



大数据应用实践

Big Data Application



张华平 博士 副教授

大数据搜索与挖掘实验室

kevinzhang@bit.edu.cn

@ICTCLAS张华平博士

2016.12



- 不务正业大乱炖（政治报告、营销、活动、撤侨、文本自动生成）
- 某大厦电力数据挖掘
- 95598 客服挖掘
- 国家电网语义智能搜索
- 国家电网头条
- 公安警情大数据挖掘





大数据挖掘：政治报告解读

2011年两会报告词频统计：发展182;建设95;加强85;经济84;社会82;改革70;推进70;提

高68;加快64;保障57;完善52;工作42;民生41;农村41;推进41;制度40;积累40;实

施40;坚持39;管理39;机制39;政策38;全面36;政府36;基本33;服务32;体制32;投资31;能力

31;创新31;文化31;城镇30;市场29;国家29;基础29

3月6日17:41 来自新浪微博 转发(15) | 收藏 | 评论(8)

2012年两会报告词频统计：发展150;经济88;加强81;建设70;改革69;社会66;提高60;推

进59;政策46;保障44;服务42;实施40;促进39;工作39;管理38;企业36;农村35;加快35;扩

大34;完善33;政府33;积累33;稳定31;文化30;农业29;制度29;水平28;教育27;关系26;体

系25;就业25;城镇25;增长25;产业25;地区24;投资24;

3月6日17:30 来自新浪微博 转发(1) | 收藏 | 评论(4)

2013年两会报告关键词：食品药品安全国际

金融危机城镇化企业退休人员房地产市场调

控战略性新兴产业基础设施现代化建设区域

发展扩大内需收入分配财政赤字改善民生保

障性住房和平发展养老保险消费价格涨幅人

民制度就业保障



大数据挖掘：政治报告解读

➔ 美国2011.01.25奥巴马总统国情咨文演说

我们87; 我20; 美国18; 人14; 人民12; 国家11;
努力10·新10·协议8·能够7·阿富汗7·开始





➔ 2008.5.20 台湾地区

我们55; 台湾47; 民主2
球11; 社会11; 中华民国
8; 未来7; 历史7; 经济7
宪政5; 国家5; 价值5;

全
平
5;

➔ 2012 马英九就职演说

我们80; 台湾42; 两岸21; 国际19; 经济19; 文化18;
产业16; 未来14; 社会13; 中华民国12; 发展12; 民
主11; 人民11; 国家10; 自由10; 成长10; 竞争力9;
和平9; 政府8; 人才8; 关系8; 人8; 创造8; 开放8;
安全7; 我国7; 推动7; 环境7; 机会7; 司法7; 支柱7;
幸福7; 合作7; 大陆6; 宪法6; 全球6; 政策6; 世界6;
大陆6; 民主6; 与政府6; 群策群力6; 投资6; 民众6; 价值6;

清华论坛
BBS.HSW.CN



大数据挖掘：解读十八届三中全会

NLPIR汉语分词系统 (又名：ICTCLAS2013版) 张华平博士出品,新增新词发现、关...

NLPIR分词



分词



用户词典



关键词提取



指纹提取

相关介绍

授权发布：中共中央关于全面深化改革若干重大问题的决定 新华社北京11月15日
日电 中共中央关于全面深化改革若干重大问题的决定（2013年11月12日中国共产党第十八届中央委员会第三次全体会议通过） 为贯彻落实党的十八大关于全面深化改革的战略部署，十八届中央委员会第三次全体会议研究了全面深化改革的若干重大问题，作出如下决定。 一、全面深化改革的重大意义和指导思想
（1）改革开放是党在新的时代条件下带领全国各族人民进行的新的伟大革命，是当代中国最鲜明的特色。党的十一届三中全会召开三十五年来，我们党以巨大的政治勇气，锐意推进经济体制、政治体制、文化体制、社会体制、生态文明

打开文件

分析

清空

关键词	权重
中国特色社会主义	n_new/15.77
社会组织	n_new/11.81
领导干部	n_new/11.38
事业单位	n_new/10.19
行政执法	n_new/9.81
非公有制经济	n_new/9.34
转移支付	n_new/9.30
资源配置	n_new/8.79
混合所有制经济	n_new/8.74
宏观调控	n_new/8.58
高校毕业生	n_new/8.21
自然资源资产	n_new/8.05
社会主义市场经济体制	n_new/7.91
税收优惠政策	n_new/7.91
高校毕业生就业	n_new/7.91
生态文明制度	n_new/7.91
一般性转移支付	n_new/7.91
收入分配	n_new/7.73

大数据挖掘：解读十八届三中全会

➤ 关键语义：

- 中国特色社会主义 15.77; 社会组织 11.81; 领导干部 11.38; 事业单位 10.19; 行政执法 9.81; 非公有制经济 9.34; 转移支付 9.3; 资源配置 8.79; 混合所有制经济 8.74; 宏观调控 8.58; 高校毕业生 8.21; 自然资源资产 8.05; 社会主义市场经济体制 7.91; 税收优惠政策 7.91; 高校毕业生就业 7.91; 生态文明制度 7.91; 一般性转移支付 7.91; 收入分配 7.73; 中国特色 7.69; 保险制度 7.59; 基础设施 7.5; 所有制经济 7.32; 发展 7.25; 社会 7.22; 企事业单位 7.06; 管理 6.84; 文化 6.82; 生态文明 6.79; 社会主义市场经济 6.73; 优惠政策 6.73; 承包经营权 6.66; 特许经营 6.66; 经济 6.62; 政府 6.61; 制度 6.6; 推进 6.58; 市场 6.57; 创新 6.49;



大数据挖掘：甄嬛体？汪峰体？

有一位网友统计了汪老师在大陆发行的9张专辑共117首歌曲的歌词

同一词语在一首歌出现只算一次。形容词，名次和动词的前十名分别是（词语后面的数字是出现的次数）：

	形容词		名词		动词
0	孤独：34	0	生命：50	0	爱：54
1	自由：17	1	路：37	1	碎：37
2	迷惘：16	2	夜：29	2	哭：35
3	坚强：13	3	天空：24	3	死：27
4	绝望：8	4	孩子：23	4	飞：26
5	青春：7	5	雨：21	5	梦想：14
6	迷茫：6	6	石头：9	6	祈祷：10
7	光明：6	7	鸟：9	7	离去：10
9	理想：6	8	瞬间：8	8	再见：9
9	荒谬：5	9	桥：5	9	埋：6

再来一个，比如某人的生日19820307：自由，桥，再见，迷惘，生命，死，孤独，鸟

润色一下：

站在通向自由的桥上

再见了，迷惘的生命

犹如死亡般的孤独

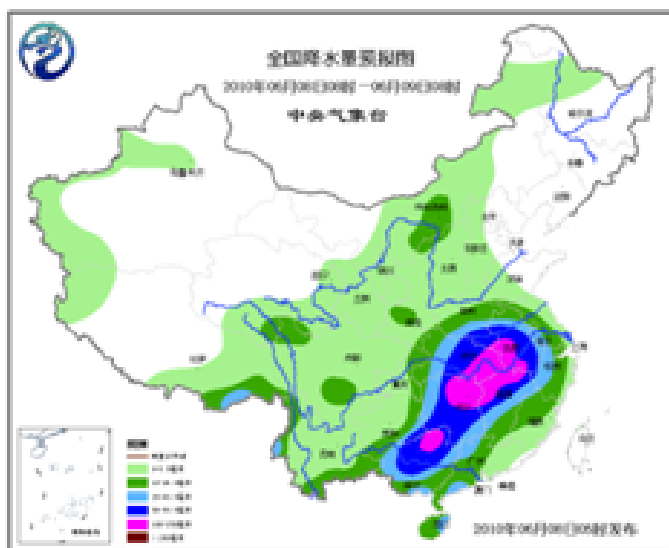
将不在桎梏这只小鸟





未来三天，青藏高原中东部及其以东大部地区自西向东、自北向南将有一次明显降雨过程，其中，湖北大部、河南南部、安徽、江苏西南部、浙江西部、福建大部、江西、湖南、广西、广东中西部、海南等地有大到暴雨，局地有大暴雨。另外，西藏东南部、云南西北部和东南部的部分地区有大雨。上述部分地区并伴有短时雷雨大风或冰雹等强对流天气。

具体预报如下：



8日08时至9日08时，新疆西部、西藏东部和西部的部分地区有小到中雨（雪）或雨夹雪，西北地区中东部、内蒙古中西部和东北部、黑龙江西部、山西、河北西部、黄淮大部、江淮、江汉、江南、华南大部以及西南地

风或冰雹等强对流天气；

区等地有小到中雨或阵雨，其中，西藏东南部、云南西北部、河南南部、湖北大部、安徽大部、江苏西南部、浙江西部、江西中北部、湖南大部、贵州东南部、广西大部、广东西部、海南东部等地有大到暴雨，其中，湖南东北部、江西北部、湖北东部、安徽西南部、广西东北部等地的部分地区有大暴雨；上述部分地区还将伴有短时雷雨大风或冰雹等强对流天气。新疆西部和北部、甘肃西部、宁夏、内蒙古中西部等地有4~6级风，北部还有5~6级偏南风。

风 = [地点]{将}有[级数]
[地点]{将}有[级数]
[地点]有[级数]
[地点]{将}有[级数]
[地点]风力[级数]

雨雪 = [地点]{将}有

气温 = {时间}[地点]
{时间}[地点]
{时间}[地点]
{时间}[地点]

能见度 = {时间}[地点]
{时间}[地点]
{时间}[地点]
{时间}[地点]

时间 = 今天白天到夜
今天下午到夜
今天白天，
今天夜间，
今天早晨，
今天早晨到上
[DAY]日开始，
[DAY]日起，
[DAY]日早晨和

大数据：某大型保险客服背后的规律

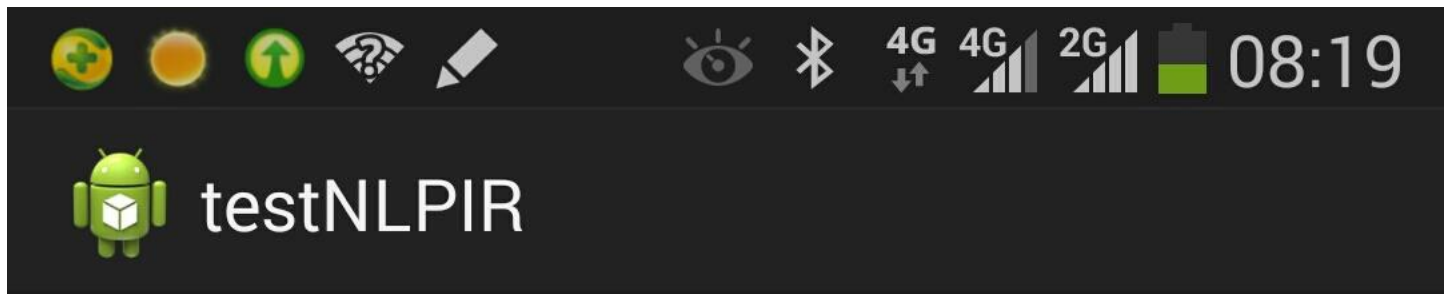
某保险一天客服文本大数据挖掘

科室	部位	病情
住院 1553	骨 376	意外 262
外科 244	手 275	摔倒 130
儿科 220	头 146	骨折 124
骨科 125	腿 108	身故 100
内科 121	瘤 105	肺炎 82
妇科 57	肺 105	发烧 75
手术 56	脚 101	疾病 74
耳鼻喉科 40	腰 91	摔伤 66
五官科 18	脑 77	癌 63
	气管 70	肿瘤 55
	支气管 67	感冒 54
	左手 67	神经 54
	右手 59	急性 46
	尿 58	气管炎 46
	肠 40	支气管炎
	肾 39	狗咬伤 40
	肝 36	扭伤 38
	左腿 32	感染 38
	面 29	呼吸 37
	子宫 27	砸伤 29
	胃 27	囊肿 28
	乳腺 26	结石 27

