基于统一搜索的信息服务平台①

朴　岩　陈远平　及俊川　(中国科学院 计算机网络信息中心 北京 100080)

摘　要: 针对目前多数 IT 系统的信息服务只具备简单数据查询且资源定位能力薄弱的现状，本文对现有信息服务模式进行了分析，提出了基于信息集成的统一搜索模式。本模式主要包括两部分内容。第一，通过信息集成将分布式异构数据进行提炼、转换和汇总，形成数据全集做为信息源。第二，基于互联网搜索引擎模式和 Compass 搜索框架，将不同类型、不同粒度和不同主题的信息以统一的方式检索与展示。它的优势在于实现了多种数据源的应用整合并提高了系统可扩展性，更为重要的是极大地提升了用户的信息定位能力，使其获得类似百度或谷歌的使用体验。此模式已被成功应用于中国科学院十一五信息化项目信息管理与服务平台的开发中，获得了良好的应用效果。

关键词: 统一搜索; 信息服务; 数据集成; 应用整合; Compass 框架

Information Service Platform Based on Unified Search

PIAO Yan, CHEN Yuan-Ping, JI Jun-Chuan

(Computer Network Information Center of Chinese Academy of Science, Beijing 100080, China)

Abstract: Nowadays, most of IT information service system are lack of resources localization ability and only provide simple data query. Therefore, I analyse the drawback in the current information service models, and propose unified search structure based on information integration framework to improve information search quality. This model consists of two components . Firstly, in order to form data collection as the source of information, distributed heterogeneous data is extracted, transformed and aggregated by the information integration module. Secondly, according to Internet Search Engine model and COMPASS framework, different types, size and themes data is queried and displayed in a unified way. This enhances system information integration capabilities and shows powerful expansibility. What is more important is our proposed system extremely promotes user's information localization ability and using experience as Baidu or Google. Finally, this model is successfully applied to the information management and service platform development, which is one of the Eleventh Five-year informationization projects of Chinese Academy of Sciences.

Keywords: unified serach; information service; data integration; application conformity; compass framework

1　引言

随着信息化建设的深入，IT 系统逐渐成为企业运营和日常工作不可或缺的重要支撑体系。目前，许多单位已经完成“基本建设”，进入“深化应用”阶段。然而，由于多种应用系统并存以及“信息孤岛”的现象，出现了数据有但查找难的情况。一般来说，应用系统的核心操作是对信息的增删改查，其中“查”占据了大约 90%的使用比例。当前互联网搜索引擎的广泛使用从侧面佐证了使用者对信息定位的强烈需求。以前的信息系统相对简单且数据量小，只需要建立一些数据展示页面并辅以筛选条件即能满足要求。现在的应用系统越来越庞大复杂，且信息量也急剧膨胀。

① 基金项目:中国科学院十一五信息化项目:中国科学院资源规划(ARP)项目二期工程(o846011104)
收稿时间:2010-03-04;收到修改稿时间:2010-04-12

此外，用户对信息的要求也不仅是简单的数据展示，而是需要提供信息整合、数据挖掘、主题分析等综合性信息服务解决方案。这促使信息服务成为未来深化IT 应用的重要推动因素。

从技术角度来看，信息服务模式主要包括信息采集方法、数据整合策略、信息定位方式和信息服务应用等。目前比较主流的模式是将应用级数据进行提取、过滤、整合形成基础数据源，一般以数据仓库形式存在。基于此全集根据需求开发不同粒度的信息服务应用，如信息查询、统计系统、报表分析、决策支持等。这种模式在信息整合方面比较成熟，但是在信息定位方面还有很大改进空间。尤其当信息量十分庞大或用户对数据路径很难了解的时候，这种弊病就突显出来。严重时会阻碍系统应用的推进与普及。因此，以搜索的方式来定位信息资源，将结构化和非结构化数据以统一的形式展现出来，提高用户定位信息的能力并提升信息服务效率是本文着力改进的方向。基于中科院信息资源的整合与服务需求，将此模式应用于中国科学院十一五信息化项目之中国科学院资源规划(ARP)二期工程中信息管理与服务平台的设计开发，取得了较好的实践效果。

