

NLPIR大数据语义分析技术的在线演示

-支持所有平台

NLPIR SDK

支持的开发语言

支持的操作系统

NLPIR SDK存放了13种组件包:

- Classify规则组件
- Cluster聚类组件
- DeepClassifier训练分类组件
- DocExtractor实体抽取组件
- HTMLPaser网站正文提取组件
- NLPIR-ICTCLAS分词组件
- JZsearch精准搜索组件
- JZSearch精准搜索客户端组件
- KeyExtract关键词提取组件
- RedupRemover文档去重组件
- Sentiment情感组件
- SentimentAnalysis情感分析组件
- Summary摘要组件

C语言
C++语言
C#语言
JAVA语言
等
几乎囊括了
市面所有主
流的编程语
言

Windows 32
位/64位操
作系统
Linux32位
/64位操作
系统
Android操
作系统
IOS操作系
统
国产红旗等



NLPIR之大数据聚类

```
<?xml version="1.0" encoding="gb2312" standalone="yes" ?>
- <LJCluster-Result>
  <clusnum>200</clusnum>
+ <clus id="0">
+ <clus id="1">
- <clus id="2">
  <feature>甲型 流感疫苗 流感病毒 流感确诊病例 流感疫情 流感病例 重症病例 接种疫苗 季节性流感</feature>
- <docs num="150">
  <doc>多国H1N1流感疫情呈严重态势 美国疫苗供不应求</doc>
  <doc>卫生部就甲型H1N1流感疫情防控及疫苗预防接种工作答问</doc>
  <doc>甲型流感疫情在各国呈现严重态势</doc>
  <doc>中国报告4例接种甲流疫苗后死亡病例(实录)</doc>
  <doc>日本拟撤销对孕妇接种流感疫苗限制</doc>
  <doc>卫生部召开甲型流感防控疫苗接种通气会</doc>
  <doc>全球甲型流感疫苗接种进入高潮</doc>
  <doc>沪上近7万人已接种疫苗 天气转冷甲流危险性将累加</doc>
  <doc>对鸡蛋过敏者不宜接种甲流疫苗</doc>
  <doc>钟南山: 甲流疫苗副作用小 学校应优先接种</doc>
  <doc>甲流疫苗答问公布:对鸡蛋过敏者不宜接种疫苗</doc>
  <doc>揭开甲型H1N1流感神秘的面纱:寻踪 变异 应对</doc>
  <doc>世卫公布甲型H1N1病毒调查:新病毒 人类无免疫</doc>
  <doc>福建、山东发生聚集性甲型H1N1流感疫情</doc>
  <doc>北京首现甲流死亡病例:北航军训团1名新生死亡</doc>
  <doc>流感疫苗对孕妇及胎儿免疫率达90%</doc>
  <doc>甲流可控可防可治</doc>
  <doc>首批甲流疫苗已运抵澳门 高危族23日起优先接种</doc>
  <doc>新疆报告1例甲流死亡病例 3类人群禁止接种疫苗</doc>
  <doc>法专家称接种甲型流感疫苗有助长期防疫</doc>
  <doc>接种疫苗仍是最有效手段</doc>
  <doc>美国甲流疫苗严重短缺 感染人数以百万计</doc>
  <doc>专家提醒:注射甲流疫苗后也应注重个人防护</doc>
  <doc>欧洲卫生专家称:甲流可能导致欧洲4万人丧命</doc>
  <doc>成都严禁以防控制甲型流感为由推销其它疫苗</doc>
```

