

文章编号: 1006-2467(2021)S2-0015-07

DOI: 10.16183/j.cnki.jsjtu.2021.S2.003

# 面向电网调度设备监控的智能搜索框架技术

翟海保<sup>1</sup>, 韩博文<sup>2</sup>, 吴国松<sup>2</sup>

(1. 国家电网有限公司华东分部, 上海 200120; 2. 北京科东电力控制系统有限责任公司, 北京 100192)

**摘要:** 电网调度监控业务发展对调度系统的数字化和智能性提出了新的要求, 结合电网调控数据和业务特征, 研究面向电网调度监控的智能搜索框架以及电力领域词库建模、电力特征分词算法、多因子相关性排序算法、电力知识图谱关联建模等关键技术. 研发了面向调控云的电网智能搜索应用, 实现电网设备模型、设备运行、监控事件、服务等搜索, 能够对海量调度设备监控运行数据进行全面、准确、快速、智能检索, 为调度设备监控业务管理及决策提供有力支撑.

**关键词:** 设备监控; 电力领域词库; 电力知识图谱; 智能搜索

中图分类号: TM 721

文献标志码: A

## Intelligent Search Framework Technology for Power Grid Dispatching Equipment Monitoring

ZHAI Haibao<sup>1</sup>, HAN Bowen<sup>2</sup>, WU Guosong<sup>2</sup>

(1. East Branch of State Grid Corporation of China, Shanghai 200120, China;

2. Beijing Kedong Power Control System Co., Ltd., Beijing 100192, China)

**Abstract:** The development of power grid dispatching and monitoring business brings forward new requirements for the digitization and intelligence of dispatching systems. In combination with the power grid regulation data and business characteristics, this paper studies the intelligent search framework for power grid dispatching and monitoring, as well as the key technologies such as power domain thesaurus modeling, power feature word segmentation algorithm, multi-factor correlation sorting algorithm, power knowledge map correlation modeling, etc. The grid intelligent search application for regulation cloud is developed to realize the search of grid equipment model, equipment operation, monitoring events, and services, which can comprehensively, accurately, quickly and intelligently retrieve the monitoring and operation data of massive dispatching equipment, and provide strong support for the business management and decision-making of dispatching equipment monitoring.

**Key words:** equipment monitoring; power domain thesaurus; power knowledge map; intelligent search

随着电网调度自动化系统的建设, 逐步形成了层次清晰、功能完备的调度自动化技术体系, 支撑了电网的安全稳定运行. 随着特高压及新能源接入等

业务的发展, 电网规模日益扩大, 电网的形态发生变化, 积累了丰富的调控运行数据, 促进了电网调度的数字化和智能化发展趋势.

收稿日期: 2021-09-01

基金项目: 国家电网有限公司华东分部科技项目(SGHD0000DKJS2100225)

作者简介: 翟海保(1978-), 男, 安徽省巢湖市人, 高级工程师, 博士, 从事电力系统自动化、电力系统运行工作.

电话(Tel.): 021-38732946; E-mail: zhai\_hb@ec.sgcc.com.cn.

同时,云大物移智等新的 IT 技术的发展,也为智能型调度转型提供了新的思路 and 手段,云平台为调度自动化系统的数字化及智能化提供基础设施,结合调控系统运行现状及发展趋势,引入云计算、大数据等先进成熟的 IT 技术,建设物理分布、逻辑集中的调控云平台<sup>[1]</sup>,实现了资源的弹性调配、数据的标准化管理、业务的服务化封装.

调控云利用云计算技术,将服务器、存储设备和网络设备等资源虚拟化,建立计算、存储和网络资源池,为调度业务的信息化提供保障.按照电网调度通用数据对象结构化,以电网对象为数据标准化的设计核心,设计了模型数据云平台、运行数据平台、实时数据云平台、大数据平台,形成了规范化、一体化、标准化的调控数据体系,为调控业务的数字化提供基础.同时,按照互联网服务化设计理念,根据调控业务特征,进行业务拆分、封装,设计了以服务化为中心的应用构建方法,形成了清晰的业务体系,提高了服务的复用性,按照快速迭代、服务共享等互联网思维,快速构建应用满足业务的发展.

