

# 治安信息五要素智能搜索引擎研究及应用

杨令省<sup>1</sup>, 唐金文<sup>2</sup>

(1. 曲靖师范学院 数学与信息科学学院, 云南 曲靖 655011;

2. 曲靖师范学院 计算机科学与工程学院, 云南 曲靖 655011)

**摘要:** 如何把分散的信息有效组织起来, 实现跨部门、跨地区的信息共享, 是金盾工程建设必须解决的一个关键问题。主要介绍如何围绕“人、地、物、事件、组织”治安信息五要素的业务模型, 利用 Web Services、UDDI 等技术创建智能搜索引擎, 实现全省范围内治安信息五要素关联查询及异构系统间的全网漫游。

**关键词:** 治安信息五要素; 关联查询; 异构系统; 数据漫游

中图分类号: TP311.52

文献标识码: A

文章编号: 1674-7720(2011)14-0062-04

## The research and application of public order information five elements intelligent search engine

Yang Lingxing<sup>1</sup>, Tang Jinwen<sup>2</sup>

(1. Institute of Mathematics and Information Science, Qujing Normal College, Qujing 655011, China;

2. Institute of Computer Science and Engineering, Qujing Normal College, Qujing 655011, China)

**Abstract:** How to effectively organize the sporadic information and realize trans-department and trans-region information share is a crucial problem must be resolved in golden shield project construction. The paper mainly introduces the method of building intelligent search engine surrounding the public order information five elements model comprising—"people, terra, material, event and organization", by using Web Services and UDDI etc. technology, realizes the public order information five elements correlation search and data roaming between heterogeneous systems in yunnan province's public security department.

**Key words:** public order information five elements; correlation search; heterogeneous system; data roaming

近年来,公安信息系统从简单的信息处理向以数据为中心的数据共享和关联的复杂应用转变,治安信息五要素智能搜索引擎就是采用先进的计算机技术如 Web Services、SOAP、UDDI、XML、WSDL、分布式技术等构建的应用系统,旨在实现全省范围内人、地、物、组织、事件的关联查询及异构系统间的全网漫游;实现分布在全省不同地方的不同系统间或系统本身的五要素关联查询,并能实现定义与发现异构数据源,能自定义搜索条件与搜索结果的展现方式,容易和其他系统对接。治安信息智能搜索引擎的成功实施推动全省公安信息化的发展。

### 1 治安信息五要素模型

#### 1.1 五要素关系模型

随着全国公安信息化建设的深入,国内各级公安在实践的基础上,总结出治安信息的五要素模型为<sup>[1]</sup>:人、

地、物、组织、事件。人指自然人;地指地理位置;物指生物或非生物;组织是一种社会实体,它是人、地、物等要素之间的一种稳态关系;事件是导致人、地、物、组织等要素关系发生变化的特殊要素,即事件使各要素从一种相对稳态关系转变为另一种相对稳态关系。事件是体现动态和关联的重要环节,只有处理好事件的表述方式,才可以将五要素信息进行有机的组织,达到体现“动态全息”的目的。五要素关系模型见图1。

理想情况下,同一实体在全省公安系统(甚至在全国公安系统)中应只有一个对应要素,所有要素可以唯一标识,要素之间的各种联系是有机和完整的。但在实际条件约束下,由于种种原因导致要素的重复创建:公安不同的应用系统可能对同一实体分别在本系统内重复创建对应要素;以地市级为中心建立治安信息系统

# 技术与方法

## Technique and Method

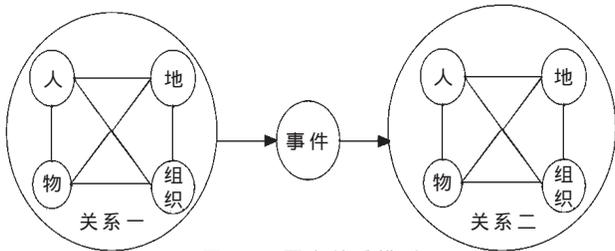


图1 五要素关系模型

时, 导致同一实体在不同城市系统内创建对应要素等。所以, 为简单起见, 这里假定治安管理信息“五要素”业务模型为: 地市内各治安业务之间做到要素唯一创建、唯一标识、末态属性唯一、有机关联; 地市之间, 要素可以重复创建, 通过唯一标识彼此识别, 进一步获取要素的各方面信息; 提供与数据库、文件服务器间的交互支持, 提供数据存储与业务对象的映射, 通过“五要素”业务模型协助解决数据库在性能与访问平衡中的问题, 实现数据库组织策略调整对业务系统的无关性。