CDA 数据分析师
www.cda.cn



大数据论坛
BigdataBBS.com



大数据搜索挖掘

I 科学的大数据观

II 文本大数据挖掘关键技术

III 大数据精准搜索关键技术

IV 大数据语义应用实践



JZSearch大数据精准搜索

- **搜索基本功能：**

- 多字段关联搜索、指定字段排序、精确搜索与模糊搜索

- **搜索特色功能：**

- 内嵌正负面情感等极性分析、语义联想搜索、临近搜索、搜索结果去重；



- 内嵌了ICTCLAS智能分词系统；

- 数据库实时同步：数据库增删改10秒内即可同步到搜索；

- **搜索维护功能：**

- 单点故障容错；支持增量索引；自动备份与恢复机制；自动缓存机制；自动优化机制；搜索屏蔽与恢复；

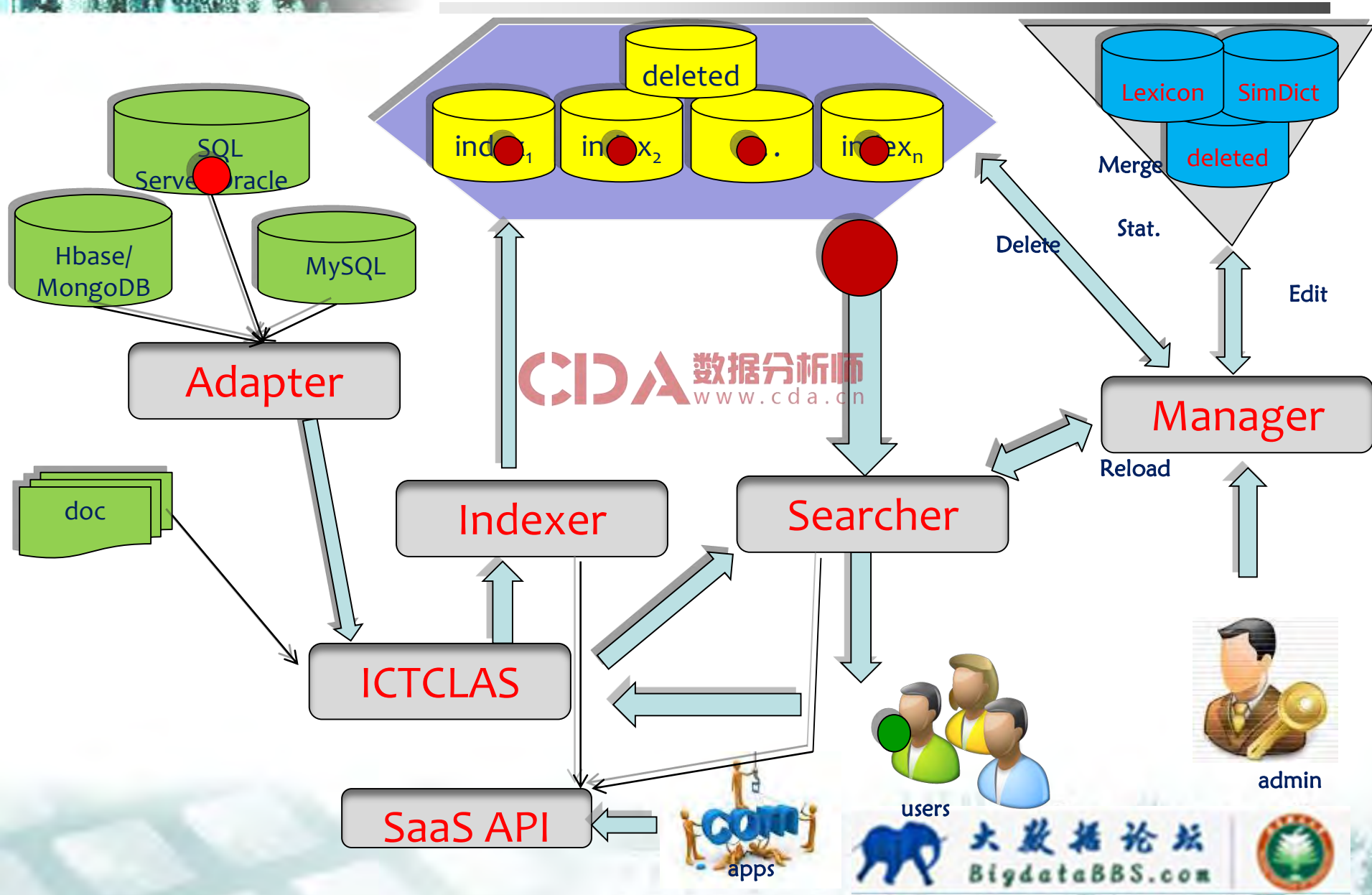


JZSearch内部语法示例

- [FIELD] * [NEAR] 尚福林 ##负面JZSearch## 12
- [FIELD] price [RANG] 1.0 9.0 [FIELD] name [AND] 牛奶
儿童
- [FIELD] name [PREF] 张
 - 姓名字段name必须以“张”作为前缀开头
- [FIELD] id [PREC] 123
 - 字段id必须以“123”精准匹配，如“1234”或者“0123”均不作为匹配结果；
- [field] content [complex] 统计局||中国统计局||CPI 骗人||
砖家 10
 - 要求content字段内，同时出现{统计局||中国统计局
||CPI }{骗人||砖家},且两者之间的距离必须在10个词内。



JZSearch Big Data Search Engine Architecture



JZSearch应用示例

少数民族语言精准搜索引擎

- 柑
- 瓜
- 农
- 豆
- 食
- 农
- 星
- 农
- 种
- 粮
- 经
- 畜
- 植
- 一
- 星
- 标
- 食
- 标
- 粮
- 标
- 蔬
- 蔬
- 烟

偏移量	长度	文件路径	标题	全文
0	11532	F:/data/MultiLang/Corpus/20100426/000@201004261	خەلق تورى -- 人民网维吾尔文	正文
57433	11532	F:/data/MultiLang/Corpus/20100426/000@201004261	خەلق تورى -- 人民网维吾尔文	正文
69320	10674	F:/data/MultiLang/Corpus/20100426/000@201004261	شىنجاڭ خەۋەرلىرى -- 人民网维吾尔文	正文
80102	9776	F:/data/MultiLang/Corpus/20100426/000@201004261	كىچىق مائى خەۋەرلەر -- 人民网维吾尔文	正文
89986	10393	F:/data/MultiLang/Corpus/20100426/000@201004261	تەسۋىرگىچە خەۋەرلىرى -- 人民网维吾尔文	正文
101320	4984	F:/data/MultiLang/Corpus/20100426/000@201004261	مۇشەبەت ئاسراش -- 人民网维吾尔文	正文
106304	3690	F:/data/MultiLang/Corpus/20100426/000@201004261	دىڭ -- 人民网维吾尔文	正文
109994	4304	F:/data/MultiLang/Corpus/20100426/000@201004261	قىزىق نۇقتا -- 人民网维吾尔文	正文
114298	4037	F:/data/MultiLang/Corpus/20100426/000@201004261	ئىدىيەسىڭىز -- 人民网维吾尔文	正文
118335	8526	F:/data/MultiLang/Corpus/20100426/000@201004261	سۈرەتلىك خەۋەرلەر -- 人民网维吾尔文	正文
127751	4385	F:/data/MultiLang/Corpus/20100426/000@201004261	خەلق تورى -- 人民网	正文