调控中心集中监控面临电网规模不断扩大,监控设备增多、监控信息数量倍增、异常故障情况愈加复杂等情况随着调控云模型、运行、管理等数据不断接入,服务体系逐步完善,业务应用逐步丰富,面临数据量大、种类复杂、服务多样、不易检索的问题,传统的门户网站不能满足业务人员快速获取数据的需求.

本文结合互联网搜索引擎相关技术,根据设备监控业务和调控云数据、服务自身特点,设计面向设备监控业务的智能搜索框架,实现监控设备模型、设备运行、监控事件、调度规程等各类数据的接入,对数据进行全面、快速、准确、智能的检索,成为调控生产运行管理人员与调度数据的新型交互方式.

## 1 设备监控搜索业务分析

### 1.1 搜索范围

搜索框架主要搜索范围是调控云存储的海量数据和构建的大量服务(见图 1).基于结构化设计标准规范,在调控云中汇集了模型、运行、电网资料等调控中心各类业务数据.其中,模型数据包括组织机构、电力容器、一次设备、自动化设备、保护设备等电网基础模型,运行数据包括量测历史数据、电量、事件、故障、告警、计划、预测、负荷等数据<sup>[2]</sup>.这些数据种类多,数据量大,台账履历、故障异常等业务数据的快速搜索能够帮助调控人员更好地利用数据,辅助调控运行管理工作.

按照服务化设计规范,在调控云中建设了大量的服务,包括公共类服务、基础类服务、模型类服务、数据类服务、计算类服务、展示类服务和交互类服务等,每个服务都是独立的运行单元,融入整合了固定的业务逻辑.针对这些服务的搜索能够辅助调控运行人员快速定位到相关的业务服务,满足对应的业务需求.

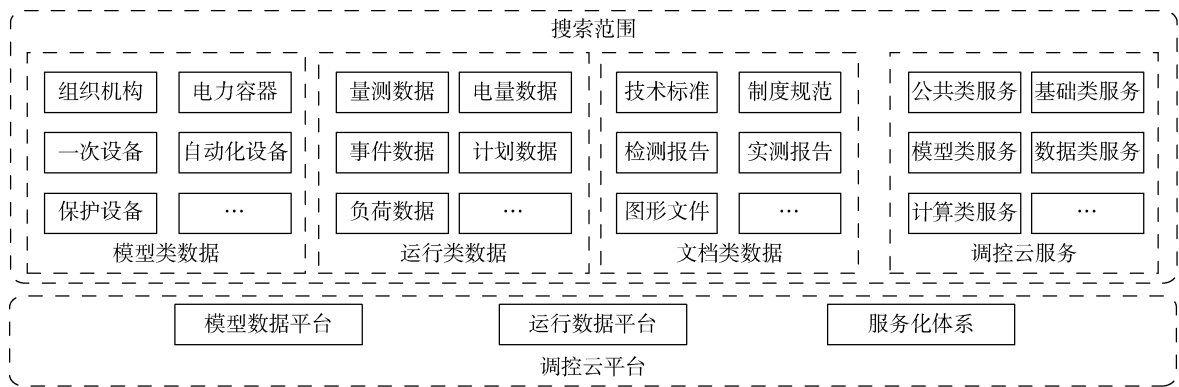


图 1 智能搜索范围分析业务架构

Fig. 1 Application architecture diagram of intelligent search range analysis

### 1.2 智能搜索的要求

随着电网业务日益复杂,调度运行工作挑战加深,基于海量的数据和服务,调度运行人员对数据和服务提出更加全面、准确、快速、智能的需求,智能搜索需要满足简化调控人员的工作复杂度,提高运行管理效率,同时利用海量数据分析加强对电网的支

撑能力,促进智能化及数字化调度发展.具体要求如下:

(1) 搜索的全面性.