2 信息服务模式的研究与改进

2.1 信息服务现状

由于历史与技术的原因，许多企事业单位 IT 系统存在着多种应用并存且不同系统无法互联互通的现象。用户需要特定主题信息时，需要登录不同系统才能取得相应数据。

一般而言，不同领域的数据存在于不同的应用系统中。信息服务的主要作用是提供强大的信息整合能力和便捷的资源定位方式，使用户能够快捷而准确的获取所需信息。目前主流的信息服务模式有以下两种：

2.1.1 基于统一认证的应用整合模式

单点登录(Single Sign On)，是目前比较流行的统一认证的实现方式。它使得用户只需在网络中主动地进行一次身份认证，即登录一次，便可以访问其被授权的所有网络资源[1]。单点登录的主要意义在于使用户只需记住一套用户名和密码就能访问所有权限范围内的系统和数据。这不仅带来了更好的用户体验，更重要的是降低了安全的风险和管理的消耗[2]。但是，单点登录在信息定位方面并没有更加便捷，用户仍然要知道所需数据在不同系统中的准确路径。如果需要找出分布在多个系统中的同一类数据，则要在不同的系统间切换并进行手动的检索。比如，用户需要人事信息，而在多个子系统中都存在不同维度的人事数据。因此，用户可能需要访问人事信息系统、报表分析系统、决策支持系统等来获取不同类型的人事信息资源。这种模式的优势在于拥有良好的扩展性，当增加新的应用时只需在认证系统中加入新的认证信息就可以快速集成到单点登录体系中。总体而言，基于统一认证的信息服务模式是从应用层面进行的整合，以提高应用效率，简化用户操作为目的。资源定位和信息整合能力还属于初级阶段。

2.1.2 基于信息集成的分类检索模式

信息集成的作用是通过集成、结构化各个异构数据源的数据并向外提供统一的访问接口[3]。目的是将来自多种异构数据源(包括结构化、半结构化和非结构化数据源)的信息进行有效的结构集成和语义集成。从技术角度来看，一般采用中间器/包装器的体系架构，将不同的数据源通过翻译、转换和集成等中间步骤生成统一的全局数据集，作为对外提供信息服务的数据源[4,5]。基于此建立一个信息检索平台，并根据信息分类建立资源目录。一般来说有两种组织方式：

(1) 按数据主题进行分类。一个目录分支包含同一领域的数据，不同级子目录代表了不同的数据层次。如人事、财务、项目作为顶级目录；“人事”下第二级目录可以包括：全体人员统计、人员报表分析、人力资源 OLAP；第三级则包括人员基本信息查询、部门人员信息等最细粒度的数据。这是一种纵向分类的方式。

(2) 按数据粒度进行分类。决策支持、报表分析、数据查询等作为顶级目录。一个目录包含了同一层次不同主题的信息。如决策支持下可以包括人力资源、财务分析、固定资产等不同模块的决策支持模型，这是最宏观和抽象的层次。报表分析则包含人力、财务、项目等不同领域的统计报表。这是一种横向分类的方式。

这两种方式各有利弊，要根据目标用户的使用习惯和查询需求来决定采用何种方式。最佳方式是提供一个默认目录结构，并提供用户自定义功能兼顾通用性与个性化。此外，根据角色的不同权限设置可访问的数据范围，并为不同用户赋予与真实职责相对应的角色权限来保证信息的安全性，避免敏感信息泄露。这种模式目前应用比较普遍。好处是将分布式的异构

