

基于 J2EE 平台的寄生虫虫种资源库的构建

沈海默, 胡薇, 陈韶红, 郑琦, 周晓农

(中国疾病预防控制中心寄生虫病预防控制所, 上海 200025)

摘要: **目的** 构建寄生虫虫种资源信息管理系统。**方法** 对系统进行功能分析和结构设计, 确定系统的模块组成、各模块之间的逻辑关系。结合 J2EE, XML, JDO, JSP, MVC 等多种技术, 运用企业级应用的分层构建原则, 在设计中应用设计模式理念, 将显示与逻辑、内容与表示相分离。**结果** 实现了跨平台数据库的构建, 建立了可供二次开发的模块、查询模块和逻辑关联, 并组合成一个完整的系统。**结论** 完成了基于 J2EE 平台具有可扩展结构的寄生虫虫种资源信息管理系统的构建。

关键词: 寄生虫虫种资源; J2EE; 数据库; 检索

中图分类号: S852.7 文献标识码: A 文章编号: 1005-0868(2007)01-0008-05

THE DEVELOPMENT OF PARASITE RESOURCE INFORMATION COMMAND SYSTEM USING J2EE TECHNIQUE

SHEN Hai-mo, HU Wei, CHEN Shao-hong, ZHENG Qi, ZHOU Xiao-nong

(Parasitic disease prevention and control center, Chinese center for disease control and prevention, Shanghai 200025, China)

Abstract: **Objective** In order to complete the construction of parasite resource register & researching system. **Methods** After we make the functional analysis and structure design of the system, and confirm the composition of modules and the logical connection among the modules, we use J2EE, XML, JSP, SOAP technique and sort other correlative techniques, to complete the functional analysis and structure design. It is available to according to the Enterprise-Level idea of bedding construction to separates "DISPLAY and LOGIC", "CONTENT and EXPRESSION". **Conclusion** The construction of expandable parasite resource register & researching system which is based on J2EE is established.

Key words: Parasite resource; J2EE; database; searching

寄生虫科技基础平台的建立, 将实现我国寄生虫虫种资源的数据和实物资源(包括基因资料)的共享。构建国家级的寄生虫虫种资源库, 并以国际互联网为依托, 最大限度地扩大资源共享范围, 为我国乃至国际科技活动提供必要虫种资源和相关信息的支撑, 对全面提升我国寄生虫病防治科研水平, 提高我国寄生虫病的预防控制水平, 保障我国人民健康和经济社会可持续发展具有重要意义。

我们使用企业级 J2EE (Java2 Enterprise Edition) 技术完成网络数据库部分的开发。在本研究项目的第一阶段, 按照国家科技基础平台要求, 我们完成了寄生虫虫种资源数字化、网络化的管理。在今后项目第二阶段里, 我们会建立资源提供者和使用者的数据库, 将用户、资源提供者相匹配, 提供安全的合作平台, 最终将寄生虫虫种资源网建成一个集各家之长并面向世界所有用户的综合科技信息通道。

1 项目背景

按寄生虫分类系统及其病原生物学特点, 制订重要寄生虫虫种资源标准化的技术规程和标准, 形

收稿日期: 2006-03-28

基金项目: 国家自然科学基金资源平台项目(2005DKA21104)

作者简介: 沈海默(1979~), 男, 上海市人, 实习研究员, 主要从事网络构建、数据库开发与整合研究工作。

成国家级寄生虫种资源保藏中心,建立寄生虫种资源共享网络体系,为科研、教学、生产提供标准化的虫种资源。在整理全国医学寄生虫、兽医寄生虫、医学节肢动物和医学贝类等种质资源名录的基础上,制定和完善寄生虫种资源的分类规范,经专家论证后实施。按照分类规范进行各类寄生虫标本的预处理、图像记录、鉴定、分类和入库保存;制定和完善各种标本保存标准,包括保存的设备、设施、管理人员及相应制度等;对一些稀有种类标本或不同正、副模,制定出不同等级保存方法、建立和完善主要寄生虫保种、传代、维持生活史循环的标准化操作规程。在对各类寄生虫资源进行数字化处理前,制定和完善不同种类标本的数据库标准,按规范标准录入数据库。

2 系统

2.1 系统要求 支撑和推动数据库共享平台技术演化的是不断更新的要求。通过对虫种资源信息管理系统的需求分析,我们认为本项目与以往的同类项目有几点重要区别:

2.1.1 项目面对的是来自全国各个研究机构的海量资源数据,各种资源之间具有各种联系。数据无法放在同一处,这就意味着需要使用大型分布式、稳定性极强的数据库。

2.1.2 最终使用者来自各行各业,他们会对同一个资源提出截然不同的信息要求,所以无法给出一个固定的资源数据预处理标准。可以预见该项目会在其他课题中被多次引用,这就要求开发者使用可重复使用的组件,提前做好二次开发的准备。

2.1.3 该项目的最终形态应该是一个规模巨大的、分布在全国各地的网络体系,由各资源提供单位的数据库共享平台组合而成。由于无法预料各单位的网络环境,因此项目应该具备对不同平台的自适应性。

2.1.4 在以后几期的项目实施中,随着资源共享规模扩大、参与人数的增加、共享方法日趋复杂,现有的 3 层结构将随着实际情况扩展到 4、5 乃至 N 层。所以该项目必须是一个开放式数据体系,开发者必须在设计时就做好扩展结构的准备。

2.2 架构的选择 虽然市面上最流行和使用最广泛的动态网站建设工具是 ASP、PHP,但是根据对项目的分析,我们认为它不适合本项目。在设计“寄生虫种资源信息管理系统”项目时,综合考虑了开发效率、代码重用、无缝连接、系统管理以及后期

维护等诸多方面的因素,在整体上采用得到广泛认可的、面向对象的、完整的 J2EE (Java2 企业版) 架构。

寄生虫种资源信息管理系统在采用 J2EE 传统的架构的基础上,应用设计模式 (Design Pattern) 的理念,结合最新的 J2EE 各层实现技术,形成灵活的、包容性的适合项目的体系架构^[1],图 1 显示了管理系统 J2EE 体系构架^[2]。

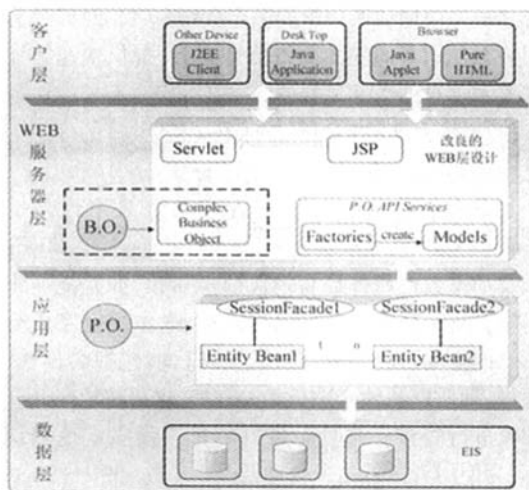


图 1 管理系统 J2EE 体系架构

2.3 数据库管理系统的选择 现在比较流行的商用数据库系统主要有 Oracle、Sybase、SQL Server 等。由于这些系统都支持标准的 SQL 语言和相关功能 (如存储过程、触发器),所以相互之间的转换比较容易。为了便于开发,我们同时采用了 SQL Server 和 Oracle 作为开发时数据库系统。在开发过程中,我们充分利用了 SQL Server 和 Oracle 直观性的特点,采用关系图模式进行数据库设计,严格规划了数据表之间的级联关系,有力保证了系统数据的完整性和一致性。

在 JSP 程序与数据库的连接问题上,本系统采用了 JDBC 连接池的模式。这不仅提高了数据库操作的执行效率,而且可以根据实际情况,方便的改用其他数据库系统,如 Sybase 等,进一步提高了系统平台的通用性。

3 系统要求分析

3.1 目标和定位 根据各类虫种标本的数据库个性数据库规范,对标本进行数字化,包括各种属性的描述数据、图像处理等,以分别建立医学寄生虫种资源数据库、兽医寄生虫种资源数据库、医学节肢

动物种质资源数据库、医学贝类种质资源数据库、实物标本数据库、寄生虫基因资源数据库、地理分布数据库、传播方式及危害数据库等。数据库将根据各类寄生虫的关联性、从属性等关系建立,以保证数据的完整性,减小数据的冗余性,确保数据的安全性,使各类数据库以图文并茂的形式存储、检索和分析等。由此目标应分为以下几点:

3.1.1 寄生虫虫种资源信息管理系统是一个集各家之长并面向所有用户的综合系统。在信息时代为用户开创随需应变的网络信息服务,提供快速智能搜索和理想的个性化服务。

3.1.2 该系统将成为科技信息交流、共享的理想信息中心。

寄生虫虫种资源信息管理系统采用最新的Web服务和Web服务技术,如:Web services, J2EE, 以及动态数据库等最新现代信息科技,如智能搜索引擎。支持智能化的快速信息检索,由关键字进行信息的查询,提供多种条件进行查询,包括名称、资源号、种质资源编号进行相关查询,也可按地区、省、国家进行查询。并且对各科研机构建立本地知识库,为使用者进行信息自动匹配。

3.2 模块的划分和大致功能

3.2.1 寄生虫虫种资源信息管理系统可能会包括中、英语言版本,所实现功能主要包括以下几点:

3.2.1.1 以中、英语言编程和显示,适应多语种显示的需要;

3.2.1.2 提供科技信息的常规查询方法,由大类-小类-分组-元素-详细信息的查询;

3.2.1.3 提供信息的录入和查询,支持在线修改其信息;

3.2.1.4 采用智能搜索引擎支持智能化的快速信息检索,由关键字进行信息的查询,提供多种条件进行查询;

3.2.1.5 具有用户定制信息的功能,对用户定制的信息进行跟踪,自动进行反馈。

3.2.2 系统整体上可以规划分为以下模块:

3.2.2.1 Search——搜索模块

说明:有两类查询方法:一般检索和高级检索。

功能:可对医学寄生虫虫种资源数据、兽医寄生虫虫种资源数据、医学节肢动物种质资源数据、医学贝类种质资源数据、实物标本数据、寄生虫基因资源数据、地理分布数据、传播方式及危害数据等信息进行模糊查询、复合查询。

根据用户的不同情况,提供的站内查找功能,主

万方数据

要分为单字节搜索和组合搜索。

一般搜索(单字节搜索):用户可以在网页左侧找到寄生虫虫种资源信息的指定条目,在下面的文本框中输入要查找的关键字,并选择所要查找的信息所属的类别,点确定按钮,系统将返回数据库中查找对应的信息内容。如没有匹配的记录,将返回找不到的信息提示。

高级搜索(组合搜索):高级搜索是提供给登录用户的一项服务,用户进入不必遵循给定的条目,而是在组合搜索框里随意键入查找的关键字,选择要查找的主题和相应的关键词与、或、非的组合方式,选择所属的字段名,点确定按钮,系统将返回数据库中查找对应的信息内容。如没有匹配的记录,将返回找不到的信息提示。

3.2.2.2 User——用户模块

说明:为各类用户划分权限和提供对应的服务。在寄生虫虫种资源信息管理系统中,我们把用户分为4种:一般用户、高级用户、数据管理员和程序管理员即网管。一般用户的需求相对简单,基本是信息的查找、参考等。所以一般用户只能按照需要完成对信息、通知、法规条例的搜索,所查询的信息只可显示少数字段。与之对应的是高级用户,他们可以看到全部的信息。数据管理用户的需求相对复杂,除了查询以外,数据管理者还可以发布其相关的信息供其他用户使用。数据管理者可以通过模块发布、修改、删除相关的信息。最后由程序管理员完成用户管理和程序维护。

用户模块部分的实现包括三个子模块:发布信息,修改信息以及个人信息。

发布信息:分为“公告信息”,“虫种信息”两个功能。分别可以发布和维护寄生虫虫种资源信息和相应的公告信息。所有发布的信息,以是否登录为分界线,分为一般浏览界面和高级浏览界面,同时以是否登录为分界线将用户划分为一般浏览用户和高级用户。一般浏览用户仅可以浏览所有发布信息的许可字段和其它许可的静态信息;高级用户即登陆用户,可以看到全部信息和全部静态信息。

修改信息:登录用户中有一小部分是数据管理员,他们可以对信息进行增加、修改、删除等操作。对于程序管理员而言,修改信息的作用就是维护程序和网站。

个人信息:每个登录用户,不管具体是属于那种登录用户性质,都有自己的个人信息,用户可以选择性的公开部分信息,用于交流。

4 系统的具体实现

4.1 客户层 主要是用户访问网站的客户端,目前网站提供对常用的客户浏览器的支持,包括支持 Java Applet。其他的包括 Java 桌面程序和无线设备的访问暂不提供。