下一页 尾页 共有 1380 条记录, 当前第 1/92 页

باش بىت قىلىڭىز ساقلاڭتەھرىر خەت ساندۇقى ئالاقىلىشىڭىز بەتمەز كەز -- خەلق تورى
 خەۋەرلىرى خەلقئارا خەۋەرلەر شىنجاڭ: #x0D; #x0D; #x0A; روھى مەملىكەت
 خەۋەرلىرى ئىقتىساد جەمئىيەتمائارى پتەنتەر بى يەمىلەت ۋە مەدەنىيەتمۇھىت ئاسراشدىنقىزىق نۇقتاشەخسلەر سۈرەتلىك
 خەۋەرلەر ماتېرىيال مەخسۇس سەھىيە پەزەھەر ۋە ئەيدىز يېڭى خەۋەرلەر مەملىكەتلىك ئەرلەر ۋاسكىتبول كەسپىي بىرلەشمە
 مۇسابىقىسىنىڭ ئومۇمىي ھەل قىلغۇچ مۇسابىقىسىدە گۇاڭدۇڭ كوماندىسى شىنجاڭ كوماندىسىنى يەڭدى شىنجاڭ
 تۇرپاندا كەم كۆرۈلىدىغان بورانلىق ھاۋا رايى كۆرۈلدى گەنسۇ جاڭيېدا قۇم بورانلىق ھاۋا رايى كۆرۈلدى 1-پەسىلدە سىرتتىن
 شىنجاڭغا سېلىنغان مەبلەغ ئوخشاش ۋاقتىتتىكىدىن 4 يېرىم ھەسسە ئاشتى روسىيەنىڭ تۇنجى تۈركۈمدىكى ياردەم
 بۇيۇملىرى يېتىپ كەلدى شىنجاڭ قۇمۇلدا قۇم بورانلىق ھاۋا رايى كۆرۈلدى ئالدىن مەلۇمات: ئادىل ھوشۇر خەلق تورىدا
 شەنجاڭ خەلق ئۇنۋەن سەنئەت ئەمەك ئەلچى تەرەققىيات تەشەببۇسىدا سەنئەت ئىلىمى بارەدە شەنجاڭغا قارىتا



