文。
5. 数据库团队同商学院开展了密切合作，针对财会、企业管理等专业的具体科研需求，以网络数据为主，尝试构建商学通用数据库，目前项目进展顺利，并针对商学院的研究生组织了多次关于数据挖掘与数据整理方面的专题培训。

(2) 研究生代表性成果（列举不超过 3 项）

简述研究生在实验室平台的锻炼中，取得的代表性科研成果，包括高水平论文发表、国际学术会议大会发言、挑战杯获奖、国际竞赛获奖等。

1. 博士生陈峻在陈跃国副教授与杜小勇教授共同指导下以第一作者完成的论文《SEED: A System for Entity Exploration and Debugging in Large-Scale Knowledge Graphs》被数据库顶级会议 ICDE 2016 (32st IEEE International Conference on Data Engineering) 录用。

2. 博士研究生孙舒扬的 How Do Citizen's Experiences and Perspectives on Personal Information Violation Change? (《公民个人信息被侵害经历与观点如何转变》) 被第十一届网络冲突与网络安全国际会议 (11th International Conference on Cyber Warfare and Security,

ICCWS 2016) 录用。孙舒扬参会并做报告。

3. 金琴老师指导的代表队在 ACM Multimedia 2016 “Video to Language” 评测比赛中获得第一名, 在 MedieEval 2016 “Movie Emotion Impact” 评测比赛中获得第一名, 在 NLP 2016 中文微博立场评测比赛获得第一名, 在 Audio-Visual Emotion Challenge (AVEC) 2016 比赛获得第二名, 在 CCPR2016 中文多态情感识别大赛获得第二名。

(3) 研究生参加国际会议情况 (列举 5 项以内)

序号	参加会议形式	学生姓名	硕士/博士	参加会议名称及会议主办方	导师
1	大会发言	丁子涵 王向女	博士	第八届档案教育与研究协会 国际会议	
2	大会发言	加小双 祁天娇	博士	第四届亚太地区档案教育研讨会/ 会/	
3	发表会议论文	范冠艳	博士	第 28 届国际机构信息管理协会 大会/	
4	发表会议论文	胡燕华	硕士	第十七届国际数字政府研讨会 议/复旦大学	
5	其他	刘学雅 李思艺	硕士	2016 年度社会、教育与技术亚 洲会议	

注: 请依次以参加会议形式为大会发言、口头报告、发表会议论文、其他为序分别填报。所有研究生的导师必须是实验室固定研究人员。

五、开放交流与运行管理

1、开放交流

(1) 开放课题设置情况

简述实验室在本年度内设置开放课题概况。
无。

序号	课题名称	经费额度	承担人	职称	承担人单位	课题起止时间

注: 职称一栏, 请在职人员填写职称, 学生填写博士/硕士。

(2) 主办或承办大型学术会议情况

序号	会议名称	主办单位名称	会议主席	召开时间	参加人数	类别
1	2016 中国 CIO 论坛	信息资源管理学院	张斌	5. 29	100	全国性
2	第七届中国电子文件管	信息资源管理学院	张斌	11. 19—11. 20	280	全国性

	理论坛					
3	2016 中国信息资源管理论坛	信息资源管理学院	冯惠玲	11.18	60	全国性
4	第五届大数据分析与管理国际学术研讨会	重点实验室	杜小勇	7.16	180	全国性
5	2016 (首届) 中国隐私保护学术会议	中国保密协会隐私保护专业委员会	孟小峰	11.7	200	全国性

注：请按全球性、地区性、双边性、全国性等类别排序，并在类别栏中注明。

(3) 国内外学术交流与合作情况

请列出实验室在本年度内参加国内外学术交流与合作的概况，包括与国外研究机构共建实验室、承担重大国际合作项目或机构建设、参与国际重大科研计划、在国际重要学术会议做特邀报告的情况。请按国内合作与国际合作分类填写。