4.2 Web 层 JSP 和 Servlet 组成的 Web 层由应用服务器中 Web 容器管理,我们采用了基于 Struts 框架构建系统,以期明确开发人员的分工,最大程度的利用复用技术,提高开发效率,增强系统的可维护性^[3]。

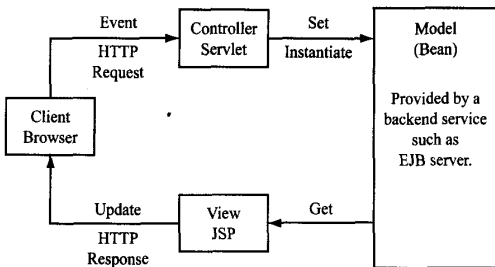


图 2 JSP Model2

来自 Apache Jakarta 项目的 Struts 框架为 J2EE 带来了 JSP Model 2 结构的一种较好实现(图 2)。Struts 力图从整体上减轻构造企业级 Web 应用的负担,并提供国际化和数据库连接池支持^[4]。

Struts 体系可以看成有两个相对独立的部分组成(见图 3):第一个部分是 StrutsAPI,用于编写支持 Struts 的应用组件;第二部分是 Struts 的 JSP 标记库,由 html、bean、logic 和 template 四个标记组成^[5]。

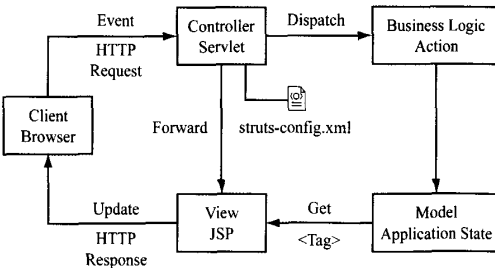


图 3 Struts 体系结构

4.3 数据层 根据多语言的要求数据层设计成可扩展的多语言的动态数据库,数据层存放了寄生虫种资源信息管理系统项目中的各种表信息。数据库的设计主要考虑多语言性、扩展性和性能三方面的关系。数据库应当符合以下要求:

- 4.3.1 数据信息的存储应该为搜索优化做好准备;
- 4.3.2 数据库有一定的扩展性;

4.3.3 数据库简单、可读性强,易维护;

4.3.4 数据库有一定的优化处理;

由此得到系统框图(见图 4)。

4.4 整体结构 寄生虫种资源信息管理系统采用三层模式,即 B/S/D(浏览器/服务器/数据库)结构(见图 5),用户浏览器发出的请求交由 weblogic 服务器解析、运行。当动态网页等程序需要调用数据库时,并不是直接建立连接,而是经过系统唯一的数据库连接池类来进行。

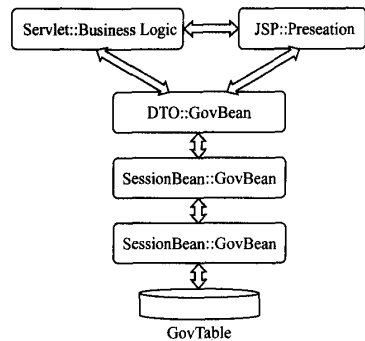


图 4 系统设计框图

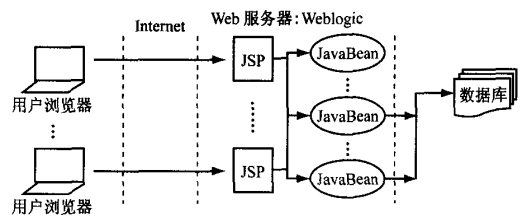


图 5 B/S/D 结构

查询部分所采用的 J2EE 总体设计方案(体系架构)。数据层存放了与查询部分业务逻辑相关的数据表格。应用层, EntityBean (CMP) 与 Session Bean 构成典型的会话门面模式(Session Facade),负责对业务逻辑相关的数据进行持久化。我们称之为 P. O. WEB 层,经过改良之后,增加了 Complex Business Logic,使得复杂的业务逻辑在这里实现;Servlet 调用 Complex Business Logic 中实现的具体业务逻辑方法和操作等,获得所需数据,并传给 JSP,最终数据到达客户端^[6]。

为不同的用户分配帐号,管理相应的虫种信息。用户具有增加、删除、修改、查找虫种信息、法律法规等的权限。具体来说,使用 CMP + Session Bean + Servlet + JSP 的模式实现上述管理模式。链接数据库时采用连接池技术。

