JZSearch大数据精准搜索关键技术

JZSearch Big Data Precise Search Key Technology

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

2016.10

纲要

大数据搜索

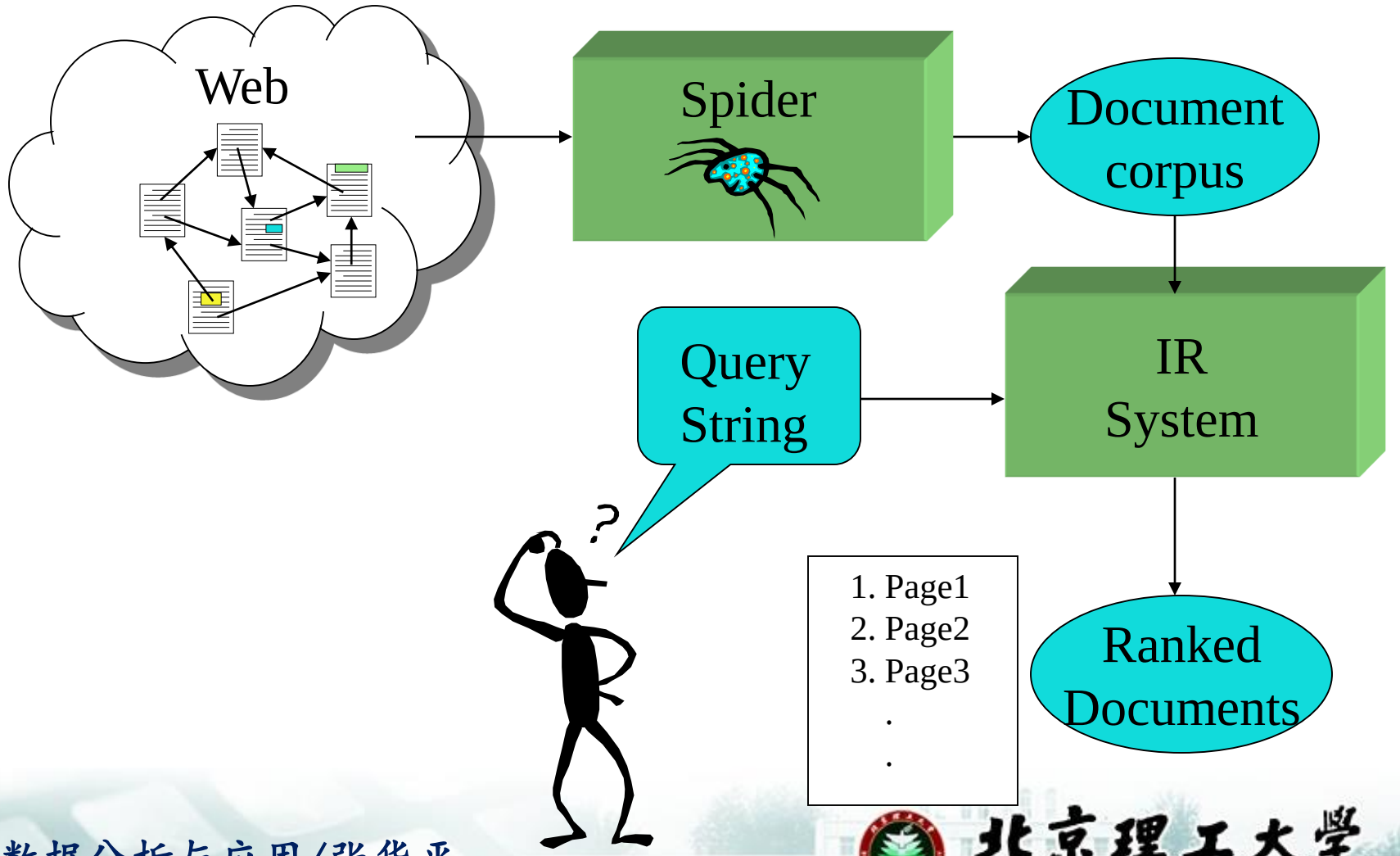
I 简单Web搜索引擎基本原理

II 当前搜索技术的主要不足

III JZSearch大数据精准搜索引擎

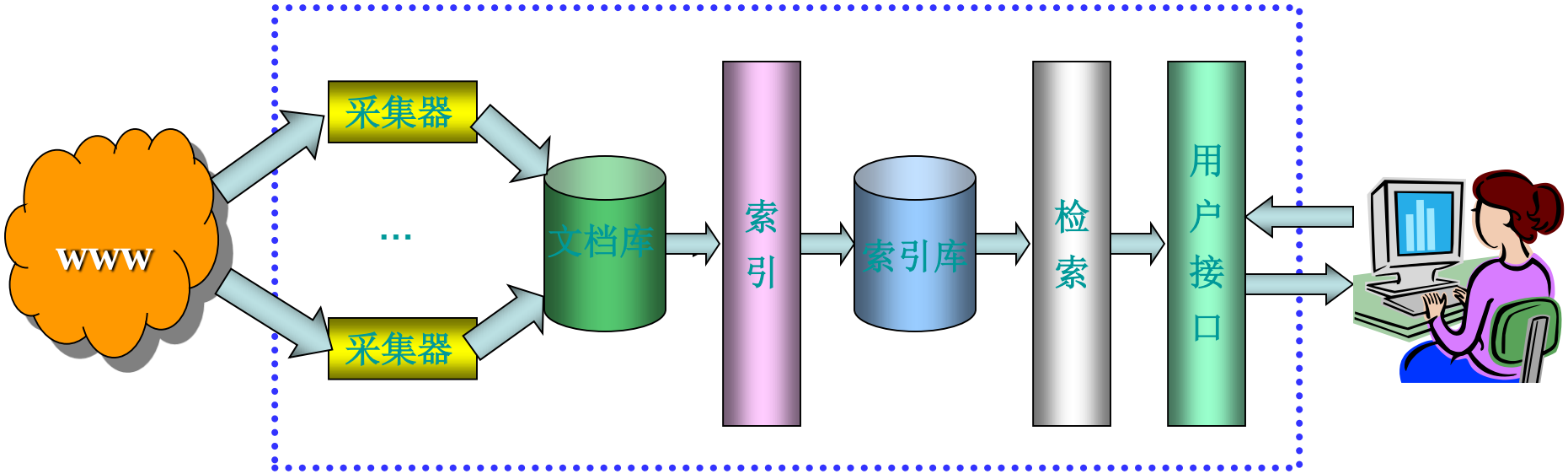
IV JZSearch大数据搜索应用案例

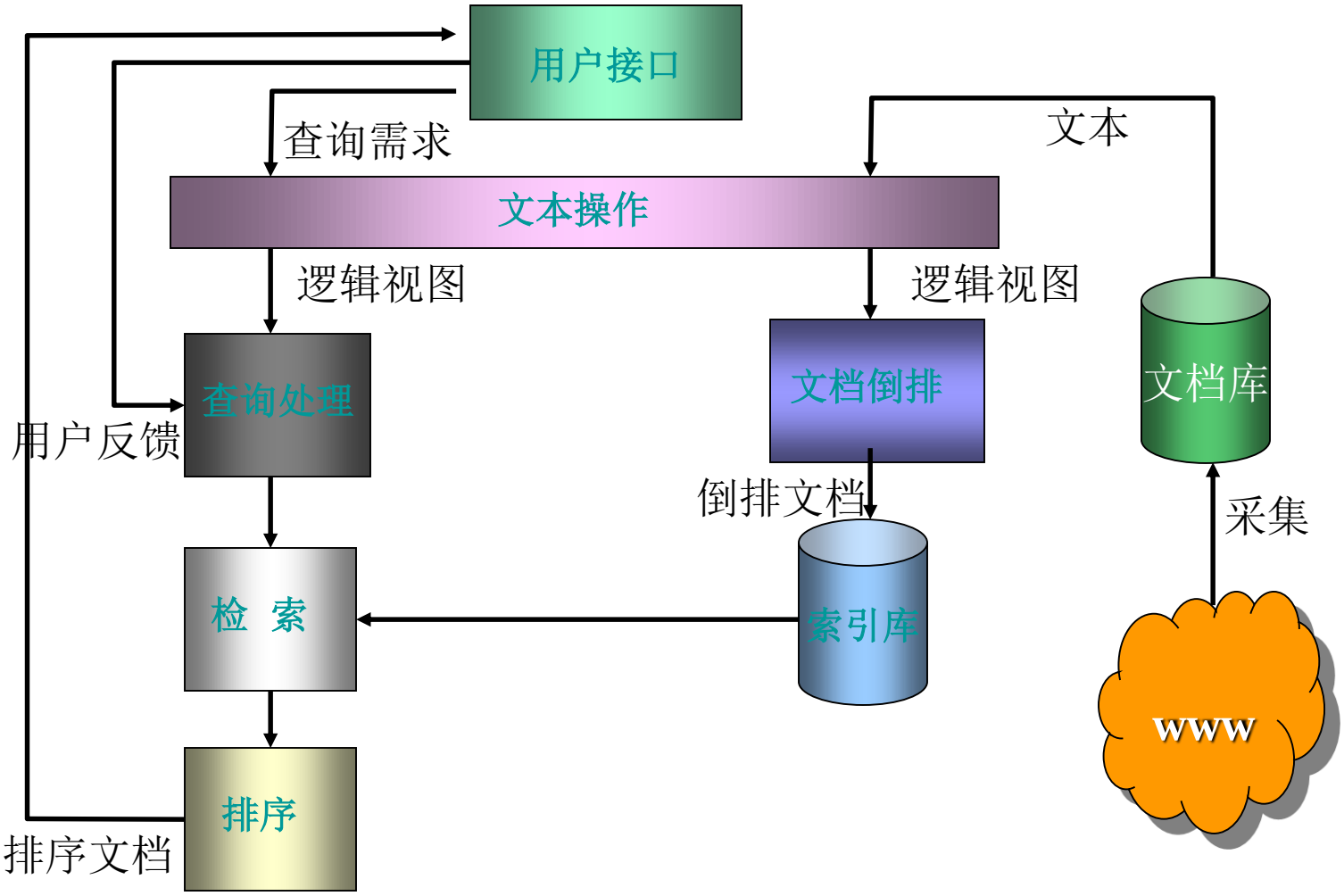
Web Search Engine Using IR



最简单的搜索引擎

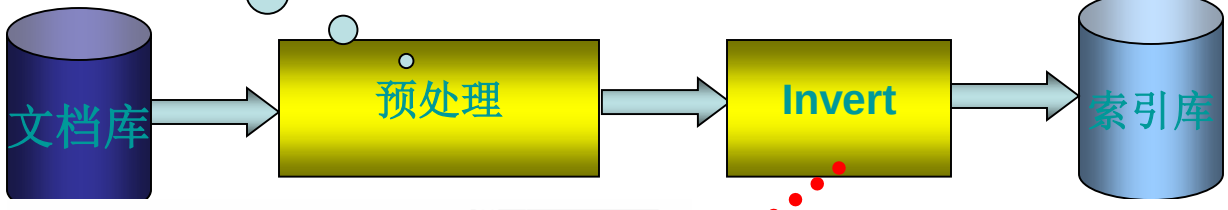
搜索引擎结构




用户接口
查询需求
文本
文本操作
逻辑视图
逻辑视图
查询处理
文档倒排
文档库
用户反馈
倒排文档
采集
检 索
索引库
www
排序
排序文档

倒排索引

文档分析，编码识别，词语切分，去停用词等



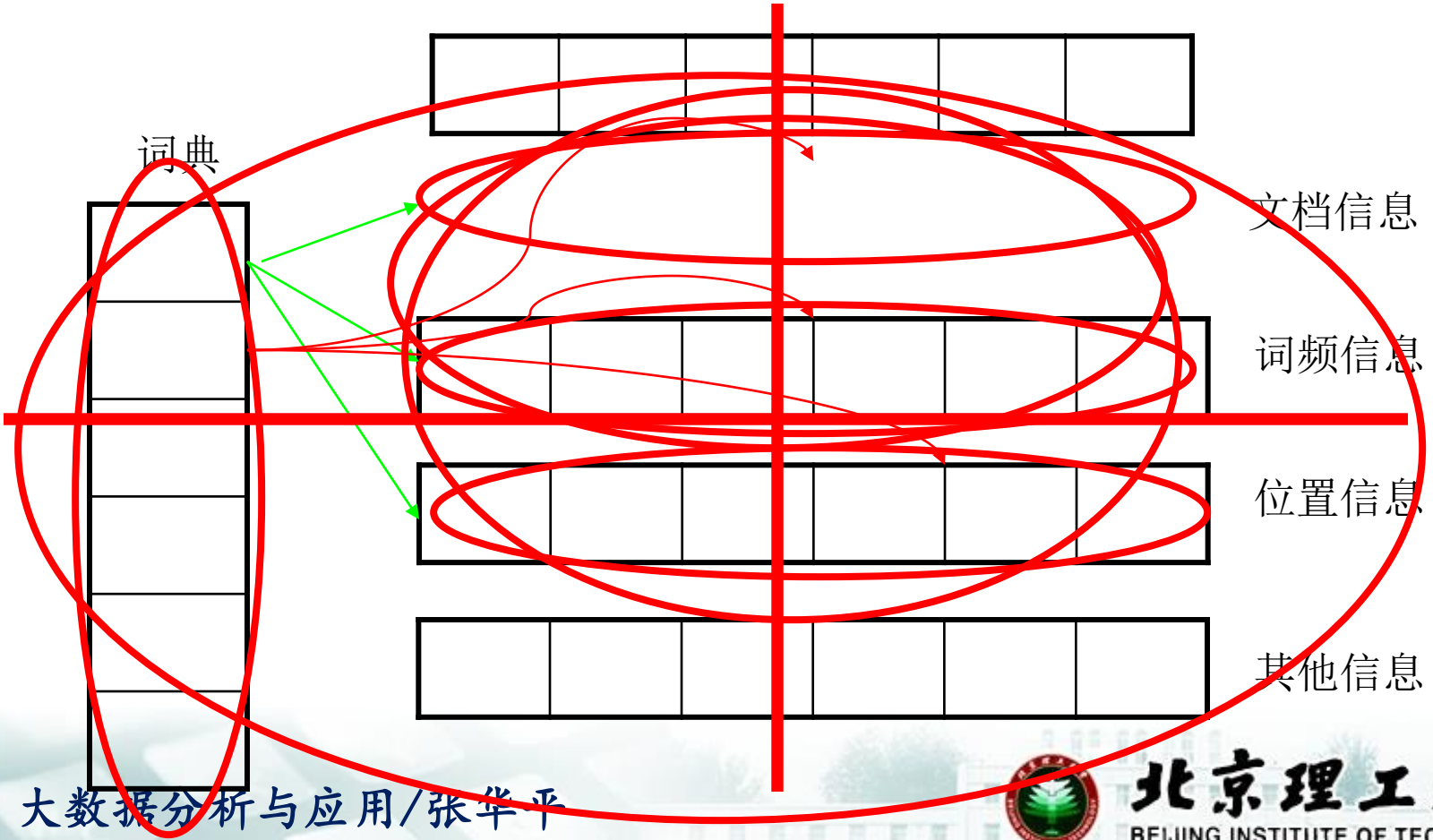
Document	Text
1	Pease porridge hot, pease porridge cold,
2	Pease porridge in the pot,
3	Nine days old.
4	Some like it hot, some like it cold,
5	Some like it in the pot,
6	Nine days old.

文档倒排，生成Inverted Files

Number	Term	(Document; Words)
1	cold	(1; 6), (4; 8)
2	days	(3; 2), (6; 2)