营销服务部 客户经理 刘先生 合理理赔 女行营业部 扣费 处理结果 泰康放心理财 在三重赔付服务 领取本金加 退保金 交费 代理机

华为大数据挖掘

- 作为华为大数据搜索与挖掘的战略合作伙伴，全面提供采集、搜索与挖掘技术组件产品，已经应用于华为手机。



分词结果为：我/rr 请/v 你/rr 吃/v 一菲烤肉/cater , /
wd 一藏寿司吧/cate 。 /wj 张华平/nr 博士/n 问候/vn
你/rr , /wd NLPIR/ICTCLAS/xu 这个/rz 是/vshi 测试/
v 而已/y ! /wt

大数据助力“利比亚撤侨”

灵玖论坛 > SOS紧急救援 > 804人待救援

主题管理

回复

发帖

返回列表



版主



IP 编辑此人 禁止此人

Elint 发表于 2011-2-25 11:03 | 只看该作者

打印 字体大小: T T 1 #

804人待救援

@乞力马扎罗的山和雪: 说: 求大家帮我转发, 我现在在利比亚两个动乱城市的中间的一个叫米苏拉塔军事重镇往南45公里左右的叫塔瓦嘎的小镇上, 我们若前往港口必须经过被禁严的米苏拉塔, 我们无法通过, 求祖国救我们! 我们截至现在未得到任何撤离或被撤离的消息, 这边连工人共有c! 求大家帮我转发! 原文转发(106)|原文评论(42) 2.25 6: 04



SIGNATURE

ELINT网络舆情实时监测利比亚国内外消息

回复 引用 编辑

评分 使用道具 管理



版主



IP 编辑此人 禁止此人

Elint 发表于 2011-2-25 16:30 | 只看该作者

3 #

亮行者: @SaigonLacoste @ICTCLAS张华平博士 //@亮行者:外交部领事中心已经确认收到“乞力马扎罗的山和雪”的求救信息! 如果有当事人的进一步联系方式请提供。前方撤离工作组会尽快核实情况, 组织撤离工作。

今天 14:26 来自新浪微博转发 | 收藏 | 评论(1)

SIGNATURE

ELINT网络舆情实时监测利比亚国内外消息

回复 引用 编辑

评分 使用道具 TOP 管理



版主

Elint 发表于 2011-2-27 12:54 | 只看该作者

4 #

@乞力马扎罗的山和雪: 目前804人已经安全到米苏拉塔港口, 等待上船

今天 04:37 来自S60客户端转发(11) | 收藏 | 评论(8)

联合国紧急救援中心牛皮记

心灵科学是研究心灵的科学现象，以及心灵与自然、社会、宇宙关系的新兴学科。我们认为心是生命的阳性系统，灵是生命的阴性系统。心灵科学研究生命的阴阳系统运动、发展、转化、平衡、升级等规律。心灵科学研究的是网米状态下的科学现象，网米

www.56.com/u98/v_MTIyNzY0Mzk5.html

56.com

导航

吃烧烤被烧

乐搜

登录 | 注册



56.com

新闻报道

2014.8.15 世界人民一届三次会议中华龙民族魂高峰论坛暨联合国紧急救援中心新闻发布会

北京

00:22 / 07:26

相关视频



筷子兄弟 小苹果 中国神曲走向世界走向二次元 筷子兄弟小苹果mv

67.2万



神曲《独角戏》盘点慈善大咖陈光标 歪歌调侃联合国授予的世界首善

41.0万



筷子兄弟 小苹果 中国神曲走向世界走向二次元 高清观看

78.4万



筷子兄弟-小苹果 中国神曲走向世界走向二次元 高清观看

19.2万



【江西】老人被困景区“一线天” 紧急救援

18.4万



迈入二次元世界的少女

68,662

首页 > 原创 > 世界人民一届二次会议中华龙民族魂高峰论坛暨联合国紧急救援中心

249次播放

热门文章

[图文]天津常务副市长杨栋...
致CCTV律师函
[图文]中国矿山紧急飞行救...

相关文章

没有相关文章

- 中国矿山安全(内参)公告
- 国家工商总局批准中国矿山安全万里行组...
- 中国三农事业发展中心回函万里行组委会
- 刘晓总担任联合国紧急救援中心基金会矿...
- 国家宗教事务局将坚永副局长会见刘晓总...

學
LOGY



联合国!

url=VxfIC_Ye0aveNmiMEDaJRgfWdmm8R_odhIsbWlftTOa--gIFxPQyCmCSwjka0SNL3eiYP4

廖锡强 搜索答案

ISO9001质

新型环保 石灰窑 专利

ISO9001质量体系认证企业

新型环保 石灰窑 专利产品

电话 40 石家庄 科技股

? 谁认识陆民一

2008yinxiang | 分类: 个人理财 | 浏览990次

有没有熟悉《纪元国际石油化工有限公司》的呢? 他们有做大宗贸易的能力
有没有对纪元国际石油化工有限公司董事长陆民一先生的呢?
不要网页上抄录的答案, 希望能得到真实的人的回答。

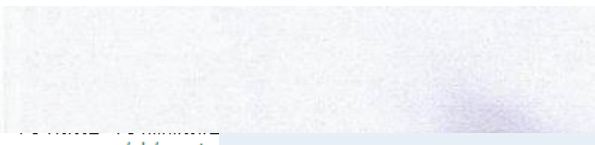
我来帮他解答

? 陆民一是谁?

AAAAABBBBBZZZZ | 分类: 电视 | 浏览876次

陆民一, 是一个只会吹牛的大骗子, 做不成什么事, 吹牛第一, 喜欢做女工

我来帮他解答



沐醒风: 揭穿“联合国”
注册地是香港, 所以国
链接<http://t.cn/8svUek>

南宁深圳等地都有, 很多国际贸易公司就是打着这个幌子在行骗

? 有谁知道广州市平驰国际贸易有限公司? 想了解一下

136****6050 | 来自手机知道 | 分类: 贸易 | 浏览72次 2014-03-08 16:57

分享到:

提问者采纳 举报 | 2014-03-08 18:31

热心网友
骗子公司, 老板叫廖锡强, 夫妻档
广东省广州市中级人民法院
刑事裁定书
(2008)穗中法刑二终字第114号
原公诉机关广东省广州市越秀区人民检察院。
上诉人(原审被告)廖锡强。因本案于2007年2月24日被羁押并刑事拘留, 同年3月27日被逮捕。现被羁押于广州市越秀区看守所(槎头)。
辩护人范奕兴, 广东煜日律师事务所律师。
广东省广州市越秀区人民法院审理广东省广州市越秀区人民检察院指控被告人廖锡强犯合同诈骗罪一案, 于2007年12月20日作出(2007)越法刑初字第947号刑事判决。宣判后, 原审被告人廖锡强不服, 提出上诉。本院受理后, 依法组成合议庭, 经过阅卷和审核证据, 听取辩护人意见, 认为本案事实清楚, 决定不开庭审理。现已审理终结。
原判认定: 2000年5月29日, 被告人廖锡强以广州市寺右新马路126-130号广信豪景大厦10楼h室为住所

本裁定为终审裁定。
审判长: 黄理 审判员: 平文林 代理审判员: 李晓刚 二〇〇八年三月廿八日 书记员: 胡忠民

提问者评价: 非常感谢! 有家里人差点上当, 准备报警

评论 | 1 | 0

大数据千人会是怎样炼成的？



人民网 people.cn

12 数字青年

“爸爸去哪儿”背后的

12版 数字青年

新闻排行榜

火影忍者 mp3

火影忍者 mp3 火影忍者 mp3

社会媒体改变世界: SMP大会的见证

2014-11-04 杨静 杨静lillian



2014年11月1-2日，我在世纪金源大饭店参加了以“社会媒体改变世界”为主题的第三届全国社会媒体处理大会（英文名称为 National Conference of Social Media Processing），业界专家将其简称为“SMP 大会”。大会由中国中文信息学会社会媒体处理专业委员会主办，北京理工大学承办，旨在构建社会媒体处理领域的产学研生态圈，成为中国乃至世界社会媒体处理的风向标。

本以为大会特邀报告会以社会媒体的数据挖掘和自然语言处理技术为主题，但周末听完两天的会，发现学术

打印 网报

人民网
三届全国社
承办，旨在
风向标。

大会联
表的新型社
社会媒体在
方式，正在

大数据千人会是这样炼成的！

➤ 大数据时代公益行销的启示；

- **社交媒体**一个半小时内构建大数据千人会社区，承载了每周一期的千人远程教育与交流
- **移动互联网**提供了无时无地的泛在学习终端
- **开放众筹**实现了零成本下的最优教育资源与体验；
- **大数据**构建了线下组织与O2O活动



- 不务正业大乱炖（政治报告、营销、活动、撤侨、文本自动生成）
- 某大厦电力数据挖掘
- 95598客服挖掘
- 国家电网语义智能搜索
- 国家电网头条
- 公安警情大数据挖掘