4.5 获取查询结果的方案讨论 对于用户的查询,

我们通过数据库的搜索来得到结果,由于数据库本身是一个比较成熟的产品,而且通过添加索引、优化表设计,查询的速度还是很快的。所以在数据查询的第一阶段,可以由数据库的搜索引擎来实现。

由于查询结果数的不定性,返回的结果数可能是 0,也可能是几千几万条,所以我们需要以分页的方式向用户返回结果,同时还必须充分考虑性能。目前比较广泛使用的分页方式是,将查询结果缓存在 HttpSession 或有状态 bean 中,翻页的时候从缓存中取出一页数据显示。这种方法有两个主要的缺点:一是用户可能看到的是过期数据;二是如果数据量非常大时,第一次查遍结果集会耗费很长时间,并且缓存的数据也会占用大量内存,效率明显下降。因此,比较好的分页做法应该是,每次翻页的时候只从数据库里检索页面大小的块区的数据。这样虽然每次翻页都需要查询数据库,但查询出的记录数很少,网络传输数据量不大,如果使用连接池更可以略过最耗时的建立数据库连接过程。而在数据库端有各种成熟的优化技术用于提高查询速度,比在应用服务器层做缓存有效多了^[7]。

5 总结

通过对系统功能的细致分析,运用企业级应用的分层构建思想,采用 J2EE, XML, XSL, JSP, SOAP 等多种技术,我们构建了一个基于 J2EE 平台的寄生虫虫种资源信息管理系统。整个系统的设计完全按照显示与逻辑、内容与表示分离的原则^[8]。系统通过数据层、业务层、表示及用户接口层四个层次的“过滤”,把数据库中存储的原始的数据转变为用户需求的信息,进而全面地完成了预定的功能。

我们还需要完成的功能有:以中、英两种语言显示,适应多语种显示的需要;提供资源单位信息的录入和查询,支持单位在线修改其信息;提供资源使用人信息的录入和查询,支持用户在线修改其信息、用户定制信息的功能;对用户定制的信息进行跟踪,自

动进行反馈,为资源发布和使用双方提供一个安全的网上的共享环境,实现网上合作配对,并提供邮件的自动发送,自动反馈。

该系统是一个集各家之长并面向所有用户的综合科技信息通道。可以预见到它将成为科研人员进行寄生虫虫种资源信息交流、共享与紧密合作的理想信息服务中心,在虫种资源信息收集、处理、寄生虫研究项目合作以及促进信息网络化等方面为用户提供捷径,实现信息的收集,学术探讨,科研项目洽谈,决策支持等功能,增强科研单位的竞争力。从信息的获取到资源共享的实现,力求在整个科技业务流程中为研究者提供尽善尽美的支持。

参考文献:

- [1] 全国信息技术标准化技术委员会教育技术分技术委员会. 网络教育技术标准 CELTS—3.3(CD1.6)——学习对象元数据实践指南[EB/OL]. [http://www.celtsc.edu.cn/download/CELTS-3.3\(CD1.6\).zip](http://www.celtsc.edu.cn/download/CELTS-3.3(CD1.6).zip). 2003.
- [2] IMS Global Learning Consorting, Inc. IMS Content Packaging BestPractice Guide version 1.1,3 Final Specification [EB/OL]. <http://www.imsglobal.org/content/packaging/cpvlplp3/imscp-bestv1plp3,HTML,2001>.
- [3] 全国信息技术标准化技术委员会教育技术分技术委员会. 网络教育技术标准 CEL FS—9.3(CD1.6)——内容包装实践指南[EB/OL], [http://www.celtsc.edu.cn/download/CELTS-9.3\(CD1.6\).zip](http://www.celtsc.edu.cn/download/CELTS-9.3(CD1.6).zip). 2003.
- [4] Erich Gamma, Richard Helm, et al. Design Patterns: Elements of Reusable Object-Oriented Software [M]. Boston: Addison-Wesley, 1995: 54-87.
- [5] Richard Monson Haefel. Enterprise Java Beans [M]. 2000: 362-368.
- [6] Ed Roman. 精通 EJB [M]. 刘晓华等译. 北京: 电子工业出版社, 2002: 293-297.
- [7] 邵维忠, 杨芙清. 面向对象的系统分析 [M]. 北京: 清华大学出版社, 1998. 11-28.
- [8] 张素智, 卢正鼎, 李春林. XML 数据库及其应用研究 [J]. 计算机工程与应用, 2002, (8): 32-36.