

科技基础条件平台空间信息服务的实现

王希友¹, 曹维芳¹, 程振林², 董 慧²

(1. 新疆哈密吐哈油田, 新疆 哈密 839009;

2. 中国科学院 计算技术研究所, 北京 100190)

摘 要:空间信息服务节点应用在国家科技基础条件平台应用服务支撑系统中, 为网络科技环境中的应用提供地图服务。文中从设计、总体架构、软件功能、实现、示范应用等方面对空间信息服务节点进行了阐述, 并着重对其中所用软件工具以及工作流程、兴趣点搜索的实现等重点进行了说明。经过近一年时间的测试运行, 空间信息服务节点运行稳定, 并在国家科技成果平台等系统中得到了应用。以空间信息服务方式提供地图功能, 减少了数据重复采集的成本与一致性维护的复杂工作, 并且能提供较好的访问效率, 被证明是高效的实现方法。

关键词:国家科技基础条件平台; 空间信息服务节点; 应用服务支撑系统

中图分类号: TP393

文献标识码: A

文章编号: 1673-629X(2009)11-0111-04

Implementation of Geographic Information Service in National Science & Technology Infrastructure

WANG Xi-you¹, CAO Wei-fang¹, CHENG Zhen-lin², DONG Hui²

(1. Tuha Oilfield Company, China National Petroleum Corporation, Hami 839009, China;

2. Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Geographic information service is applied to National Science & Technology Infrastructure Development program. It provides services to applications in the science and technology network environment. This article gives an overview of the design, architecture, implementation and application of the geographic information service system, focusing on the software tools and workflow for the implementation, the retrieval of point of interest. Geographic information service has been tested and applied in the national science & technology achievement platform.

Key words: national science & technology infrastructure; geographic information service; application supporting environment

0 引言

国家科技基础条件平台建设计划旨在运用现代信息技术, 以资源共享为核心, 对科技基础条件资源进行合理布局、战略重组和系统优化, 创新管理体制和运行机制, 促进全社会科技资源的科学配置和有效利用, 为建设创新型国家提供有力支撑。该平台包括大型仪器和实验基地平台、自然科技资源共享平台、科学数据共享平台、科技文献共享平台、网络科研环境平台、成果转化服务平台等^[1]。其中国家科技基础条件平台应用服务支撑系统是网络科技环境平台中的核心之一。

作为应用服务支撑系统的一部分, 空间信息服务

节点为网络科技环境中的应用提供地图服务。人类获得的信息中 80% 包含空间成分 (80 percents of the information collected has a geographic or spatial component to it)。科技资源和数据 (包含科研工作、科研管理中产生的数据) 也包含空间成分。要对这些科技基础条件资源进行基于地图的展现、查询、浏览、统计分析与挖掘等功能, 都离不开空间信息服务。

空间信息服务节点的建设可以为整个科技基础条件平台提供空间信息相关的服务, 应用在诸多的行业和领域^[2]。譬如, 研究实验基地和大型科学仪器设备的空间分布和聚集程度展示; 自然科技资源中动物、植物种质资源的地域分布、人类遗传资源的空间特征展示等; 各个行业、领域 (包括气象、农业、海洋、地质、能源、交通、科技等) 的科学数据的基于地图的统计、分析、规划等。在全国科技信息服务网及县市科技信息服务平台中, 作物或水产等的地理分布、气象情况的展

收稿日期: 2009-04-20; 修回日期: 2009-07-22

基金项目: 国家科技基础条件平台项目 (2005DKA3900); 中国科学院知识创新项目 (20036020)

作者简介: 王希友 (1974-), 男, 山东人, 硕士, 研究方向为大型信息系统开发。

示、土壤资料的展示以及基于土壤资料的作物种植专家系统等都可以用到空间信息服务节点。在科技成果信息服务平台中科技成果、试验结构、专家的地域分布、统计查询等,也可以基于地图服务来实现。

1 系统实现

1.1 总体结构

空间信息服务节点可以分为软件、硬件和相关的支撑与安全环境三大部分,见图 1。其中,软件环境建设又分为空间数据库系统建设、空间信息服务(地图服务)系统建设、辅助工具(包括地图发布工具等)以及空间信息服务节点示范应用(基于地图的统计图表系统、基于地图的科技信息发布系统)等内容。空间信息服务节点提供 JavaScript 类库(通过 Ajax 方式与后台服务器进行交互)进行二次开发^[3,4]。