### 1.2 五要素数据分布

目前全省范围内治安信息系统分布在省厅、各州、各县区和各派出所。有些系统是使用省中心库, 如省级人口信息系统; 有些是使用地市集中库, 如重点人口系统; 有些是工作数据库分布在派出所, 再将数据库集中到区县库和地州库, 如现有的实有人口管理系统。五要素数据分布网络结构模型见图2。

从图2可以看出, 五要素数据可能分布在不同的地理位置和存储于不同的系统中, 公安机关各职能部门通常都有自己的应用系统, 而且建设模式基本上覆盖了

省、地市、区县、派出所四级应用。智能搜索引擎就是要实现跨部门、跨区域和异构数据的漫游查询, 以 Web Service 的方式为治安工作人员提供统一的操作界面, 屏蔽平台的差异性。

### 2 治安信息五要素智能搜索引擎体系结构

从上述模型来看, 如此庞大的治安管理信息系统, 各自的实现方式是不同的, 如何有效地控制调用这些不存在主从关系的子系统、事务和模块, 如何组织动态的链接、模块之间如何通信等, 是智能搜索引擎需解决的主要问题, 因此提出如图3所示的体系结构。

整个搜索引擎分为3个层次: 用户层、应用层和数据层。其中, 用户层只和用户交互, 不进行任何业务处理, 实际上就是只有客户端。应用层包含系统中业务逻辑的实现和数据访问控制。应用层由 .net 程序及 XML Web Services 实现, 并使用 SOAP 协议与客户端连接。应用层中的事务日志、用户状态管理、五要素数据源与数据表现定义、管理工具、主搜索引擎、五要素资源注册中心等都在省中心应用服务器上运行, 搜索数据库分布在省中心数据库服务器上运行。智能搜索引擎中, 应用层的主要功能如下<sup>[1]</sup>。