3	hot	(1; 3), (4; 4)
4	in	(2; 3), (5; 4)
5	it	(4; 3, 7), (5; 3)
6	like	(4; 2, 6), (5; 2)
7	nine	(3; 1), (6; 1)
8	old	(3; 3), (6; 3)
9	pease	(1; 1, 4), (2; 1)
10	porridge	(1; 2, 5), (2; 2)
11	pot	(2; 5), (5; 6)
12	some	(4; 1, 5), (5; 1)
13	the	(2; 4), (5; 5)

倒排索引结构

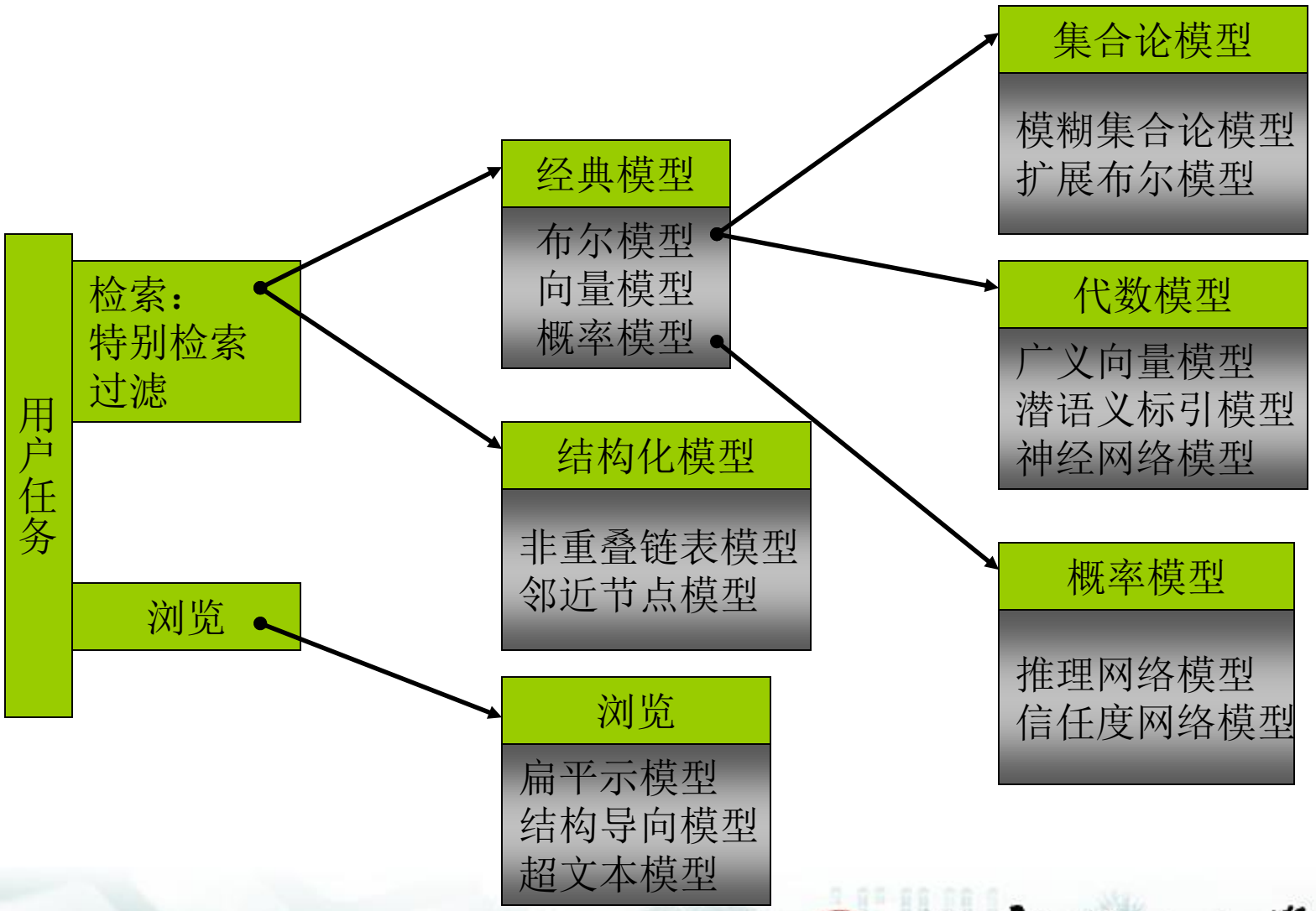
- 索引文件结构：不管怎么变化基本都由这几部分组成



倒排索引的挑战

- Fast：如何快速构造倒排索引？
- Fast：如何构造索引使检索尽可能的快？
- Minimized：如何使索引尽可能小？
- Online Indexing：如何构造动态文档集的索引(增量，差量索引和索引更新)？
- Scalable：如何在资源有限的情况下构造海量数据的索引？

信息检索

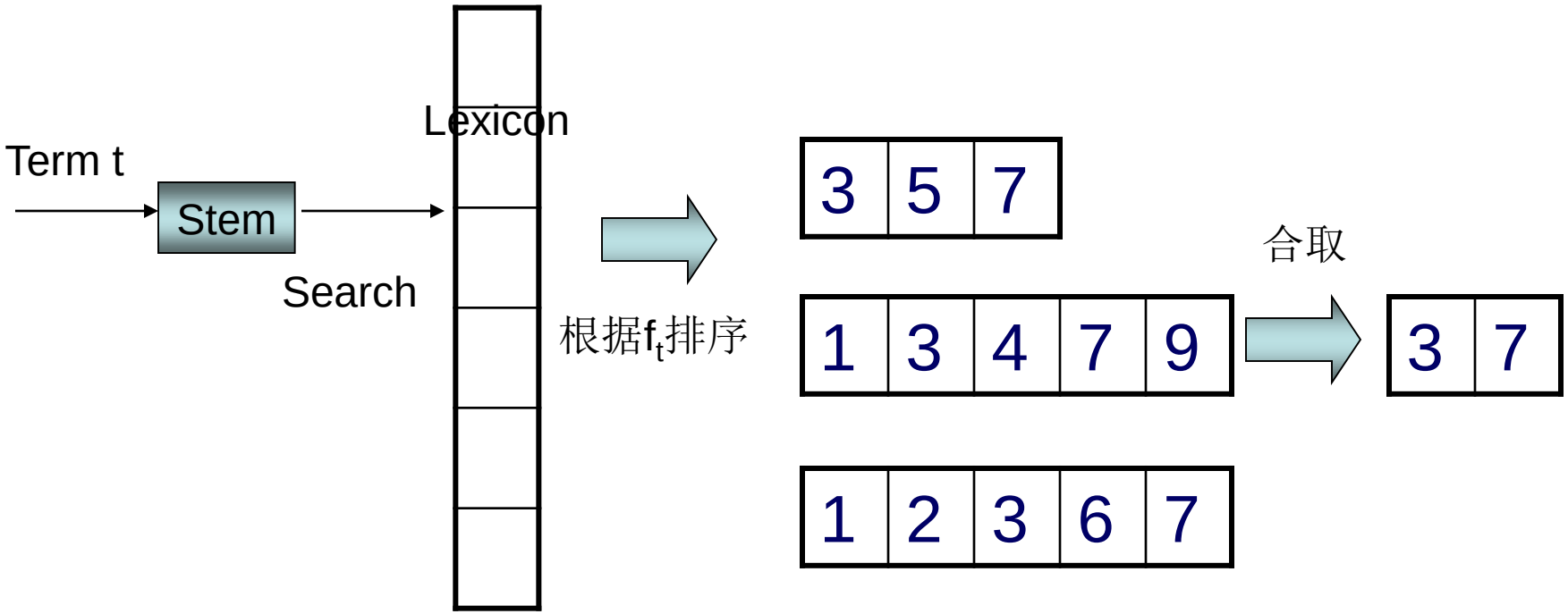


信息检索模型分类

布尔查询

- 一种简单的检索模型，建立在经典的集合论和布尔代数的基础上。
- 遵循两条基本规则： 每个索引词在一篇文档中只有两种状态：出现或不出现，对应权值为 0或1。
- 查询是由三种布尔逻辑运算符 and，or，not 连接索引词组成的布尔表达式。

合取布尔查询处理



Ranking和信息检索

- 经典布尔模型能精确判断文档是否出现某一查询，但并不能给出相关性排序
- 信息检索是一个查询Q和文档D_d相似度计算过程：

$$M(Q,D_d)=Q\bullet D_d=\sum_{t=1}^{n} wq,t\bullet wd,t=\sum_{t\in Q} wq,t\bullet wd,t$$