空间数据库系统为空间信息服务节点提供数据,是空间信息服务节点生命的源泉。地图服务系统是空间信息服务节点建设的核心;辅助工具用来维护空间信息服务节点、帮助用户生成基于地图的应用。

硬件环境建设主要包括网络环境建设和统一存储资源环境建设等内容。网络环境建设为空间信息服务节点提供安全、高效的 Internet 和 Intranet 服务;统一存储资源环境建设将为空间信息服务节点提供统一的数据存储与访问服务;支撑与安全环境建设为空间信息服务的节点的运行支撑、管理提供支持,并为空间信息服务节点提供严格统一的安全机制和容灾备份机制。

表 1 软件工具列表

| 软件工具 | 所用功能 |
|------------|---------|
| MapBuilder | 地图发布 |
| CoCatalog | 空间数据管理 |
| CoMap | 空间数据编辑 |
| VegaSymbol | 空间数据符号化 |

以上工具基于织女星地理信息系统(VegaGIS)平台。织女星地理信息系统(VegaGIS)的技术核心是中科院计算技术研究所空间信息处理技术实验室在国家十五期间国家 863 项目“面向网络海量空间信息的大型 GIS”支持下,在国家火炬计划、国家信息产业部电子信息产业发展基金、国家科技基础条件平台等国家项目的大力推动下,研发的具有自主知识产权、支持分布式计算的大型地理信息系统支撑平台。VegaGIS 在科技部组织的国家 GIS 基础软件平台中表现优良。其符号库系统(VegaSymbol)和支持分布式异构数据库的空间数据引擎^[5](CoSDE)多次获得表彰。这些成果已经广泛应用在国家大型灌区信息化示范工程建设、国家科技基础条件平台建设、某军区情报系统建设、科技部信息中心和多个省市信息中心建设中。

Mapbuilder 系统需要实现的主要目标是提高矢量地图发布的自动化程度^[6],应对大规模数据的要求,可在拥有完整数据配置信息(.xml)和对应的工程(.cmp)的情形下,直接打开 xml 和 cmp,进行添注记、铺符号、平移要素、平移注记,可以更新 cmp,以及直接发布图片。