(1) 单点登录和用户状态管理: 用户只需一次登录系统, 就能操作权限范围内的功能菜单, 并为每一个登录的用户保存状态信息和记录用户的查询操作信息。

(2) 五要素数据源配置与数据表现自定义管理: 对新接入的系统进行数据源配置, 系统不用修改程序代码便可挂接新系统, 用户可自定义查询结果展现方式。

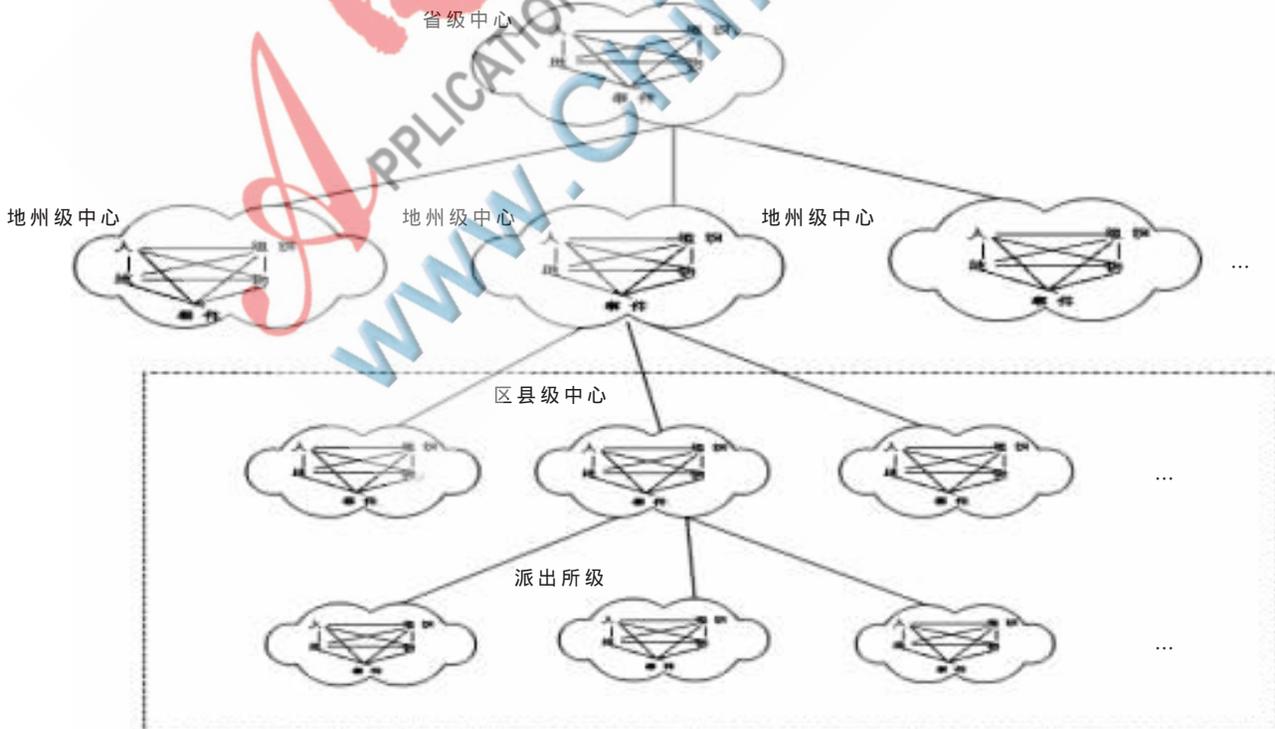


图2 五要素数据分布网络模型图

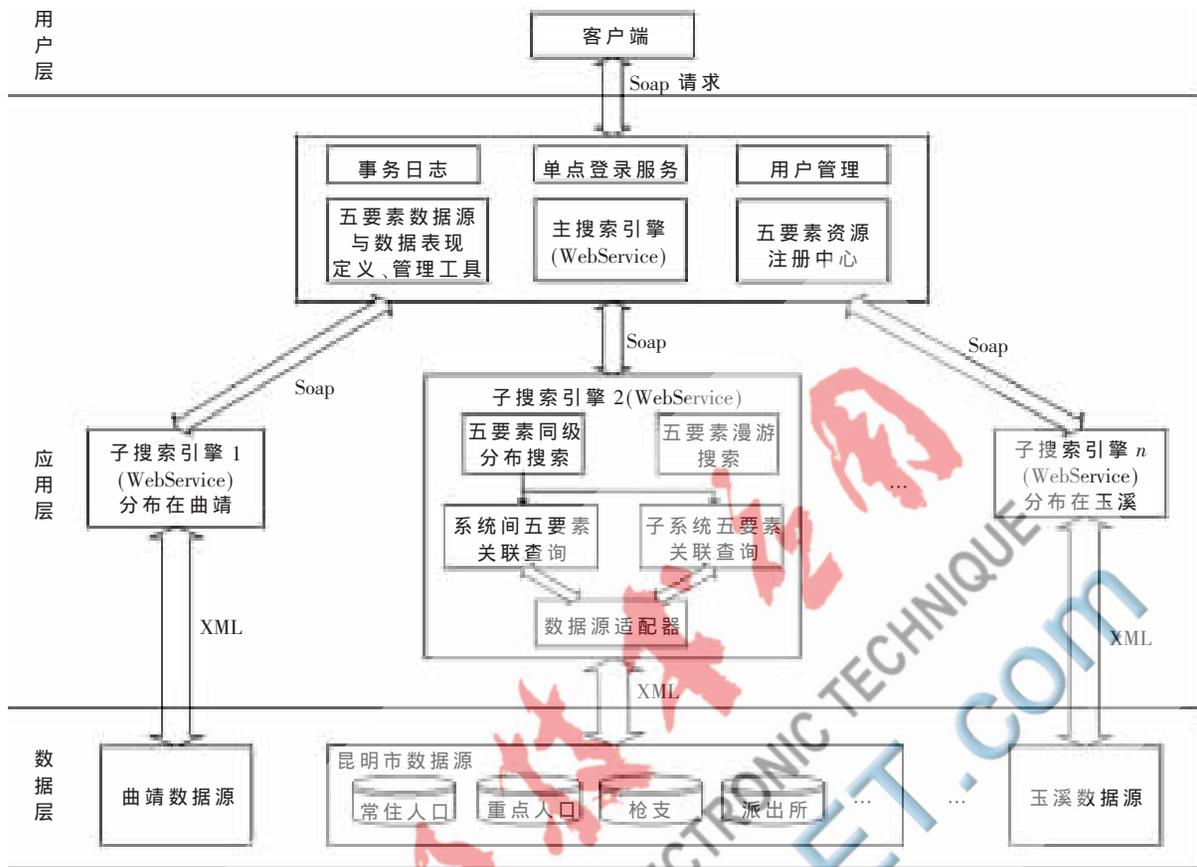


图3 五要素智能搜索引擎体系架构图

(3)分析处理器:接收并分析用户的查询请求,访问UDDI注册中心,调用搜索引擎,并返回数据。

(4)主搜索引擎:搜索任务分配的中心调度器。系统从客户端接收到查询请求后,主搜索引擎便对查询请求进行分析,把任务分发到对应的子搜索引擎,从子搜索引擎接收数据,并返回客户端数据。主搜索引擎与各子搜索引擎定时互发路由消息,以便主搜索引擎掌握各子搜索引擎的路由情况。