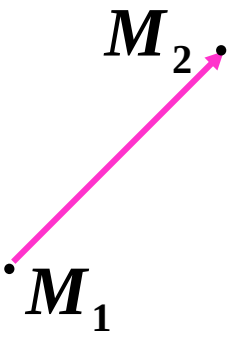
- 存在一个问题：当Q包含常用词t时，那些包含比较多t的文档总是排在前面，其他的非常用词根本不起作用，所以需要根据inverse document frequency (IDF)计算Term的权重 w_t：

$$w_t=\frac{1}{f_t}$$

f_t是包含term t的文档数

向量

$\vec{x}$　　$\overrightarrow{M_1M_2}$



- 向量(矢量，vector)：既有大小又有方向的量，通常用有向线段表示，记作　或者
- 考虑从空间坐标系原点出发(其他向量可以平移到原点出发)的向量 $\vec{x}$，终点坐标为 $<x_1,x_2,\cdots,x_n>$，我们称之为一个n维向量
- 向量的运算：加、减、倍数、内积

$$\vec{x} \pm \vec{y} =< x_1 \pm y_1, x_2 \pm y_2, ..., x_n \pm y_n >$$

$$\lambda\vec{x} =< \lambda x_1, \lambda x_2, ..., \lambda x_n >$$

$$\vec{x} \bullet \vec{y} = \sum_{i=1}^{n} x_i y_i$$

向量的模、距离和夹角

$$|\vec{x}| = \left\|\vec{x}\right\| = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$$

- 向量的模(大小):

- 向量的(欧氏)距离

$$dist(\vec{x}, \vec{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2}$$

- 夹角α

$$\cos\alpha = \frac{\vec{x} \bullet \vec{y}}{|\vec{x}| \times |\vec{y}|}$$

向量空间模型

- 向量空间模型(Vector Space Model，VSM)是康奈尔大学 Salton等人上世纪70年代提出并倡导，原型系统SMART*
- term独立性假设：term在文档中的出现是独立、互不影响的。
- 查询和文档都可转化成term及其权重组成的向量表示，都可以看成空间中的点。向量之间通过距离计算得到查询和每个文档的相似度。

文档-标引项矩阵(Doc-Term Matrix)

n篇文档，m个标引项构成的矩阵$\boldsymbol{A}_{m*n}$，每列可以看成每篇文档的向量表示，同时，每行也可以可以看成标引项的向量表示。

$$A_{m*n} = \begin{array}{c} \\ t_1 \\ t_2 \\ \vdots \\ t_m \end{array} \begin{array}{c} \begin{array}{cccc} d_1 & d_2 & \dots & d_n \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \end{array}$$

一个例子

- 查询q：(<2006,1>,<世界杯,2>)
- 文档d_1：(<2006,1>,<世界杯,3>,<德国,1>,<举行,1>)
- 文档d_2：(<2002,1>,<世界杯,2>,<韩国,1>,<日本,1>,<举行,1>)

	d_1	d_2	q
2002	0	1	0
2006	1	0	1
世界杯	3	2	2
德国	1	0	0
韩国	0	1	0
日本	0	1	0
举行	1	1	0

一个例子(续)

- 查询和文档进行向量的相似度计算：
  - 采用内积：
    - ⑩文档d_1与q的内积：1*1+3*2=7
    - ⑩文档d_2与q的内积：2*2=4
  - 夹角余弦：
    - ⑩文档d_1与q的夹角余弦：$\frac{7}{\sqrt{12\times 5}} \approx 0.90$
    - ⑩文档d_2与q的夹角余弦：$\frac{4}{\sqrt{5\times 8}} \approx 0.63$

典型的静态信息检索

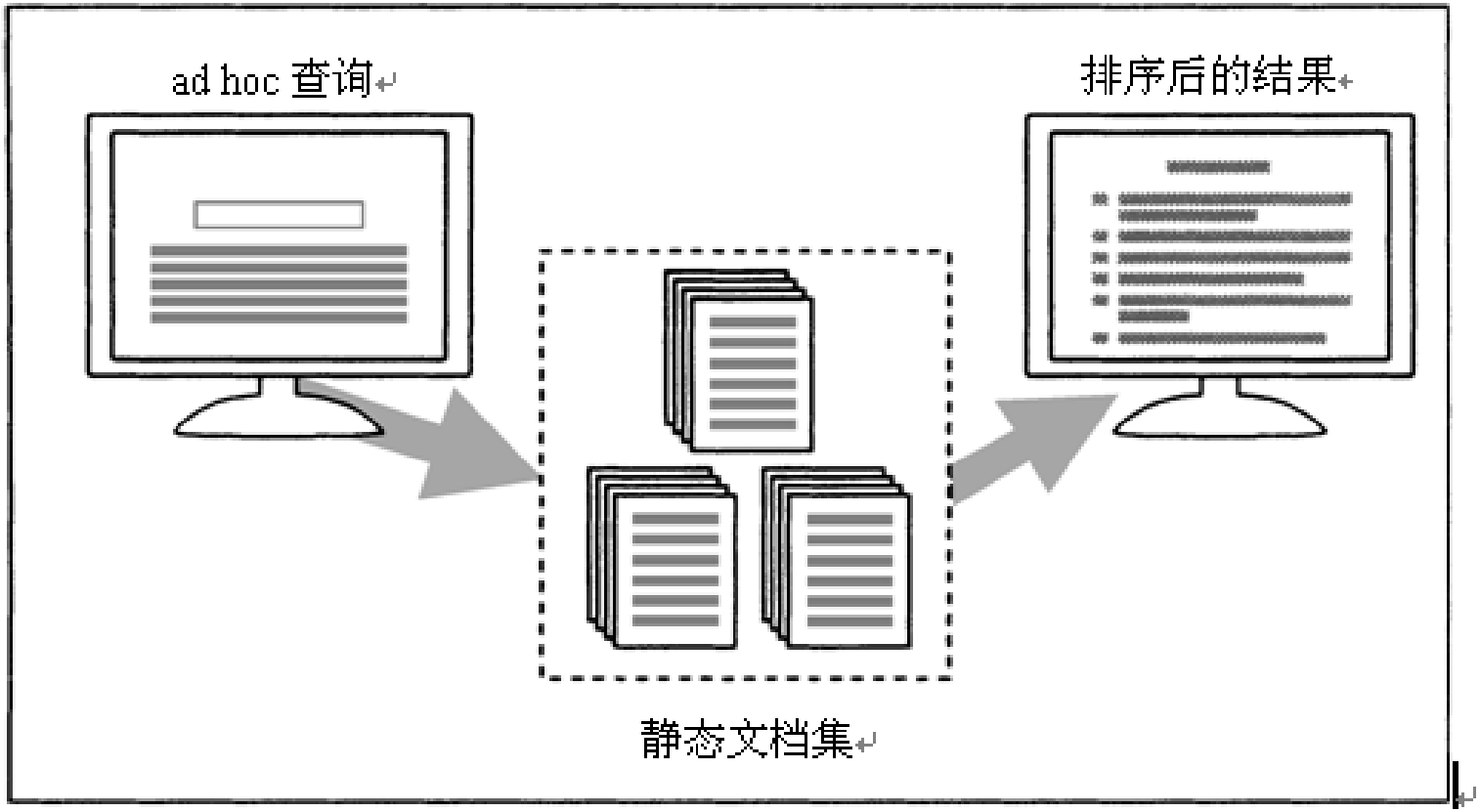
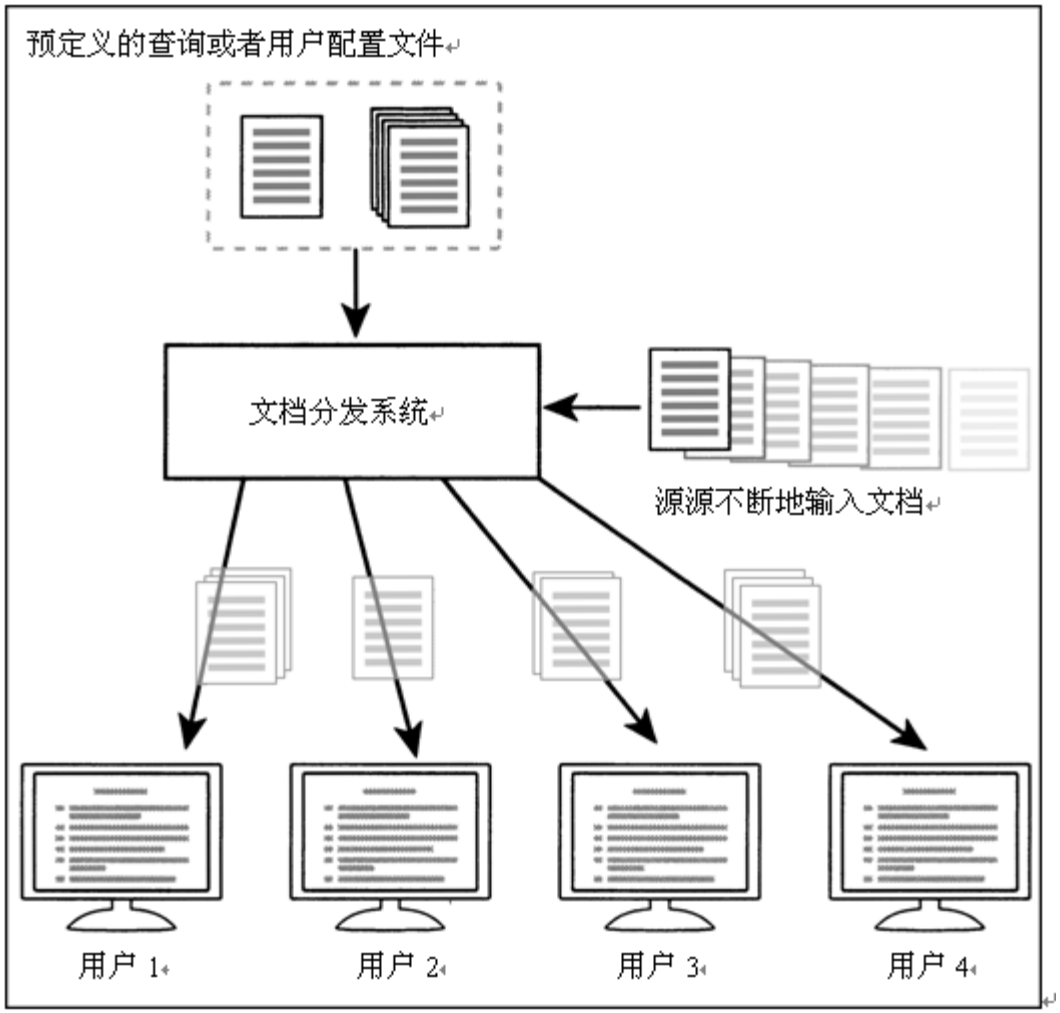


图 1-1　文档检索

典型的动态信息流过滤过程



信息检索vs.信息流过滤的差异

- 文档集：
  - 规模有限，相对静态 vs. 规模巨大、动态变化
- 用户需求：
  - 多变多样，海量用户 vs. 相对稳定、用户受限
- 技术方法：
  - 对文档进行倒排索引并检索 vs. 查询生成自动机，进行快速扫描过滤

信息检索vs.信息流过滤的共同

- 同一理论支持；
- 相同的支持技术：自然语言处理、机器学习、大规模架构设计
- 小规模的信息安全可以直接规约为信息检索问题。

信息检索需研究的五大问题

- 搜索排序问题:信息检索模型，解决搜索的根本问题;
- 效果提升问题:信息检索策略，提高效果的策略与技巧;
- 性能问题：倒排、索引压缩;
- 架构问题：并行、分布式、云计算;
- 拓展问题：多语言、跨语言等。