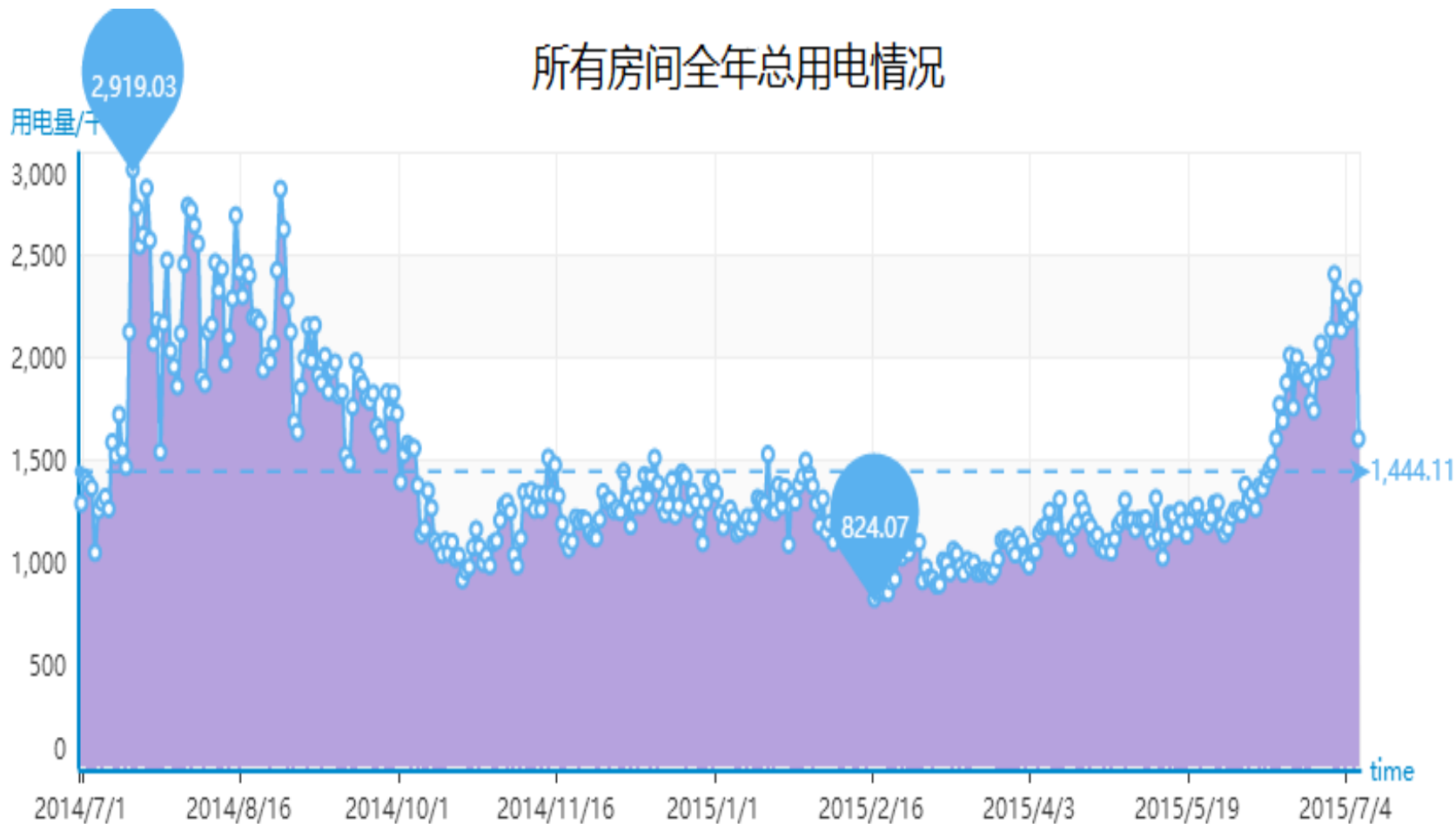
研究数据介绍

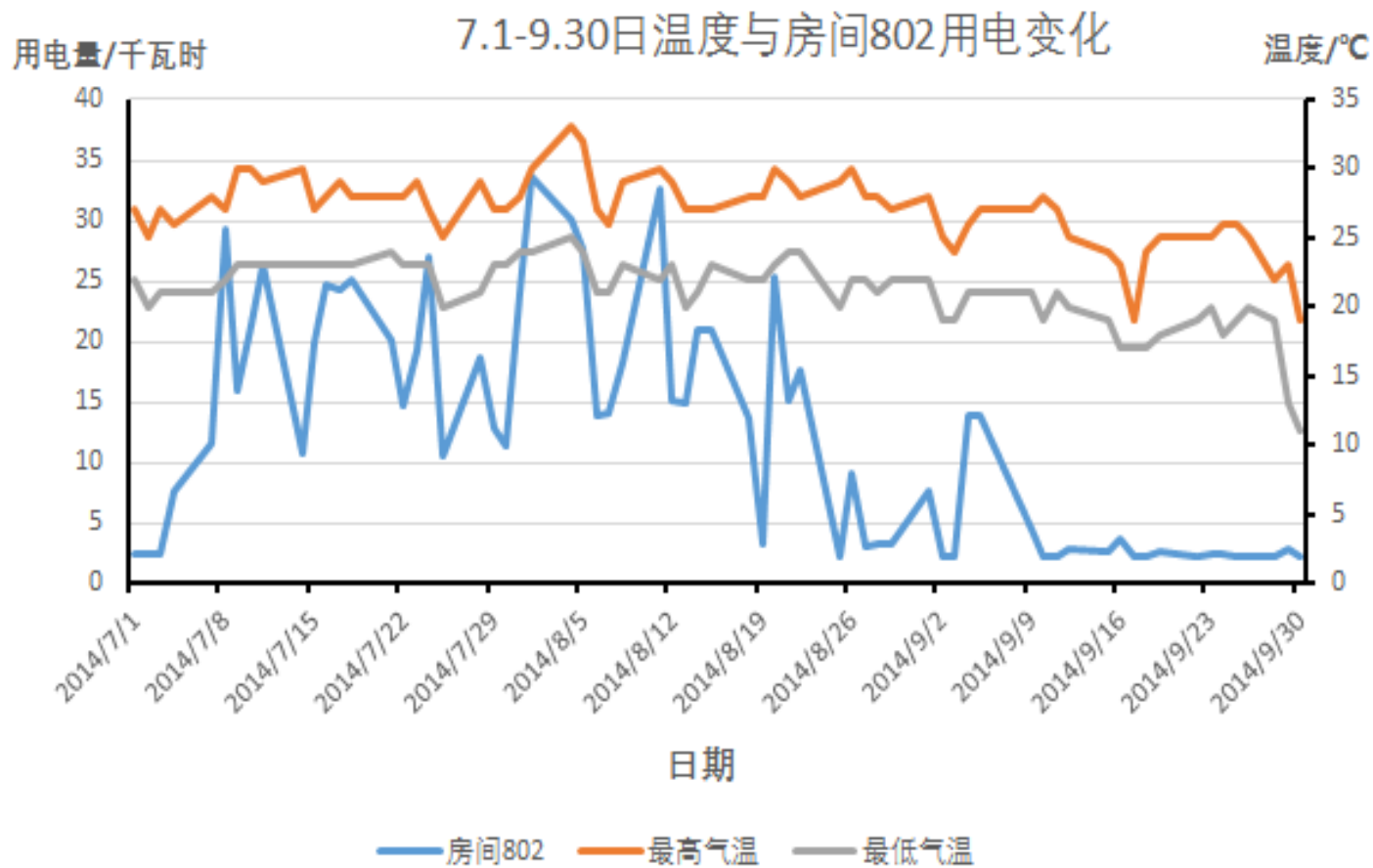
- 共**238**个房间
- 2014.07.01~2015.07.08，共**373**天
- 期间工作日共**256**天
- **单日**用电量





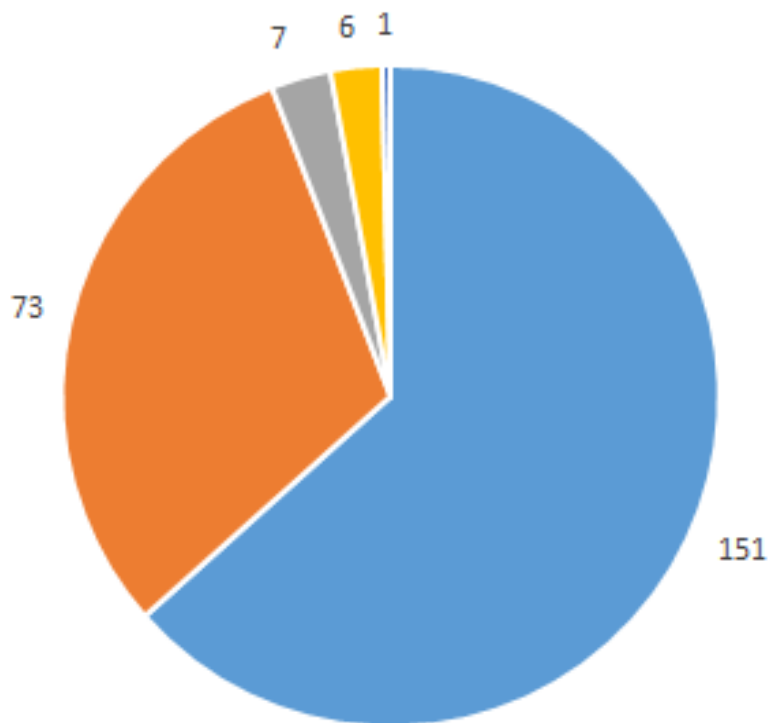
所有房间全年总用电情况







房间平均用电量聚类



■ <5千瓦时 ■ 5-15千瓦时 ■ 25-50千瓦时 ■ 15-25千瓦时 ■ >100千瓦时





计算空置率的意义

- 空置房间对正常使用房间的分析产生干扰
- 空置率是判断经济发展状况的一个重要指标






空置房间标准

- 日均用电量低于1千瓦时
- 用电量低于1千瓦时的天数分布相对分散
- 用电量低于1千瓦时的天数占到总工作日的 $\frac{1}{3}$





计算结果如下：共**76**个房间

702	703	704	705	707	709	710	711	712	805
807	809	810	811	812	815	906	910	915	1001
1011	1107	1111	1201	1209	1211	1215	1301	1309	1310
1408	1409	1411	1504	1507	1509	1803	1806	1904	1905
2101	2104	2201	2203	2205	2206	2404	2405	2503	2505
2804	2805	2806	2901	2906	3002	3005	3006	3201	3203
3204	3205	3301	3304	3402	3501	3504	3601	3604	3801
4002	4101	4201	4401	4501	4601				