针对电网发生的复杂事件,监控运行人员需要融合各类信息进行分析,作出调度运行的决策.因此,搜索框架要支持模型、运行、流程、文档等各类数

据的全面搜索,并且将各类信息关联展示,推送给监控运行人员进行分析判断。

(2) 搜索的准确性.

互联网通用搜索更多的是将匹配的内容展示给用户,用户在搜索结果的前两页自己判断所需要信息.但是,电网是精准性的物理系统,对信息的准确性要求高,如何从多种多样的数据与服务中,根据搜索词准确地检索到信息是智能搜索框架的重点.

(3) 搜索的快速性.

面对海量的模型、运行、资料等数据,需要构建高效的搜索模型,快速地实现数据的搜索,按照搜索引擎的传统标准,将搜索时间控制在 2 s 以内.

(4) 搜索的智能性.

依托电网模型拓扑,挖掘出模型中的全部附属对象及相关属性,实现由“单一数据检索”智能驱动到多维度立体化“全景信息模型”.

(5) 搜索框架的扩展能力.

随着调控云的建设,各类数据的不断接入,以及服务的逐步建设,要求智能搜索框架应该具备较好的扩展性,从而快速实现数据的接入,服务的调阅,满足调控业务的快速发展以及响应搜索需求的不断变化.

2 智能检索关键技术研究

本文研究的智能搜索框架主要包括电力领域词库建模、电力特征分词算法、多因子相关性排序、知识图谱关联建模等关键技术,为面向设备监控智能搜索框架的构建提供技术基础.

2.1 电力领域词库构建技术

搜索引擎的分词算法、排序算法、知识图谱的构建都以文本切词为基础,英文单词以空格区分词语,按照空格来拆分词语,中文的词语都是连续性的,需要构建中文词库,提取出中文词语.电网行业作为展业领域,除了包含常规的中文词语外,电网设备名称更是由行业专业词汇构成,其中也不乏地名等常规预语料的结合,形成综合性复杂名称.

图 2 所示为电力领域词库建模.本文在通用词库的基础上扩充电力系统专业术语,如“兴化电厂”“金陵变”“励磁涌流”“断路器”“自动化”等,按照层次法对词语进行建模.利用层次建模法,包含业务子库、词语标签、词语项目 3 个层次,确保词库管理的规范性、扩展性、通用性、可推广性,建模以后可以在各个电网部门,特别是各级调度内部进行重复使用,协同扩展.

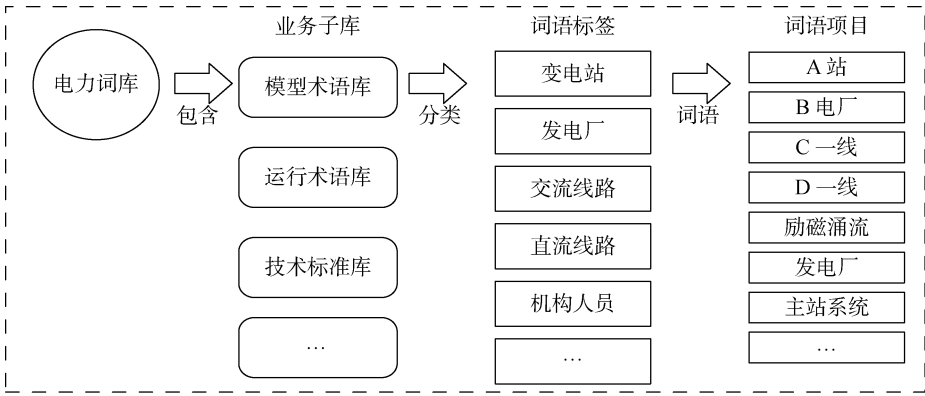


图 2 电力领域词库建模

Fig. 2 Thesaurus modeling of power domain

词语的提取主要采用两种方式:① 利用已有的发电厂、变电站、交流线路、直流线路等调控云模型表中存储的对象,进行歧义字剔除、去重、分类,存储到对应的业务子库中;② 从调度规程、稳定限额规程、技术标准、规范等非结构化文件中,利用 TextRank 算法提取关键字<sup>[3]</sup>,然后进行分类,存储到对应的业务子库中,无法分类的放到技术标准子库中.