纲要

大数据搜索

I 简单Web搜索引擎基本原理

II 当前搜索技术的主要不足

III JZSearch大数据精准搜索引擎

IV JZSearch大数据搜索应用案例

当前不足初探

信息重复冗余

关键字: 白蜡　搜索类型: 按全文搜索　栏目类别: 全部类别

地区: -地区-　行业: -请选择-　日期: 近一月　搜 索

当前位置: 中国搜标网 >>普通搜索　共16条记录　没有查到所需的信息？

信息标题	地区	日期	收藏
关于驻马店市园林管理局绿地养护项目采购公告	河南省	2012-09-17	收藏
关于驻马店市园林管理局绿地养护项目招标公告	河南省	2012-09-17	收藏
询价公告（绿化苗木、12-14-37）	新疆	2012-09-17	收藏
神新能源公司2012年秋季绿化工程招标公告	新疆	2012-09-14	收藏
滦河东路绿化景观改造工程（六标段）滦河东路绿化景观改造工程（六标段）招...	河北省	2012-09-11	收藏
滦河东路绿化景观改造工程（六标段）招标公告	河北	2012-09-11	收藏
黄骅港联轴节、维修车间维修工具采购-采购公告	河北	2012-09-07	收藏
兴辰道道路及配套管线工程监理招标公告	天津市	2012-09-07	收藏
宜白路道路及配套管线工程监理招标	天津市	2012-09-07	收藏
宁夏银川市贺兰县农业综合开发办秋季造林苗木竞争性谈判招标公告	宁夏	2012-09-06	收藏
宁夏银川市贺兰县农业综合开发办秋季造林苗木竞争性谈判招标公告	宁夏	2012-09-06	收藏

当前不足初探

- 白腊-石蜡 是专业同义词，缺乏专业知识关联

当前不足初探

搜索禁用语禁而不止，搜索结果完全无关

通用搜索 vs. 大数据专业搜索

- 召回率不是通用搜索的考量范围，只关注P@10, P@30：
  - 体量大，一大遮百丑
  - 查询：中国证监会所有的负面信息；
- 体量大，但返回给用户的有限；
  - 返回给用户的信息不足2000条
  - 查询：新疆极端势力的动态情报；
- 专业垂直整合缺乏
  - 专业业务的逻辑不在通用搜索的业务范畴内；
  - 中关村东路80号5楼1203室的邮编是多少？联想集团的信息汇总

关于作业：
最近三年(ACL, SIGIR,CIKM,WWW, SMP)
最经典的Tutorial

纲要

大数据搜索

I 简单Web搜索引擎基本原理

II 当前搜索技术的主要不足

III JZSearch大数据精准搜索引擎

IV JZSearch大数据搜索应用案例

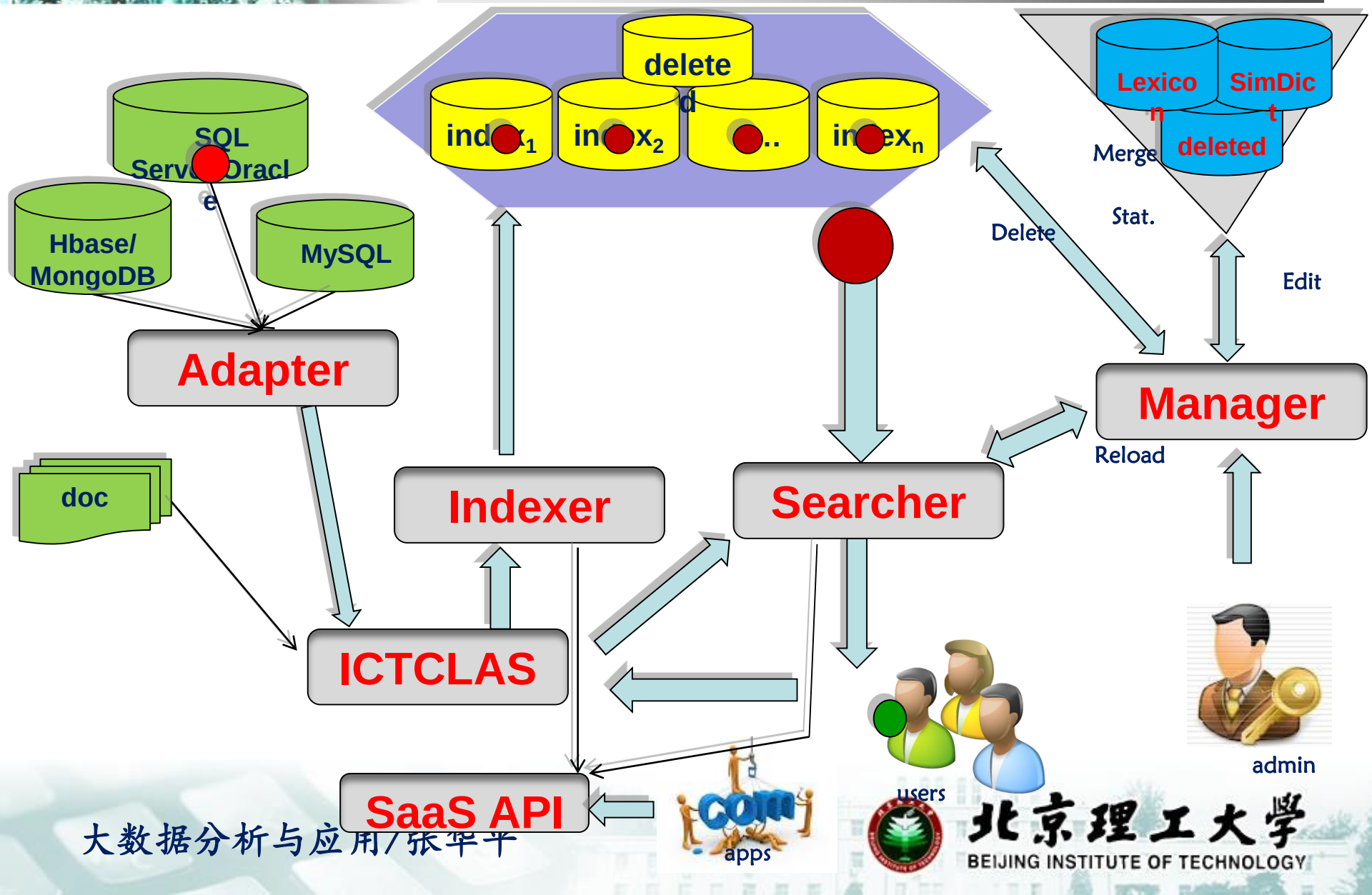
JZSearch大数据精准搜索

- 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；

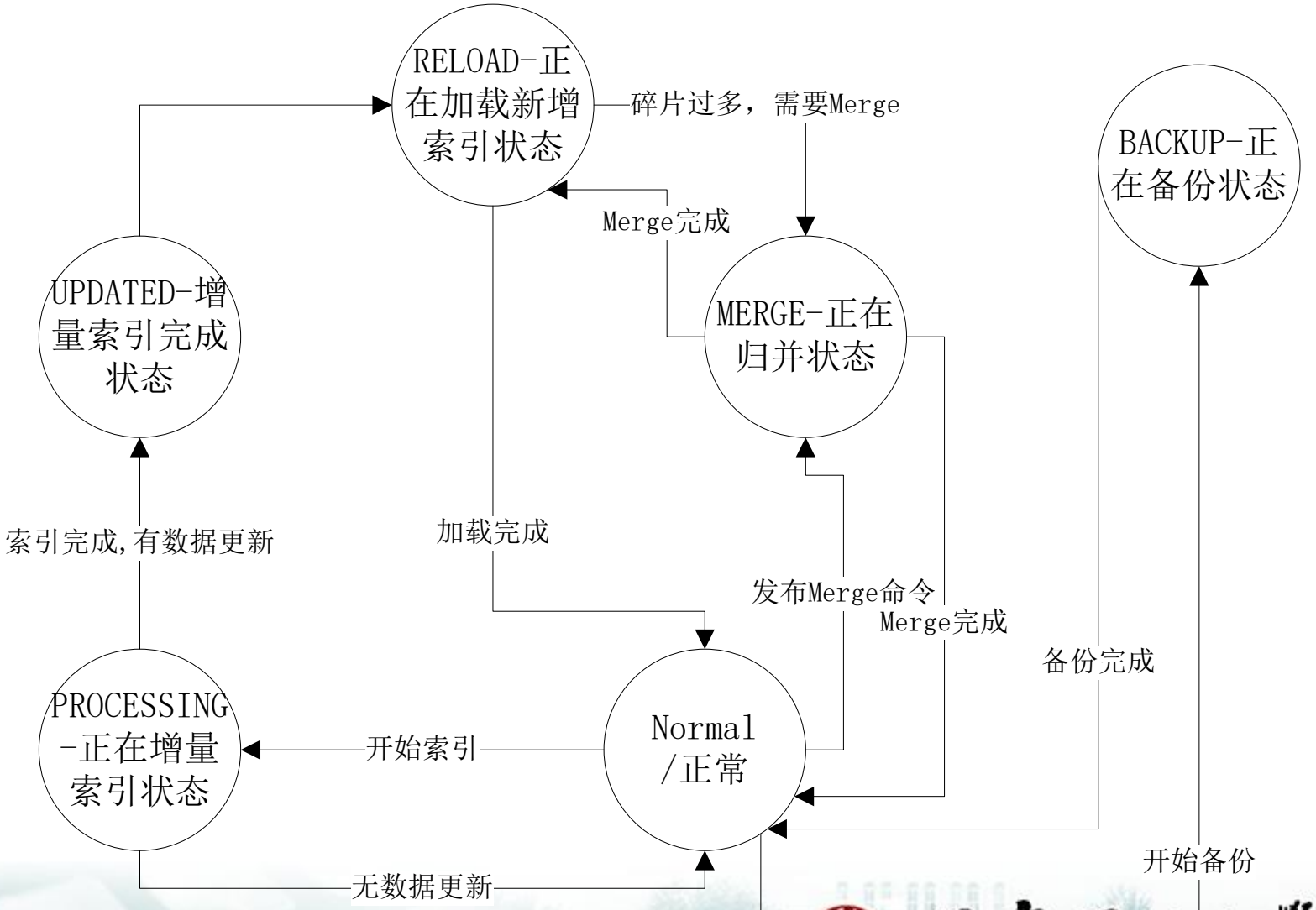
JZSearch内部语法示例

- [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 状态转移图



索引压缩技术

- 倒排索引中倒排表的大小可以通过齐普夫分布来估计。
- 它描述了自然语言中每个词的词频分布情况：如果把词出现的频率按由大到小排列，那么每个词的词频与它排序序号的乘积是一个常量。
- 常量取1时齐普夫分布中前5个词的词频分布如下 ：

排序序号	频率	常量
1	1.00	1
2	0.50	1
3	0.33	1
4	0.25	1
5	0.20	1

索引压缩:哈夫曼编码

- 在英文中，e的出现概率很高，而z的出现概率则最低。当利用哈夫曼编码对一篇英文进行压缩时，e极有可能用一个位(bit)来表示，而z则可能花去25个位（不是26）。用普通的表示方法时，每个英文字母均占用一个字节（byte），即8个位。二者相比，e使用了一般编码的1/8的长度，z则使用了3倍多。倘若我们能实现对于英文中各个字母出现概率的较准确的估算，就可以大幅度提高无损压缩的比例。

JZSearch索引压缩

- 字节对齐压缩（Byte-Aligned）
  - 0-63　　00xxxxxx
  - 64-(16K-1)　01xxxxxx xxxxxxxx
  - 16K-(4M-1)　10xxxxxx xxxxxxxx xxxxxxxx
  - 4M-(1G-1)　11xxxxxx xxxxxxxx xxxxxxxx xxxxxxxx
  -
  - 0　　00000000
  - 1　　00000001
  - …　　…
  - 63　　00111111
  - 64　　01000000 01000000
  - 65　　01000000 01000001
- 通过使用字节边界来实现BA压缩，运行时会付出些许代价，但可以获得一定的压缩比。BA算法易于实现，而且压缩比比较高（使用停用词后，压缩文件大小占未压缩索引文件大小的比例大约是15%）。

索引压缩

- 固定长度的索引压缩示例
- 给定任意一个索引词*t*1，我们考虑其索引的条目，假定*t*1在文档1、文档3、文档7、文档70和文档250中出现。
- BA压缩使用最高的两个比特位来表明存储该值需要的字节数。对于前4个值，只需要一个字节就可以表示；对于最后一个值——180，需要两个字节来表示。需要注意的是：我们只需要计算倒排表中每一项的差。最后的差值为：250-70＝180。我们需要计算所有的最终数值。这些数值和压缩后的相应比特字符串见下表。（字节对齐方式压缩）