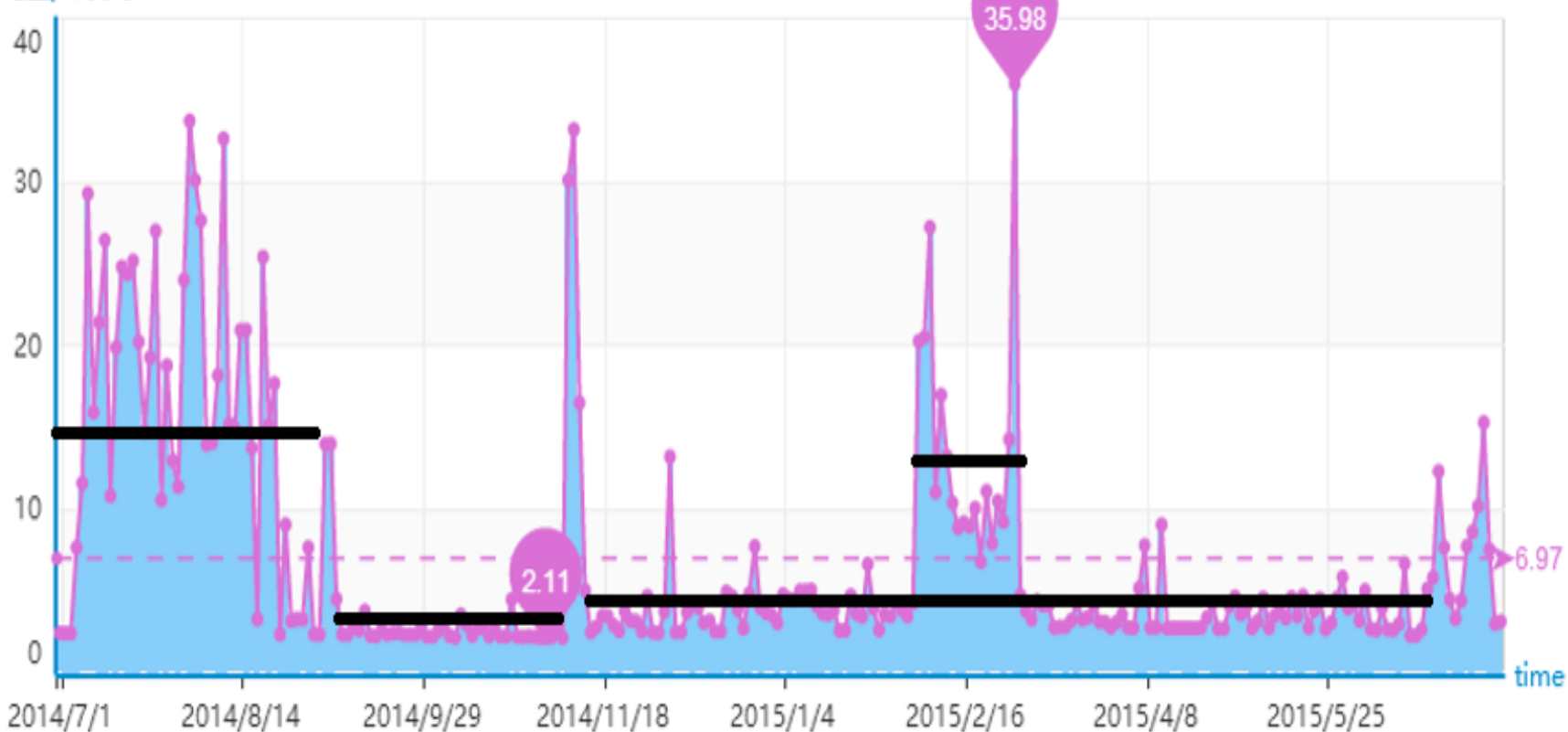
空置率=控制房间/总房间=**31.9%**





房间802工作日用电情况

用电量/千瓦时





将房间802认为是正常的写字楼房间

按电脑分析：

一台日常使用的台式计算机的功率约**300瓦**。按照每天开机**8**小时计。

9月~10月间，日均用电量约**3**千瓦时，可供约**1**台电脑工作。

从**11月**开始，日均用电量约**5**千瓦时，共两台电脑工作。说明房间内电脑加。

空调分析：

台**1P**的空调功率制冷一般为**1000**

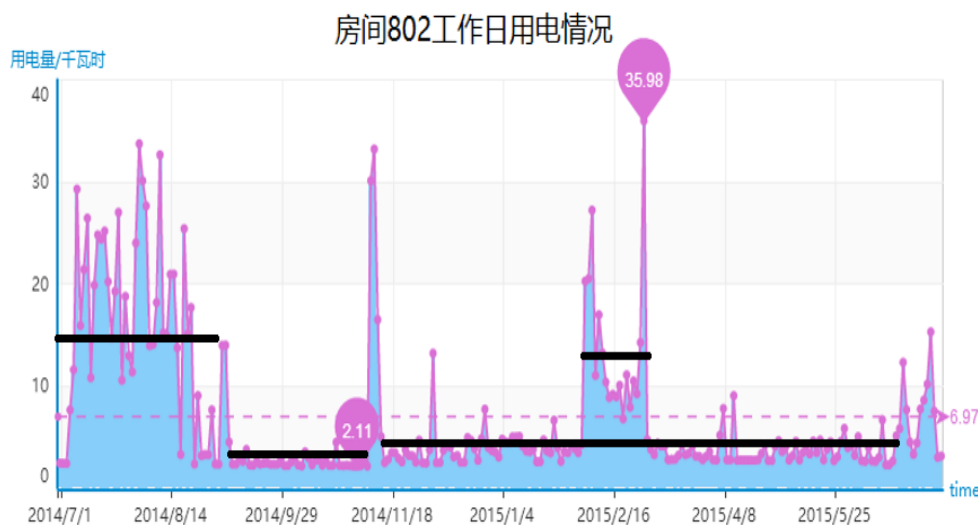
制热约**1100瓦**；**2P**制冷为**1600**

制热为**1700瓦**。按照每天开机**8**

小时计。

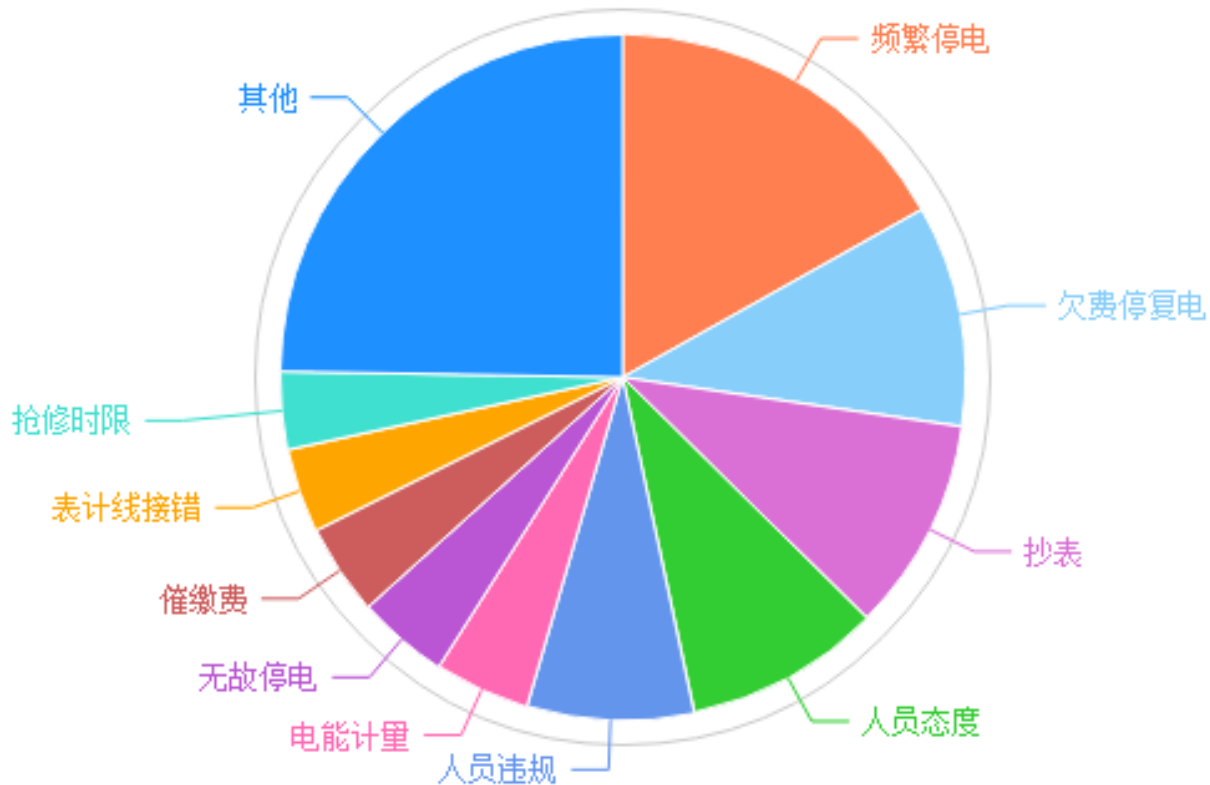
7~8月间，日均空调用电约**15**千瓦时，可供**2P**空调工作。

2月中，按照**2P**空调计算，约耗电**14**千瓦时，与图中所示基本吻合。





故障类型



故障投诉类型包括72种，侧重于频繁停电、欠费停复电、抄表、人员态度、人员违规的管理

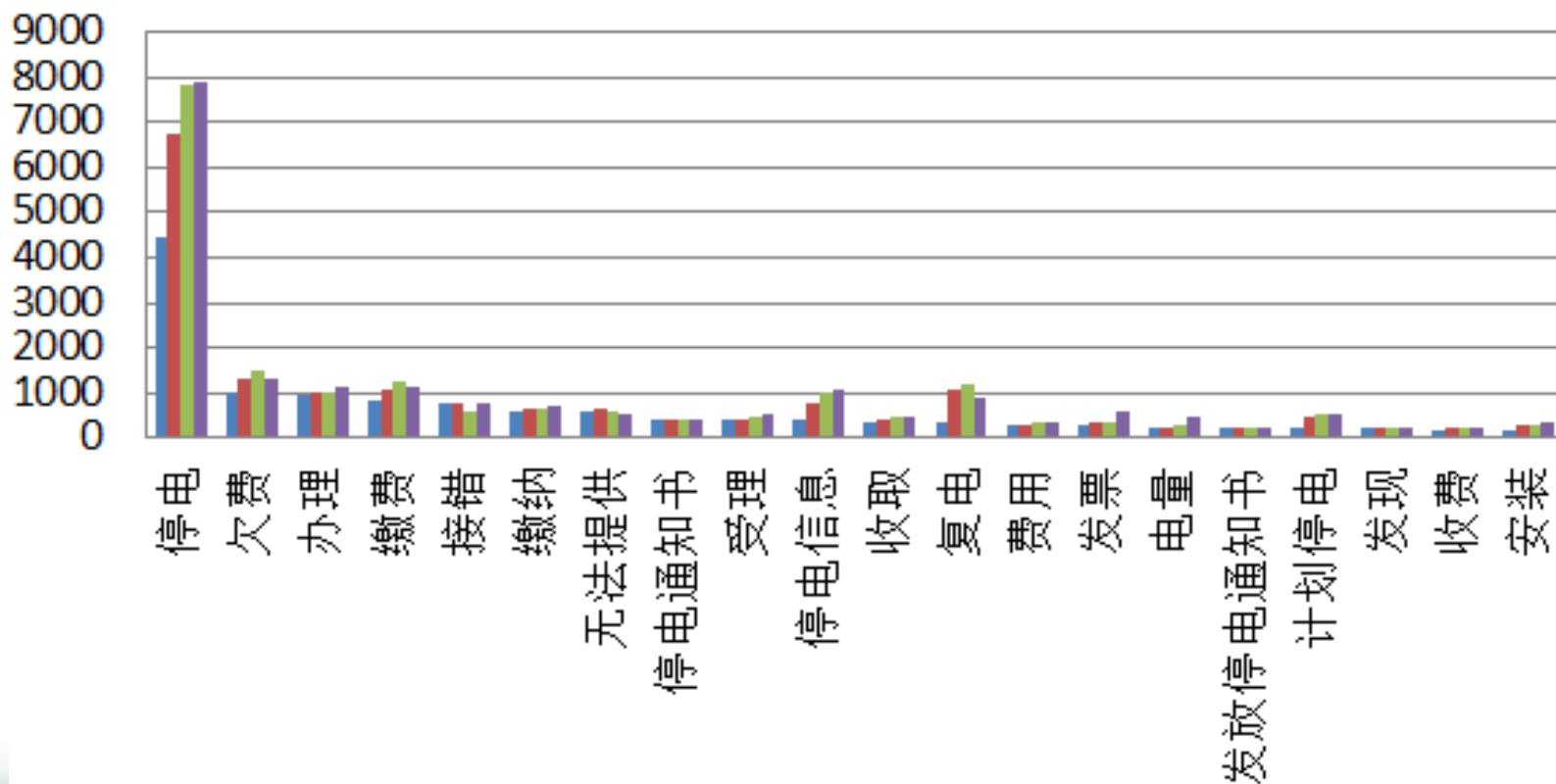


- 不务正业大乱炖（政治报告、营销、活动、撤侨、文本自动生成）
- 某大厦电力数据挖掘
- 95598 客服挖掘
- 国家电网语义智能搜索
- 国家电网头条
- 公安警情大数据挖掘



