

文章编号:1671-251X(2019)03-0103-06

DOI:10.13272/j.issn.1671-251x.2018110071

# 全文搜索引擎在煤矿运维系统中的应用

孟峰<sup>1</sup>, 荆诚<sup>2,3</sup>

(1. 国家能源集团公司 煤炭产业中心, 北京 100011;

2. 中煤科工集团常州研究院有限公司, 江苏 常州 213015;

3. 天地(常州)自动化股份有限公司, 江苏 常州 213015)



扫码移动阅读

**摘要:**针对目前已有的煤矿运维系统功能单一、自动化程度低等问题,将全文搜索引擎应用到煤矿运维系统中,通过搜索引擎从故障解决方案数据库中搜索出匹配的解决方案。介绍了煤矿运维系统流程及故障解决方案来源、数据结构和录入流程;通过对比选型,选择 Elasticsearch 作为煤矿运维系统的全文搜索引擎,并给出了其具体应用方法。实例分析及测试结果表明,该搜索引擎提供的搜索结果与告警内容所反映的故障相关度较高,能够为解决故障提供有效帮助。

**关键词:**煤矿安全;煤矿运维系统;全文搜索引擎;Elasticsearch;故障解决方案

中图分类号:TD67 文献标志码:A

网络出版地址:<http://kns.cnki.net/kcms/detail/32.1627.TP.20190226.1344.001.html>

Application of full-text search engine in coal mine operation and maintenance system

MENG Feng<sup>1</sup>, JING Cheng<sup>2,3</sup>

(1. Coal Industry Center, China Energy Group Corporation, Beijing 100011, China;

2. CCTEG Changzhou Research Institute, Changzhou 213015, China;

3. Tiandi (Changzhou) Automation Co., Ltd., Changzhou 213015, China)

**Abstract:**In view of problems of single function and low degree of automation of existing coal mine operation and maintenance system, full-text search engine was applied to coal mine operation and maintenance system, and matching solutions were searched from fault solution database through the search engine. Flow of coal mine operation and maintenance system and source, data structure and entry process of fault solution were introduced. By comparing and selecting, Elasticsearch was selected as the full-text search engine of coal mine operation and maintenance system, and its specific application method was given. Example analysis and test results show that search results provided by the search engine are highly correlated with the faults reflected by the alarm content, which can provide effective help for solving the faults.

**Key words:**coal mine safety; coal mine operation and maintenance system; full-text search engine; Elasticsearch; fault solution

## 0 引言

煤矿运维工作是一项日常且长期的工作,对于

保障煤矿安全生产具有重要作用。随着传感监测、工业控制、供配电、通信等技术的发展,煤矿安全监控系统、无线通信系统等相关系统也不断升级换代,

收稿日期:2018-11-30;修回日期:2019-01-28;责任编辑:胡娴。

基金项目:天地(常州)自动化股份有限公司研发项目(2018GY002)。

作者简介:孟峰(1969—),男,山西临汾人,高级工程师,硕士,主要研究方向为矿山信息化、无线通信、矿山机电,E-mail:10000335@chnenergy.com.cn。

引用格式:孟峰,荆诚.全文搜索引擎在煤矿运维系统中的应用[J].工矿自动化,2019,45(3):103-108.

MENG Feng, JING Cheng. Application of full-text search engine in coal mine operation and maintenance system[J]. Industry and Mine Automation, 2019, 45(3): 103-108.

对系统维护提出了更高的要求<sup>[1-2]</sup>。煤矿运维工作的主要内容是依靠系统告警和人工巡检发现故障,并由有相关经验的软硬件运维人员排除故障<sup>[3-4]</sup>。

目前已有的煤矿运维系统功能较为单一,只具备最基本的故障上报和简单的故障处理流程跟踪功能,其余大量运维工作还需要人工操作<sup>[5-6]</sup>。由于故障排查是一项十分依赖经验的工作<sup>[7]</sup>,为了使经验得以传承,同时自动化、规范化运维流程,本文将全文搜索引擎应用到煤矿运维系统中。首先,打通各种煤矿系统的数据通道,通过数据共享和信息推送收集各软硬件系统的运行状态。其次,通过搜索引擎分析各系统反馈的告警信息关键词,从故障解决方案数据库中搜索出匹配的解决方案。