值	压缩的比特串
1	00 000001
2	00 000010
4	00 000100
63	00 111111
180	01 000000 10110100

索引压缩

- 固定长度的索引压缩示例
- 在没有压缩之前，倒排表中的每条记录需要4字节，5条记录总共需要20个字节。这些数值和它们对应压缩前的比特字符串详见下表。（基准线：压缩前）

值	未压缩的比特串
1	00000000 00000000 00000000 00000001
3	00000000 00000000 00000000 00000011
7	00000000 00000000 00000000 00000111
70	00000000 00000000 00000000 01000110
250	00000000 00000000 00000000 11111010

- 结论：在这个例子中，未压缩的数据需要160比特，使用BA压缩后仅需要48比特。

索引词的处理

- 索引词典动态设计：
  - 优点：可以快速索引新词
  - 缺点：动态词典组织效率远低于静态词典

- JZSearch采用静态词典，主要包含：
  - 汉语常用词
  - 英文常用词：词形自动转换
  - 英文未知词表示：n-gram拆分
  - 数字组合：n-gram拆分

纲要

大数据搜索

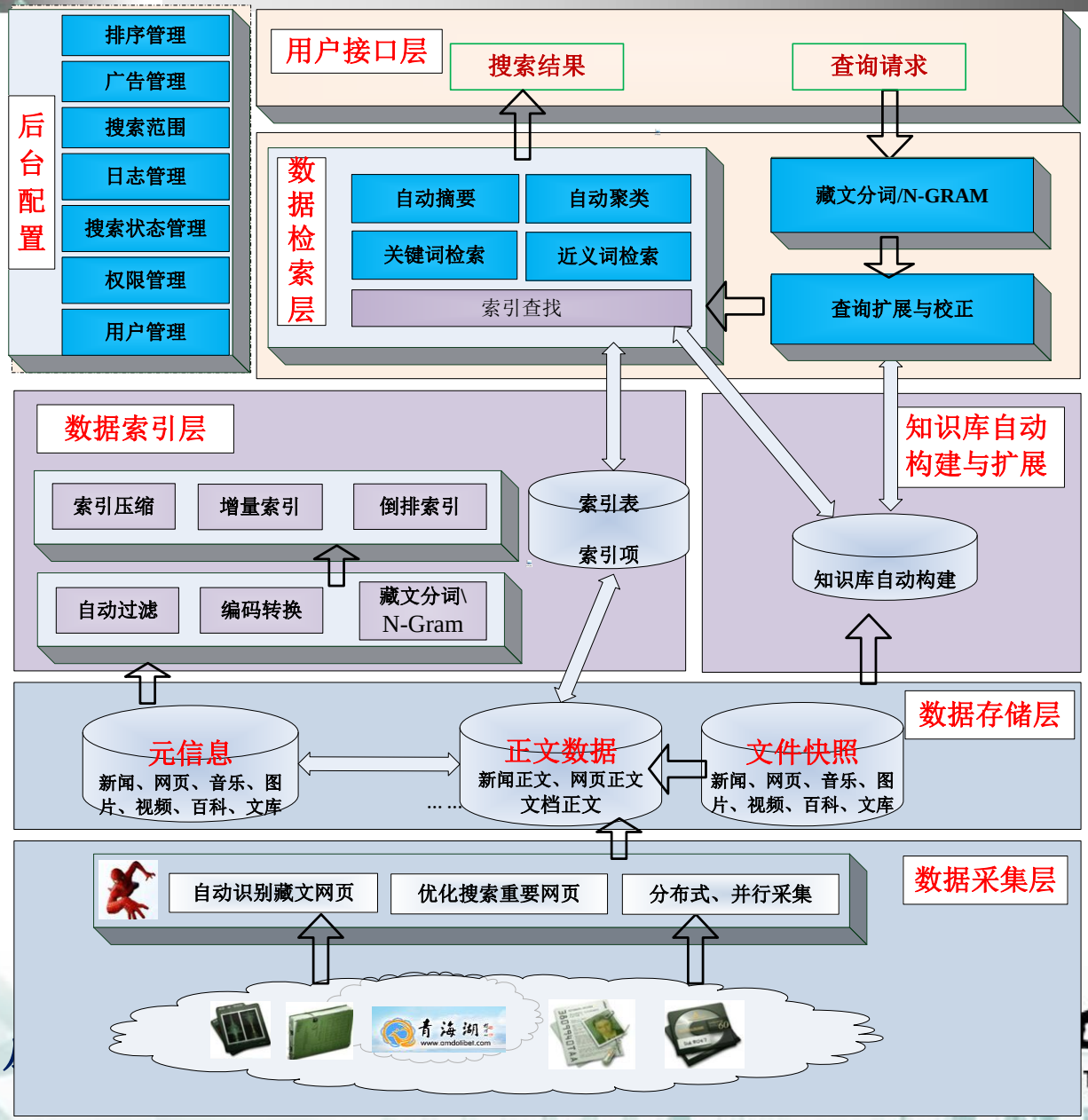
I 简单Web搜索引擎基本原理

II 当前搜索技术的主要不足

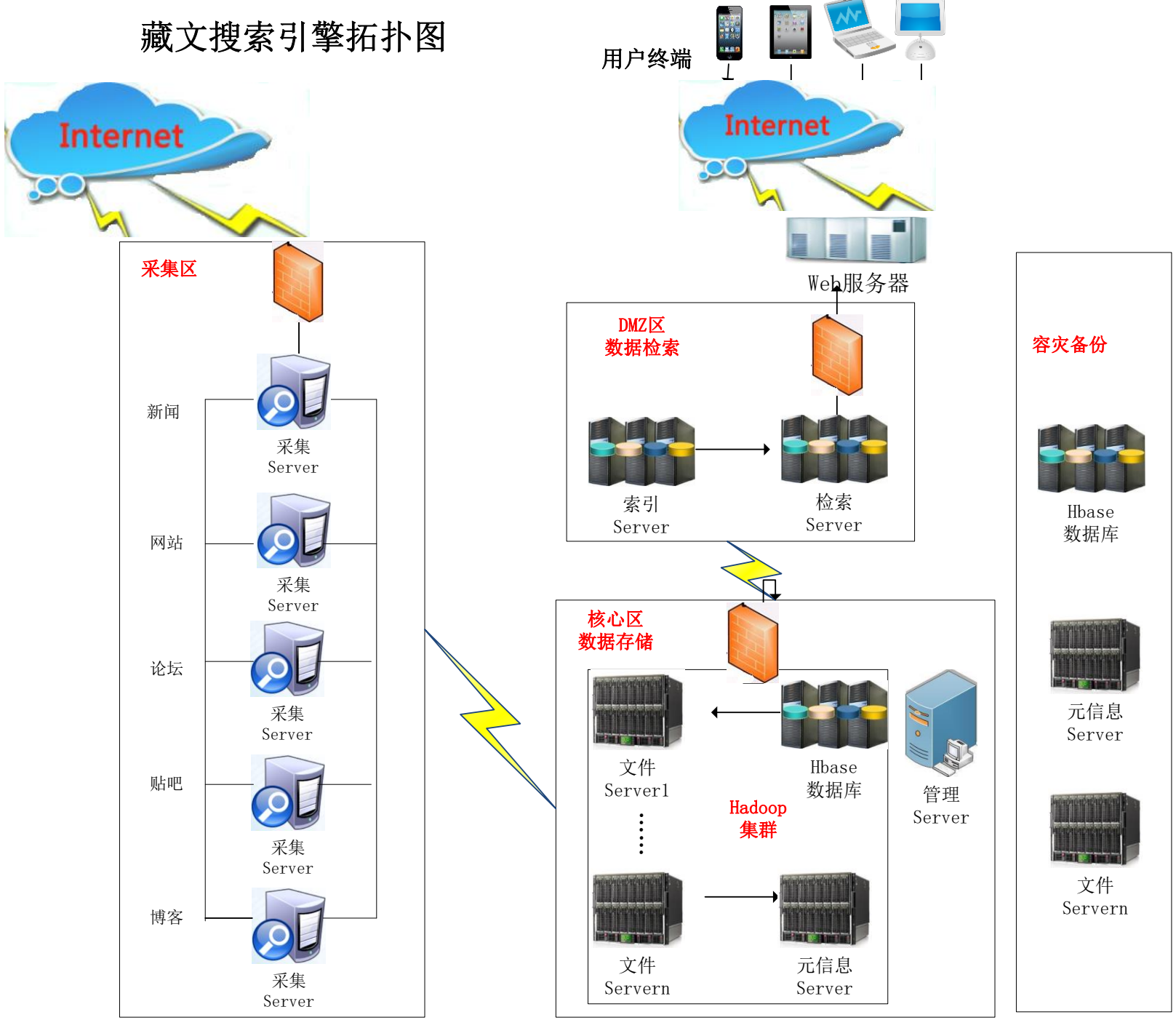
III JZSearch大数据精准搜索引擎

IV JZSearch大数据搜索应用案例

软件架构


后台配置
排序管理
广告管理
搜索范围
日志管理
搜索状态管理
权限管理
用户管理
用户接口层
搜索结果
查询请求
数据检索层
自动摘要
自动聚类
关键词检索
近义词检索
索引查找
藏文分词/N-GRAM
查询扩展与校正
数据索引层
索引压缩
增量索引
倒排索引
自动过滤
编码转换
藏文分词\ N-Gram
索引表
索引项
知识库自动构建与扩展
知识库自动构建
数据存储层
元信息
新闻、网页、音乐、图片、视频、百科、文库
正文数据
新闻正文、网页正文
文档正文
文件快照
新闻、网页、音乐、图片、视频、百科、文库
数据采集层
自动识别藏文网页
优化搜索重要网页
分布式、并行采集