数据通过信息集成形成全数据集，便于综合查询和权限控制。同时避免用户切换不同系统的困扰。缺点是信息定位的能力还有待提高。

2.2 基于信息集成的统一搜索模式

统一搜索即通过统一输入域接收用户的搜索关键词，以搜索引擎的方式检索基于信息集成所形成的全数据集并获得所有权限内的匹配结果，将结构化与非结构化的数据以统一的方式展现给用户。这种方式有两大优势。第一，具有强大的信息整合能力。通过信息集成将分布式和异构的数据进行有效提炼与集成，形成包含全部业务信息的资源全集，成为信息服务的基础数据源。此外，通过提供外部接口使得系统具备不断扩充的能力。第二，提供强大便捷的资源定位能力。通过提供类似百度搜索的操作方式和使用体验，更易上手也更加便捷。用户无需在庞杂的目录树中手动寻找需要的信息。而且，检索不同业务信息无需填写具有多个输入域的查询表单，即通过一个输入框接收关键词就可以自动搜索不同类型业务实体的各类字段，如标题、类型、描述、正文等。避免了一般的查询系统可能需要多个表单输入域的困扰。

模式改进后的优点：(1)在保留了原有模式优点的同时极大地增强了定位资源的能力，提供统一搜索方式。(2)更加强大的信息集成能力，能够将结构化与非结构化数据统一管理与展现。(3)强大的可扩展能力。提供外部资源接口可以不断增加资源，包括结构化与非结构化信息。通过配置可以将新增系统的信息纳入统一搜索。

3 系统架构与应用设计

3.1 中国科学院信息管理与服务平台概述

中科院资源规划项目(Academia Resource Planning，简称 ARP)，是实现中科院科学的资源规划的信息系统工程[6]。ARP 系统为院所两级架构的星型分布式系统。中科院总部机关部署一套院级系统，分布于全国的 120 多个研究所各自部署一套结构相同的所级系统。所级系统包括 ERP 和自主开发办公系统两个部分，ERP 管理维护核心数据，包括财务、人力、资产、项目等。办公系统包括电子公文、网上报销、统一门户等应用。每个所级系统包括多个相对独立的模块和子系统。

中科院信息管理与服务平台是以院级系统为中心节点，所级系统为分支节点的星型架构，通过数据交换机制将所级数据汇总到院级数据仓库，形成统一的有机整体并对各级用户提供不同层次的信息服务。本平台是 ARP 系统的信息资源中心，将各研究所结构化与非结构化信息通过集成汇总为全院的数据全集，并基于此提供决策支持、报表分析、信息查询和资源检索等的信息服务功能，为科学管理与决策提供强有力的信息支持。

3.2 系统总体架构

本平台主要分为两大部分：信息集成与信息服务。信息集成通过数据交换适配器来实现。应用集成中间件 TongIntegrator(简称 TI)作为适配器为数据交换平台提供了与异构系统共享数据的能力。通过 TI 统一将所级系统中 ERP 和办公系统中的数据定期提取到所级数据缓冲区(简称所级 IRC)。各研究所节点到院中心节点的数据传输采用成熟的消息中间件 TongLINK/Q(简称 TLQ)，它提供可靠、可扩展和可监控的分布式信息传输机制。通过 TLQ 将 120 多个所级 IRC 中的数据传输到院中心数据缓冲区(简称院级 IRC)。再通过一系列的转换和加载过程形成院级主题库，形成全院数据全集，作为信息服务的数据源。

信息服务应用了基于信息集成的统一搜索模式。使用 Spring 来作为系统框架的支撑和管理容器，使用 Ibatis 作为系统的 ORM 框架。最重要的搜索功能基于 Compass 框架，它可以将 Spring、Hibernate 框架整合，为企业级开发提供搜索服务。Compass 对 Lucene 框架进行了封装，避免手工编写大量代码实现索引和搜索的功能，并且还隐藏了 Lucene 多线程的模型避免了同步问题[7]。首先，定义一个统一的用于搜索的搜索对象(Compass Entity)，它抽象出搜索的范围和并声明具体的属性，如名称(Name)、描述(Description)、内容(Content)、类别(Type)等。第二，声明此对象与搜索引擎内在数据结构(Lucene Document)的映射，这个数据结构是用来索引可搜索内容的抽象数据对象。第三，从不同的业务实体对象中提炼出用于搜索的属性并与搜索对象的属性相对应，比如用户类使用姓名对应搜索对象的名称，使用个人简介对应内容，使用角色对应类别；而文档类使用标题对应名称，正文对应内容，格式对应类别。通过编写 SQL 提取每一个实体对应的字段生成用于搜索的数据子集。最后，以后台线程调用的方式每隔一段