故障投诉内容随时间变化分析-相同词汇

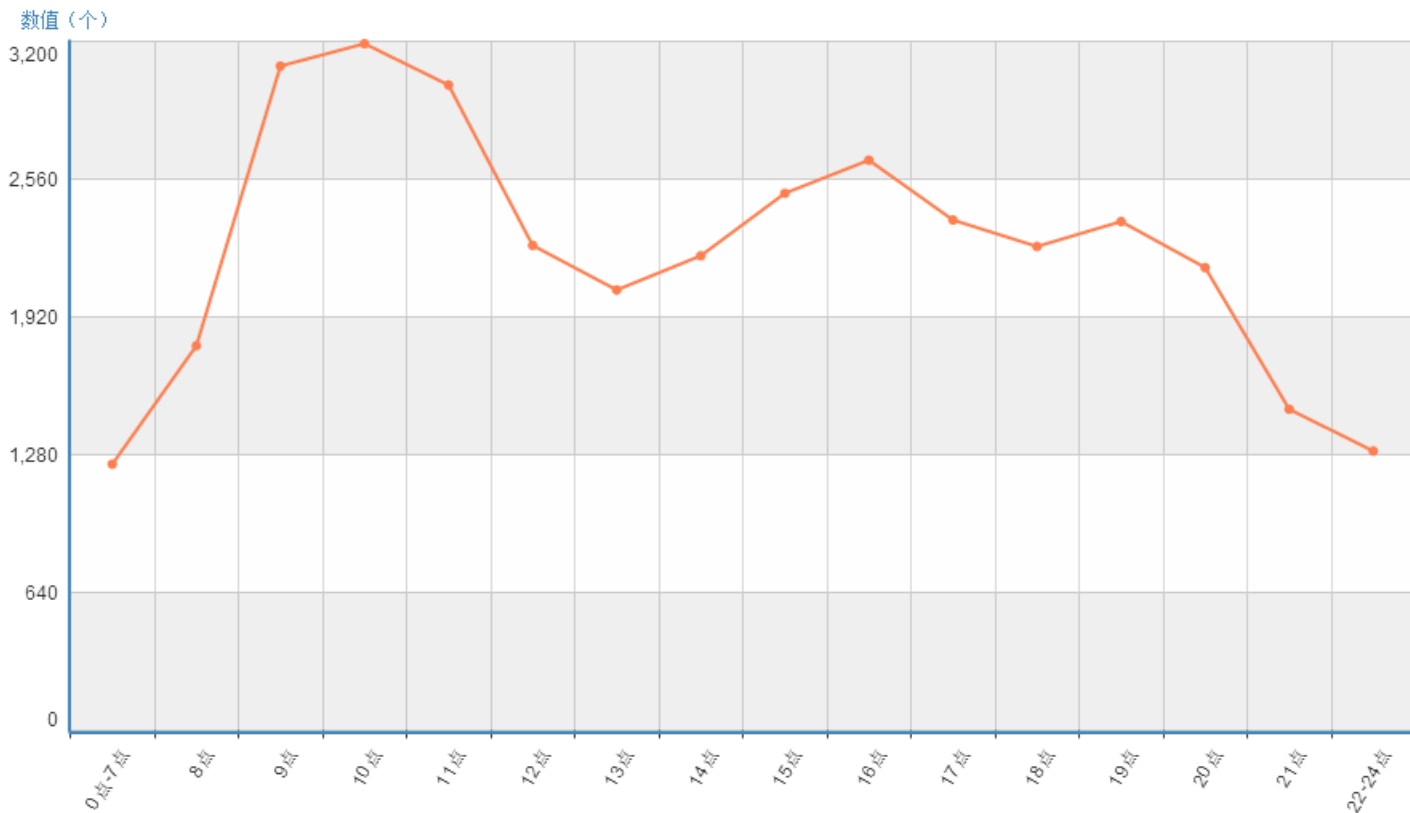
关键词3、4、5、6月词频信息





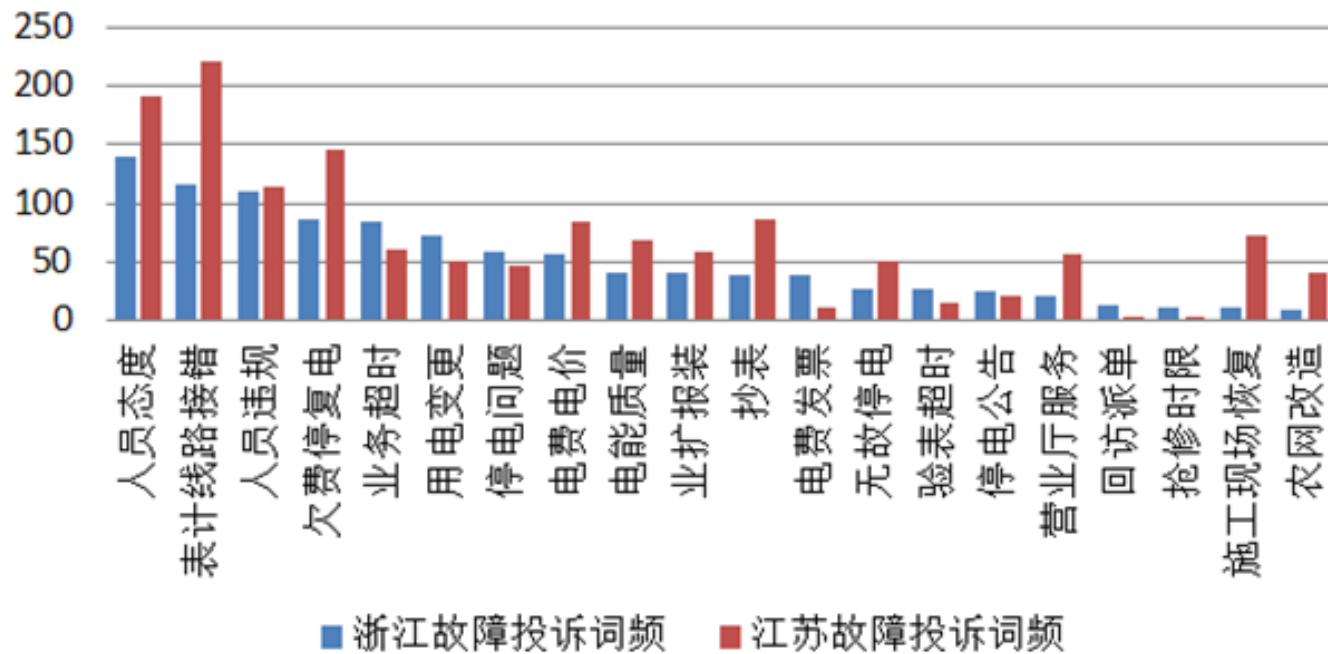
故障投诉随时间变化的挖掘分析

故障投诉与时间关系



➤ 每天投诉的高峰期为10点和16点





- 不务正业大乱炖（政治报告、营销、活动、撤侨、文本自动生成）
- 某大厦电力数据挖掘
- 95598 客服挖掘
- 国家电网语义智能搜索
- 国家电网头条
- 公安警情大数据挖掘



JZSearch语义精准搜索引擎

个人中心



综合搜索 组织结构 电网报 图片 电力地图 统计分析

刘振亚是谁

搜索

热搜词条:

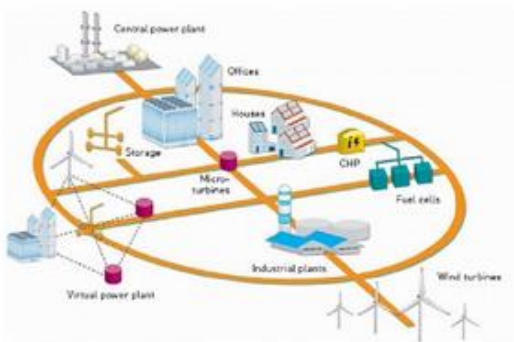
刘振亚是谁

UHV是什么

智能电网

配电网

全球能源互联网



最新消息

国网大事记

- 严格依法治企 坚持以...
- 台风“灿鸿”肆虐 供...
- 奋发有为 扎实工作 ...
- 连续抢修3天
- 风狂雨骤显担当
- 一定把损失降到最低

大数据分析与应用/张华平



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



JZSearch语义精准搜索引擎

聚类结果

特高压

教育实践活动

智能电网

公司

全球能源互联网

学习实践活动

经济社会发展

安全生产

语义结果

询问对象：刘振亚，

刘振亚，男，汉族，1952年8月生，山东鄒城人，1984年加入中国共产党，1971年参加工作，山东工学院电力系电力系统及自动化专业毕业，大学学历，山东大学电气工程学院电气工程及其自动化专业硕士研究生毕业，电气工程硕士，教授级高级工程师，享受国务院政府特殊津贴。现任国家电网公司董事长、党组书记。

语义统计分析

分析对象：刘振亚 分析

智能搜索

返回检索结果约1906个结果...

1. 刘振亚总经理调研韩国STX集团大连造船厂项目的用电情况

来自栏目：无设定栏目 板块：要闻 发布时间：2007/05/29 00:00:00 作者：杜平

27日，刘振亚总经理(前排左四)一行在大连市市长夏德仁(前排左三)的陪同下，调研韩国STX集团大连造船厂项目的用电情况。

关键词：刘振亚 命名实体--人物：刘振亚#夏德仁#

2. 刘振亚分别会见花旗集团和通用电气高层

来自栏目：无设定栏目 板块：要闻 发布时间：2009/04/28 00:00:00 作者：岳文

语义自动计算

相关新概念发现

沙捞

公司党组

电力集团公司

能源互联网

相关人物聚类

习近平/61

李克强/22

舒印彪/17

李荣融/12

郑宝森/11

温家宝/10

俞正声/9

曹志安/9

帅军庆/7

高培/7

相关作者聚类

姚雷/295

陶思遥/108

张超义/54

江莹/41



JZSearch语义精准搜索引擎

语义统计分析
语义统计分析
分析对象：刘振亚 **分析**

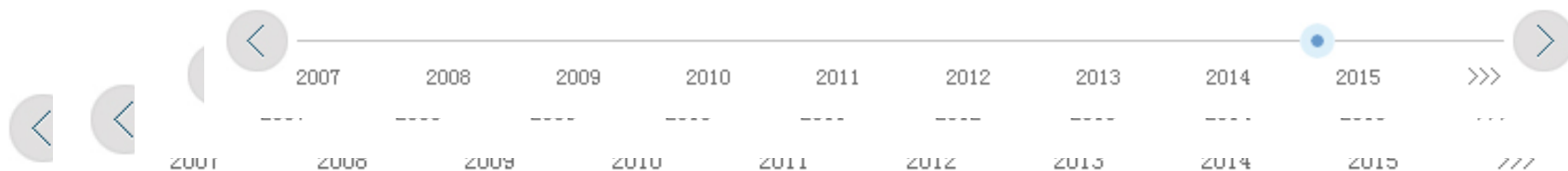
分析对象：刘振亚是谁

时间：2007-2015

分析机构：国家电网公司

数据来源：10年国家电网报

说明：左图为关键词随着时间的变化，关键词也在发生变化。



JZSearch语义精准搜索引擎

3. 刘振亚会见马来西亚沙捞越州首席部长

来自栏目: 无设定栏目 板块: 要闻 发布时间: 2010/04/14 00:00:00 作者: 姚雷

总经理刘振亚在公司总部会见了到访的马来西亚沙捞越州首席部长泰益玛目一行, 双方进行了亲切友好的交谈, 并就发挥各自优势加强合作深入交换了意见。刘振亚对马来西亚客人的来访表示热烈欢迎, 并详细介绍了中国电力工业尤其是电网的发展情况。他希望双方能进一步增进了解和信任, 达成合作共识, 并在更广泛领域积极开...

关键词: 刘振亚 命名实体--人物: 刘振亚#哈吉·拉旺#杜至刚#

4. 刘振亚会见东方电气集团董事长

来自栏目: 头条2 板块: 要闻 发布时间: 2009/07/16 00:00:00 作者: 姚雷

刘振亚对王计一行的到访表示热烈欢迎, 对于东方电气广大干部员工在特大地震灾难面前顽强不屈的精神和恢复重建取得的成绩表示钦佩。刘振亚说, 东方电气集团公司是我国最大的发电设备制造企业之一, 国家电网公司一直十分关注东方电气集团灾后重建和发展, 帮助东方电气集团恢复重建是国家电网的社会责任所在。经过大...

关键词: 刘振亚 命名实体--人物: 刘振亚#栾军#

5. 刘振亚会见法国电力集团公司董事长

来自栏目: 无设定栏目 板块: 要闻 发布时间: 2010/12/09 00:00:00 作者: 姚雷

总经理刘振亚在公司总部会见了到访的法国电力集团公司董事长兼首席执行官亨利·普格里奥一行。双方表示, 将巩固合作基础, 加强交流, 进一步拓展双方在电力领域的合作, 实现未来共同进步。刘振亚对亨利·普格里奥一行的来访表示热烈欢迎。他说, 国家电网公司非常重视与法国电力同行的友好关系, 双方已有的合作是双赢...

相关概念词发现



相关人物计算



基于语义的自动学习计算

- 不务正业大乱炖（政治报告、营销、活动、撤侨、文本自动生成）
- 某大厦电力数据挖掘
- 95598 客服挖掘
- 国家电网语义智能搜索
- **国家电网头条**
- 公安警情大数据挖掘



电网头条应用实践

➔ “电网头条”：全媒体个性化智能推荐平台

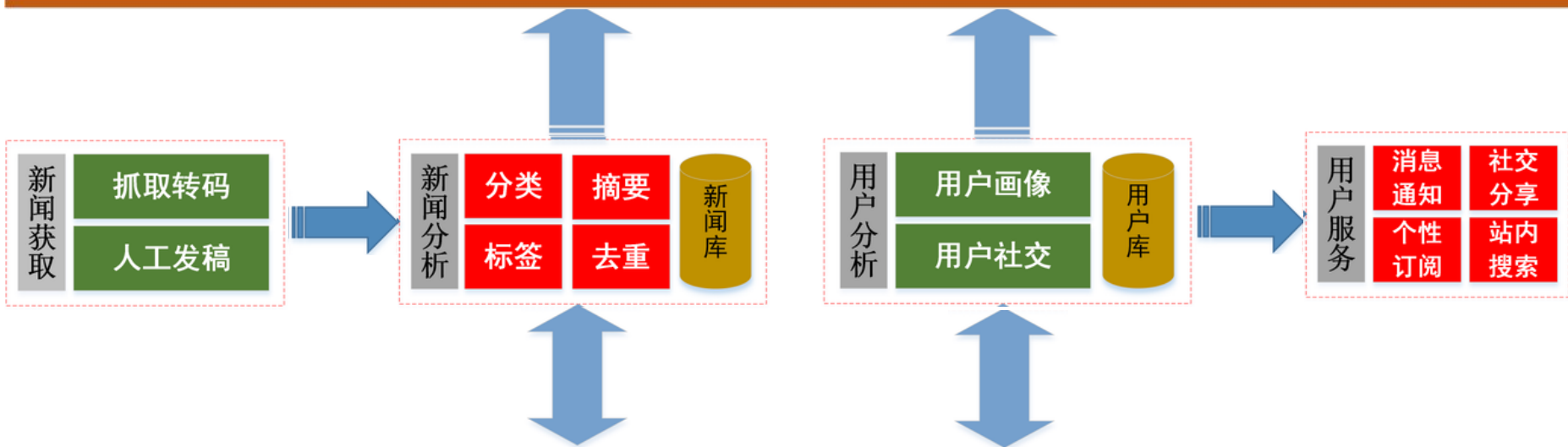
- 全媒体：多维一体、富媒体
- 云应用：用户、编辑、审核、管理员等SaaS
- 个性化：因时因地因人而变
- 采编发审中央厨房