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



大数据论坛
BigdataBBS.com



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



大数据论坛
BigdataBBS.com



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**

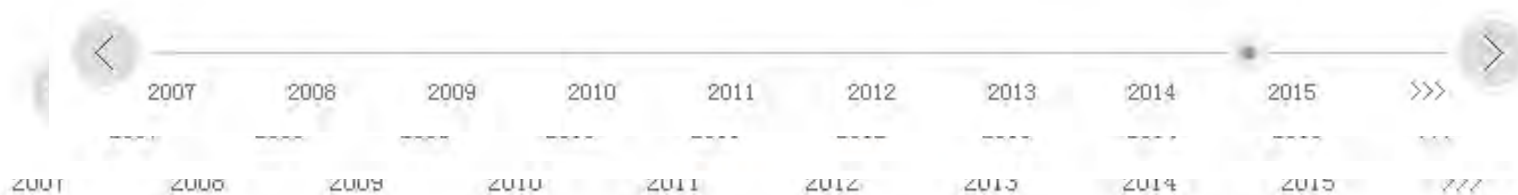
分析对象：刘振亚是谁

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



大数据论坛
BigdataBBS.com



大数据搜索挖掘

I 科学的大数据观

II 舆情分析的困局与迷思

III 大数据搜索与挖掘关键技术

IV 社交网络大数据语义应用实践





大数据语义应用实践

➔ 某大厦电力数据挖掘

➔ 95598 客服挖掘

➔ 国家电网头条

CDA 数据分析师
www.cda.cn

➔ 公安某局的探索



大数据论坛
BigdataBBS.com





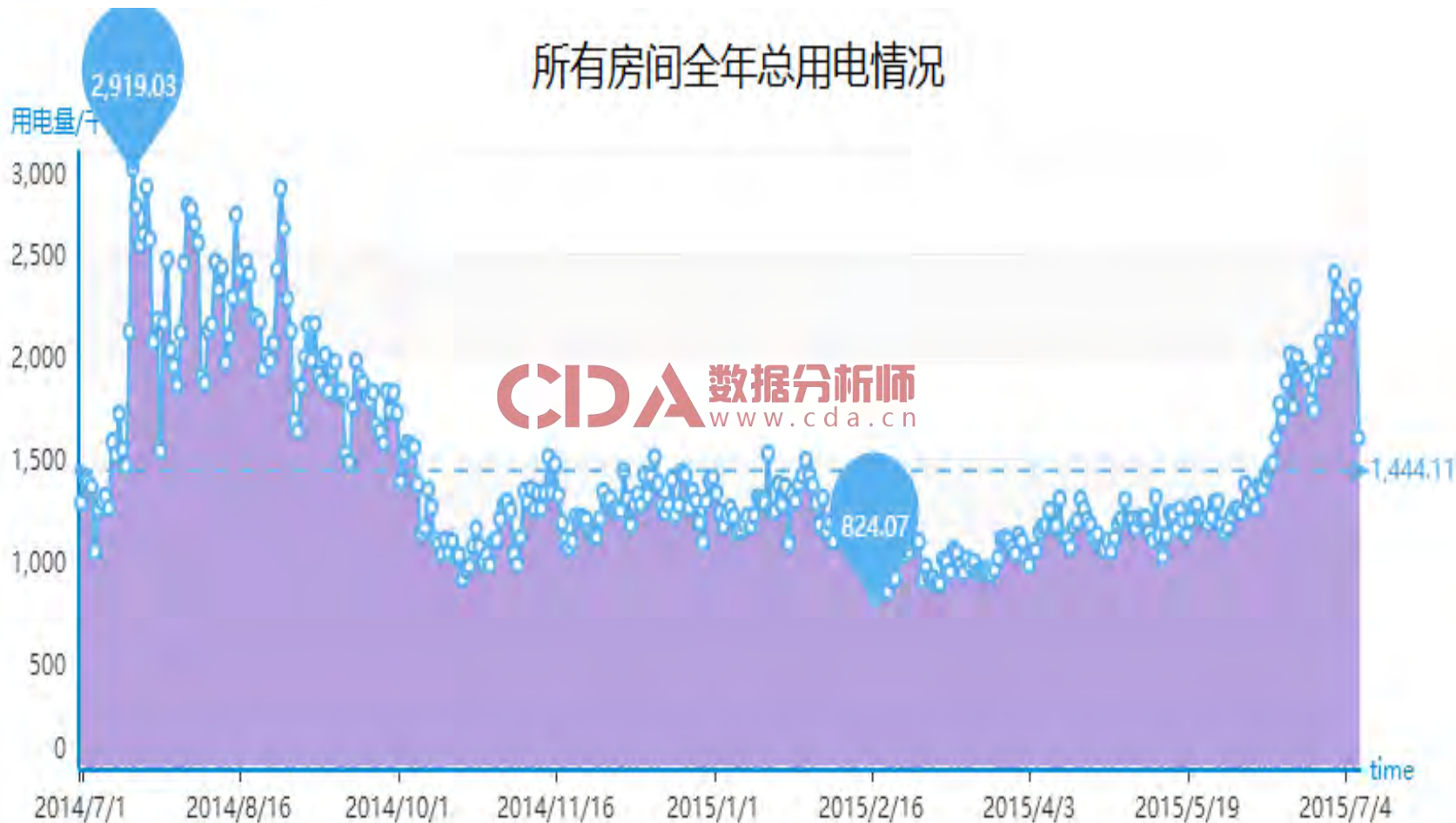
研究数据介绍

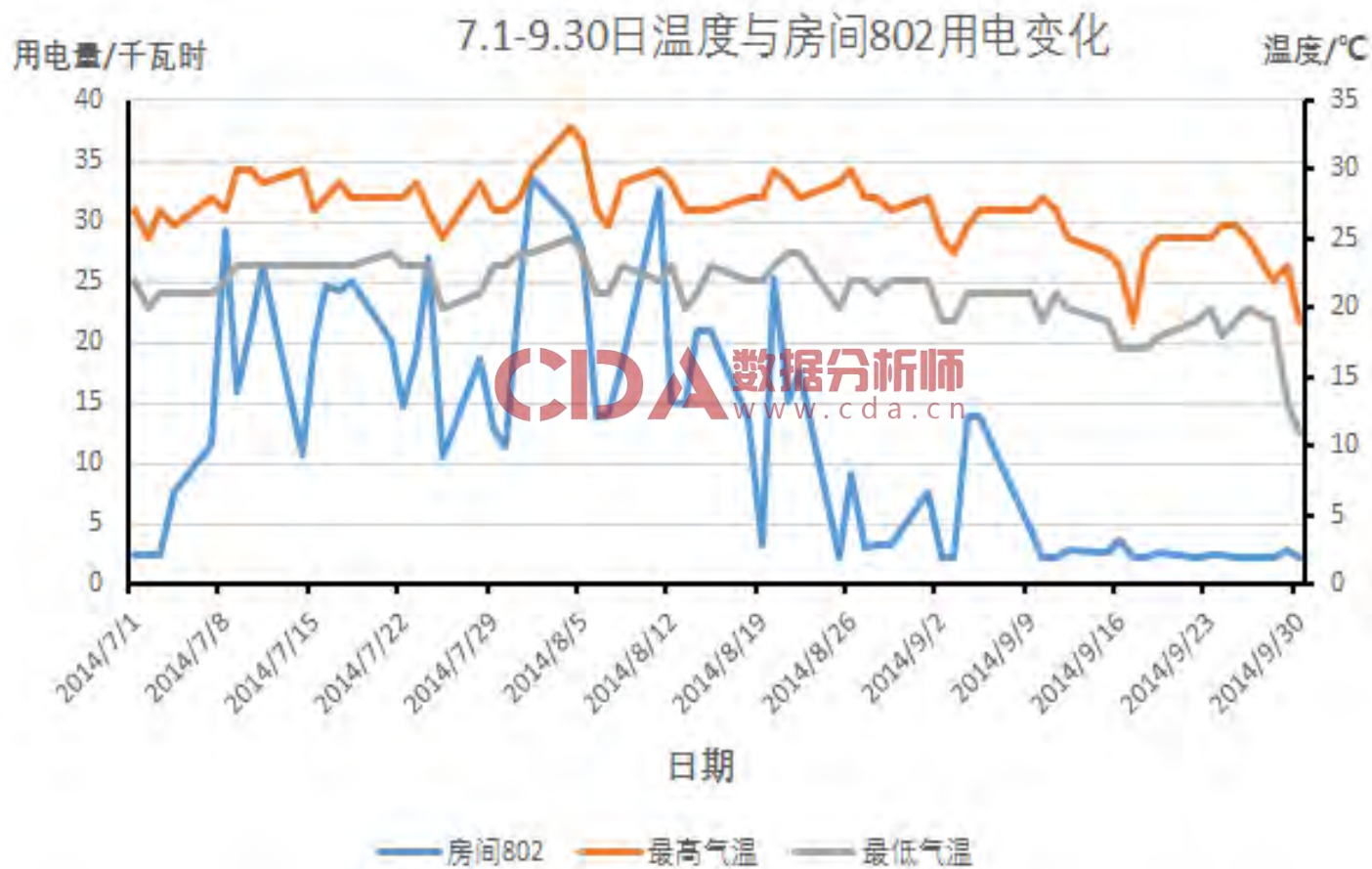
- 共**238**个房间
- 2014.07.01~2015.07.08, 共**373**天
- 期间工作日共**256**天
- **单日**用电量