CoSDE 系统实现的主要目标就是以可视化的方式

来统一管理空间数据。主要操作对象为空间数据库中的数据,支持本地数据。主要功能包括:数据源管理、数据管理、数据浏览、空间数据的浏览、数据用户和权限管理等。

CoMap 主要用来完成空间数据编辑,实现各种地物要素如矢量数据、栅格数据、注记和属性数据的编辑、显示。

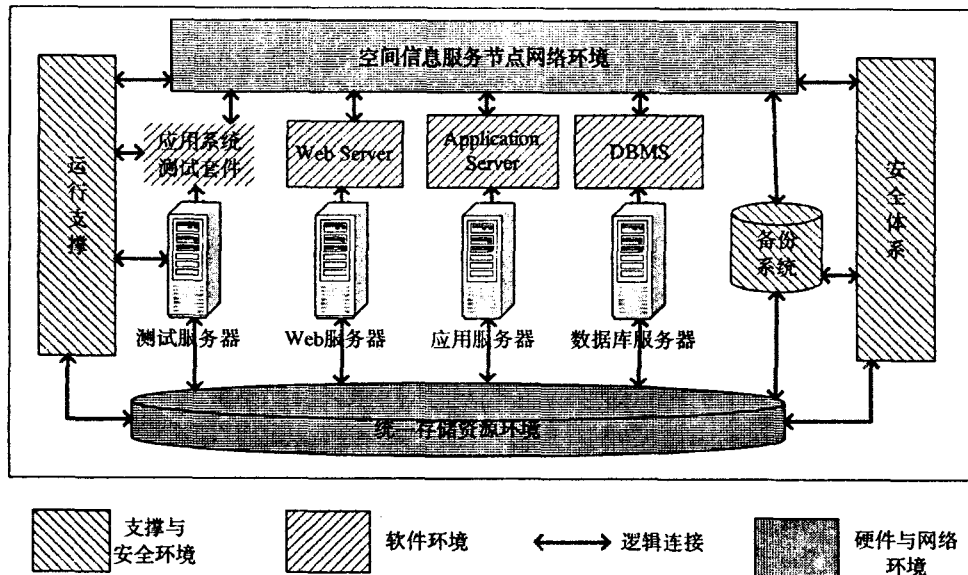


图 1 空间信息服务节点系统结构

1.2 所需软件平台及工作流程

空间信息服务节点建设中的软件工具如表 1。

VegaSymbol 是用来制作地图符号的工具。空间信息节点建设中主要用到其如下功能:符号库的编辑、

符号的设计、符号系统库的管理、符号库的输出、颜色选择与编辑、编辑专色等。

空间数据库系统以全国范围的基础地理数据为基础,集数据的存储、更新、融合与发布等功能于一体的综合性空间数据管理系统。需要管理的数据包括:

(1)发布前的矢量数据:覆盖中国各个省/直辖市、300 多城市的 SHP 文件,文件包含教育机构、科研院所、大中院校、教育培训等近 30 个大类,530 个子类的 193 万个 POI 信息点。

(2)发布后的栅格数据:CIMGE 格式的栅格图像,分 14 个级别,大小为 165G 左右空间数据的整理与发布按如下流程、规则进行:

* 阶段性文件存放位置和个人文件存放位置的目录管理。

* 发布比例尺,进行数据分层。

* 使用 VegaSymbol 软件进行点符号制作。

* 使用 Comap 软件进行数据整理。

* 使用 CoCatalog 软件进行数据入库。

* 使用 CoCatalog、CoSDE 软件在数据库建索引、字段。

* 使用 Comap、Mapbuilder 软件设置图层颜色与符号。

* 使用 Mapbuilder、Comap 软件将数据库数据存为 cmp 文件。

* 使用 Mapbuilder、Comap 添加注记。

* 使用 Mapbuilder 发布图层。

1.3 兴趣点搜索的实现

POI(Point Of Interest)搜索,即地图兴趣点搜索,在技术上与网络搜索一脉相承,支持用户搜索与地理位置相关的兴趣点,如大学、科研机构等。POI 搜索引擎是基于 POI 数据进行搜索的垂直搜索引擎,在逻辑功能上包括分词、索引、搜索、搜索结果排序等,主要由分词、文本索引文本检索以及缓存等模块组成。POI 搜索引擎流程如下:

* 利用已有的 POI 数据集建立 Berkeley DB 数据库。

* 对每个 POI 的名称字段进行分词,建立倒排索引。

* 系统启动时,将整个倒排索引读入内存。

* 接受用户输入查询请求,首先查找缓存,如果命中,则直接返回查询结果;如果不命中,则对请求的文本进行分词,对得到的分词结果分别查找倒排索引,然后对查询结果文本排序,将排序后的结果返回给用户,同时把排序得到的前 50 个结果存入缓存。

本系统中的分词工具采用中科院计算所的 ICT-

CLAS 分词工具^[7],该工具采用了层叠隐马尔可夫模型(Hierarchical Hidden Markov Model)^[8],将汉语词法分析的所有环节都统一到了一个完整的理论框架中,获得最好的总体效果。

POI 搜索引擎中的文本索引采用倒排索引。倒排索引被大多数系统(如百度、Google)所广泛采用。做法如下:将 POI 的名称字段进行分词得到多个词(term),对每个 term 建立倒排索引,索引项是 term,数据项是包含这个 term 的所有 POI 的相关信息,例如,每个 POI 的 ID,该 term 在该 POI 中出现的次数,该 POI 包含的 term 的个数。这样设计也是为查询结果排序算法服务。