1 煤矿运维系统

1.1 系统流程

煤矿运维系统收到告警信息后,创建运维服务单,同时智能匹配相关故障的解决预案并推送给运维负责人,运维负责人确认之后将运维服务单推送至运维人员手中。随着将来需求的清晰,人工智能系统的不断更新升级,任务指派模式将从手动指派优化为智能辅助人工指派,最终发展为系统自动派单加人工审核模式。

煤矿运维系统简化流程如图 1 所示。完整的运维软件中包含了许多复杂的环节,监控系统的告警推送和运维软件的派单流程将会需要与其他系统进行数据交互。在此简化了相关的流程,将数据流转过过程作为重点。煤矿运维系统流程:① 监控系统推送故障告警至运维系统;② 运维系统将告警信息输入搜索引擎,进行解决方案搜索。③ 搜索引擎返回最佳匹配方案至运维系统。④ 运维系统进入派单流程,将解决方案和故障信息附加至新创建的运维服务单,派单至运维人员。⑤ 故障解决后,运维人员对解决方案进行评价,相关解决方案的采纳次数更新,以帮助搜索引擎提升搜索结果准确率。

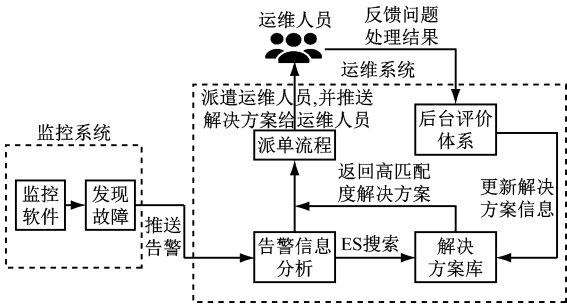


图 1 煤矿运维系统简化流程

Fig. 1 Simplified process of coal mine operation and maintenance system

1.2 故障解决方案来源

为了完善初期的故障解决方案数据库,通过咨询运维人员和文档记录的方式,搜集煤矿软硬件系统各种常见故障的现象、原因和解决方法,并将这些信息录入数据库。在煤矿运维系统上线之后,故障信息及解决流程和结果将被记录下来,通过人工筛选将成功案例收录至数据库,从而不断扩大和完善故障解决方案数据库,提高预案搜索的成功率。

在解决预案数据库不断完善的过程中,搜索引擎的搜索结果也会越来越准确,可为运维人员解决问题提供技术支持。

1.3 故障解决方案数据结构

将故障解决方案作为数据源,在搜索引擎中建立关键词索引,索引结构设计为设备类型、相关系统、相关硬件、故障现象、故障原因、解决方案、采纳次数这几项。对已有的解决方案建立这种数据结构的索引。在解决方案提供后,通过运维人员的反馈来更新相关数据。对于完整的数据,排序逻辑会适当增加采纳次数的权重,因为采纳次数高意味着这条解决方案是常用且有效的。在采纳次数缺失的前期阶段,通过分析告警信息内容,最终推送的解决方案排序逻辑主要参考相关硬件、相关系统、故障现象这 3 条属性,根据这 3 条属性的相关性来推荐适用的解决方案。解决方案的录入流程如图 2 所示。

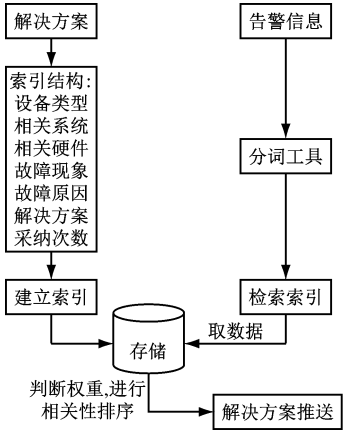


图 2 故障解决方案的录入流程

Fig. 2 Entry process of fault resolution

2 搜索引擎对比及选型

Lucene 是一种应用广泛的搜索引擎,不过因为其复杂性,给普通用户带来了很大的学习成本。于是出现了很多基于 Lucene,但隐藏其复杂性并可提供更加便捷调用方式的搜索引擎,例如 Elasticsearch, solr, Katta, Hadoop contrib/index, Lucandra, HBase 等<sup>[8-10]</sup>。