系统组成方案

➔ 系统总体架构

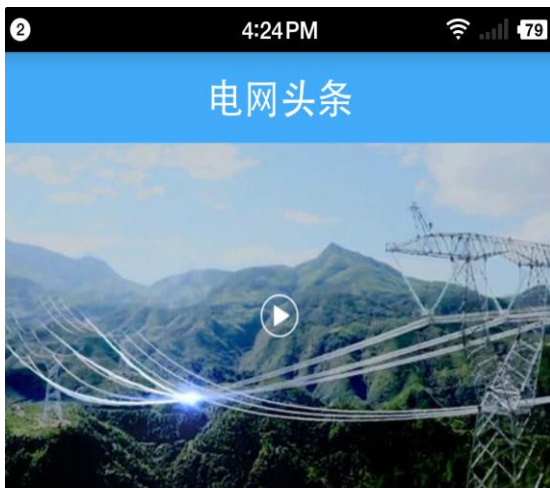
精准匹配及新闻推送系统



采编控制平台



全媒体：多元化富媒体展示



全球能源互联网大会今日启幕，震撼视频...
来自电网头条



全球能源互联网高端访谈之杜祥琬
来自电网头条

多元化展示

图文混排、视频播放

多元化内容

微信、相关新闻、热门评论

多元化互动

喜欢、收藏、投票、发表评论





个性化-因时因地因人而变

The image consists of three vertical panels illustrating digital technology and user experience:

- Left Panel (Teal background):** Titled "云计算 用户画像" (Cloud Computing User Profile) with the subtitle "想你所想 为你而来" (Think what you think, come for you). It features icons of clouds, gears, a laptop, a smartphone, and a person's silhouette, representing data analysis and user profiling.
- Middle Panel (Dark Blue background):** Titled "精准推送" (Precision Push) with the subtitle "知识分享的平台" (Knowledge Sharing Platform). It shows a hand touching a smartphone screen, with various social media and utility icons (like a globe, Wi-Fi, heart, thumbs up, and location pin) connected to it, symbolizing targeted content delivery.
- Right Panel (Dark Blue background):** Titled "汇聚电网达人" (Gathering Grid Experts) with the subtitle "你原创的就是头条" (Your original content is the headline). It depicts a central globe surrounded by circular profiles of people, with icons for social media, location, and sharing, representing a community of experts.



个性化：用户画像应用案例

➔ 利用每日新增的网络访问日志数据500G，采用大数据挖掘技术，提供商情精准营销增值服务。



大数据分析

电网头条的应用实践

器 8.1

http://www.sgcctop.com/pcnewsview/

电网头条

推荐 热点 GEI 电网

中国移动 4G 10:36

北京 电网头条

推荐 热点 GEI 电网 服务 科技 +

大厂县供电公司打造电动汽车示范区

浏览: 1

国家电
17:5

全球能源互联网大会视频图文报道



置顶

电网头条

评论 0



晋江市供电公司推进农网改造工程

浏览: 0

国家电
17:4

GEI是什么都不知道？你OUT啦！



置顶

国家电网报

评论 4



盐山县供电公司检查春灌供水

浏览: 0

国家电
17:4

家电耗电“琅琊榜”，您猜谁是第一名？



主页

视频

达人

我的



大数据分析与应用/张华平

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

电网头条-采编发审中央厨房

电网头条采编管理系统

欢迎您! 编辑301 [账号设置] [注销]

电网头条采编管理系统

欢迎您! guo123 [账号设置] [注销]

欢迎您! guo123 [账号设置] [注销]

[注销]

[注销]

系统审查

日志查看

在线人数

日志统计分析

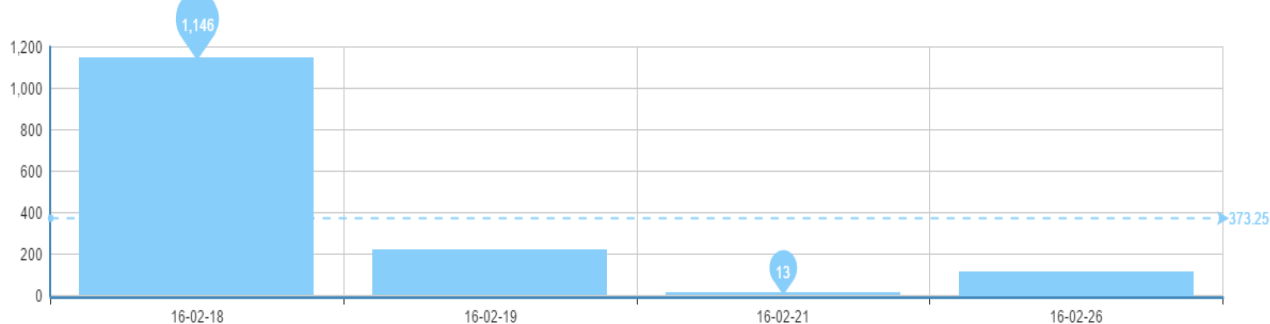
审计日志等级

审计通知设置

业务事件统计

业务事件

刷新 删除 打印 导出 刷新



刷新 按照时间升序 按照时间降序 按照事件名升序 按照事件名降序

ID	操作时间	用户名	用户属性	账户类型	IP地址	操作类型	操作状态
1	2016-02-18 01:05:17	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无
2	2016-02-18 01:05:17	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无
3	2016-02-18 01:05:17	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无
4	2016-02-18 01:05:35	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无
5	2016-02-18 01:05:35	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无
6	2016-02-18 01:05:35	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无
7	2016-02-18 01:05:37	guo123	系统审计员	查看操作日志	127.0.0.1	业务事件	无

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

[预览]

编辑 [预览]

编辑 [预览]

- 不务正业大乱炖（政治报告、营销、活动、撤侨、文本自动生成）
- 某大厦电力数据挖掘
- 95598客服挖掘
- 国家电网语义智能搜索
- 国家电网头条
- 公安警情大数据挖掘

公安大数据-串并案

```
item_id:1
doc_id:2267
keywords:号院;西经;<font color="red">三轮车</font>;电动;天桥;停放;发票;被盗;民警;电动车
address:null
time:T273;T72;T10;T48;T43;T273;T72;T2;T25;T5;T92;T91;T273;T72;T10;T48;
people:P1532;P1532;
content:将电动<font color="red">三轮车</font>（红色红旗牌电动<font
color="red">三轮车</font>, ;T5;#;T92;#购买, ;T91;#价值人民币3100元, 有发票）停放在西经路2号院2号楼1单元楼道内
#许, 事主发现车不见了, 于是报警。...
```

```
item_id:2
doc_id:3315
keywords:电动车;人因;停放;<font color="red">三轮车</font>;被盗;北门;医院;发现
address:A0;A1;A7;
time:T273;T56;T50;T145;T14;T56;T50;T81;
people:P2252;
content:;T273;#;T56;#;T50;#;T145;#;T14;#许, 报警人;P2252;#将电动车<font
color="red">三轮车</font>停放在;A0;#;A1;#;A7;#人民医院北门门口, ;T56;#;T50;#;T81;#50许发现电动车被盗, 报警人
```

```
item_id:3
doc_id:168
keywords:<font color="red">三轮车</font>;快递;电动;价值;主将;下楼;大厦;广义;不见;人民币
address:null
time:T116;T33;T78;T37;T26;T56;
people:null
content:;T116;#;T33;#;T78;#;T37;#;T26;#, 事主将电动<font
color="red">三轮车</font>车停在广义街祥龙大厦门口, 上楼去送快递, ;T56;#下楼事主发现电动<font color="red">三轮
color="red">三轮车</font>价值人民币3000元左右, 车里还有60多份快递价值一万余元。
```

```
item_id:4
doc_id:2271
```





公安大数据-诈骗案语义聚类

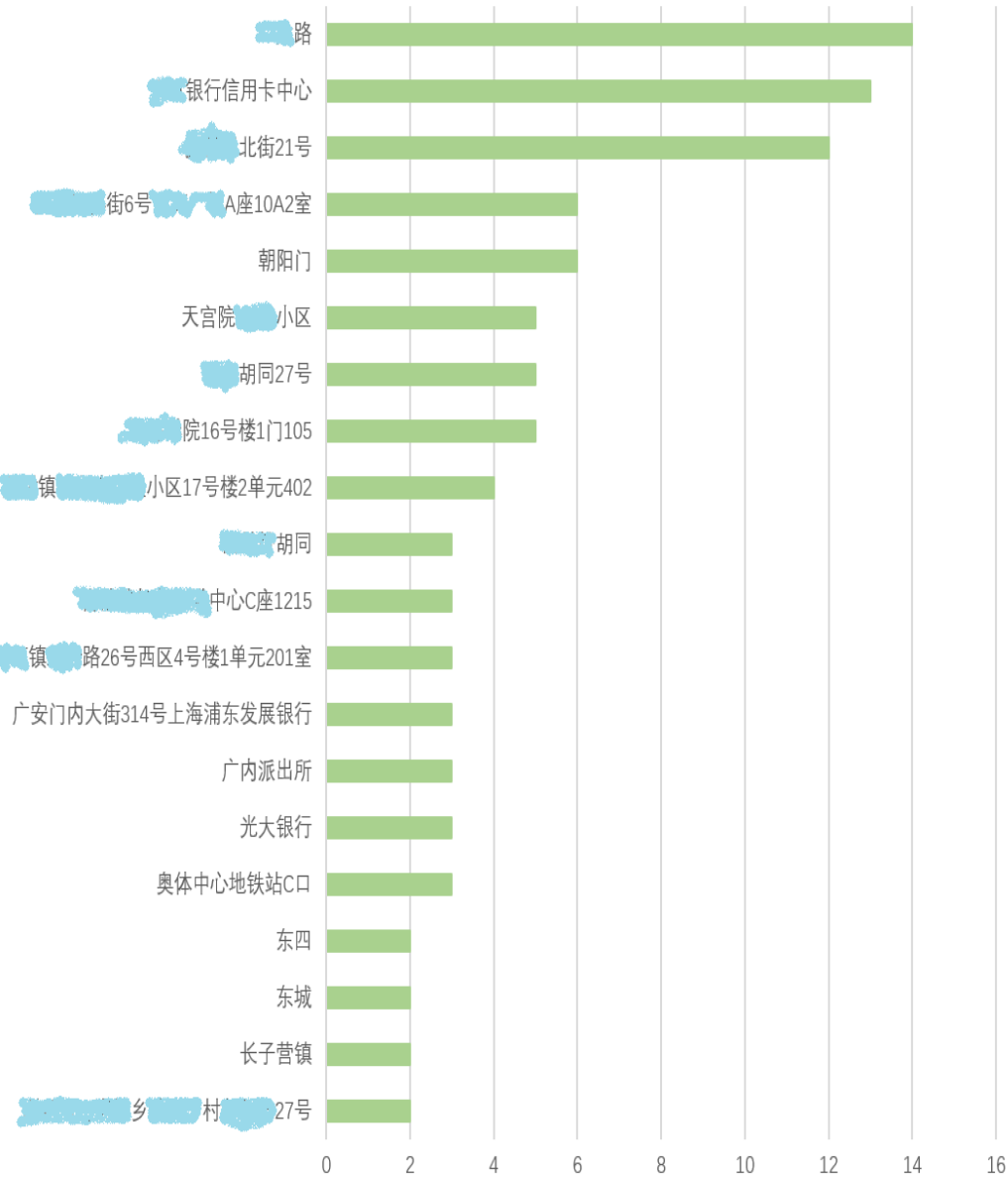
- Cluster 0: 天作 使用费 非法占有
- Cluster 1: 伪造 交通事故 赔偿金 保险公司
- Cluster 2: 博亿 房屋买卖 虚构
- Cluster 3: 鲁迅文学奖 作家协会 诗词
- Cluster 4: 骗走 人民币 平房 公墓 灵泉 塔位 灵塔 地葬位
- Cluster 5: 押金 天纺 柯东 冒用 丰体
- Cluster 6: 为名 培训 假借 团伙 考证 帮助