1. 左美云教授团队和派择科技有限公司联合成立“派择科技-中国人民大学商业数据科学联合实验室”。派择科技是国内最大的数字营销集团华扬联众的大数据业务全资子公司，致力于提供最好的程序化广告产品及全网大数据营销服务。联合实验室的研究项目按照长期研究和短期研究项目相结合的方式设立。其中，长期研究主要资助对行业有示范性的研究成果，短期研究主要资助与派择当年重点发展业务相关的研究成果，由双方根据具体研究议定当年项目经费。同时，人民大学将派择科技确定为实习实训基地。
2. 中国人民大学—华为数据科学联合实验室新一批合作项目立项工作顺利开展。双方将进一步围绕软硬件结合的数据库优化技术展开合作研究，尤其是在 GPU 查询优化、SSD 缓存优化和基于 NVM 的事务处理优化等方向上。
3. 重点实验室同金仓信息技术股份有限公司针对 NVM 这种新型硬件在对传统的数据库系统进行改造和优化方面持续开展合作。
4. 人大—国双大数据科学联合实验室继续围绕国家网络安全和信息化需求，依托大数据技术，充分发挥双方优势，共同开展战略性、前瞻性研究，为政府和社会机构开展基于大数据的挖掘服务，及为决策数据可视化应用提供支撑。
5. 华东师范大学—中国人民大学—印孚瑟斯数据科学联合实验室成立于 2015 年，三方主要在数据科学和工程领域广泛开展相关理论研究以及软件系统和商业解决方案的研究与开发工作，并积极推动研究人员的交流与合作。2016 年，联合实验室的各项工作顺利进行，印孚瑟斯首席执行官 Vishal Sikka 专门访问重点实验室，并就相关的科研进展情况进行了调研。
6. 2016 年 5 月 30 日至 6 月 1 日，冯惠玲教授应邀在“2016 年韩国广泛无纸化会议(U-Paper less Korea Conference 2016)”上做了题为《中国电子文件管理组合策略》的特别报告。
7. 2016 年 11 月 9 日-10 日，谢丽教授受邀参加第 28 届国际机构信息管理协会大会(International Business Information Management

Association Conference), 就论文” The Concept of Records Disposition and its Application to Information System: the case of the United States” 作特别报告。

(4) 科学传播

简述实验室本年度在科学传播方面的举措和效果。

1. 包括杜小勇教授、文继荣教授在内的多位教师参加中国人民大学 2016 年中央和国家司局级干部研修班、教育部机关及直属单位干部专题研修“互联网、舆情分析与大数据应用”专题班的教学工作, 共计 300 余名干部参加学习。
2. 杜小勇教授参与中纪委网络公开课《大数据与治国理政》的录制工作。
3. 张斌教授团队提交的《基于领导决策的档案知识库建设研究的报告》获国家档案局长李明华批示。
4. 张斌教授、安小米教授和徐拥军副教授完成的《全国企业档案事业发展“十三五”规划》通过专家验收, 被国家档案局采用为行业发展指南。
5. 安小米研究团队 2016 年完成的《深圳市政府开放数据行动计划研究报告》和《深圳市政府大数据应用研究报告》被深圳市经济贸易和信息化委员会采纳, 所起草的《深圳市促进大数据行动计划(2016-2018)》得到深圳市政府批准发布。
6. 安小米教授团队提交并发表的“云治理时代的政务数据管理转型——当前我国档案事业发展的的问题与建议[J]. 人民论坛·学术前沿, 2015, (16):72-84. 人大复印资源档案学, 2015 (06): 3-10, 全文转载, 被国家局用纳, 所提建议被写入《全国档案事业发展“十三五”规划纲要》。
7. 与中国计算机学会企工委联合组织中国计算机学会微课题沙龙活动。包括启明星辰集团、太一星晨、天融信、Infosys China、华为、人大金仓、英特尔中国研究院、腾讯、卫士通、联索科技、山石网科、北方计算机研究所、石化盈科、公安三所、360、汤森路透、拓尔思等在内的知名企业技术专家参与了交流活动, 并对 20 个优秀系统展开评比。
8. 组织萨师煊大数据研究中心 2016 年暑期研究夏令营。本校研究生和青年教师 50 余人, 以及 3 名国外著名高校的研究生、8 名国内其它院校的本科生和研究生、2 名香港高校的研究生参加夏令营活动。科研夏令营从 7 月 4 日开始, 于 7 月 29 日结束, 每周定期由来访海外教授举行