算法 1 TextRank 关键词提取算法.

Input: 输入调度规程等文本数据  $T$

Output: 电力专业关键词语

1. 把给定的文本  $T$  按照完整句子进行分割
2. 形成句子数组:  $T = \{S_1, S_2, \dots, S_n\}$
3. 对  $S$  分词,过滤停用词,保留名词
4. 得到  $S_i$  的分词结果关键词:  $S_i = \{t_{i,1}, t_{i,2}, \dots, t_{i,m}\}, i \in [1, n]$
5. 构建结果关键词图  $G = (V, E)$ ,其中  $V$  为节点集,由  $S_i$  产生的候选词组成, $E$  为节点的边,将节点间相同词语个数定义为  $M$ ,表示窗口大小
6. 利用 TextRank 公式,迭代传播计算各节点的权

重,直至收敛

7. 根据权重倒序排列节点,获得排序最高的  $T$  个单词

8. 将获取的词语进行分类,存储到对应的子库中

## 2.2 电力特征分词算法

一般两种情况下涉及到分词算法<sup>[4]</sup>,一种是针对搜索词先进性分词,然后去索引里匹配;另一种是针对模型、运行、文档等数据,进行分词后建立搜索索引.基于电力领域词库以及电力分词的准确性要求,本文采用最大正向匹配分词方式,实现基于电力特征的分词算法.

目前 IT 界常用分词中的实体识别方式包括基于词匹配、词信息、词向量和词属性<sup>[5]</sup>.但是,由于电力词汇的领域特异性,使得基于统计学方式的实体识别在电网领域的识别准确率较低,故本文采用基于领域词库的电力特征分词方法.

电力分词的整体流程是:① 先对输入词进行虚词过滤,过滤到无用的语气词、形容词等;② 利用电力词库和最大正向匹配算法,对句子字符串拆分,获取分词结果;③ 形成分词列表返回,如针对“A 电厂 6 号机组的发电量是多少”这句搜索,其中“A 电厂、机组、发电量”都是电力词库中特殊词汇,最终拆分结果为数组(A 电厂、6 号、机组、发电量).整体流程如图 3 所示.

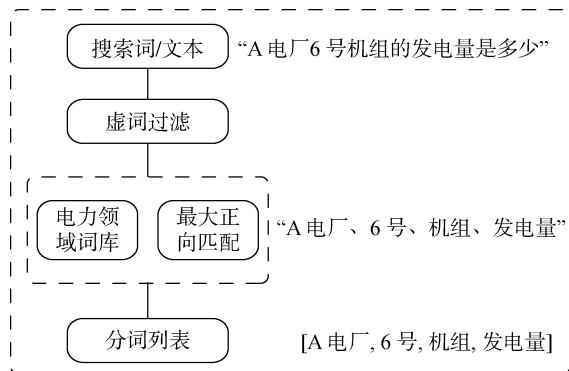


图 3 电力特征分词流程

Fig. 3 Power feature word segmentation process

该流程中,核心的步骤是最大正向匹配算法,该算法运算效率较高,适合需要频繁调用的分词算法,时间复杂度为  $O$ ,具体算法流程如下.