新词发现、类别关键词提取
等



公安大数据-串并案分析

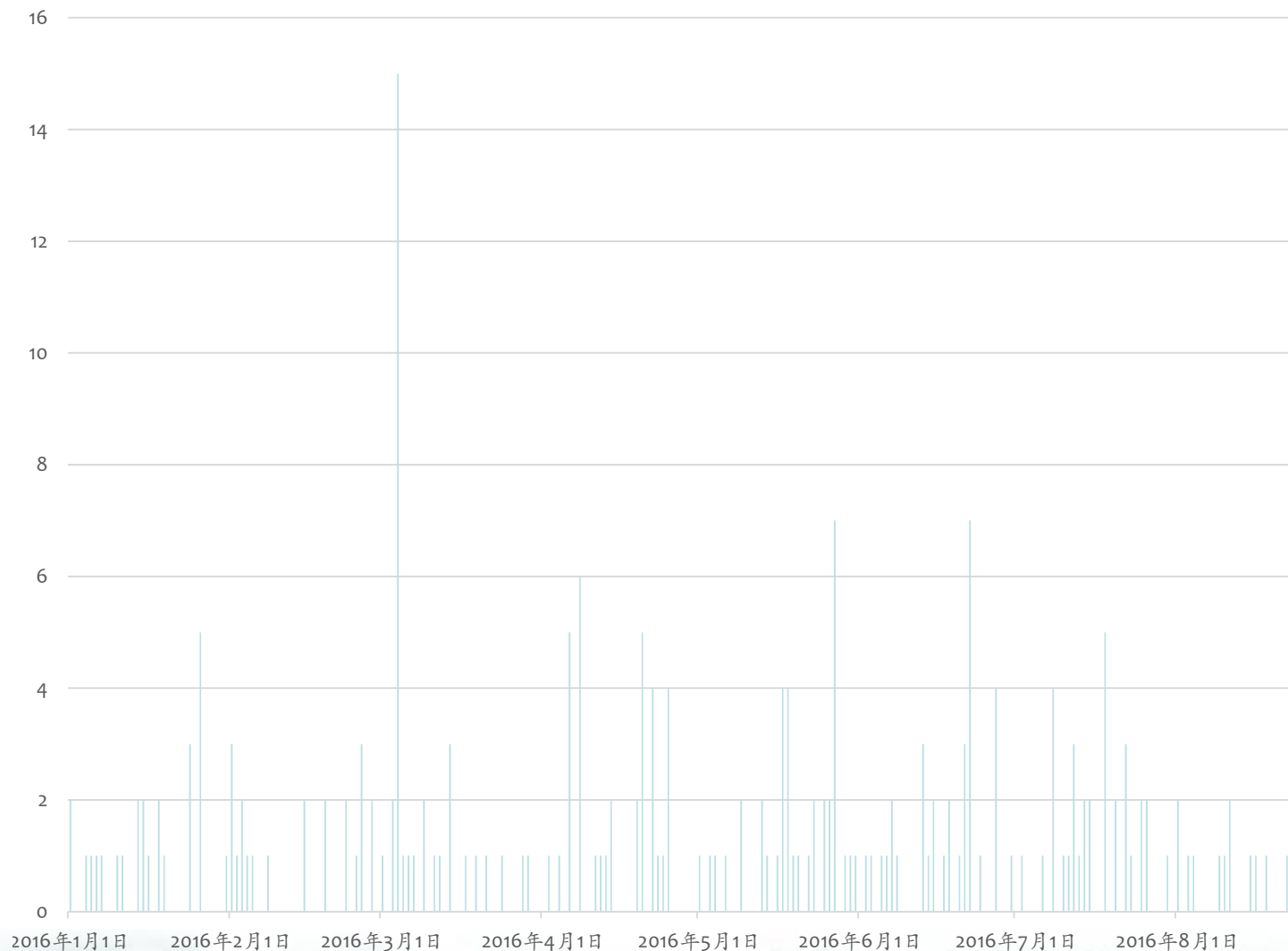
案发地址





公安大数据-串并案分析

案发详细时间



大数据分析与应用/张华平

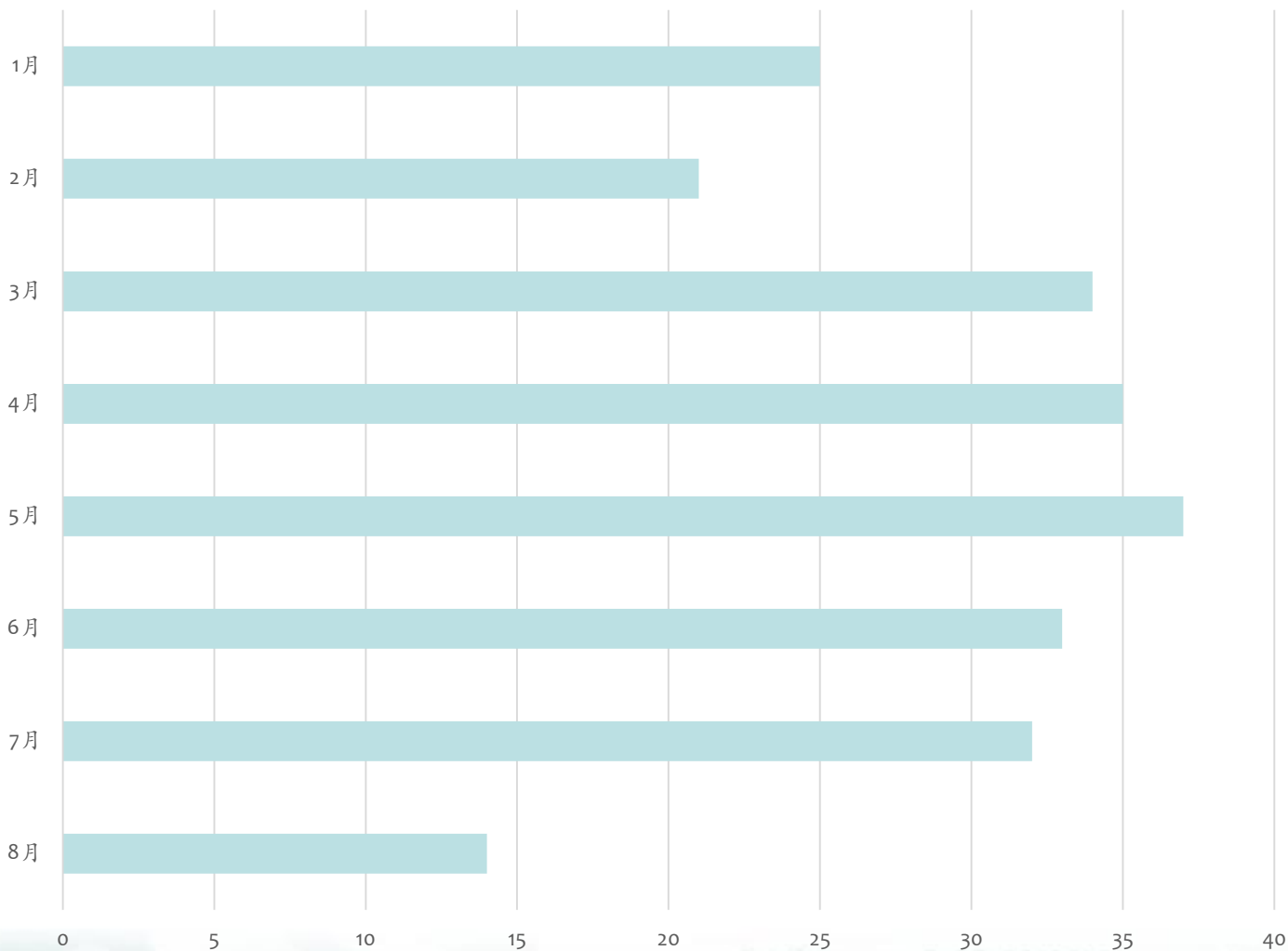


北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



公安大数据-串并案分析

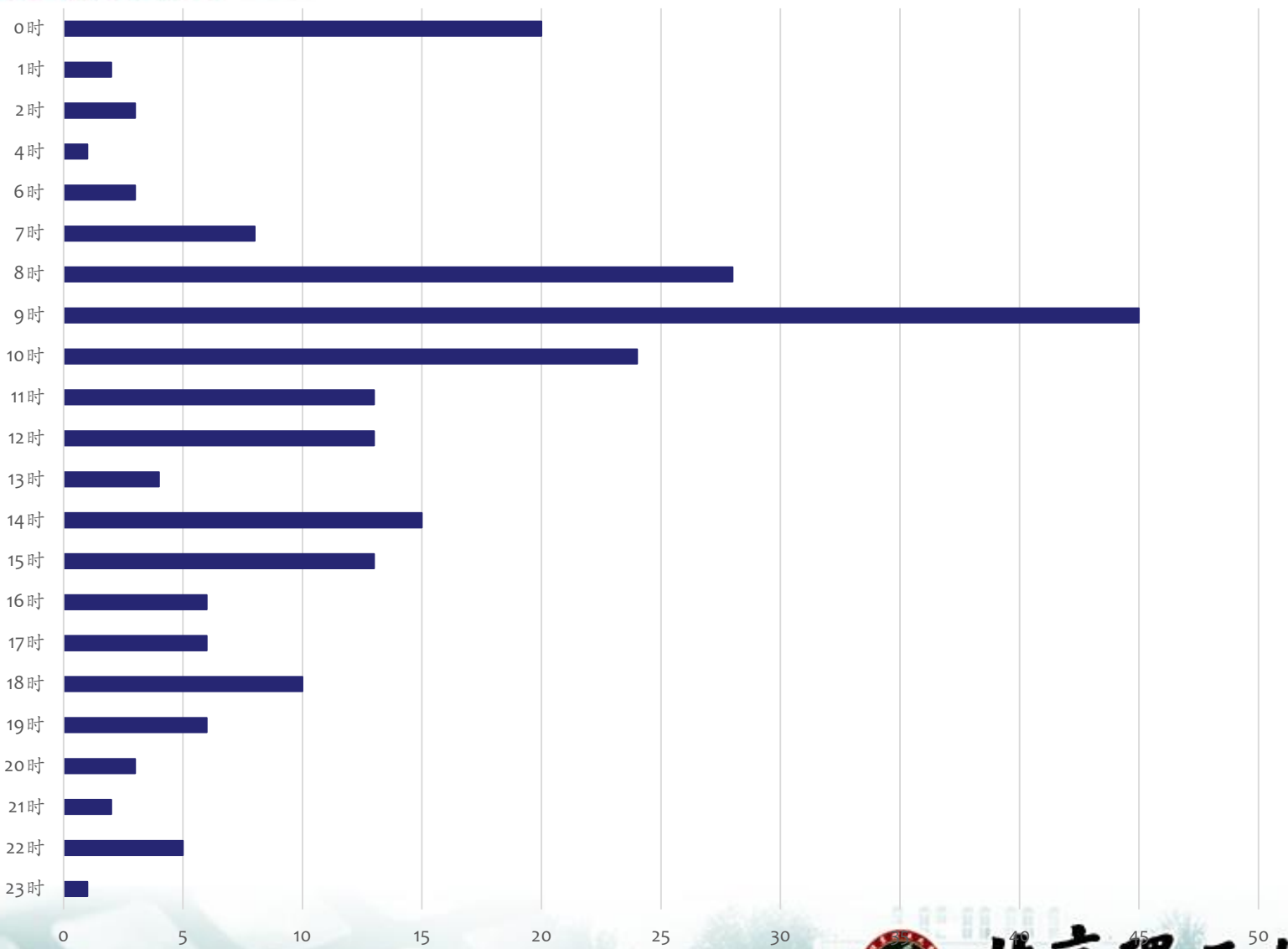
案发月份





公安大数据-串并案分析