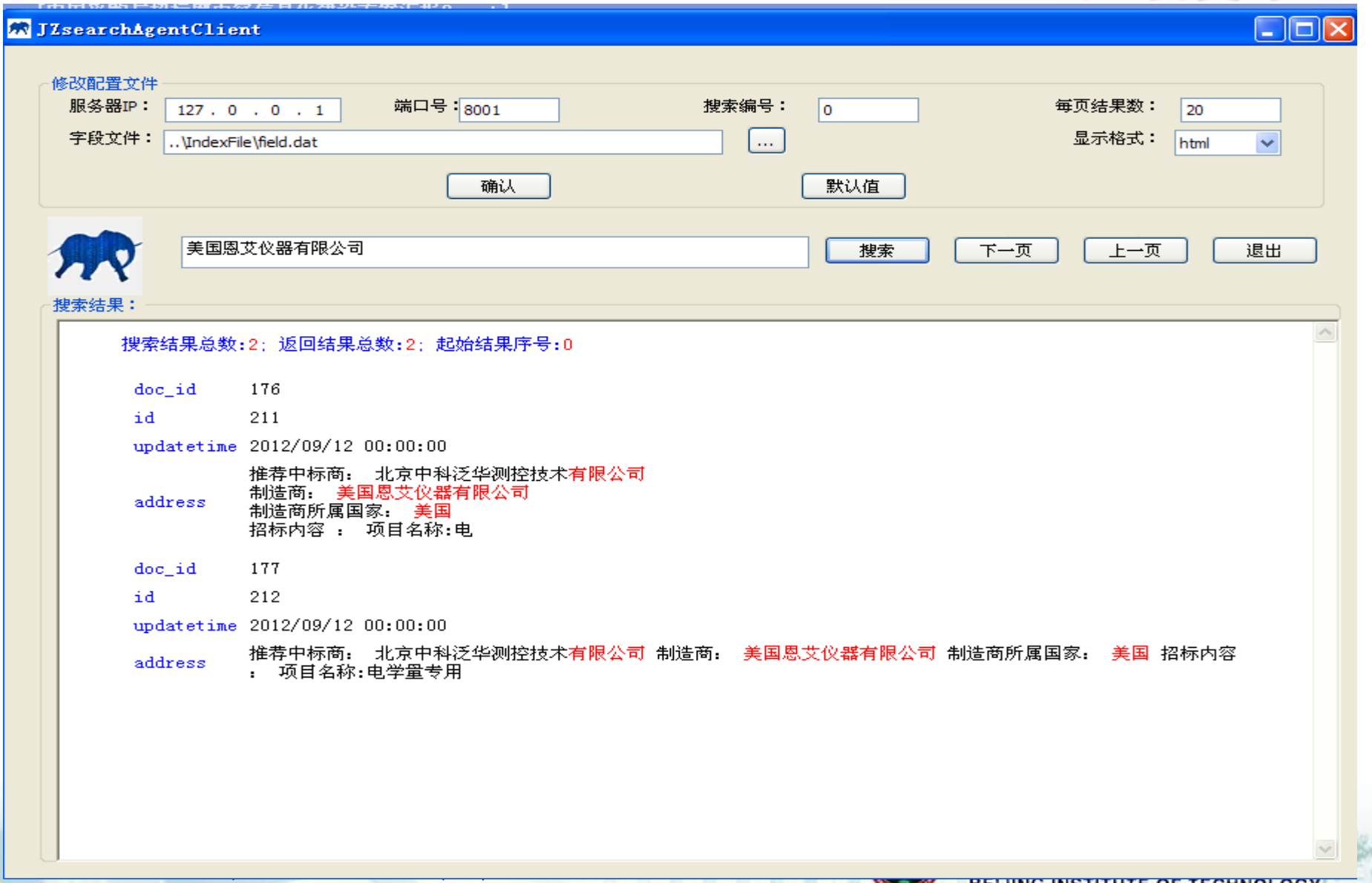
藏文搜索引擎拓扑图


用户终端
Internet
Internet
Web服务器
采集区
新闻
采集
Server
网站
采集
Server
论坛
采集
Server
贴吧
采集
Server
博客
采集
Server
DMZ区
数据检索
索引
Server
检索
Server
核心区
数据存储
文件
Server1
Hbase
数据库
Hadoop
集群
管理
Server
文件
Servern
元信息
Server
容灾备份
Hbase
数据库
元信息
Server
文件
Servern

搜索引擎-爬虫技术


ELINT网络舆情挖掘系统
最新消息
舆情热点
舆情追踪
舆情报告
系统设置
舆情预警
检索条件
舆情监控系统 - 傲游(Maxthon) 2.5.15
http://192.168.1.102:8080/elint/main.html
舆情追踪
我的追踪
城管
群体事件
腐败
追踪管理
实体追踪
人物追踪
机构追踪
网络舆情挖掘系统
1 天心在云 一伙城管暴徒，殴打人民教师，当局莫视？ //@李骁:分享李骁的博文：李骁：暴力拆迁是滋生黑社会的土壤 http://t.cn/bOYqO //@于建峰:这样的事情为何一再发生？根本的原因是由于公权力过于强大，而执掌公权力的人又无任何惧怕。 来源:新浪微博- 评论(2)- 转发(0)-发布时间：2010-12-23 09:35:42
2 寒小妖2010 真他妈狗血的又见城管，一大早长沙雨花区城管队，发生在红旗区暴力执法，欧打农村卖菜妇女，殴打水果商，欧打家禽商，真心觉得如此执法没有天理良心，被路人直指‘‘土匪’’！ 来源:Android客户端- 评论(1)- 转发(0)-发布时间：2011-12-03 11:52:38
3 百科全书V 【揭露城管暴力执法一幕视频】暴力，血腥，残暴，冷血，没人性！这是一群畜生而非人类，却被披上城管制服；这群畜生涉嫌故意伤害罪，寻衅滋事罪，当地警方必须立案查处；请网友们看见此贴也都转发一次吧。 http://t.cn/hDXQj3 来源:新浪微博- 评论(0)- 转发(0)-发布时间：2011-05-30 15:16:11
4 澳門新聞眼Lam 【兩岸消息】穗男子拍城管暴力執法遭毆：廣州白雲區1名男子在本月24日晚圍觀城管對街邊小販的執法行動時，因見到城管暴力執法便用手機拍下照片，但卻突然遭一群男子圍毆，手機亦被強行格式化，最後更要寫檢討。當局解釋事件由誤會引起，會對事主作出道歉及經濟賠償。 来源:新浪微博- 评论(0)- 转发(1)-发布时间：2011-11-30 17:14:39
5 观海知心远 求助一次爱心二次生命 ！父亲为了给白血病的哥哥的求助，每天早起就到“黄兴步行南街”去乞讨，今天却被保安及城管的拖拉抢夺，还差一点被打成重伤上医院，爸爸本来就身体差，可爸爸说为了哥哥就是被打也还会继续去乞讨求助，希望社会爱心人士能帮帮我家，求转发下 http://t.cn/SbQpVf 来源:微博相册- 评论(0)- 转发(0)-发布时间：2011-12-03 11:56:50
完成
840M
缩放:100%
网络舆情挖掘系统
21 故宫建福宫被指成富豪会所 工作人员：不能留宿 新华社北京分社 2011-05-13 09:52 日报 查看

公司搜索

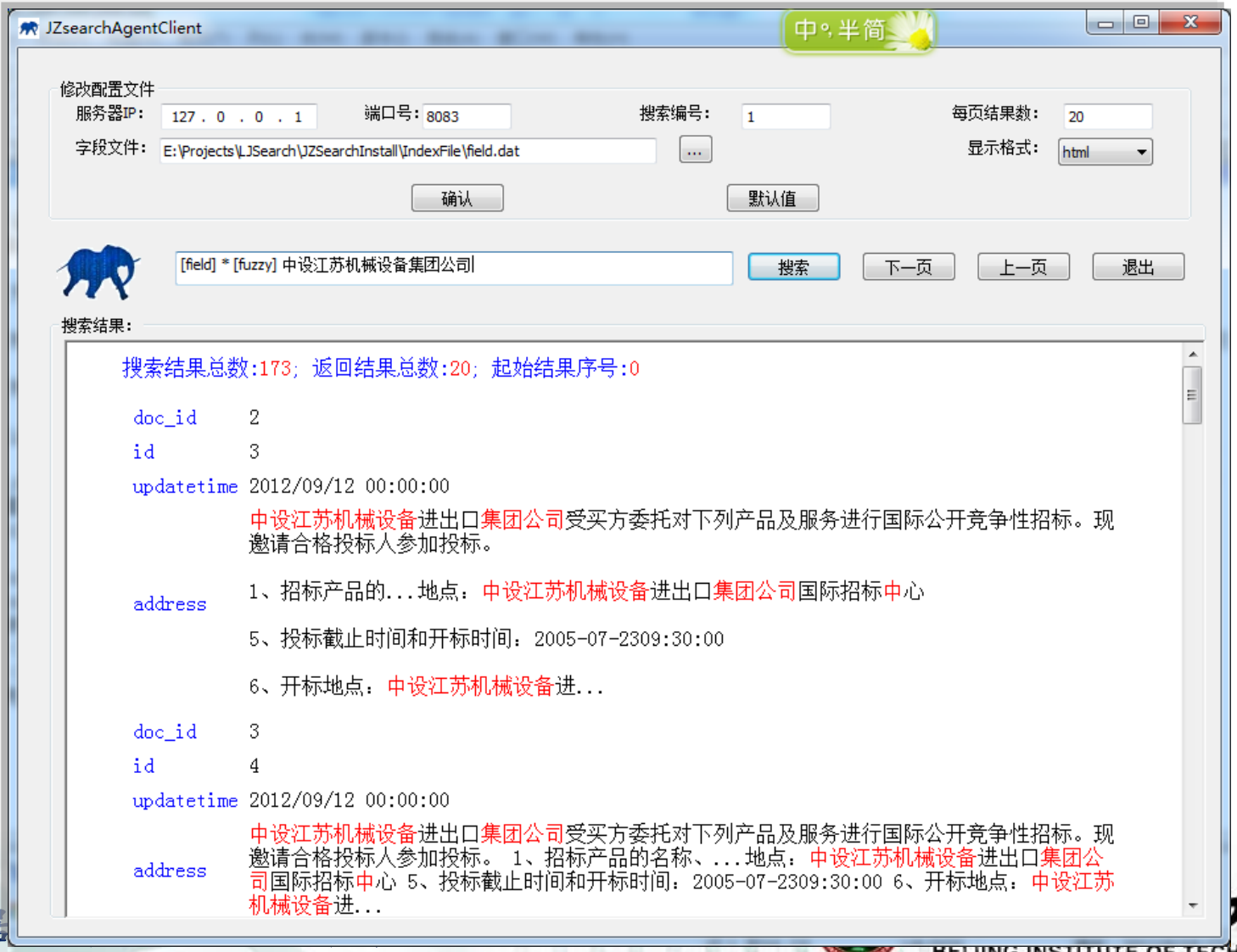

JZsearchAgentClient
修改配置文件
服务器IP：
127 . 0 . 0 . 1
端口号：
8001
搜索编号：
0
每页结果数：
20
字段文件：
..\IndexFile\field.dat
...
显示格式：
html
确认
默认值
美国恩艾仪器有限公司
搜索
下一页
上一页
退出
搜索结果：
搜索结果总数:2; 返回结果总数:2; 起始结果序号:0
doc_id 176
id 211
updatetime 2012/09/12 00:00:00
address 推荐中标商： 北京中科泛华测控技术有限公司
制造商： 美国恩艾仪器有限公司
制造商所属国家： 美国
招标内容 ： 项目名称:电
doc_id 177
id 212
updatetime 2012/09/12 00:00:00
address 推荐中标商： 北京中科泛华测控技术有限公司 制造商： 美国恩艾仪器有限公司 制造商所属国家： 美国 招标内容 ： 项目名称:电学量专用

长句搜索


JZsearchAgentClient
修改配置文件
服务器IP： 127 . 0 . 0 . 1
端口号： 8001
搜索编号： 0
每页结果数： 20
字段文件： ..\IndexFile\field.dat
...
显示格式： html
确认
默认值
或法人委托书到汉寿县城关镇小南门55号（汉寿县建设局办公楼4楼）购买资格
搜索
下一页
上一页
退出
搜索结果：
搜索结果总数:2; 返回结果总数:2; 起始结果序号:0
doc_id 94
id 118
updatetime 2012/09/12 00:00:00
address r>常德工程招标代理有限公司受汉寿县土地开发整理中心委托，对其汉寿县围堤湖土地整理（二期工程）项目进行国内公开招标。凡具有建设行政管理部门颁发的房屋建筑工程施工总承包叁级（...00时（北京时间），逾期不予办理。各投标人在网上报名后凭介绍信或法人委托书到汉寿县城关镇小南门55号（汉寿县建设局办公楼4楼）购买资格预...
doc_id 95
id 119
updatetime 2012/09/12 00:00:00
address 常德工程招标代理有限公司受汉寿县土地开发整理中心委托，对其汉寿县围堤湖土地整理（二期工程）项目进行国内公开招标。凡具有建设行政管理部门颁发的房屋建筑工程施工总承包叁级（...00时（北京时间），逾期不予办理。各投标人在网上报名后凭介绍信或法人委托书到汉寿县城关镇小南门55号（汉寿县建设局办公楼4楼）购买资格预...