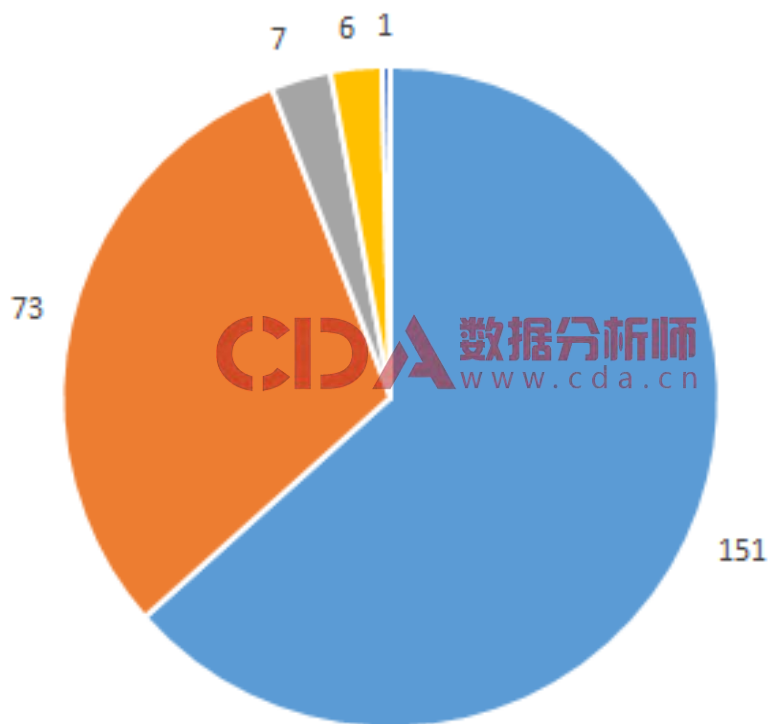
所有房间全年总用电情况







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



大数据论坛
BigdataBBS.com





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

CDA 数据分析师
www.cda.cn



大数据论坛
BigdataBBS.com





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1/3

CDA 数据分析师
www.cda.cn



大数据论坛
BigdataBBS.com





计算结果如下：共**76**个房间

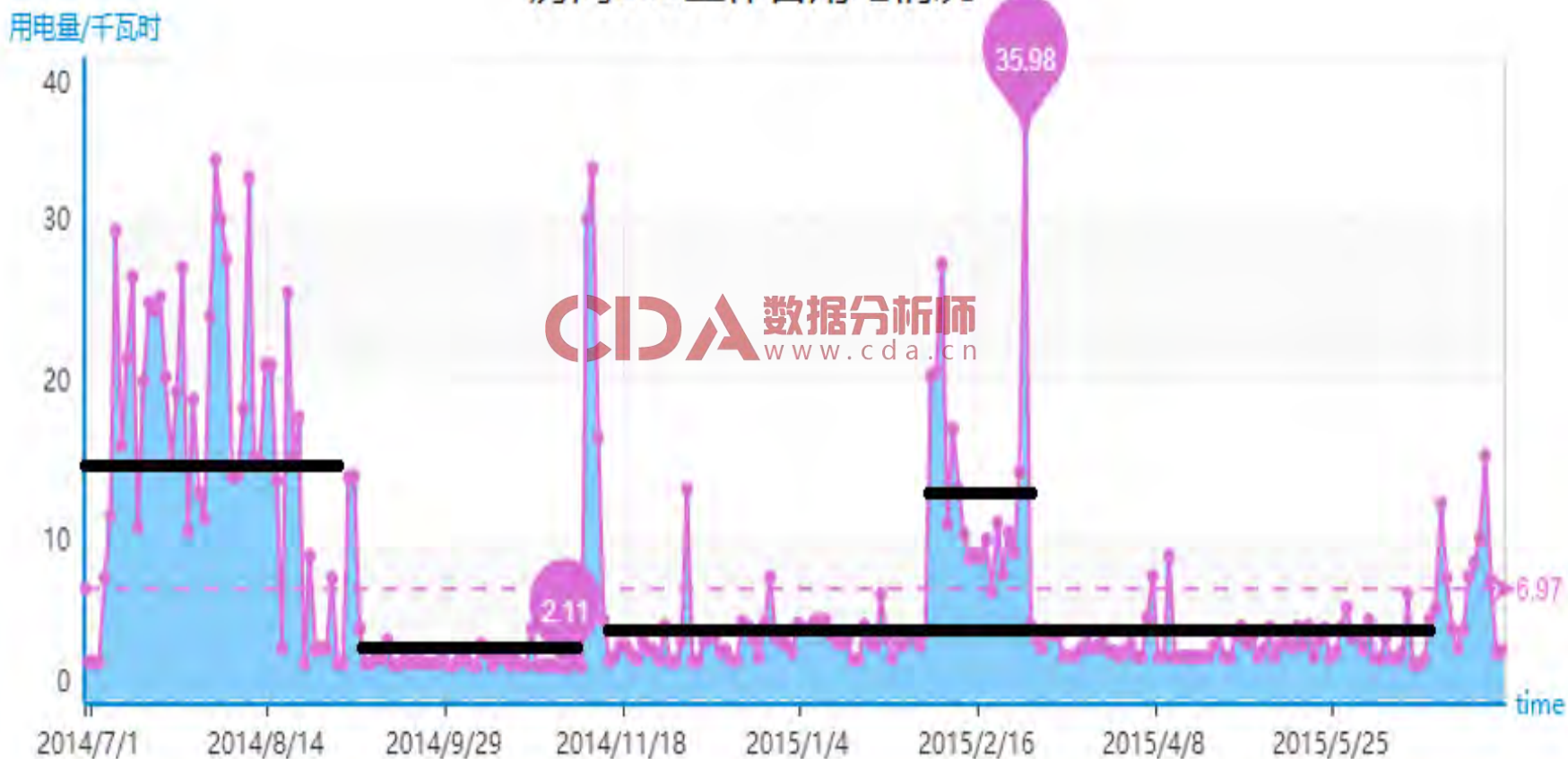
702	703	704	705	707	709	710	711	712	805
807	809	810	811	812	815	906	910	915	1001
1011	1107	1111	1201	1209	1211	1215	1301	1309	1310
1408	1409	1411	1504	1507	1509	1803	1806	1904	1905
2101	2104	2201	2203	2205	2206	2404	2405	2503	2505
2804	2805	2806	2901	2906	3002	3005	3006	3201	3203
3204	3205	3301	3304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