时间生成一次全部资源的索引文件，以此来实现统一搜索。由 Compass 来管理生成索引用于统一检索，并与 ORM 框架配合来完成业务实体与数据库、业务模型与 Lucene 模型的映射。然后，配置查询结果的排序模式，即根据搜索结果的相关性和重要性来进行排序。本平台中按照报表分析、信息查询、决策分析和资源库的权重顺序来对结果集排序。此外，针对非结构化数据提供了批量导入和用户上传接口，针对每种方式都需要将出用于搜索的几个匹配属性中信息填写完整，这样就可以用于生成索引。最后，针对具体业务信息的查询则使用 ORM 框架 Ibatis 来根据具体的对应关系动态生成查询表单来对应其每一个字段，实现最细粒度的数据检索。总体架构见图 1。

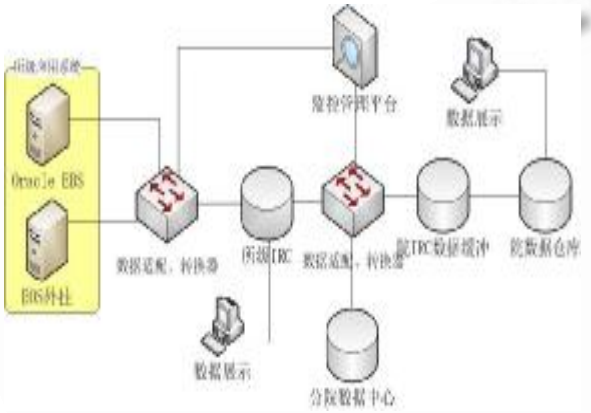

图 1 系统架构图

本平台的体系结构分为五层：首先，通过“信息交换与传输”的数据交换、自动采集和资源上载模块完成分布式异构数据的采集。其次，通过“信息存储与管理”的配置形成不同层次的数据集，包括所级管理库、院级主题库和元数据库以及非结构化数据的资源库。这两个模块共同完成了数据集成和管理的工作。基于此基础数据源，“信息处理与发布”完成数据建模、信息提取和资源发布的工作，对数据进行第一步提炼并具备初级信息服务功能，并为日后开发保留扩展性。“信息共享与服务”包含各级用户直接使用的功能点和访问的信息资源，也是最终与用户交互的接口。界面使用目录树配合选项卡的方式整合各类资源并以统一形式展现，依靠搜索引擎的资源定位方式提供便捷的信息检索功能。此模块是用户能够使用的信息服务全集。实际上除了超级用户外大多数用户都不能访问全部信息资源，需要根据角色权限过滤不能访问的信息。此外，基于信息服务提升出个性应用，如年终统计、个人看板等。这些模块本质上是信息服务的子集，由于策略限制和特殊需求将其单独开发成为独立模块或功能选项供特定用户使用。因此以“产品定义与订阅”方式为日后的功能扩展保留接口。平台总体体系结构如图 2 所示。