### 算法 2 最大正向匹配算法.

**Input:** 带分词的文本或搜索词,电力词库

**Output:** 分词结果数组

1. 计算输入字符串字符长度
2. 记录匹配起始位置
3. while 记录的起始位置小于字符串长度 do
4. 记录正向最大长度的单词
5. if 该单词已经与词库中某个单词匹配 then
6. 输出该单词并将指针向下一位移动
7. else
8. 按照单个字切分,输出单个字并将指针指向下一位
9. return 设备名称分次数组

## 2.3 多因子相关性排序

词频(TF)-逆文本频率指数(IDF)是一种信息检索中常见的加权排序算法,用于评估某一个词语对于一个文件或数据的匹配程度<sup>[6]</sup>,可以计算出匹配得分. TF-IDF 核心有两点:① 如果一个词语在文件中出现的频率越高,则该词语越重要,评分越高,用 TF 表示;② 如果一个词语在很多文件中出现,则该词语区分文件的能力较低,评分也越低,用 IDF 表示.

$$T-I=TI \quad (1)$$

式中:  $T$  为词频;  $I$  为逆文本频率指数.

$T$  的计算公式对于在一个文件  $d_j$  里的词语  $w_i$  来说,  $w_i$  的词频可表示为

$$W_{ij} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (2)$$

式中:  $n_{i,j}$  为词语在文件  $d_j$  中出现的次数.

某一特定词语的 IDF 数值可以由总文件数除以包含该词语的文件数,再将得到的商取对数处理后得到:

$$I_i = \lg \frac{|D|}{|\{j: w_i \in d_j\}| + 1} \quad (3)$$

式中:  $|D|$  为语料库中所有文件总数;  $|\{j: w_i \in d_j\}|$  是包含词语  $w_i$  的所有文件总数,因为文件数可能为 0,所以在分母上加 1. TF-IDF 算法得出的排序结果针对于网页的搜索排序相对准确,网页基本上都是文本长度有限,其中包含的关键词也较少,用 TF-IDF 进行网页的区分通常结果较好.可是,针对本文的调控云数据搜索,数据的特征较为不同,主要体现在 3 个方面:

(1) 存在大量的小数据,即包含文本较少的数据,比如一个发电厂的模型参数,其中包含的文本信息主要包括电厂名称、电压等级、所属电网、调管范围等,如果单纯按照 TF-IDF 算法,就会导致 IF 的评分较低,模型参数信息本身较为重要,排名就会比

较靠后。

(2) 存在一部分的大数据,即包含文本较为长、关键词个数多的数据,主要是非结构化的文件。电网的非结构化文件主要是调度规程、技术标准、技术规范等文件,该类文件具有篇幅较长、信息量较大的特征,如调度规程中,会存在检修、计划、故障、机组等很多电力特征词语,并且这些词语会频繁出现,会导致 TF 值较高,同时由于调控云搜索引擎的数据量多会导致 IDF 值较低,于是很多时候文档资料把其他更专业的信息覆盖掉。

(3) 除了 TF-IDF 得出的匹配度评分,影响电网搜索的因素还有数据类型与搜索词的匹配度、数据的时效性、用户所属专业或电网等因素。

因此,在 TF-IDF 评分的基础上,本文引入了改进的多因子相关性排序算法,包括 TF-IDF 评分、数据类型评分、时效性评分、所属电网评分,通过加权系数,形成统一的相关性计算公式,给调度用户返回更关心的信息。

$$S = (R - I) + gR + mP + eU \quad (4)$$

式中:  $S$  为得分;  $R$  为数据类型评分;  $g$  为系数,比如用户搜索词是一个电厂名称时,模型数据类型的评分较高;  $P$  为时效性评分;  $m$  为计算系数,当搜索历史数据时,最新的数据评分进行放大;  $U$  为用户专业及电网评分;  $e$  为计算系数,如用户检索电网故障时,用户所在电网发生的故障优先排名展示。