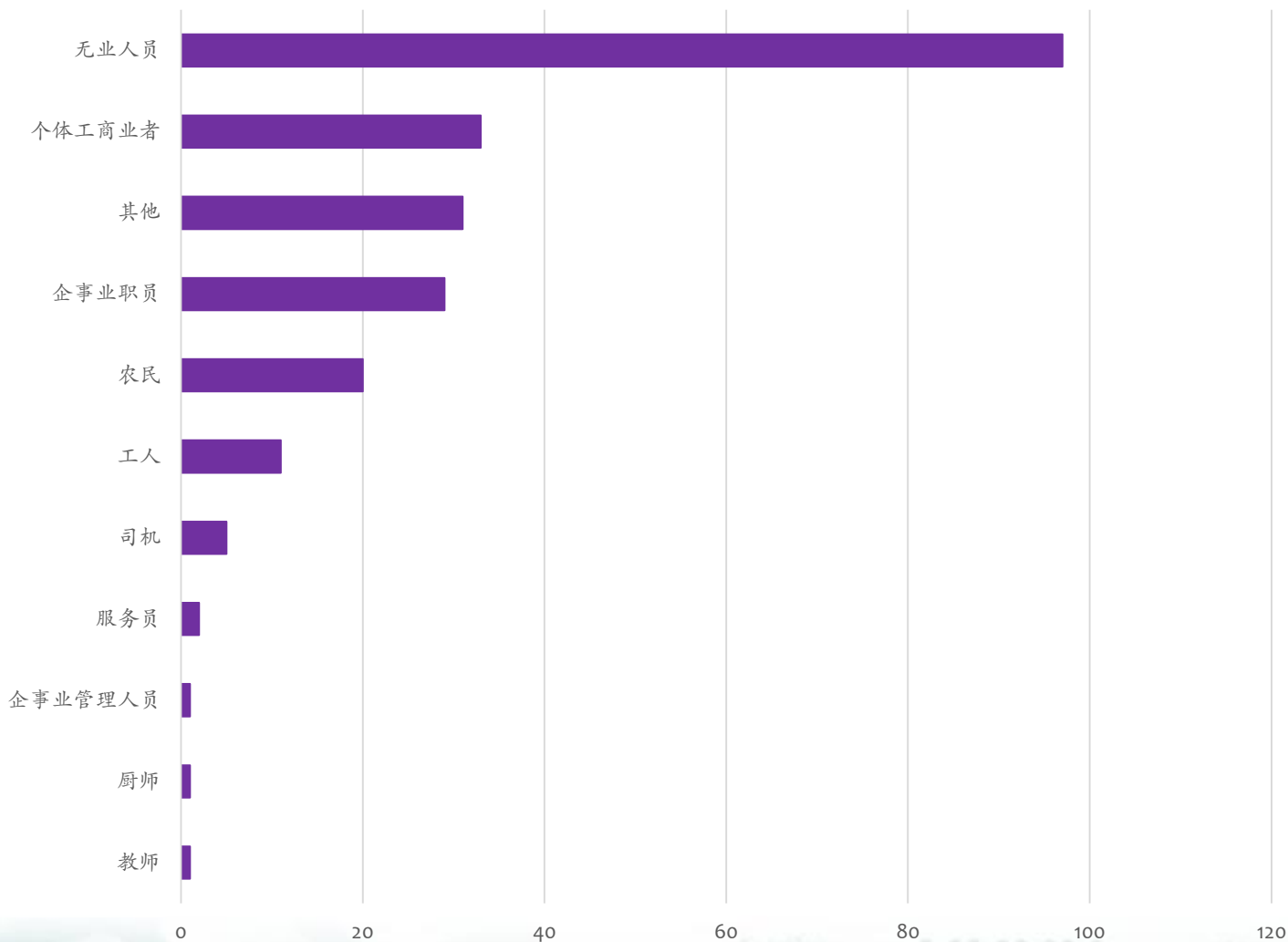
案发时间(小时)





公安大数据-串并案分析

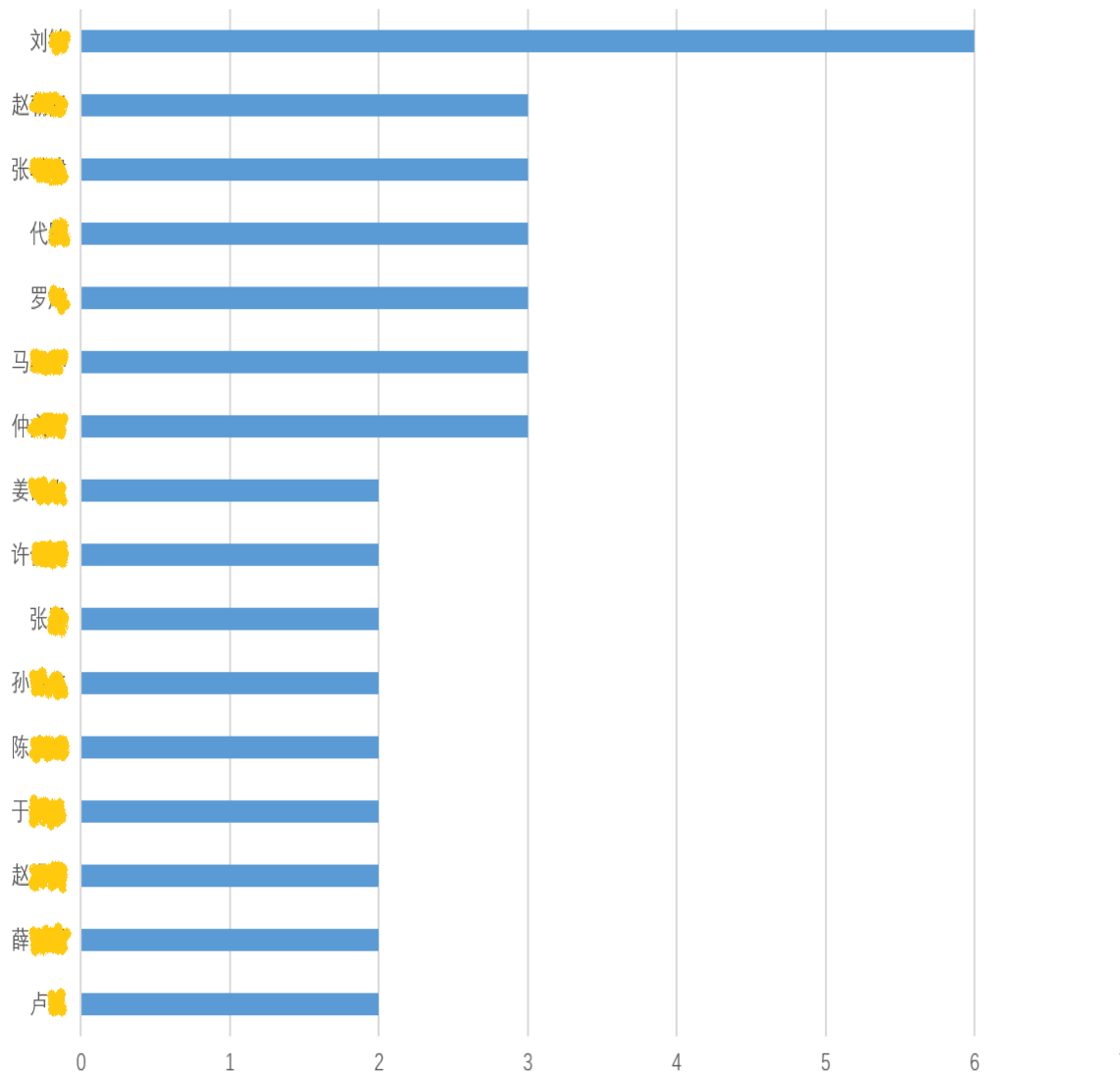
身份





公安大数据-串并案分析

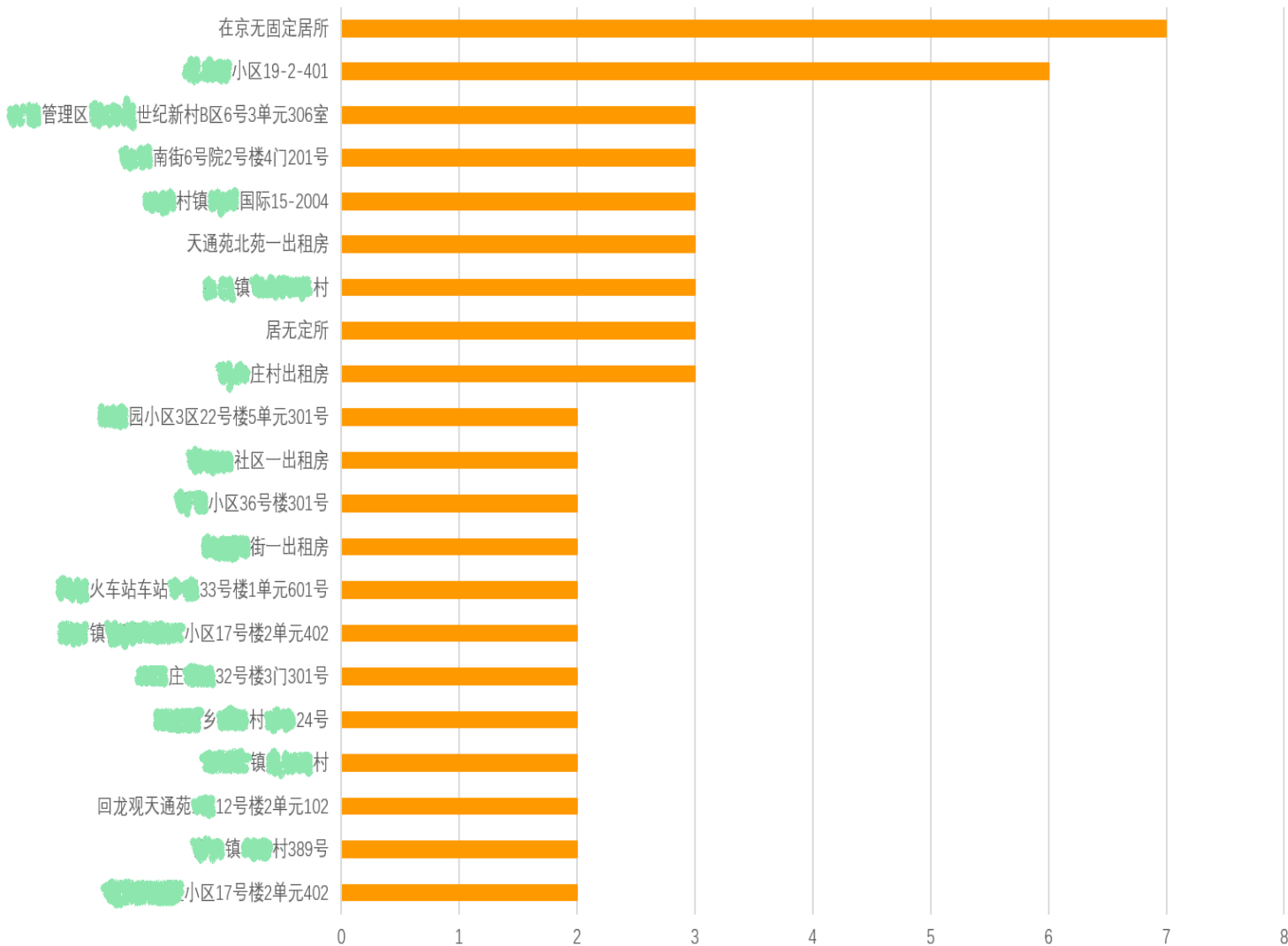
人名统计





公安大数据-串并案分析

居住地





公安大数据-犯罪地图



- 1 时间
- 2 地点
- 3 人
- 4 ...



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