访问文本索引获取匹配的查询结果集后利用 TF-IDF 方法计算查询请求和结果集中每个文本的相似度,根据相似度进行排序生成排序列表^[9]。

词条频率(TF)表示某个 term 在某个 POI 的名称字段中总共出现的次数。

$$tf = \begin{cases} 0 & \text{if } \text{freq}(d, t) = 0 \\ 1 + \log(1 + \log(\text{freq}(d, t))) & \text{otherwise} \end{cases} \quad (1)$$

$$idf = \log \frac{1 + |d|}{|d_t|} \quad (2)$$

逆文档频率(IDF)定义为:

$$TF - IDF = tf * idf \quad (3)$$

其中 $\text{freq}(d, t)$ 表示 term t 在文档 d 中出现的次数, $|d|$ 表示所有的文档总数, $|d_t|$ 表示包含 term t 的文档总数。实现中要计算查询请求的 TF-IDF 向量和查询结果集中每一个 POI 名称字段的 TF-IDF 向量,然后计算两者的余弦相似度。

2 基于空间信息服务节点的示范应用

2.1 基于地图的科技信息发布系统

基于地图的科技信息发布系统的研发目标是:通过网页发布、中文分词、空间信息服务等技术,将用户发布的科技信息中的所包含的地理信息,以地图这种更加直观的方式展现出来。也可以将科研机构与其关联的科技基础条件平台资源通过地图关联在一起,实现双向互动查询、展示、浏览。

应用示例如图 2 所示。

2.2 基于地图的统计图表系统

功能目标是用户可以方便地利用此系统从自己的数据源获得关心的或者调查研究所得科技资源数据,通过可视化的数据选取窗口将地域名称与空间信息服务结合,并选择相应的图表类型,定制数据统计规则,然后通过在地图上的直观选择查看相应地域的统计分

