

基于地理信息的智能搜索在电网GIS平台中的研究和应用

随着国网北京市电力公司电网GIS平台应用的不断深入以及使用范围的不断扩大,相关业务部门纷纷对其提出了更高的应用要求。采取全文检索技术,通过开发电网资源搜索引擎,实时建立数据索引,实现基于地理信息的电网资源智能化搜索功能,能够在电网GIS平台中快速查找、定位电网资源和设备,进一步提升电网GIS平台的信息挖掘能力,大幅提高电网GIS平台数据检索的方便性、易用性和综合展示效果,在公司日常生产管理中工作中得到广泛的应用。

恽珺 叶妍 尹康 / 国网北京市电力公司信息通信分公司



恽珺/工程师

关键词/Keywords

GIS ·
电网资源 ·
搜索引擎 ·
输入助理 ·
索引 ·

0 引言

目前,国网北京市电力公司(以下简称“北京公司”)现有的电网地理空间信息系统(Geographic Information System, GIS)平台已形成了贯通输、变、配、用和通信等业务环节的一体化电网模型及其模型维护机制,能够以面向服务体系架构^[1](Service-Oriented Architecture, SOA)的方式为其他业务应用系统提供电网GIS图形及分析服务,并成功应用在北京公司日常生产管理中。但是,随着GIS应用的不断深入以及使用范围的不断扩大,相关业务部门纷纷提出了更高、更严的要求,因此需要对现有的电网GIS平台作进一步扩展。

随着北京公司电网GIS平台应用的不断深入,电网空间信息数据也开始呈现爆炸性增长的趋势,用户要在信息海洋里查找自己所需的信息,就像大海捞针一样。在现有电网GIS技术中,对电

网设备空间信息的查询主要由空间数据库搜索担任,设备属性信息则由关系型数据库搜索完成,信息搜索机制没有充分考虑利用电网设备之间的空间关联关系,其信息搜索的操作效率会大大降低,并且在电网资源的智能化信息检索以及显示方面,其功能仍然比较薄弱。

搜索引擎技术恰好解决了这一难题。搜索引擎是指电网GIS平台上专门提供检索服务的一类功能,电网GIS平台中大量的信息经过加工处理建立信息数据结构和数据索引,从而对用户提出的各种检索进行响应,并提供用户所需的信息。智能搜索是指根据查询关键字,从电网GIS平台的各类设备属性数据、矢量地图中的地名和道路名称数据中查询包含关键字的数据记录,并能够在电网GIS平台中进行快速定位。

因此,通过开发新型电网资源搜索引擎,进一步提升电网GIS平台的信息挖掘能力,实现基于地理信息的快速、智能化信息搜索功能,使电网GIS平台的应用更加便捷和高效。

1 智能搜索技术方案

原有电网GIS平台采用关系型数据库对设备的属性信息进行管理。在关系型数据库中使用模糊匹配查找时,无法通过建立索引来提高查询性能,搜索过程类似于逐行对数据库进行记录,对于模糊查询的数据库操作,尤其是对多个表、多个字段进行模糊查询时,其查询性能非常低,无法满足快速搜索的要求。因此,为保证电网GIS平台智能搜索

的查询效率,本文采取全文检索技术^[2],重新建立数据索引来实现快速、智能的电网资源信息搜索功能。

全文检索是一种将文件中所有文本与检索项匹配的文字资料检索方法。该方法由计算机程序通过扫描文章中的每一个词实现,对每一个词建立一个索引,指明该词在文章中出现次数和位置。当用户查询时,根据建立的索引进行查找,类似于通过字典检索字表查字的过程,目前主流的搜索引擎都采用了该技术。其核心技术点包括数据索引、数据搜索和中文分词等内容。

1.1 数据索引

(1) 有下列两条原始数据

原始数据1

唯一标识:

EA8AD9AC8CC34C17840CD1B7C4BC6B8B

户号: 0000310262

用户名称: 崇文门菜市场

现用名称: 崇文门菜市场物美店

联系人: 张三

用户地址: 崇外大街甲2

原始数据2

唯一标识:

61E9ABB6CA1409F0E0440017A4762D5E

变电站编码: 120

变电站名称: 崇文门站

变电站地址: 崇文门外大街

(2) 通过以下两个步骤,建立全文检索所需的数据和索引

1) 将所有设备、设施、地名和路名等信息复制到一张数据索引表中,即设备全表,如表1所示,该表需根据记录号建立索引。

表1 设备全表

记录号	设备类型	设备唯一标识
1	用户	EA8AD9AC8CC34C17840CD1B7C4BC6B
2	变电站	61E9ABB6CA1409F0E0440017A4762D5E

2) 提取关键字,并建立倒排索引表,即关键字索引表,如表2所示,该表需根据关键字建立索引。

表2 关键字索引表

关键字	记录号
崇文	1
崇文门	1
菜市场	1
物美	1
崇外大街	1
甲	1
甲2	1
崇文	2
崇文门	2
站	2
外	2
外大街	2

1.2 数据搜索

在建立表1和表2后,可根据查询关键字从这两张表中进行关联数据查询。

1) 对于单个关键字,可直接根据该关键字进行准确的匹配查询,其结构化查询语言

(Structured Query Language, SQL) 语句如下select a.记录号,

a.设备类型, a.设备唯一标识 from 表1 a, 表2 b where a.记录号 = b.记录

号 and b.关键字 = keyword order by a.设备类型

2) 对于两个或多个关键字,可分别根据多个关键字进行准确的匹配查询,并将查询结果中多次匹配关键字的记录显示在最前面,其SQL语句如下

select * from 表1 a, (select 记录号, count(*) count from (select 记录

号, 关键字 from 表2 where 关键字 = keyword1 union select 记录号, 关键字

from 表2 where 关键字 = keyword2)group by 记录号) b where a.记录号 =

b.记录号order by b.count desc, a.设备类型

1.3 中文分词

在上述过程中,如何从一个较长的名称或地址中找到所有单词(即分词),是实现智能搜索功能的重点。英文单词由于用空格分隔,比较容易处理;而中文单词间是连在一起的,其分词难度相对比较大,本文采用机械匹配法、特征词库法等方法进行处理。

1.3.1 机械匹配法

机械匹配法的基本思想是:事先建立一个单词库,其中包含所有可能出现的单词,对于给定的待分词的汉字串S,按照某种确定的原则切取S的子串,若该子串与词库中的某词条相匹配,则该子串是词,继续分割剩余的部分,直到剩余部分为空;否则,该子串不是词,转上重新切取S的子串进行匹配。具体原理如图1所示。

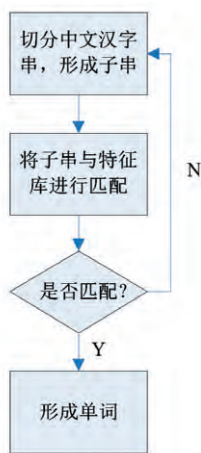


图1 机械匹配法基本原理

1.3.2 特征词库法

特征词库法的基本思想是：事先建立一个特征单词库，其中包含各种具有切分特征的词，对于给定的待分词的汉字串S，首先根据特征词库将S分割成10个较短的子串，然后对每个子串分别采用机械匹配法进行切分。由于每个子串都比S短，所以切分速度较快。具体原理如图2所示。

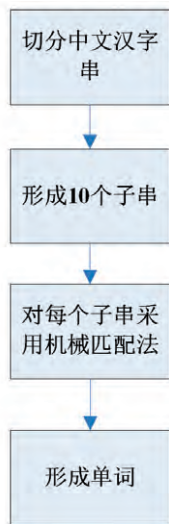


图2 特征词库法基本原理

2 主要功能设计

2.1 建立索引表

实现智能搜索的关键技术是建立设备全表和关键字索引表，这两张表中的数据将随着设备内容的变化而变化，电网GIS平台能够提供定时维护索引和实时自动维护索引两种方式。

定时维护是指根据数据库中设备、地名和路名等

全部内容，定期重新建立设备全表和关键字索引的过程；实时自动维护是指根据设备等数据内容的变化增量，实时对设备全表和关键字索引自动维护的过程。

2.2 智能搜索功能

根据设备类型和关键字等搜索条件在电网GIS平台中查询相关数据，查询结果可根据数据记录中该关键字出现频率、所属设备类型等进行排序，能够支持对多个关键字同时进行数据搜索的功能。