众多搜索引擎中, Elasticsearch<sup>[11]</sup> 和 solr<sup>[12-13]</sup>

是目前应用比较广泛的 2 种。为了选择符合需求的搜索引擎,对这 2 种搜索引擎进行对比分析,结果见表 1。

表 1 2 种搜索引擎对比

Table 1 Comparison of two kinds of search engine

| Elasticsearch        | solr                     |
|----------------------|--------------------------|
| Java 编写              | Java 编写                  |
| 提供 RESTful API       | 提供 HTTP/XML 和 JSON 的 API |
| 分布式和实时分发             | 分布式的搜索和索引                |
| 维护群体较小,民间维护者         | 更大更成熟的用户、开发者和贡献者社区       |
| 多种交互方式               | 外部配置丰富                   |
| 不适合 Index Warmup API | 实时搜索效率不如 Elasticsearch   |

用 solr 搜索已有数据时,速度更快,而需要实时建立索引时,Elasticsearch 具有明显优势。就目前需求来看,已有的处理预案数据量不大,Elasticsearch 和 solr 表现都很优秀。考虑到后期自动派单功能需要自动收集案例,实时建立索引,因此选择 Elasticsearch 作为煤矿运维系统的全文搜索引擎。

3 Elasticsearch 工具分析

3.1 核心概念

Elasticsearch 是一个兼有搜索引擎和 NoSQL 数据库功能的开源系统<sup>[14]</sup>,图 3 中将 Elasticsearch 与关系型数据库结构进行对比。

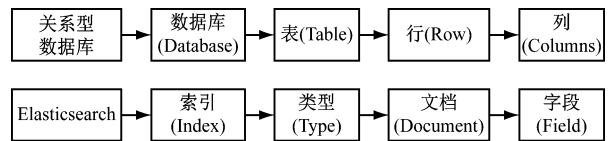


图 3 Elasticsearch 与关系型数据库结构对比  
Fig. 3 Comparison of structure of Elasticsearch and relational database

索引(Index): 索引类似于关系型数据库实例。它只是一个逻辑命名空间,指向一个或多个分片,在其内部使用 Lucene 来实现数据读写,通过引用索引名称来完成相关文档的增、删、改、查操作。

文档类型(Type): 相当于数据库中的 table 概念。在 Elasticsearch 中每个文档都必须设定其类型。文档类型使得在同一个索引中存储结构不同的文档时,只需要依据文档类型就可以找到对应的参数映射(Mapping)信息,方便文档存取。

文档(Document): 文档类似于数据库中的行数据,是索引中的基本数据单位,文档内容是以 JSON

格式存储的,一个索引中可以存储多个文档,且文档在第 1 次存储时就指定了类型。

3.2 倒排索引

Elasticsearch 引擎使用了一种称为倒排索引(Inverted Index)的数据结构<sup>[15-16]</sup>,与普通索引面向文档不同,倒排索引是面向分词的。通常的文档和词条索引结构见表 2,以文档 ID 为键,对应的值是文档中出现的关键词。采用倒排索引的数据结构见表 3,以分词为键,文档 ID 及其出现次数为对应的值。这是一种适用于快速全文搜索的数据结构。

表 2 文档对应词条索引

Table 2 Document corresponding index

| 文档 ID | 词条          |
|-------|-------------|
| 1     | 运维 监控       |
| 2     | 大数据 智慧矿山 运维 |
| 3     | 运维 AI       |

表 3 倒排索引

Table 3 Inverted index

| 分词   | 出现次数 | 文档 ID |
|------|------|-------|
| 监控   | 5    | 1     |
| 运维   | 3    | 1     |
| 智慧矿山 | 2    | 4     |

3.3 分词工具

由于煤矿运维系统所需搜索信息多为文本信息,其中汉字出现的频率很高,为了搜索一段文字中的关键字,需要借助中文分词器来进行分词。Elasticsearch 试用后发现,其默认的中文分词插件是单字切词,无法将一段话切分成词语,不能满足正常的汉语分词要求。因此,需要额外安装中文切词插件,对比调研后决定使用 IK 中文分词器<sup>[17-18]</sup>插件。