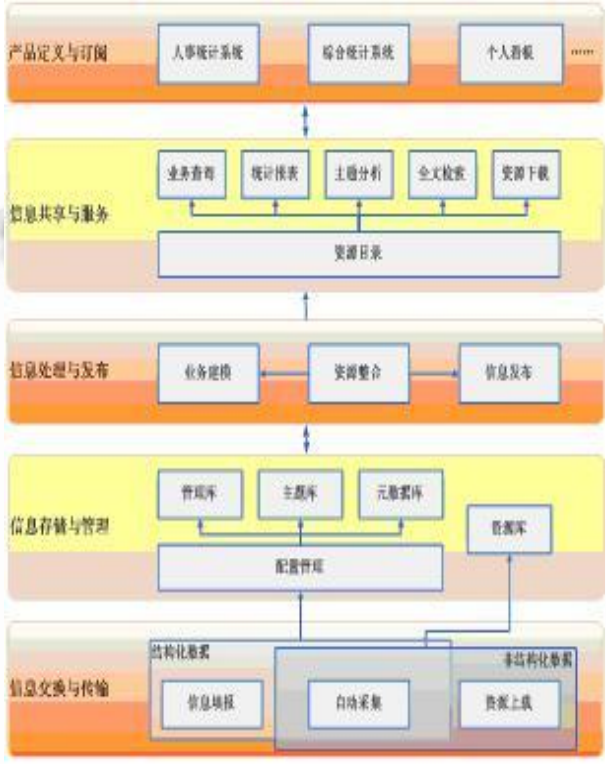

图 2 平台体系结构

3.3 实现关键技术

3.3.1 异构信息集成方式

本平台目标是涵盖中国科学院所有的信息化资源，旨在通过统一入口就能够访问不同系统和不同类型的数据。因此，将不同系统、不同结构和不同粒度的信息资源有效整合是重要而基础的环节。

这分为两个层面。一是将分布式异构结构化数据的采集，二是结构化数据与非结构化数据的整合。结构化数据包括三大类：(1)业务应用数据。如财务、人事、资产、项目等存储于 ERP 中的最细粒度信息和电子公文、网上报销等辅助办公系统中的业务数据；(2)统计分析报表。如人力资源类、财务基建类、科研产出物类等基于详细数据提炼出的汇总信息；(3)决策支持分析。如科研基尼系数模型、学科分布有效性分析

等在更高层面对管理和决策提供信息化支持的数据挖掘。非结构化数据主要包括多媒体资料和全文文档，如图片、视频、Flash、Word、PDF 等。

数据采集基于东方通的应用集成中间件 TongIntegrator(简称 TI)和消息中间件 TongLINK/Q(简称 TLQ)完成。根据中科院院所两级的星型分布式架构和数据单向传输的应用需求，即数据从研究所应用系统到研究所缓冲区，各研究所缓冲区到中科院中心缓冲区。首先，针对不同的应用模块编写设计需要提取的数据范围并编写相应的存储过程；第二，为不同模块配置相应的策略。如果是全删全插模式，则需要每次清空原有数据并插入新的全数据集，如人事模块；如果是增量提取模式，则需要比对上一次抽取的时间戳，如果数据的操作时间晚于上一次抽取时间则说明这是修改或新增数据，需要抽取否则无需提取，如财务模块。第三，使用 TI 的 SQL 嵌套方式管理存储过程的执行及向所级缓冲区的数据加载。这一步是在各所局域网的内部系统中完成的。当数据到达所级缓冲区后，使用 TLQ 消息传输中间件完成数据从所级系统传输到院中心数据缓冲的过程。这是通过互联网的数据交换。在每个所配置发送队列，每个模块对应一个队列，共配置 9 个发送队列。在院级缓冲区为每个所配置一个接收队列，共有 124 个接收队列。通过 TLQ 的可靠传输完成数据的交换和汇总。当数据接收完成后开始转换、清洗和加载等一系列过程形成全院数据仓库。

数据集成方法为针对不同模块分别提炼出通用的核心属性或字段，如调用链接、标题、路径、权限、描述、分类等信息并统一存储到通用数据表中，此表包含了全部信息资源的抽象汇总。开发与此数据表对应的搜索对象模型，用于生成索引文件。其次，制作不同的展现页面接收运行结果，在统一的展示容器中动态嵌套不同的结果页面完成信息展示。前台使用这个通用数据表来展现具体的目录结构，使得不同类型资源从用户角度看来是一致的。业务数据的调用链接是不同的数据源，以网页通用表单的方式展现。如查询“个人基本信息”，点击此链接后台就会访问存储个人信息的数据表并通过权限控制取出登录者的个人信息并以 Html 网格页面的方式展现，见图 3。