执行智能搜索后，其查询结果采用列表和图形结合的方式进行分页显示，每页可显示固定条目的搜索结果，并能够在图形窗口中对显示搜索结果的位置进行标注。

2.3 输入助理

在搜索输入框中输入查询关键字时，可提供相关关键字下拉框的方式协助用户进行输入。如当用户在搜索框中输入“崇文”时，会出现包括崇文区、崇文门及崇文门站等内容的下拉列表框，以方便用户输入。

2.4 综合信息展示

电网GIS平台能够提供互动性强的综合操作面板，通过操作面板可快速执行空间定位、属性资料查询和统计分析等功能。具体功能如表3所示。

表3 综合信息展示功能

序号	设备类型	功能描述
1	站房（变电站、开关站和配电室）	居中显示、详细资料、接线图、站房设备和供电范围统计等功能
2	输电线路/配电线路	居中显示、详细资料和查看专题图等功能
3	高压用户	居中显示、详细资料、接线图、站房设备和供电范围统计等功能

2.5 智能搜索服务

将电网GIS平台中的智能搜索功能进行服务封装，并且以SOA的方式集成到企业服务总线（Enterprise Service Bus，ESB）中，使其他业务应用系统无须复杂开发，只须通过服务调用的方式即可具备完整的智能搜索功能^[3]。

3 关键技术与创新点

本文涉及的关键技术和创新点主要有以下几方面：

1) 智能搜索引擎技术。采用全文检索技术，构建基于地理信息的电网资源智能化搜索引擎，能够在电网GIS平台中快速查找、定位电网资源。

2) 索引表维护。通过预先建立属性数据索引表和关键字索引表,提升电网GIS平台模糊搜索查询效率,并且支持定时维护索引表和实时维护索引表两种方式。

3) 中文分词。采用机械匹配法和特征词库法相结合的方法,对输入的中文进行分词,达到计算机自动识别语句含义的效果。

4) 模糊查询功能。以描述文句为搜索输入项,进行智能化的关键字模糊搜索,并且支持多关键字查询。

5) 智能化搜索服务。将智能搜索功能以SOA的方式进行服务发布,使其他业务应用系统通过服务调用的方式即可获得智能搜索功能。

4 应用效果

在电网GIS平台的输入框中输入关键字“城子”,左边的综合操作面板会显示所有符合条件的结果,包括变电站、配电室和高压用户等设备类型,搜索结果采用列表和图形结合的方式进行分页显示。

在电网GIS平台中显示的搜索结果并不是孤立的,而是和电网设备的拓扑、地理位置密切相关,可以实时进行关联互动,如显示搜索结果的地理位置、查询设备的属性资料及分析设备的供电范围等,具体效果如图3所示。

5 结束语

本文采用全文检索技术,构建了基于地理空间信息的电网资源智能化搜索引擎,能够在电网GIS平台中实现对电网资源的快速查找、定位和互动,使电网GIS平台在数据检索的方便性、易用性及综

合展示效果等方面获得跨越性提升,大幅提高了信息搜索的效率和便利性,使北京公司各级管理人员、一线生产人员能够方便地查找、定位电网设备和资源,在公司内部获得了广泛的应用效果。



图3 电网GIS平台智能化搜索效果

参考文献

- [1] Thomas Erl. SOA服务设计原则[M]. 郭耀,译. 北京:人民邮电出版社,2009.
- [2] 苏新宁. 信息检索理论与技术[M]. 北京:科学技术文献出版社,2010.
- [3] 邴冬燕. SOA构架电网GIS应用的研究与实践[J]. 信息化应用,2009:349-353.

(收稿日期:2015-05-26) EA

作者简介

恽珺(1984-),女,工程师,硕士研究生,主要研究方向为智能电网信息化建设、一体化平台应用研究、应用运维管理。

叶妍(1963-),女,高级工程师,硕士研究生,主要研究方向为信息系统建设、技术管理。