IK 插件将 Lucene IK 分析器集成到 Elasticsearch 中,支持自定义字典。分词器的工作流程:第 1 步,切分关键词,将语句切分成英文单词和中文词语;第 2 步,去除停用词,英文停用词包括 a、an、the、of 等,中文停用词是包括“的”、“了”、“着”、“是”、标点符号等。对于英文单词,还需要有第 3 步,即把所有字母转为小写,因为搜索时不区分大小写。

以“应用于煤矿系统的软件系统和硬件都得到了巨大的提升”这段话作为例子,IK 插件分词结果为“应用于,煤矿,系统,的,软件,系统,和,硬件,都得,到了,巨大,的,提升”。其中“都得到了”这部分由于“都得”和“得到”都属于合理切分,所以结果与本来含义有了偏差,这种情况属于切词中的正常情

况。可见 IK 插件可以很好地将一段中文语句切分成含义相对准确的词语,便于 Elasticsearch 建立索引,提高搜索的速度和命中率。

4 实例分析

4.1 查询实例

用 Elasticsearch 本身提供简单的 RESTful API 来进行相关操作,通过简单的 POST、GET 命令即可录入数据,并进行查询。按照前文设计的数据结构将数百种各系统的故障解决方案录入搜索引擎,通过关键词搜索可得到匹配的结果。

按照表 4 所示的数据结构建立索引,将已有解决方案录入搜索引擎。

表 4 故障解决方案数据结构

| Table 4 Data structure of fault solution |      |         |
|------------------------------------------|------|---------|
| 参数名称                                     | 参数含义 | 参数类型    |
| device_type                              | 设备类型 | text    |
| system_type                              | 系统类型 | text    |
| related_hardware                         | 相关硬件 | text    |
| fault_phenomenon                         | 故障现象 | text    |
| fault_reason                             | 故障原因 | text    |
| content                                  | 解决方案 | text    |
| accept_times                             | 采纳次数 | integer |

当煤矿运维系统获得告警信息后,通过调用 Elasticsearch 提供的 API 接口进行搜索工作,并通过配置分词插件进行语句查询。当告警信息为“电源关闭”时,共得到 13 条匹配结果,排得越靠前的结果匹配度越高,其中前 3 条结果如图 4 所示。以第 1 条结果为例,其具体内容如图 5 所示。

|                                                                                                                                                                                                                                                                                                                                            |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ▶related_hardware:电源fault_phenomenon:1.闭锁控制器的所有指示灯不亮2.电源指示灯亮,但闭锁控制器工作不正常3.显示屏不亮4.备用电池不能充电fault_reason:1.无电源电压或电源电压严重不足2.电源开关未打开3.电源线未接好4.保险管烧坏5.备用电池管理系统损坏content:1.检查线路及各接线端子是否接好2.检查电源电压是否正常3.电压的极性是否正确4.检查电源开关是否闭合5.更换保险管6.更换备用电池管理系统device_type:传感器system_type:监控系统accept_times:1_id:5_type:solution_index:cari_result_score:4.277     |
| ▶related_hardware:闭锁控制器,电源fault_reason:1.闭锁控制器地址设置错误2.电源电压波动过大或受到干扰3.闭锁控制器电源故障4.通信协议不对或串口设置错误content:1.检查闭锁控制器地址与监控主机呼叫的地址是否一致或地址设置是否有重复2.重新接线3.检查主机传输协议是否与闭锁控制器一致4.重新给闭锁控制器上电5.测量电源电压,确保电压正常6.去除线路干扰源7.检查RS485总线阻抗是否匹配device_type:控制器system_type:监控系统fault_phenomenon:不能通信accept_times:1_id:6_type:solution_index:cari_result_score:3.173 |
| ▶device_type:开关related_hardware:开关fault_phenomenon:开关不能合闸fault_reason:1.高压开关合闸机构故障2.合闸出口继电器故障3.无合闸工作电源4.合闸回路断线5.失压线圈不动作content:1.检查机构,排除故障2.更换继电器3.检查更换电源插件或整流模块4.合闸回路中间串有门到位、断路器到位等行程开关,合闸回路断线时要检查这些行程开关是否已经闭合。5.失压线圈环或失压线圈回路断线。system_type:监控系统accept_times:1_id:3_type:solution_index:cari_result_score:2.64                            |