(5)子搜索引擎:实现用户定义系统的接入,并把相关信息发送到中心;承担主搜索引擎发送来的查询任务,并返回主搜索引擎所需数据;定时发送路由信息给主搜索引擎,以便主搜索引擎掌握各子搜索引擎的路由情况。

(6)五要素同级分布搜索:对该子搜索引擎管辖范围内的同一地区级的数据库进行搜索。

(7)系统间和子系统五要素关联查询:系统与系统间的关联查询,统一系统内五要素的关联查询,并返回数据。

(8)数据源适配器:调用与目标数据库对应的驱动程序,直接与数据库打交道,向数据库发送查询命令,并取得数据的模块。

(9)五要素漫游搜索:有异步返回数据功能,当子搜索引擎接收到漫游查询后,启动五要素漫游搜索线程,在足够多的时间内搜索目标数据库,将搜索所得的数据

放入漫游数据队列,并返回给主搜索引擎。

(10)五要素资源注册中心:对主搜索引擎注册,使得主搜索引擎通过注册中心被系统程序发现和调用。

采用Web Services<sup>[3,4]</sup>技术将功能组件封装成一个整体的服务,以UDDI的方式注册发布。通过五要素引擎把所有的程序组织起来,把物理位置分布不同、运行环境不同、功能相对独立的系统信息统一起来,模块之间的搭线连接由搜索引擎去实现;系统对用户而言以浏览器为表现方式,网络就是一个面向服务的大计算机,这就突破了传统的分布式计算模型在通信、应用范围等方面的限制。这就是五要素智能搜索引擎“无所不在的计算”的核心思想。

### 3 智能搜索引擎关键技术介绍

#### 3.1 分布式数据适配器

治安信息智能搜索引擎涉及的业务数据量大,而且这些数据广泛分布在不同的应用系统中,每个系统采用的数据库可能不相同,要访问这些数据库,不可能为每种数据库分别编写数据访问代码,而且搜索引擎有时需要同时访问多种数据库。为此,专门设计实现了一个数据适配器,用于访问分布式的异构数据库<sup>[5]</sup>,其实现框架如图4所示。