### 2.4 知识图谱关联建模

本文以上的关键技术主要解决搜索全面、搜索准确、搜索快速的问题。现有互联网领域搜索引擎提供的自然语言的分析检索以及智能音箱提供的多轮交互问答等新型人机交互方式背后都是知识图谱提供的支撑。因此,本文利用调控云结构化设计具有较为规范、数据质量不断提高的优势,依据静态资源维度、计划维度、过程维度数据的分类原则,将调控搜索数据划分为 3 类<sup>[7]</sup>,并引入知识图谱关联建模,构建电网调度领域知识图谱<sup>[8]</sup>,提供“问答式”的搜索体验。主要建模及使用方式如图 4 所示。

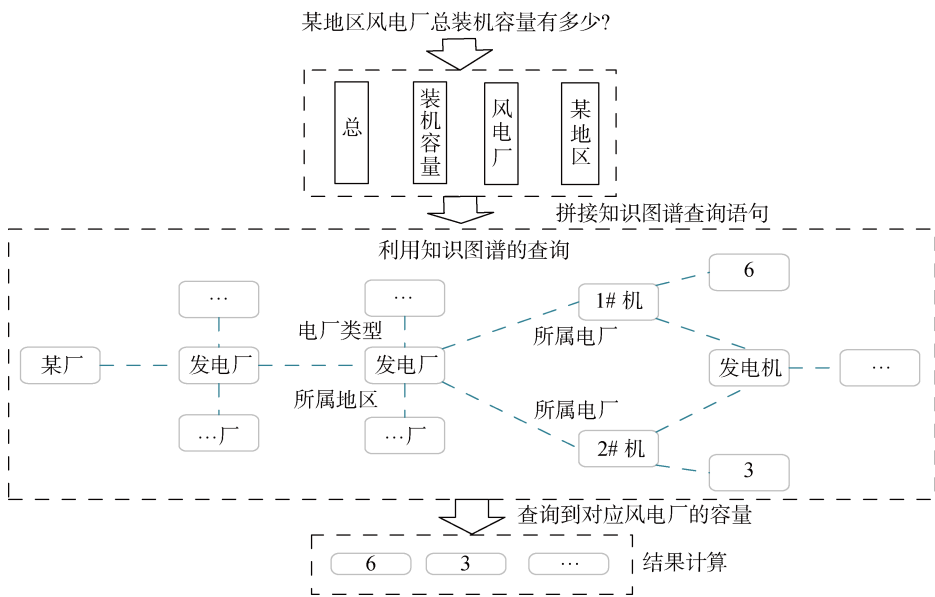


图 4 知识图谱建模及应用

Fig. 4 Modeling and application of knowledge map

如图 4 所示,智能搜索基于结构化的电网调控模型,利用本体的建模理论,分析各类电力对象实体(发电厂、发电机等)之间的关系,并建立电力对象与对象属性、对象事件之间的拓扑关系,构建形成电力本体知识图谱模型。首先确定本体对象(如火电厂),然后以这个本体对象为中心,确定中心本体与其他附属本体(如电厂电量、电厂出力)的关联范围及拓扑关系,逐步形成以中心本体为核心、其他相关本体为枝叶的本体模型,最后对相关底层属性数据

(如当日煤耗量、发电量)进行钻取,形成一套完整电力知识图谱,为智能推理及问答搜索提供基础。如:询问“某地区风电厂总装机容量有多少?”此类问题可直接利用知识图谱将结果计算并返回给用户。