图 4 当告警信息为“电源关闭”时得到的前 3 条匹配结果  
Fig. 4 Top 3 match results when alarm information is "power off"

从搜索结果来看,该搜索引擎提供的搜索结果与告警内容所反映的故障相关度较高,将这些预案推送给运维人员能帮助他们有效解决故障。

煤矿运维系统相关人员得到的告警信息推送界面如图 6 所示。在填写预案时,软件会默认将后台反馈的匹配度最高的预案填入,以供告警处理人员

参考。

|                    |                                                                                                                                                                                                                                                                                                                                       |
|--------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| source             | related_hardware:电源fault_phenomenon:1.闭锁控制器的所有指示灯不亮2.电源指示灯亮,但闭锁控制器工作不正常3.显示屏不亮4.备用电池不能充电fault_reason:1.无电源电压或电源电压严重不足2.电源开关未打开3.电源线未接好4.保险管烧坏5.备用电池管理系统损坏content:1.检查线路及各接线端子是否接好2.检查电源电压是否正常3.电压的极性是否正确4.检查电源开关是否闭合5.更换保险管6.更换备用电池管理系统device_type:传感器system_type:监控系统accept_times:1_id:5_type:solution_index:cari_result_score:4.277 |
| Table              | JSON                                                                                                                                                                                                                                                                                                                                  |
| t_id               | 5                                                                                                                                                                                                                                                                                                                                     |
| t_index            | cari_result                                                                                                                                                                                                                                                                                                                           |
| #_score            | 4.277                                                                                                                                                                                                                                                                                                                                 |
| t_type             | solution                                                                                                                                                                                                                                                                                                                              |
| #accept_times      | 1                                                                                                                                                                                                                                                                                                                                     |
| t_content          | 1.检查线路及各接线端子是否接好2.检查电源电压是否正常3.电压的极性是否正确4.检查电源开关是否闭合5.更换保险管6.更换备用电池管理系统                                                                                                                                                                                                                                                                |
| t_device_type      | 传感器                                                                                                                                                                                                                                                                                                                                   |
| t_fault_phenomenon | 1.闭锁控制器的所有指示灯不亮2.电源指示灯亮,但闭锁控制器工作不正常3.显示屏不亮4.备用电池不能充电                                                                                                                                                                                                                                                                                  |
| t_fault_reason     | 1.无电源电压或电源电压严重不足2.电源开关未打开3.电源线未接好4.保险管烧坏5.备用电池管理系统损坏                                                                                                                                                                                                                                                                                  |
| t_related_hardware | 电源                                                                                                                                                                                                                                                                                                                                    |
| t_system_type      | 监控系统                                                                                                                                                                                                                                                                                                                                  |

图 5 当告警信息为“电源关闭”时得到的第 1 条故障解决方案的具体内容  
Fig. 5 Specific content of the first fault solution when alarm information is "power off"

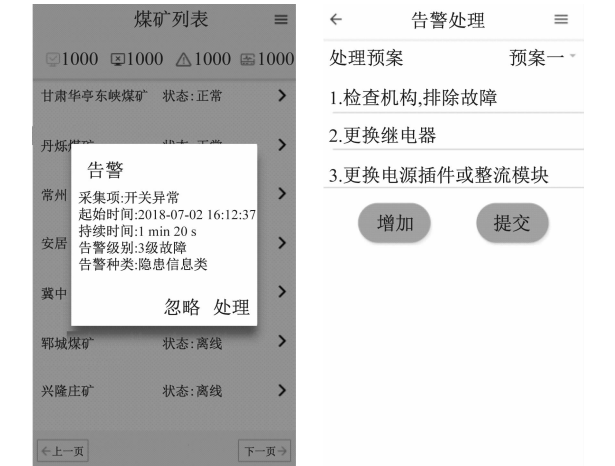


图 6 告警信息推送界面  
Fig. 6 Push interface of alarm information

4.2 存储空间

Elasticsearch 插入数据时是先写入内存,然后采用队列的形式写入磁盘。从理论上讲,并发写入应该优于传统的关系型数据库的写入方式。Elasticsearch 在存储数据时会进行压缩,这对于各种文本信息来说,其占用的存储空间会有所降低。