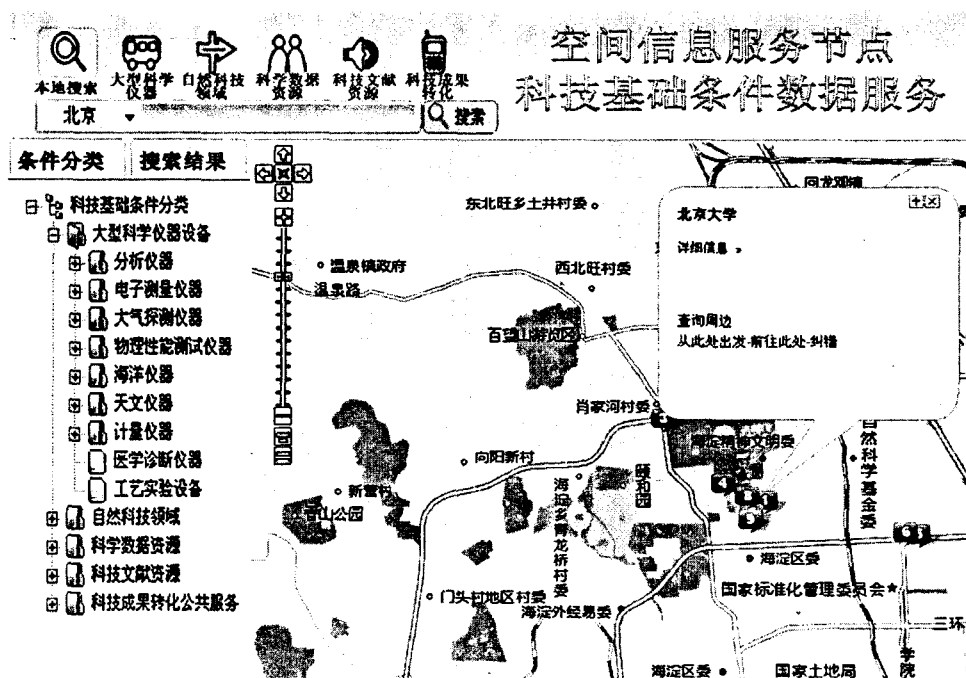


图 2 空间信息服务节点地图服务示例

析图表,最后将结果以动态内容形式导出,便于用户以后使用。系统运行示例如图 3 所示。

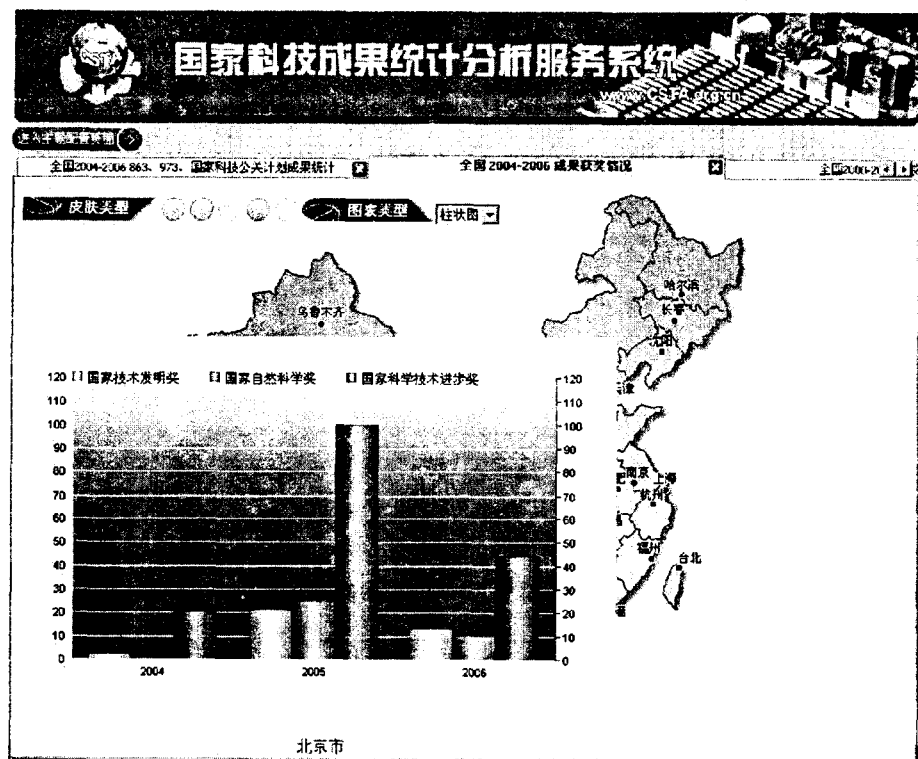


图 3 基于地图的统计图表系统应用示例图(鼠标互动到某省市时动态划出此地区的统计图表)

3 结束语

目前空间信息服务节点已经应用在国家科技基础条件平台应用服务支撑系统中,并正式上线运行。空

间信息服务节点为网络科技环境中的应用服务提供覆盖全国各级各类教育科研机构、近 30 个图层、一站式的空间信息服务。利用空间信息服务可以对植物种质资源进行按地域分布的展示,对各类科技成果进行基于地图的统计、分析、规划,对科技信息增加地图展示等。

对空间信息服务节点下一阶段的完善规划主要包括两方面:

(1)增加 POI 信息点,完善并扩充现有的

POI 数据,使空间信息服务节点包含更多的空间信息;

(2)挖掘更广阔的在科技领域的示范应用,进一步

推广空间信息服务节点在科技基础条件平台的应用。

参考文献:

- [1] 科学技术部等四部委.“十一·五”国家科技基础条件平台建设实施意见.科技部网站 [EB/OL]. 2005. http://www.most.gov.cn/gkjctjpts/zhxx/200508/t20050810_23844.htm.
- [2] Zimmermann R, Ku Wei-Shinn, Wang Haojun, et al. A Distributed Geotechnical Information Management and Exchange Architecture [J]. IEEE Computer Society, 2006 (10):26-33.
- [3] 刘刚,邓飞其,杨长海. AJAX 在 WebGIS 异步数据交互中的研究[J]. 计算机技术与发展, 2009, 19(1):53-56.

- [4] Sayar A, Pierce M, Fox G. Integrating AJAX Approach into GIS Visualization Web Services[J]. IEEE Computer Society,

(下转第 166 页)

```
<? xml version="1.0" encoding="UTF-8"? >
<Config xmlns="http://org.apache.tuscany.das.rdb/config.
xsd">
<!-- ConnectionInfo 描述数据源的连接信息,包括数据源类
型、全局名、用户名、密码等信息 -->
<ConnectionInfo>
<ConnectionProperties
driverClass="com.mysql.jdbc.Driver"
databaseURL="jdbc:mysql://localhost/user"
userName="root"
password="root"
loginTimeout="600000"/>
</ConnectionInfo>
<!-- Command 描述 SQL 操作的名称、内容、类型等信息 -->
<Command name="AllUsers" SQL="select * from user" kind
="Select"/>
<!-- Table 描述了表信息,包括表名、主键等 -->
<Table tableName="USER">
<Column columnName="ID" primaryKey="true"/>
</Table>
</Config>
```