图 3　个人基本信息

报表和决策支持的运行引擎在后台分别部署于单独的 Jboss 应用服务器。此外，将每个报表和分析模型的调用路径、标题、所属目录、归属角色和描述等信息提取并存储于通用数据表。如点击“按部门的正式职工年龄情况”见图 4，后台会将此报表对应调用路径发送给报表运行引擎，并将运行结果接收后反馈在用户的页面上，通过这种方式用户无需登录报表分析平台即可查询报表。而且平台还集成了报表管理功能，包括报表的新建、修改、授权和移动等功能，使管理员也无需切换系统就能够完成报表的管理工作。决策支持的集成方式类似于报表。本平台采用了按不同数据粒度定义目录结构的方式，包括决策支持、报表分析、信息查询和资源库。

图 4　报表分析示例

3.3.2 统一搜索索引机制

搜索引擎原理是基于将数据分词后生成索引文件，当用户输入搜索关键词后在此索引文件中进行匹配并返回。一般来说，一个索引文件对应着一类应用对象实体。因为不同类型对象模型的属性互不相同，所以在一个大型系统中如果要索引全部对象和属性理论上需要多个索引文件。但是这又带来一个弊端，如要搜索不同类型的应用对象就要对应不同的索引文件和不同的属性字段。即要对不同资源构造不同的搜索页面，这样无法提供一个类似百度的统一入口来检索所有资源。因此，我们通过提炼构造一个用于搜索的通用业务对象，而这个对象只包含所有类型资源均可以映射的属性。这样，只有一个索引文件就囊括了全部资源，提供了关键词在全部可检索属性的查询功能，以实现统一搜索。

Lucene 本身一个基于 Java 的全文索引工具包，它并不是一个完整的搜索引擎，它可以方便的嵌入到各种应用中实现全文索引与检索功能。Compass 扩展了核心 Lucene，增加了事务控制功能和快速更新，包括在生成索引文件，关键词搜索等功能。将原有 Lucene 的核心检索功能进行封装提高了开发的效率。

目前的应用系统开发中包括领域模型、实体对象及数据库中的 ER 模型，它们的意义是将现实世界中的实体抽象成为计算机世界模型，这也是面向对象开发的主要出发点之一。在搜索引擎中也有其特有的对象模型，它是基于此模型来构造索引文件，实现全文检索等功能。因而，定义应用程序对象与搜索引擎对象的映射关系是实现统一索引的重要步骤。

Compass 提供了三种从应用程序模型到搜索引擎的声明式映射。我们使用了 OSEM(Object/Search Engine Mapping)，即对象搜索引擎映射[8]。它把应用定义的实体对象映射到搜索引擎。我们定义了如图 5 的搜索实体，使用注释(Annotation)对用户实体进行了 OSEM 定义。注释能够避免配置文件的繁琐同时让代码的语义性更为直观。编写存储过程完成不同模块的数据提取并联结(Union)为一个结构相同的通用数据表，将每个字段的别名命名为统一的名称，并通过 XML 配置文件建立通用数据表字段与搜索实体的对应关系。如图 6。

```
@Searchable
public class CompassResultMap extends Entity {
    @SearchableId
    private Long id;
    /** 资源内容 */
    private String resourcesBody = "";
    /** 资源名称 */
    private String resourcesName = "";
    /** 资源类别 */
    private String resourcesTypeName = "";
    /** 链接地址 */
    private String url = "";
    /** 资源描述 */
    private String resourcesDesc = "";
    /** 角色级别 */
    private String roles = "";
    private Long directoryIndex;
    private Long indexItemIndex;
    private String indexValue;
    @SearchableProperty(name = "directoryIndex")
    public Long getDirectoryIndex() {
        return directoryIndex;
    }
    @SearchableProperty(name = "roles")
    public String getRoles() {
        return roles;
    }
    @SearchableProperty(name = "resourcesType")
    public Long getResourcesType() {
        return resourcesType;
    }
```