测试内容第 1 项是对比数据在磁盘中占用的空间,对比数据库使用 mysql。在 Elasticsearch 和 mysql 中,采用相同的数据结构,分别记录 5 000, 15 000,65 000 条数据占用的磁盘大小,测试对比数据见表 5。从表 5 可以看出,同样的数据结构,Elasticsearch 占用的空间约为 mysql 的 1/9。监控系统中的日志、监控数据都将占用大量磁盘空间,Elasticsearch 可以很好地缓解磁盘空间不足的问题。

4.3 性能测试

首先进行 Elasticsearch 批量写入测试。程序用 Python 编写,使用生产者 and 消费者模式,从 mysql 数据库读取数据,1 000 条数据写入 1 次,测试设备是 16 GB 内存, I7 处理器, 256 GB 固态硬盘

表5 Elasticsearch 和 mysql 存储空间对比

Table 5 Comparison of storage space of Elasticsearch and mysql

| 数据量/行  | Elasticsearch<br>占用空间/MB | mysql 占用<br>空间/MB |
|--------|--------------------------|-------------------|
| 5 000  | 21                       | 183               |
| 15 000 | 48                       | 551               |
| 65 000 | 257                      | 2 386             |

盘。测试结果:单进程,5 个写入线程,100 万行数据,500 s;单进程,20 个写入线程,100 万行数据,500 s。

借助压力测试工具 Esrally 进行数据库压力测试,测试数据为 3 GB 的 json 数据。测试结果见表 6,从表 6 可以看到,创建索引时间为 32 min,占用磁盘空间 3 GB,索引大小为 5 GB。目前解决方案的数据量远远低于 3 GB,后期新增的解决方案单次录入量也不会很大,因此,数据库容量可以很好地满足目前的需求。

表6 压力测试结果

Table 6 Stress test results

| 性能参数名称          | 参数值       |
|-----------------|-----------|
| 索引建立总耗时/min     | 32.660 30 |
| 单分片索引建立最少耗时/min | 6.424 22  |
| 单分片索引建立最大耗时/min | 6.735 25  |
| 数据大小/GB         | 3.082 53  |
| 索引大小/GB         | 5.708 97  |

测试数据库的查询性能见表 7,10 000 条以下的数据 Elasticsearch 搜索时间为 5~20 ms,10 万条数据搜索时间为 15~30 ms。因此,该数据库查询性能表现也可以很好地满足目前的查询需求。

表7 数据库查询性能

Table 7 Query performance of database

| 数据量/条   | 查询时间/ms |
|---------|---------|
| 1 000   | 1~20    |
| 110 000 | 15~30   |

5 结语

介绍了一种全文搜索引擎 Elasticsearch 在煤矿运维系统中的应用,搜索引擎通过分析告警信息,快速匹配出相关性较高的解决预案,为运维人员提供故障解决思路,提高运维工作效率。测试结果表明,Elasticsearch 搜索结果匹配度较高,可以满足正常数据量的解决预案的搜索工作。

参考文献 (References):

[ 1 ] 贺耀宜. 煤矿本质安全支撑体系平台研究[J]. 工矿自动化,2012,38(11):5-8.  
HE Yaoyi. Research of support system platform of intrinsic safety of coal mine[J]. Industry and Mine Automation,2012,38(11):5-8.

[ 2 ] 王莉. 智慧矿山概念及关键技术探讨[J]. 工矿自动化,2014,40(6):37-41.  
WANG Li. Study on concept and key technologies of smart mine[J]. Industry and Mine Automation,2014,40(6):37-41.

[ 3 ] 张旭平,赵甫胤,孙彦景. 基于物联网的智慧矿山安全生产模型研究[J]. 煤炭工程,2012,44(10):123-125.  
ZHANG Xuping,ZHAO Fuyin,SUN Yanjing. Study on safety production model of intelligent mine base on Internet of things [J]. Coal Engineering, 2012,44(10):123-125.

[ 4 ] 吕鹏飞,何敏,陈晓晶,等. 智慧矿山发展与展望[J]. 工矿自动化,2018,44(9):84-88.  
LYU Pengfei, HE Min, CHEN Xiaojing, et al. Development and prospect of wisdom mine [J]. Industry and Mine Automation,2018,44(9):84-88.