### 3 智能检索框架设计及应用

本文基于电网领域词库、电力特征分词、多因子相关性排序、调控知识图谱等关键技术的研究成果,结合调控云服务总线,设计了一套满足搜索引擎框

架全面、准确、快速、智能、可扩展性的整体框架。

图 5 所示为面向调控云的电网智能搜索框架, 主要包括数据来源、基础服务、业务服务 3 个部分。

(1) 数据来源: 主要利用调控云的模型数据平台、运行数据平台、大数据平台获取结构化数据, 利用文件服务获取非结构化电网资料。

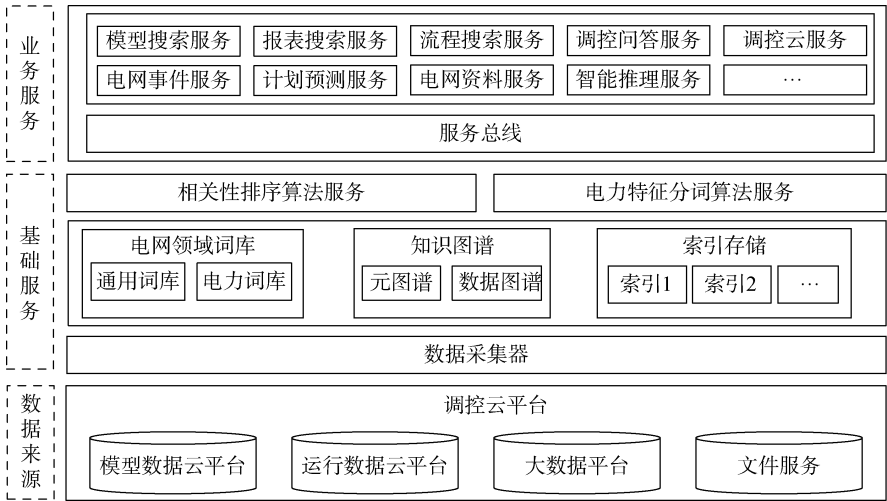


图 5 面向调控云的电网智能搜索框架

Fig. 5 Grid intelligent search framework for regulation cloud

(2) 基础服务: 包括数据采集器<sup>[9]</sup>、电网领域词库、知识图谱、数据索引、排序服务、分词服务。其中, 数据采集器提供了配置管理工具, 能够动态扩展数据来源, 实现模型、运行等各类数据的快速接入, 满足智能搜索全面性、扩展性的要求。索引存储实现各类采集数据按照倒排索引进行存储, 倒排索引能够提高数据检索速度, 满足搜索快速性的要求。电网领域词库、知识图谱、排序算法、分词算法实现了搜索的准确性和智能性。

(3) 业务服务: 利用调控云服务总线<sup>[10]</sup>, 利用基础服务, 实现模型数据服务、电网事件服务、电网资料服务、智能推理服务等, 实现数据的搜索。利用服务化的扩展性为后续业务的变化提供扩展性。在调控云服务中接入公共类服务、数据类服务、计算类服务等调控云规范服务, 实现服务的搜索。

图 6 所示为电网智能搜索应用效果, 为智能搜索框架应用效果, 可评价验证智能搜索的效果。根据智能搜索的要求, 选取验证的搜索词, 定义了搜索速度以及前 10 条搜索结果准确率两个评价指标, 验证结果见表 1。

表 1 智能搜索评价结果

Tab. 1 Evaluation result of intelligent search		
搜索词	搜索速度/s	准确率/%
金陵变电站	1.2	100
江苏电网	1.1	90
江苏电网发电量	1.3	90



图 6 电网智能搜索应用效果

Fig. 6 Application effect of grid intelligent search

经过本文验证基于电网调控领域各类结构化、半结构化、非结构化数据资源, 构建电力调控智能搜索引擎“调控百度”, 实现电力调控数据的全面、准确、快速、智能搜索。

4 结语

面向电网调度设备监控的电网智能搜索框架技术研究将调控云的模型、运行、文档等各类结构化、非结构化数据进行快速、全面检索, 并且支持调控云服务的检索, 将电网调度设备监控领域专业术语进行梳理形成电力词库, 提高了数据检索的准确性。根据模型拓扑关系形成的调度知识图谱为数据的智能

化检索提供基础,为后续的推理、问答等应用场景提供了技术保障.该套框架的设计更好地将调控云规范化的数据与服务开放出来,为调度监控运行及管理人员提供一个便捷、统一的数据访问入口,提高调控系统的数据应用水平,让员工自由方便地获取数据.同时,也为后续泛在物联网对外部用户开放业务数据服务提供了基础,能够探索调控运行领域的泛在物联网应用需求.