图 5 搜索实体对象定义

图 6 搜索实体与数据视图的映射

3.3.3 搜索结果过滤与排序

搜索引擎能够将匹配搜索输入的全部数据筛选出来，但是要对结果进行权限控制。这是信息服务系统与互联网搜索引擎的主要区别。对于百度等搜索引擎而言，它们视角中的用户是相同的，不同人搜索相同关键词得到的结果是相同的，使用者的角色也是一致的。而信息服务系统是对内的应用系统，必然要服务于不同角色的用户。统一的外观但内涵却不尽相同，需要根据角色权限对信息范围进行控制，这也是基本的信息安全手段之一。在分类目录的信息服务模式中，可以根据功能点或目录对不同角色进行授权，再将不同用户与角色绑定，形成两级权限控制机制。在本平台中，使用了两级过滤机制：即第一级在搜索对象中加入每个资源对应的角色列表，直接生成到索引文件中。在搜索时将结果集的角色级别与登录用户的角色权限进行比对，如果资源不在用户的权限范围内则不予显示。这是资源级的权限分配机制，也是最细粒度的控制方式，便于细化管理。第二级为对资源目录进行统一授权，将目录整体赋予有权限的角色。在用户角色表中存储有权限的目录列表及其子目录，在搜索时将结果集中的资源所属目录集合与用户有权限的目录集合进行比对，剔除用户无权限的资源。如将“信息处理与发布”目录授于“超级用户”。这样即使一般用户搜索这个目录下的内容也不会在结果中显示。这是较大粒度的控制方式，便于分类管理。

搜索结果的排序是影响搜索效果的重要因素之一。在信息量很大的情况下良好的排序能够极大的提升用户体验，使用效率得到提高的同时也体现了搜索服务的专业性。因而，在本平台的实践中也要对搜索结果进行排序。从技术角度来说，排序是按照两级模式进行。首先，对不同类型的资源整体排序，具体来说是根据资源的重要性和使用频率进行权值设置并把同一类资源放在一起显示，便于用户查找。如报表是日常工作中使用较多的资源，则赋予“10”的权值，公文公告类资源赋予“20”的权值，普通资源默认赋予“90”的权值。并在提取数据生成通用数据表时按不同模块的权值排序，这样在搜索结果集中无需重排。其次，在同一模块内按照匹配相关性和字符顺序进行排序，直接将排序固化到索引文件中，提高搜索效率。

4 平台应用效果

信息管理与服务平台于 2009 年 12 月在中国科学院几个试点研究所进行部署测试。平台将原有 ERP 系统、公文系统和查询系统等所有 ARP 系统信息进行了统一整合，并为每一个系统内员工都分配了账号。因而，用户的范围不在局限于行政部门而是扩大到每个员工。通过数据集成和资源发布可以半实时的将新增信息和外部资源添加到平台中，形成相对个性的应用效果。统一搜索能够便捷的将各类信息呈现给各级用户，改变了以往系统建设多年用户缺乏体验的情况。同时，这种信息服务模式降低了使用门槛，提高了操作友好度，切实提升了信息化应用水平和工作效率。

5 结束语

基于统一搜索的信息服务模式继承了原有信息服务方式的优点并且极大的提高了资源定位能力，解决了异构系统综合查询困难的普遍问题。通过搜索将结构化与非结构化的资源以统一形式呈现的方式降低了使用门槛，提升了信息化应用效果，推进了“深化应用”的目标。

参考文献

1 贾宗星.单点登录系统的研究与分析[硕士学位论文].西安:西安建筑科技大学, 2007.

2 吴茂传,郭阳,胡昌平.基于 web 的单点登录技术在企业集成中的应用.淮海工学院学报, 2008,17(1):29－32.

3 金晓磊,闫红漫,翁之浩,尉大光.基于虚拟数据库的信息系统集成研究.计算机技术与发展, 2009,19(6):87－90.

4 粟湘,吴沛.异构信息集成模型架构研究.情报理论与实践, 2006,29(5):612－614.

5 徐进,马欢.基于 XML 的异构数据库信息集成模型研究.科技资讯, 2008(13):104－106.

6 任玉平,关戎,潘亚南.关于 ARP 系统建设的思考.中国科学院院刊, 2003,(1):54－57.

7 王鹃,洪承煜,沈哲.基于 Compass 框架的电子商务网站搜索引擎设计.现代计算机, 2009,301(2):144－145.

8 全俊林,杨开英.基于 Compass 的快速建立企业全文检索.福建电脑, 2007,3:158－185.