[ 5 ] 谭章禄,陈晓. 煤炭企业信息化建设现状及发展对策探讨[J]. 工矿自动化,2016,42(7):63-65.  
TAN Zhanglu,CHEN Xiao. Discussion on situation and development countermeasures of coal enterprise informatization construction[J]. Industry and Mine Automation,2016,42(7):63-65.

[ 6 ] 王泽安. 煤炭企业信息化现状及发展趋势研究[J]. 煤炭工程,2015,47(2):146-148.  
WANG Ze'an. Research on current situation and development trend of coal enterprise informatization [J]. Coal Engineering,2015,47(2):146-148.

[ 7 ] 马昆,高顶,包慧玲. 煤矿信息化现状及其发展探讨[J]. 煤矿机械,2011,32(3):10-12.  
MA Kun, GAO Ding, BAO Huiling. Discussion on coal informatization condition and development [J]. Coal Mine Machinery,2011,32(3):10-12.

[ 8 ] 邹燕飞,于成尊,赵亮. 基于 Lucene 的文本搜索引擎的设计和实现[J]. 计算机与现代化,2011(9):40-42.  
ZOU Yanfei,YU Chengzun,ZHAO Liang. Design and implementation of text search engine based on Lucene [J]. Computer and Modernization,2011(9):40-42.

[ 9 ] 于天恩. Lucene 搜索引擎开发权威经典[M]. 北京:中国铁道出版社,2013.

[10] 周敬才,胡华平,岳虹. 基于 Lucene 全文检索系统的设计与实现[J]. 计算机工程与科学,2015,37(2):252-256.  
ZHOU Jingcai,HU Huaping,YUE Hong. Design and

- implementation of Lucene-based full-text retrieval system[J]. Computer Engineering & Science, 2015, 37(2):252-256.
- [11] 张建中,黄艳飞,熊拥军. 基于 Elasticsearch 的数字图书馆检索系统[J]. 计算机与现代化, 2015(6):69-73.  
ZHANG Jianzhong, HUANG Yanfei, XIONG Yongjun. Digital library retrieval system based on Elasticsearch [J]. Computer and Modernization, 2015(6):69-73.
- [12] 王小森. 基于 Solr 的搜索引擎的设计与实现[D]. 北京:北京邮电大学, 2011.
- [13] 邹敏昊. 基于 Lucene 的 HBase 全文检索功能的设计与实现[D]. 南京:南京大学, 2013.
- [14] 王伟,魏乐,刘文清,等. 基于 Elasticsearch 的分布式全文搜索系统[J]. 电子科技, 2018,31(8): 56-59.  
WANG Wei, WEI Le, LIU Wenqing, et al. Design and implementation on distributed full-text search system based on Elasticsearch[J]. Electronic Science and Technology, 2018,31(8):56-59.
- [15] 郑榕增,林世平. 基于 Lucene 的中文倒排索引技术的研究[J]. 计算机技术与发展, 2010,20(3):80-83.  
ZHENG Rongzeng, LIN Shiping. Research of Chinese full texts inverted index based on Lucene [J]. Computer Technology and Development, 2010,20(3): 80-83.
- [16] 马咏,蔡冰. 大数据环境下 Lucene 性能优化方法研究[J]. 南京理工大学学报(自然科学版), 2015,39(3): 260-265.  
MA Yang, CAI Bing. Performance optimization method of Lucene in big data[J]. Journal of Nanjing University of Science and Technology, 2015,39(3): 260-265.
- [17] 义天鹏,陈启安. 基于 Lucene 的中文分析器分词性能比较研究[J]. 计算机工程, 2012,38(22):279-282.  
YI Tianpeng, CHEN Qi'an. Comparison research of segmentation performance for Chinese analyzers based on Lucene[J]. Computer Engineering, 2012,38(22): 279-282.
- [18] 常建秋,沈炜. 基于字符串匹配的中文分词算法的研究[J]. 工业控制计算机, 2016,29(2):115-116.  
CHANG Jianqiu, SHEN Wei. Research on Chinese word segmentation algorithm based on string matching[J]. Industrial Control Computer, 2016,29(2):115-116.