学术报告，其他时间分组研究和讨论。

9. 举办第二届学术节“大数据之人大力量”。Infosys 首席运营官刘宏、IBM 中国工商银行首席软件架构师李纪华、中国人寿研发中心开发三部副总经理张新宇、阿里巴巴数据库高级专家张文亮、人大金仓大数据业务部总监白芸、蚂蚁金服生物及安全智能部总监陈继东、华为数据库首席架构师文继军、滴滴出行大数据决策支持组负责人李佩和百度基础检索架构负责人范彪以及 UIUC（美国加州大学香槟分校）张振杰博士作为特邀嘉宾，进行了专题报告。

2、运行管理

(1) 学术委员会成员

序号	姓名	性别	职称	年龄	所在单位	是否外籍
1	陆汝钤	男	教授	81	中科院数学所	否
2	王珊	女	教授	73	中国人民大学	否
3	于戈	男	教授	55	东北大学	否
4	冯惠玲	女	教授	64	中国人民大学	否
5	王晓阳	男	教授	57	复旦大学	是
6	李建中	男	教授	67	哈尔滨大学	否
7	应明生	男	教授	53	清华大学	否
8	刘梦赤	男	教授	55	武汉大学	是
9	杜小勇	男	教授	54	中国人民大学	否
10	周立柱	男	教授	70	清华大学	否
11	周志华	男	教授	44	南京大学	否
12	周晓方	男	教授	54	昆士兰大学	是
13	周傲英	男	教授	52	华东师范大学	否
14	赵国俊	男	教授	63	中国人民大学	否
15	梅宏	男	教授	54	上海交通大学	否

(2) 学术委员会工作情况

请简要介绍本年度召开的学术委员会情况，包括召开时间、地点、出席人员、缺席人员，以及会议纪要。

2016年7月15日，重点实验室于中国人民大学组织召开了2016年度学术委员会全体会议。学术委员会全体成员（除应明生、刘梦赤、梅宏三位教授请假）、重点实验室111引智基地的海外专家代表、中国人民大学查显友副校长以及学校财务、理工科、人事、科研、设备、资产、规划等职能部门的负责人参加了会议。

会议的主要议题包括三项。一是审议杜小勇主任就重点实验室2015年度的工作情况所做的工作报告；二是审议新加入重点实验室工作的年轻教师的学术报告；三是讨论重点实验室下一年度的工作规划及工作重点。

杜小勇教授在报告中指出，2015年，重点实验室在很多领域取得了重要成绩，包括：（1）再次引进副教授3人，研究人员中具有副高级及以上职称人员的比例提升至82%。（2）获得北京市教学名师1人。（3）863重点项目顺利结项，获得基金重点项目1项。（4）获得ICDE最佳论文1篇，CCF科学技术奖一等奖一项。（5）“大数据管理与分析方法研究实验室”被正式认定为北京市重点实验室等。杜小勇教授还重点就实验室负责建设的数据工程与知识工程创新引智基地5年来的工作情况进行了回顾：（1）随着科研团队的不断壮大，重点实验室承接科研项目尤其是重大科研项目的能力逐年提升。尤其是2012年立项的国家863计划项目《开放环境下海量web数据提取、集成、分析和管理系统平台与应用》，是第一次由中国人民大学牵头承担的国家重大科研项目。（2）2012年至2015年，重点实验室研究人员共发表EI/SCI收录论文156篇次。其中，发表CCF A类论文34篇，高水平论文发表能力得到大幅度提升。（3）越来越多的科技企业同重点实验室建立了合作关系，通过设立横向科研项目、建立联合实验室等形式，在高性能数据库、数据库一体机、互联网大数据管理、数据科学和工程等领域开展了广泛、深入的校企合作，在推进重点实验室科研成果转化的同时，进一步提升了重点实验室的科研实力。杜小勇教授最后提出，重点实验室需要进一步加强科研团队的建设工作，探索人员引进与流动的合力机制，同时应进一步整合资源，在目前工作的基础上，进一步尝试申报国家级重要科技奖项。