空置率=控制房间/总房间=**31.9%**





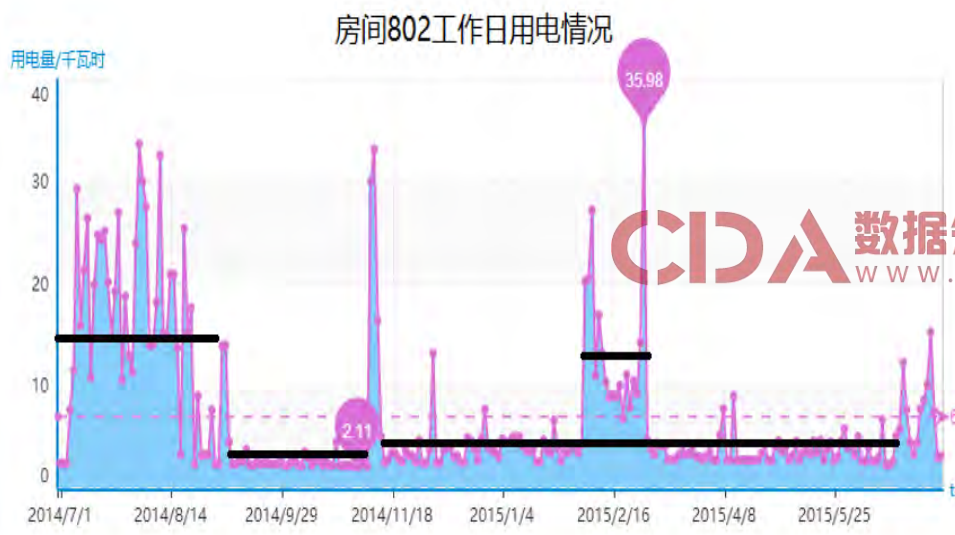
房间802工作日用电情况





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。

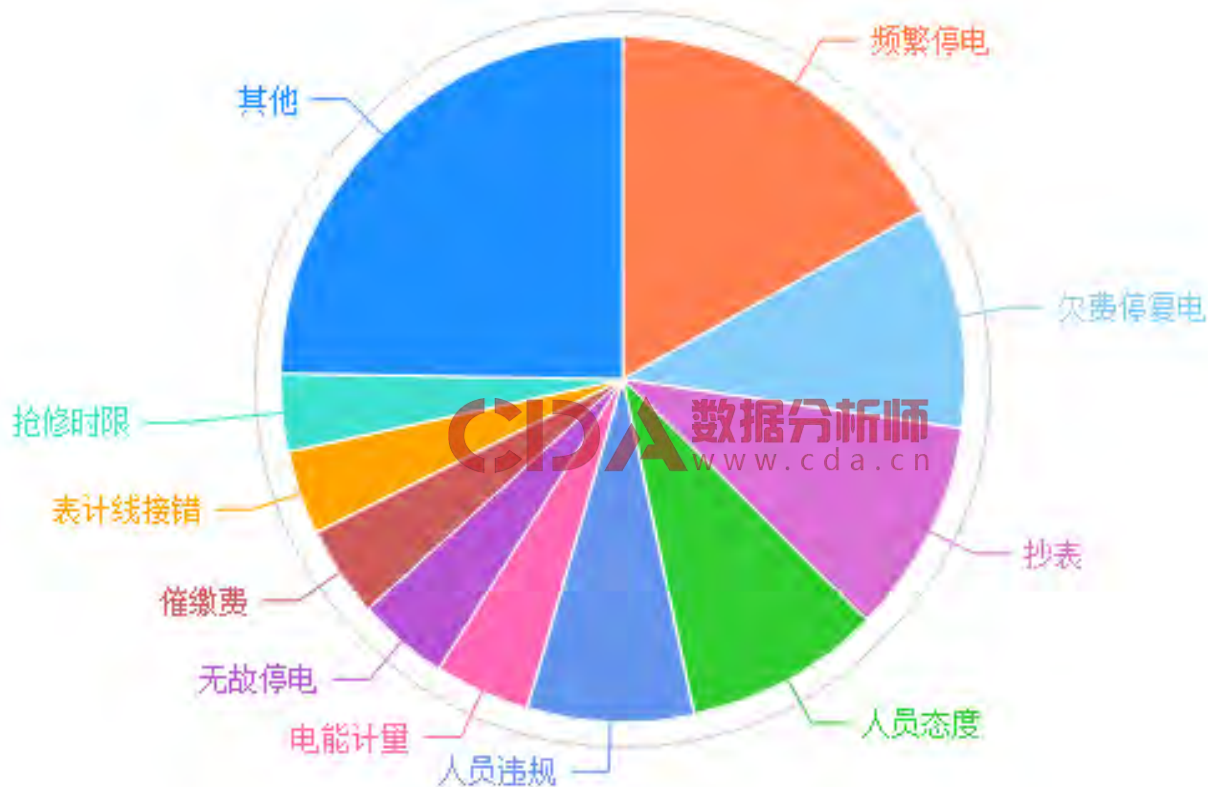


空调分析：
一台1P的空调功率制冷一般为1000瓦，制热约1100瓦；2P制冷为1600瓦，制热为1700瓦。按照每天开机8小时计。

~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





故障投诉内容中的关键词分析

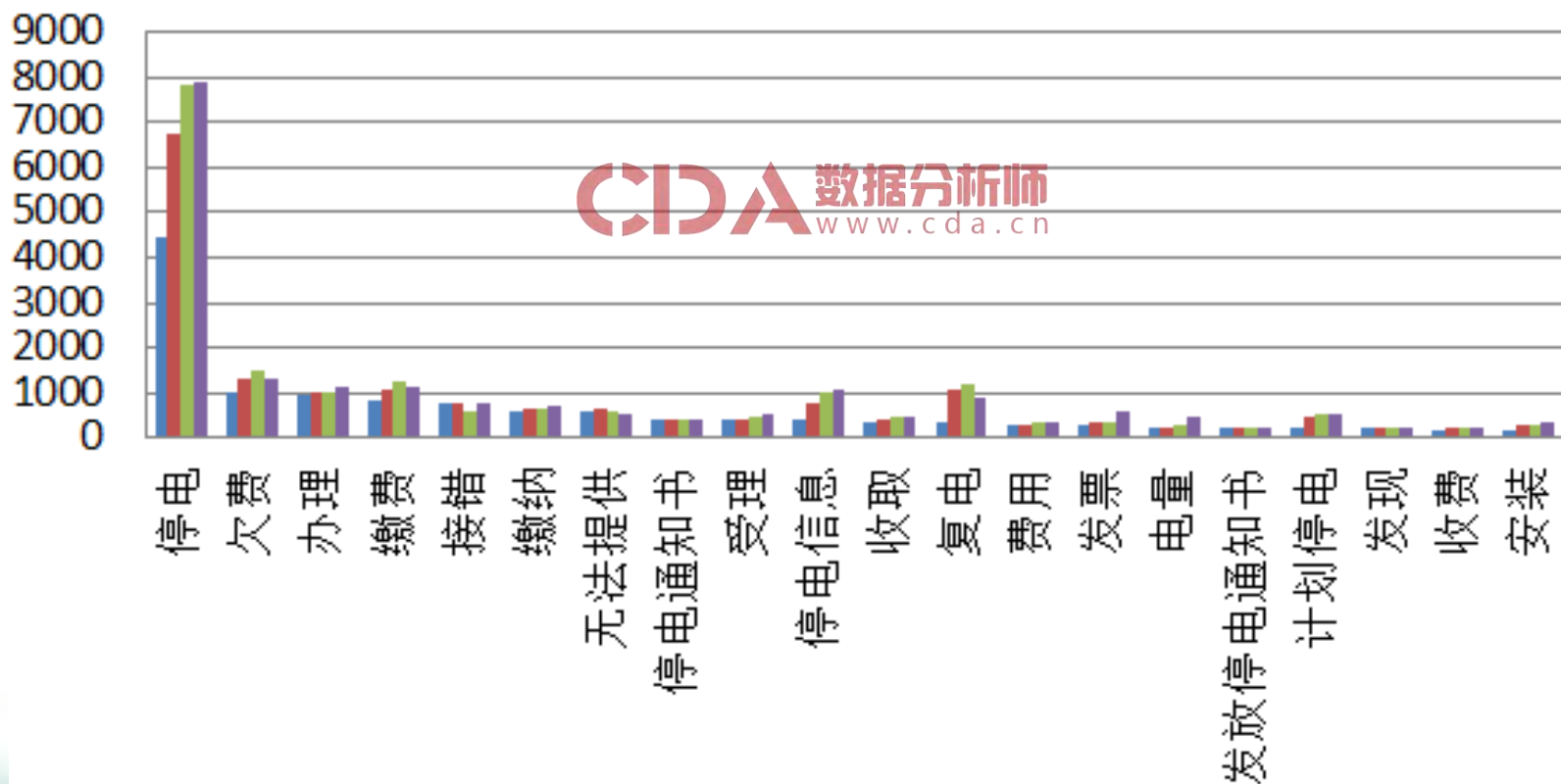


缴费
用电
复电
停电信息
告知客户
接错
缴纳
无法提供
客户无法
计划停电
停电通知书
收取
表员
费用
安装
电量
无故停电
施工
承诺时限
询问



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