分布式数据适配器综合利用线程技术、Web Service技术和搜索引擎提供的统一数据访问器和路由表服务

# 技术与方法

## Technique and Method

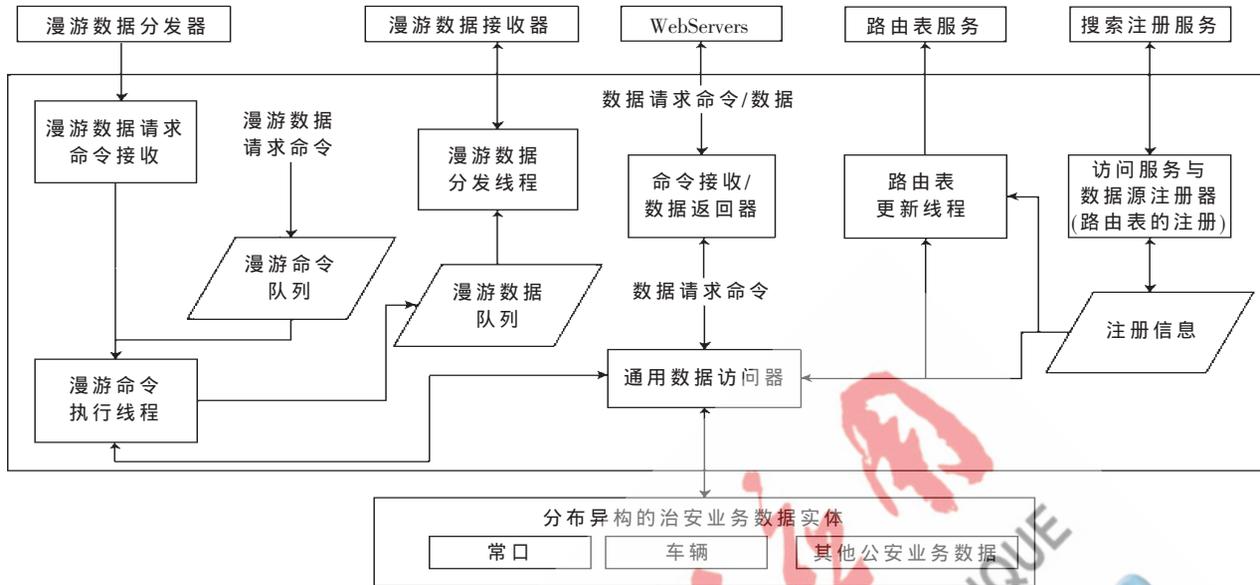


图4 五要素分布式数据适配器框架图

实现五要素信息的漫游查询。

### 3.2 系统之间接口标准和五要素元数据标准

系统把各种异构数据按五要素模型标准建模,由各种异构数据自身体现五要素的内外联系,使得搜索引擎以统一的接口处理各种异构业务数据。首先,将各种系统按五要素分类,并以某些关键字的组合关联起来;再次,将系统中的各种表信息按五要素分类,并以系统内部的某些关键字关联起来;最后,用特定的标准描述各种系统需要表现的具体信息,并予以定义。

### 3.3 分布式数据库处理技术

搜索引擎提供通用数据访问适配技术解决分布式异构数据库访问难题,本技术可以有效、安全地访问部署在防火墙后的各种异构数据库,并支持统一的标准数据库语言,使系统在统一数据库语言标准下,从而简化了开发难度。采用同步和异步 Web Service 提取数据的技术来实现。

### 3.4 搜索算法

搜索引擎算法主要有:关键字搜索、智能搜索路径、智能搜索深度优化算法等。关键字搜索就是利用系统间的关联关键字对相关系统进行分布式抓取数据,并在主搜索引擎重组数据并表现;智能搜索路径是基于系统之间的关联,如常口系统与派出所综合系统的关联,派出所综合系统与重点人口的关联等,也就是说可以形成很多关系链,而这些关系链也许是闭合的,也许是开放的,错综复杂,所以采用智能搜索路径,从主搜索系统开始,排列不同的分支系统,并按顺序搜索数据;智能搜索深度优化根据系统的多少和执行快慢,自由定义分支系统的搜索路径深度,因为搜索引擎的搜索是按照一定的搜索路径搜索的。

### 3.5 治安五要素信息 UDDI 注册中心

使用微软的 UDDI 注册中心作为治安五要素信息注册中心来公布主搜索引擎服务接口。通过该接口,其他

系统认证后可以发现并使用搜索引擎的服务并得到所需要的信息。

鉴于全省治安信息数据的分布特性和各系统数据标准的不统一性,本文分析了五要素数据模型和元数据标准,提出基于 Web Service 技术的治安信息五要素智能搜索引擎体系结构,并利用线程、UDDI、SOAP 等技术实现五要素信息的关联查询和漫游查询,实现全省范围内治安信息的“一站式”搜索。智能搜索引擎的设计和实现充分考虑系统的灵活性,把业务组件封装成 Web Service 与可重用性构件,这样系统具有易扩展性。目前系统已投入运行,并取得了一定的成效。但怎样定义元数据标准才能使查询更有效还有待进一步研究,对多线程的管理控制也有待改进。由于是基于 Web Service 技术实现的,如果把 SOAP 的思想用到设计中,系统是否更健壮等是下一步工作的内容。

### 参考文献

- [1] 云南省公安厅.五要素引擎可行性研究报告[R].昆明:2003.
- [2] 昆明世科网络技术有限公司.五要素系统需求分析[R].昆明:2004.
- [3] PAPAOGLOU M.Web Services:principles and technology [M].Prentice Hall, 1 editon(September 23,2007).
- [4] HONZLER S.XML 完全探索[M].陶阳,嵩敬波,王锋,译.北京:中国青年出版社,2000.
- [5] COYLE F P.XML、WEB 服务和数据革命[M].袁勤勇,吴青,译.北京:清华大学出版社,2003.

(收稿日期:2011-01-21)

### 作者简介:

杨令省,男,1972年生,硕士研究生,高级工程师,主要研究方向:软件工程,数据库系统等。

唐金文,男,1971年生,硕士,副教授,主要研究方向:软件工程,数据库等。