随后，新加入实验室工作的范举副教授、王永才副教授、刘家俊副教授分别向学术委员会汇报了自己的学术研究相关情况。

最后，参会人员重点围绕实验室发展过程中所面临的问题与困难进行了讨论。

在讨论中，学术委员会主任陆汝钤院士着重指出，重点实验室已经具备了较好的建设基础，但是需要再进一步凝练学术研究方向，集中力量，重点突破，尤其是，在学科交叉、文理结合方面，要更好的利用中国人民大学在数据库领域的传统优势与人文社会科学领域的深厚基础，争取早日在国家级奖项申报方面取得进展。

中国人民大学查显友副校长在发言中强调，作为中国人民大学的唯一的教育部重点实验室，数据工程与知识工程教育部重点实验室的建设与发展一贯受到学校的高度重视，学校也尽可能的从各个方面基于大力的支持，包括设置了独立机构、单独审批预算、单独核定人员编制等，学校肯定近十年来重点实验室所取得的成绩，也希望重点实验室能够在学术委员会的指导下，取得更多的成果。

很多专家还结合如何更好的将创新引智基地的工作与实验室的发展结合

起来谈了自己的意见，包括（1）要更好地建立引进海外智力的长效机制，更深入的推动引智基地的各项工作。（2）进一步扩大人员交流与合作的范围，丰富合作模式等问题。

（3）主管部门和依托单位支持情况

简述主管部门和依托单位本年度为实验室提供实验室建设和基本运行经费、相对集中的科研场所和仪器设备等条件保障的情况，在学科建设、人才引进、团队建设、研究生培养指标、自主选题研究等方面给予优先支持的情况。

2016年，中国人民大学先后向重点实验室下达专用的运行经费80万元，以及111创新引智基地的配套工作经费90万元，双一流建设经费90万元以及并将重点实验室办公、实验区域的环境改善工程列入新一年的修购专项计划，组织进行了前期的方案论证工作。

为重点实验室所属的“人大行云平台”专门改善了用电环境，进一步提高了服务器机房的用电载荷。

3、仪器设备

重点实验室的“人大行云”云平台，经过去年的硬件升级改造后，进一步提升了性能和硬件资源的利用率，可运行在云平台上的虚拟机的数量也成倍增加。2016年“人大行云”平台为近3000名本科生提供每周5天的远程虚拟实验室开放，人均每周上机时间4.5小时。以“人大行云”平台为依托，支持学生参加国内外各类学科竞赛，参加各类竞赛的学生人数每年不少于400人次，在ACM/ICPC（国际大学生程序设计竞赛）、中国大学生计算机设计大赛、美国大学生数学建模竞赛等中，都取得了优异成绩。重点实验室还基于云平台开展本科学生科研训练营计划，设立若干学生课题，使学生科技创新能力明显提高。

重点实验室还运用“云存储”技术对我国现存的互联网资源进行结构性的保存，开发、定制实验教学资源数据库，建设学科比较齐全、可进行深度挖掘的数据资源平台，为文科专业虚拟仿真实验教学中心专门构建了海量的数据教学资源环境。目前已经建立了一个包括金融网络数据、新闻热点数据、农产品价格数据等在内8个大型网络数据集，并组织了专门的团队，负责对这些数据进行持续性的更新与维护。

“人大行云”云平台进一步加强对外开放共享，除墨尔本大学、新南威尔士大学、北京大学外，还新增了昆士兰大学、华东师范大学等高校使用“人大行云”云平台，联合开展科研项目。

六、审核意见

1、实验室负责人意见

实验室承诺所填内容属实，数据准确可靠。

数据审核人：
实验室主任：
(单位公章)
年 月 日

2、依托高校意见

依托单位年度考核意见：

(需明确是否通过本年度考核，并提及下一步对实验室的支持。)

依托单位负责人签字：
(单位公章)
年 月 日