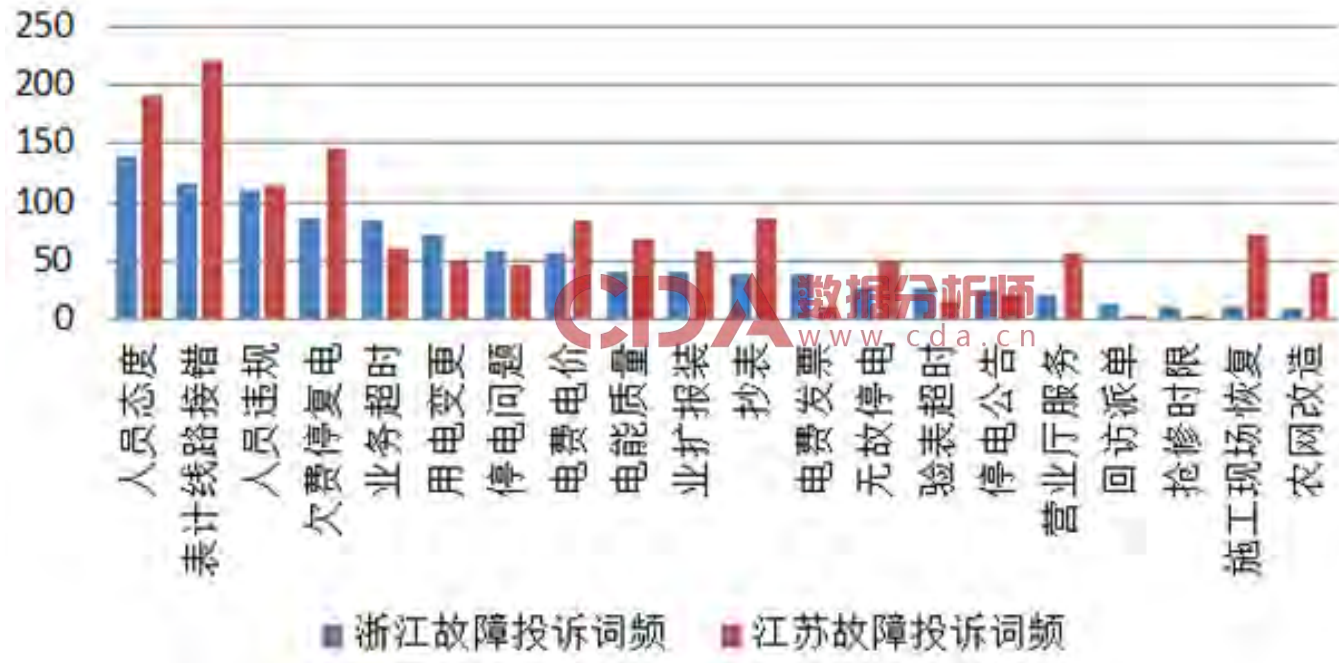


➔ 每天投诉的高峰期为10点和16点



大数据论坛
BigdataBBS.com





➔ “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

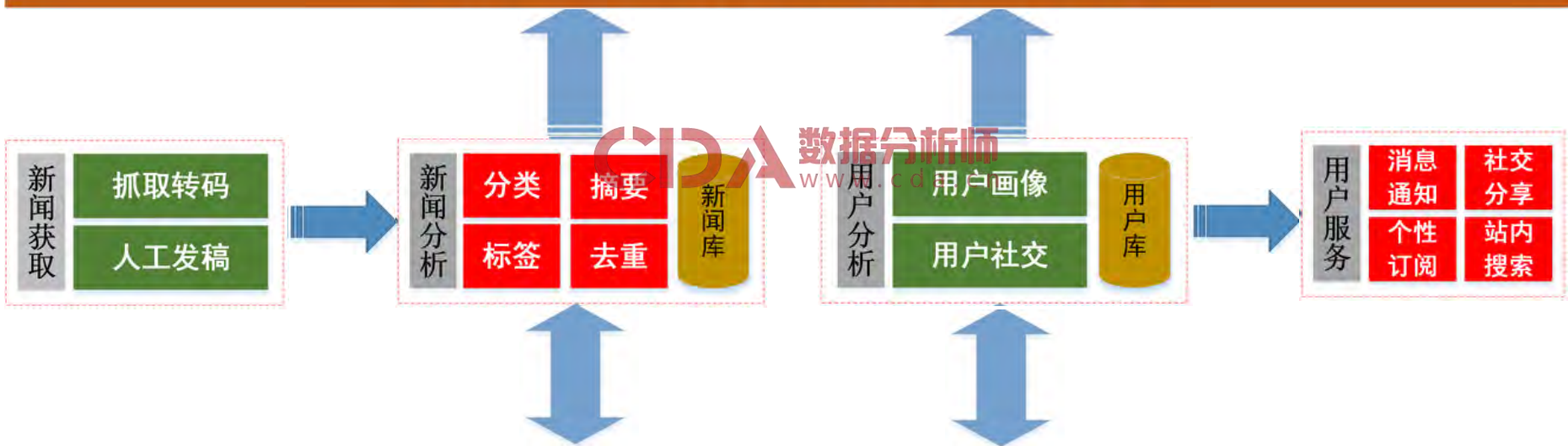
CDA 数据分析师
www.cda.cn



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



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



大数据论坛
BigdataBBS.com





个性化-因时因地因人而变

云计算 用户画像
想你所想 为你而来

精准推送
知识分享的平台

汇聚电网达人
你原创的就是头条



大数据论坛
BigdataBBS.com



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。