(3) 根据配置信息创建 DAS。

由 DAS 的工厂类中的 DAS.FACTORY.createDAS()方法创建。

(4) DAS 与数据库的连接。

DAS 根据配置信息获取数据源的连接信息,然后与相应的数据源进行连接,由 UserDatabaseInitializer (String configFile)方法和 MySQLSetup (String databaseInfo)实现。本例中数据库连接的代码:

```
if(this.platformName.equals(MYSQL)){
databaseURL = databaseURL + "? user = " + userName +
"&password = " + password + "&createDatabaseIfNotExist = " +
"true"; //databaseURL, user, password 三个变量由 java.util.
StringTokenizer 中的 nextToken()方法获得
connection = DriverManager.getConnection(databaseURL);
}
```

(5) 对数据库的操作。

DAS 与数据库连接后就可以对数据源中的数据进行增加 (set 方法),删除 (delete 方法),修改 (set 方

法)和查询 (getUser 方法)等操作了。

3 结束语

文中将 SDO 规范用于数据集成,提出了新的异构数据集成解决方案,并将其应用到实际项目的开发中。SDO 提供了统一的数据应用开发框架,统一了对多种企业信息系统 (EIS) 的数据访问,通过使用 SDO 这种独特而简单的模型,应用程序能够摆脱使用多种 API 和框架进行数据访问的复杂工作,在开发过程中只需使用一种 API (SDO API)便可操作各种异构数据源。该文提出的基于 SDO 的异构数据集成系统,还可以方便地进行数据源的添加、修改和删除,具有很好的扩展性,因此将 SDO 用于异构数据集成具有一定的理论和实践意义。

参考文献:

- [1] 周运,牟占生,徐久成.基于 XML 虚拟数据库的异构数据源集成模型研究[J].计算机技术与发展,2008,18(4): 84-87.
- [2] Fong J, Wong H K, Cheng Z. Converting relational database into XML documents with DOM[J]. Information and Software Technology, 2003,45(4):335-355.
- [3] 钟华,冯文澜,谭红星,等.面向数据集成的 ETL 系统设计与实现[J].计算机科学,2004,31(9):87-89.
- [4] 袁晓洁,于士涛,李志梁.基于 Mediation 的异构数据集成系统 HDIS 设计与实现[J].计算机工程与应用,2006,42(1):162-165.
- [5] 倪志刚,洪玫,刘佳.基于服务数据对象的异构系统数据集成方案研究[J].计算机应用,2007,27(6):21-23.
- [6] Service Data Objects, WorkManager, and Timers[EB/OL]. 2003-11. <http://www.ibm.com/developerworks/library/specification/j-commonj-sdowmt/>
- [7] Service Data Objects For Java Specification, Version 2.1.0 (pdf)[EB/OL]. 2006.11. <http://download.boulder.ibm.com/ibmdl/pub/software/dw/specs/ws-sdo/SDO-Specification-Java-V2.01.pdf>.
- [8] 陈洋,罗四维.异构数据库数据集成的研究与实现[J].计算机技术与发展,2006,16(7):192-194.

(上接第 114 页)

2006(2):169-170.

- [5] 何雄,方金云,唐志敏.织女星地理信息系统 VegaGIS 中的空间数据引擎 CoSDE[J].计算机应用,2005,25(7): 1587-1589.
- [6] 杨超伟,李琦.Web 空间信息发布研究[J].北京大学学报:自然科学版,2001,37:413-419.
- [7] 汉语分词系统 ICTCLAS[EB/OL]. 版本 3.0. 2006-04.

<http://ictclas.org/index.html>.

- [8] Baeza-Yates R, Ribeiro-Neto B. Modern Information Retrieval[M]. [s.l.]: Addison-Wesley, 1999.
- [9] 潘明远,董刊生.织女星地理信息系统 VegaGIS POI 搜索引擎的设计与实现[R].北京:中科院计算所技术报告, 2007:4-8.