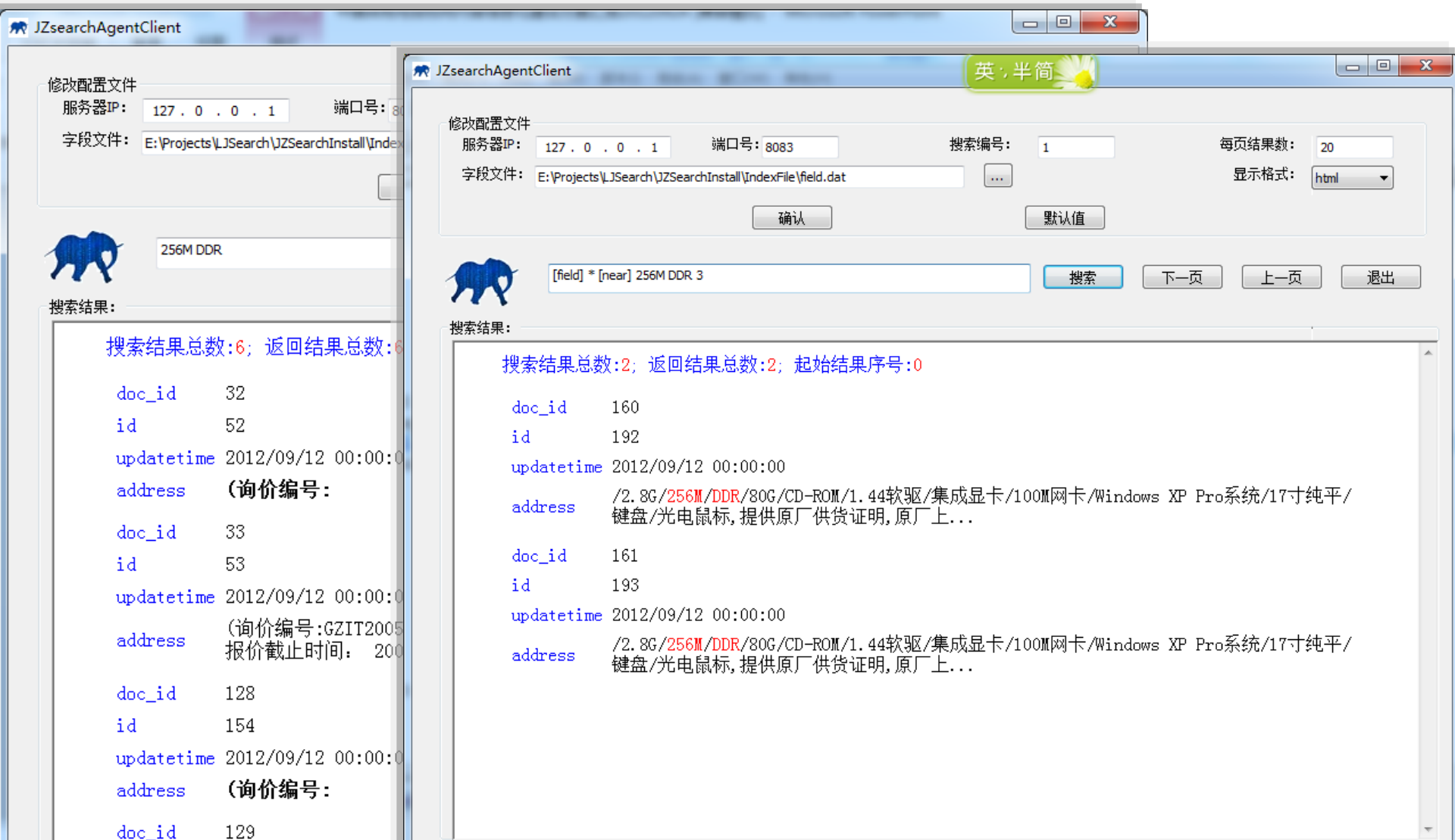
模糊搜索

公司名称表述不准确



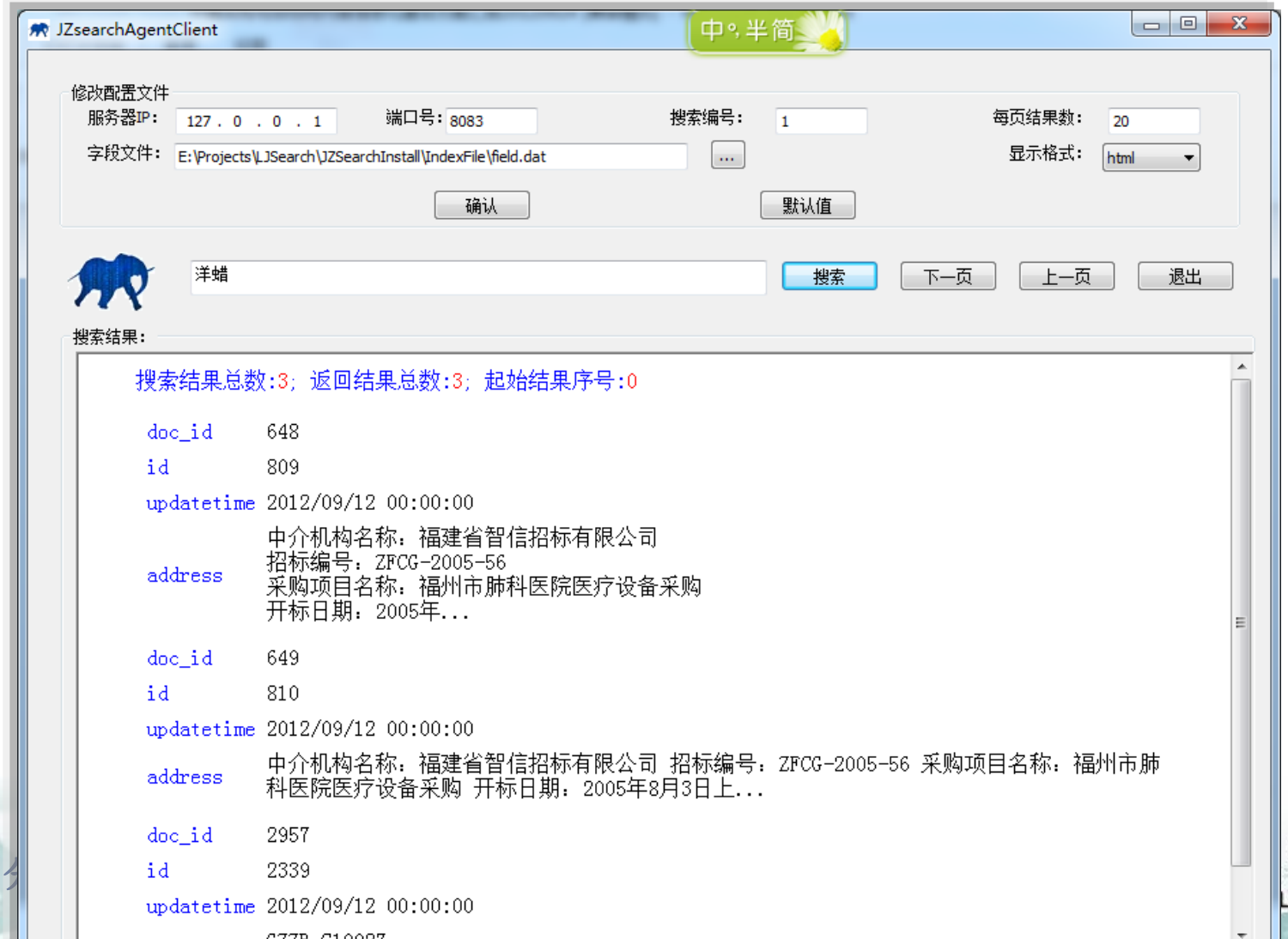
设备参数精准搜索

- 内存 256M的参数要求不等同于两者同时出现



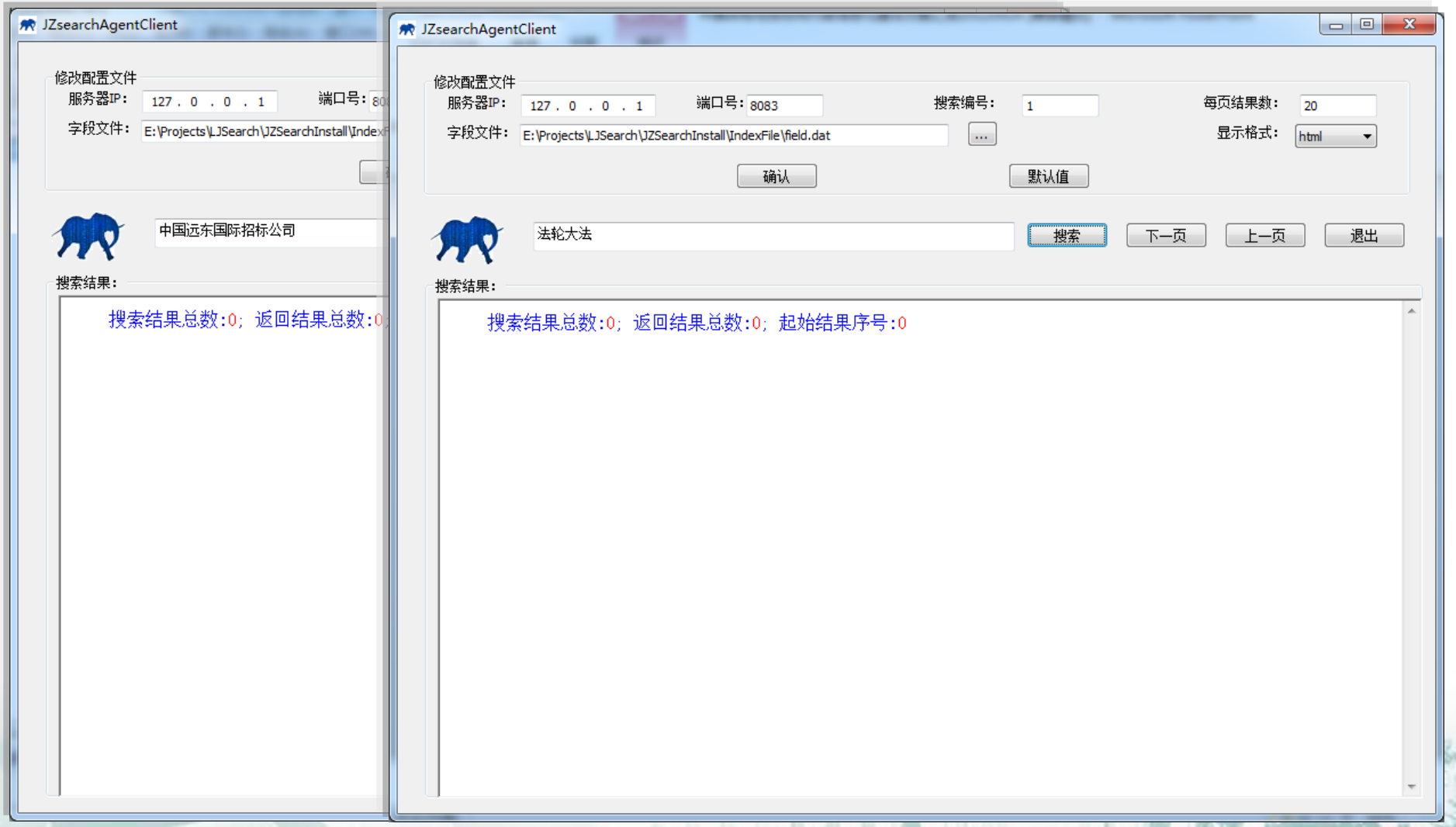
语义关联搜索

- 俗语洋蜡搜索到专业术语石蜡相关的内容



黑名单搜索屏蔽

- 可以屏蔽禁止的公司、产品或反动信息搜索



正文、附件联合查询结果样例

标题：关于陕西某公司采购设备的公告

正文：

... 陕西法士特齿轮有限公司20万台变速器技术改造 3、数控高效滚齿机 1台

附件：

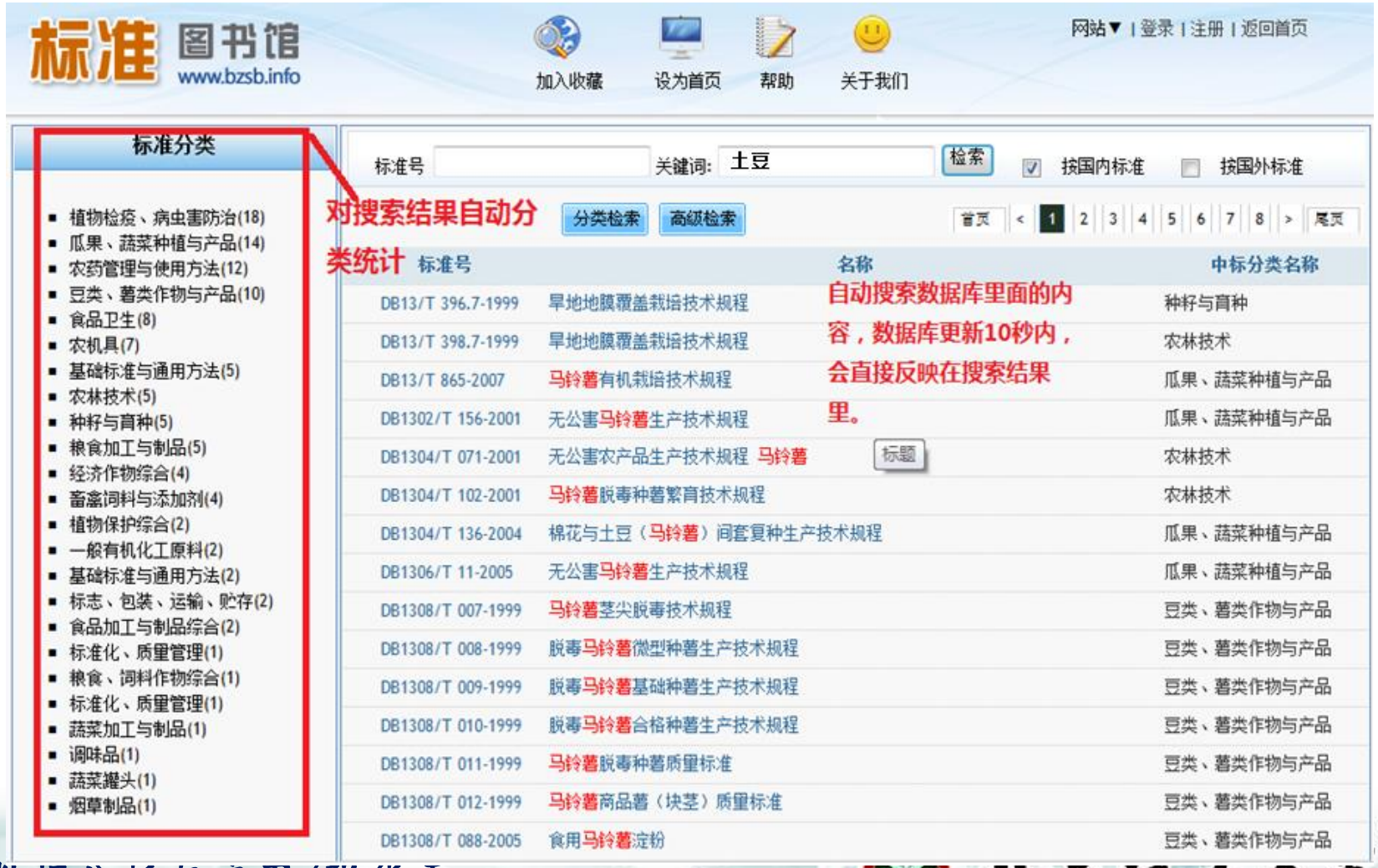
065.txt

... 西安法士特汽车传动有限公司出口齿轮生产线技术改造项目 1、数控滚齿机11台

131.txt

... 陕西法士特齿轮有限公司20万台变速器技术改造 1、数控滚齿机 2台

搜索结果自动分组统计


标准 图书馆
www.bzsb.info
加入收藏
设为首页
帮助
关于我们
网站▼ | 登录 | 注册 | 返回首页
标准分类
植物检疫、病虫害防治(18)
瓜果、蔬菜种植与产品(14)
农药管理与使用方法(12)
豆类、薯类作物与产品(10)
食品卫生(8)
农机具(7)
基础标准与通用方法(5)
农林技术(5)
种籽与育种(5)
粮食加工与制品(5)
经济作物综合(4)
畜禽饲料与添加剂(4)
植物保护综合(2)
一般有机化工原料(2)
基础标准与通用方法(2)
标志、包装、运输、贮存(2)
食品加工与制品综合(2)
标准化、质量管理(1)
粮食、饲料作物综合(1)
标准化、质量管理(1)
蔬菜加工与制品(1)
调味品(1)
蔬菜罐头(1)
烟草制品(1)
标准号
关键词: 土豆
检索
按国内标准
按国外标准
对搜索结果自动分类统计
分类检索
高级检索
首页 < 1 2 3 4 5 6 7 8 > 尾页
标准号
名称
中标分类名称
DB13/T 396.7-1999 旱地地膜覆盖栽培技术规程 种籽与育种
DB13/T 398.7-1999 旱地地膜覆盖栽培技术规程 农林技术
DB13/T 865-2007 马铃薯有机栽培技术规程 瓜果、蔬菜种植与产品
DB1302/T 156-2001 无公害马铃薯生产技术规程 瓜果、蔬菜种植与产品
DB1304/T 071-2001 无公害农产品生产技术规程 马铃薯 农林技术
DB1304/T 102-2001 马铃薯脱毒种薯繁育技术规程 农林技术
DB1304/T 136-2004 棉花与土豆（马铃薯）间套复种生产技术规程 瓜果、蔬菜种植与产品
DB1306/T 11-2005 无公害马铃薯生产技术规程 瓜果、蔬菜种植与产品
DB1308/T 007-1999 马铃薯茎尖脱毒技术规程 豆类、薯类作物与产品
DB1308/T 008-1999 脱毒马铃薯微型种薯生产技术规程 豆类、薯类作物与产品
DB1308/T 009-1999 脱毒马铃薯基础种薯生产技术规程 豆类、薯类作物与产品
DB1308/T 010-1999 脱毒马铃薯合格种薯生产技术规程 豆类、薯类作物与产品
DB1308/T 011-1999 马铃薯脱毒种薯质量标准 豆类、薯类作物与产品
DB1308/T 012-1999 马铃薯商品薯（块茎）质量标准 豆类、薯类作物与产品
DB1308/T 088-2005 食用马铃薯淀粉 豆类、薯类作物与产品
自动搜索数据库里面的内容，数据库更新10秒内，会直接反映在搜索结果里。
标题

我们的工作：JZSearch精准搜索引擎

北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

我们的工作：JZSearch精准搜索引擎

我们的工作：JZSearch精准搜索引擎

少数民族语言精准搜索引擎

搜索

偏移量	长度	文件路径	标题	全文
0	11532	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--خەلق نۇرى	正文
57433	11532	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--خەلق نۇرى	正文
69320	10674	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--شىنجاڭ خەۋەرلىرى--خەلق نۇرى	正文
80102	9776	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--تېخنىكا خەۋەرلىرى--خەلق نۇرى	正文
89986	10393	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--مەدەنىيەت خەۋەرلىرى--خەلق نۇرى	正文
101320	4984	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--مۇقىم تاسىراش--خەلق نۇرى	正文
106304	3690	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--رەسىم--خەلق نۇرى	正文
109994	4304	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--قۇرۇق نۇقا--خەلق نۇرى	正文