参考文献:

[1] 许洪强. 调控云架构及应用展望[J]. **电网技术**, 2017, 41(10): 3104-3111.  
XU Hongqiang. Architecture of dispatching and control cloud and its application prospect[J]. **Power System Technology**, 2017, 41(10): 3104-3111.

[2] 许洪强. 面向调控云的电力调度通用数据对象结构化设计及应用[J]. **电网技术**, 2018, 42(7): 2248-2254.  
XU Hongqiang. Structured design and application of power dispatching universal data object for dispatching and control cloud[J]. **Power System Technology**, 2018, 42(7): 2248-2254.

[3] 徐立. 基于加权 TextRank 的文本关键词提取方法[J]. **计算机科学**, 2019, 46(Sup.1): 142-145.  
XU Li. Text keyword extraction method based on weighted TextRank[J]. **Computer Science**, 2019, 46(Sup.1): 142-145.

[4] 成于思, 施云涛. 面向专业领域的中文分词方法[J]. **计算机工程与应用**, 2018, 54(17): 30-34.  
CHENG Yusi, SHI Yuntao. Domain specific Chinese word segmentation[J]. **Computer Engineering and Applications**, 2018, 54(17): 30-34.

[5] 张辉, 马少平. 基于用户意图的搜索结果文本突显方法[J]. **上海交通大学学报**, 2020, 54(2): 117-125.  
ZHANG Hui, MA Shaoping. Intent-aware search snippet text highlighting method[J]. **Journal of Shanghai Jiao Tong University**, 2020, 54(2): 117-125.

[6] 陈张帆, 庞帆, 张婷, 等. 电力设备信息智能理解与匹配的方法研究及应用[J]. **电脑知识与技术**, 2019, 15(15): 250-251.  
CHEN Zhangfan, PANG Fan, ZHANG Ting, *et al.* Method research and application of intelligent understanding and matching of power equipment information[J]. **Computer Knowledge and Technology**, 2019, 15(15): 250-251.

[7] 宋邓强, 周彬, 申兴旺, 等. 面向船舶分段制造过程的动态知识图谱建模方法[J]. **上海交通大学学报**, 2021, 55(5): 544-556.  
SONG Dengqiang, ZHOU Bin, SHEN Xingwang, *et al.* Dynamic knowledge graph modeling method for ship block manufacturing process[J]. **Journal of Shanghai Jiao Tong University**, 2021, 55(5): 544-556.

[8] 官赛萍, 靳小龙, 贾岩涛, 等. 面向知识图谱的知识推理研究进展[J]. **软件学报**, 2018, 29(10): 2966-2994.  
GUAN Saiping, JIN Xiaolong, JIA Yantao, *et al.* Knowledge reasoning over knowledge graph: A survey[J]. **Journal of Software**, 2018, 29(10): 2966-2994.

[9] 赵海波, 黄勇理, 陈胜, 等. 基于电厂运行数据库的专业智能搜索引擎[J]. **电力自动化设备**, 2003, 23(8): 25-28.  
ZHAO Haibo, HUANG Yongli, CHEN Sheng, *et al.* Professional intelligent search engine based on running database of power plant[J]. **Electric Power Automation Equipment**, 2003, 23(8): 25-28.

[10] 尚学伟, 赵林, 范泽龙, 等. 基于调度数据网的广域数据总线体系架构和关键技术[J]. **电力系统自动化**, 2018, 42(11): 109-114.  
SHANG Xuewei, ZHAO Lin, FAN Zelong, *et al.* Architecture and key technologies of wide-area data bus based on dispatching data network[J]. **Automation of Electric Power Systems**, 2018, 42(11): 109-114.

(本文编辑:陈晓燕)