114298	4037	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--شەخسىلەر--خەلق نۇرى	正文
118335	8526	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--سۈرەتلىك خەۋەرلەر--خەلق نۇرى	正文
127751	4385	F:/data/MultiLang/Corpus/20100426/000@20100426	文尔吾维网民人--خەلق نۇرى--amp;nbsp;amp;nbsp;سەھىيە amp;nbsp;مۇسى	正文

下一页 尾页 共有 1380 条记录，当前第1/92 页

باش بەت قىلىڭبەت ساقلاڭتەھرىر خەت ساندۇقىنالاقىلىشىڭباش بەتمەركەز文尔吾维网民人--خەلق نۇرى
خەۋەرلىرىخەلقنارا خەۋەرلەرشىنجاڭ روھىمەملىكەت #x0D; #x0D; #x0A;
خەۋەرلىرىنىقىتىسادجەمئىيەتمانارىپتەنتەربىيەمىللەت ۋە مەدەنىيەتنۇھىت ناسراشدىنقىزىق نۇقتاشەخسلەرسۇرەتلىك
خەۋەرلەرماتېرىيالمەخسۇس سەھىيەدەھەر ۋە ئەيدىزىيېڭى خەۋەرلەر مەملىكەتلىك ئەرلەر ۋاسكىتبول كەسپىي بىرلەشمە
مۇسابىقىسىنىڭ ئومۇمىي ھەل قىلغۇچ مۇسابىقىسىدە گۇاڭدۇڭ كوماندىسى شىنجاڭ كوماندىسىنى يەڭدى شىنجاڭ
تۇرپاندا كەم كورولىدىغان بورانلىق ھاۋا رايى كورولدى گەنسۇ جاڭيېدا قۇم بورانلىق ھاۋارايى كورولدى 1-پەسىلدە سىرتتىن
شىنجاڭغا سېلىنغان مەبلەغ نوخشاش ۋاقىتتىكىدىن 4 يېرىم ھەسسە ئاشتى روسىيەنىڭ تۇنجى تورگومدىكى ياردەم
بۇيۇملىرى يېتىپ كەلدى شىنجاڭ قۇمۇلدا قۇم بورانلىق ھاۋا رايى كورولدى ئالدىن مەلۇمات: ئادىل ھوشۇر خەلق نۇرىدا

网上通用搜索功能

2013北京旅游攻略_北京景点线路游记_百度旅游
最新北京旅游攻略，百度旅游为您准备了北京景点、线路、交通、美食、住宿等旅游攻略信息和实用精美游记，供赴北京旅游的亲们参考。
lvyou.baidu.com/

首都之窗-北京市政务门户网站
北海公园 图说北京 魅力公园 北京印象图集 “绿色出行”图片征集 园博会图片征集 2013年北京冬季图片征集 热点关注[会议]市委十一届三次全会于12月22日至12月23...
www.beijing.gov.cn/ 2013-11-27 - 百度快照

北京旅游攻略_北京北京旅游景点_北京旅游网
北京欣欣旅游网,提供北京北京旅游景点推荐、12月北京旅游攻略、北京旅行社、北京旅游线路、北京酒店预订、北京旅游地图等出行指南及旅游服务●欣欣旅游网 CNCN.com ...
beijing.cncn.com/ 2013-12-07 - 百度快照

北京汽车_北京车市_北京汽车报价_北京汽车网_汽车之家
汽车之家北京站为您提供北京汽车报价,北京汽车行情,北京汽车经销商推荐,最精彩的北京汽车新闻、评测、导购及二手车信息,是提供北京汽车信息最快的汽车网站
www.autohome.com.cn/beijing/ 2013-12-07 - 百度快照

2013北京旅游攻略,北京自助游攻略,蚂蜂窝北京出游攻略游记 - 蚂蜂窝
胡同是地道北京人的居所,这里充满着丰厚的生活气息,保留着许多文物古迹。当今的胡同,又进驻了许多潮流元素,这种传统与现代的结合,是北京最有魅力的地方之一。 ...
www.mafengwo.cn/travel-scenic-spot/m... 2013-12-06 - 百度快照

北京网-首都城市综合信息服务平台-北京生活、旅游、交通、文化、...
北京网是北京市政府主导建设,为公众提供首都城市综合信息服务的公益性网站。提供北京地图、交通、旅游、住房、餐饮、医疗、演出、教育、就业等领域的便民服务信息。
www.beijing.cn/ 2013-12-07 - 百度快照

JZSearch语义精准搜索引擎

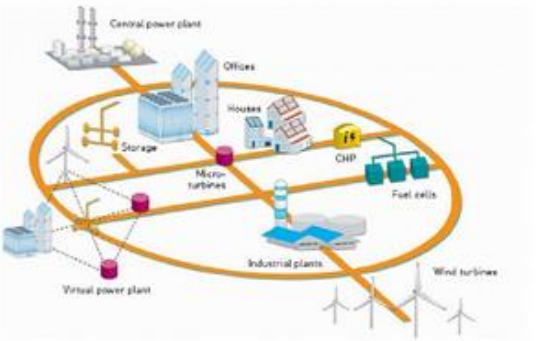
个人中心

国家电网公司 STATE GRID CORPORATION OF CHINA

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

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，山东郯城人，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。

2007 2008 2009 2010 2011 2012 2013 2014 2015 >>>

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

基于语义的自动学习计算

相关人物计算

基于语义的自动学习计算

北京理工大学 BEIJING INSTITUTE OF TECHNOLOGY

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