

Web 数据库技术和实例分析

Web Database Techniques and Case Studies

孟小峰 谷明哲 编译 TP393

(中国人民大学信息学院数据与工程研究所 北京 100872)

Internet 尤其是 WWW 提供了在 Internet 和 Intranets 上可获得各种数据和信息的有效途径。但是,所获得的信息作成 HTML 静态网页就不太适于已存在于 DBMS 中庞大的数据。潜在的巨大数据量必须双重管理,数据更新难于处理,并且查询能力有限。另一局限是标准的 WWW 技术不允许把从不同服务器的部分结果合为单一的结果页面。

当人们开发和使用 WWW 软件时,上述的这些限制对他们来说就越来越明显。本文介绍了解决这些问题的几种技术:即万维网(WWW),面向对象编程语言 Java,对象管理组织(OMG)的公用对象请求代理程序体系结构(CORBA),概述了在网上访问数据库的不同方法,此外,提出了比较这些不同方法的标准。给出了来自 FZI 项目的几个案例分析。

1 基本技术

1.1 万维网(WWW)

WWW 对分散的文档提供了显式访问,它遵循严格的客户/服务器体系结构,无国界的超文本传输协议(HTTP)用于客户(浏览器)与 WWW 服务器之间的通讯。网上最通用的显示信息的格式是超文本标记语言(HTML),它是一种创建格式化的超文本文档的标准语言,文档中甚至可包括可执行的内容。

这些文档可通过统一定位方法进行访问,此外,对这样的文档如静态 HTML 网页,服务器可提供象对其他信息源如 DBMS 一样的显式访问。为此 Web 服务器通过公用网关接口 CGI 与应用程序通讯,这一机制允许客户转而委托服务器端由 Web 服务器和数据库服务器处理。

1.2 Java

Java 是一种面向对象的程序语言,尤其用于不同种计算机网络上(独立于平台)的设计。Java 既能单独使用又可与 WWW 连接使用,小的 Java 程序被称作 Applets,是利用 HTTP 以字节代码的形式从 Web 服务器传输到浏览器,这些 Applets 由运行于浏览器端的虚拟机来执行,实时的编译提高了效率。甚至对通过 Internet 下载的 Applets 执行方式也能确保达到一个相当好程度的安全性。

1.3 公用对象请求代理程序体系结构(CORBA)

公用对象请求代理程序体系结构(CORBA)是由对象管理组织(OMG)定义的开放分布系统的标准。它定义了对象和客户在分布环境下的交互方式。

CORBA 详述的主要特征是有一个核心的对象模型、定位透明(客户和服务不需要了解它们各自位置如不同的主机),并且独立于程序语言。这些通过为对象提供接口定义语言(IDL)和动态更新接口(DII)来实现。CORBA 的 IIOP 使得来自不同厂商的对象请求代理能够相互通讯。

此外,CORBA 的服务和 CORBA 的便利是特定的。CORBA 的服务包括安全性、传输处理、命名等必要的服务,假如由系统提供这些服务,那么人们一定会抱怨 CORBA。CORBA 提供诸如桌面管理和非托管的定时操作。

1.4 复合技术

为了把 Java 在客户端的实力和 CORBA 在服务端端实力的益处结合在一起,一些供应商目前正把 CORBA 和 Java 合成一体。IIOP 可用于 Java Applets 作为 CORBA 客户和 CORBA 对象之间的通讯。这种结合的例子之一是基于 Java 的 VisiBroker,这是一个连接 CORBA 和 WWW 的对象请求代理。基于 Java 的 VisiBroker 是基于 CORBA 实现 VisiBroker for C++,并且用 Java 写的,对象可由 Java 或 C++ 实现。

2 Web 数据库

2.1 方法

一些技术和大量工具使得从 Web 上访问数据库成为可行。根据下面的原则可以将其分成三类。服务器端方法:依赖标准的 Web 浏览器、HTML、HTTP 和

CGI,它们主要扩充服务器端。可执行内容的方法:受到有效客户端计算能力的激励,是最近出现的。前两种方法集中于访问单一数据库。中间层方法:允许从不同的数据源合并数据。下面我们讨论这些方法的一些典型实现。图 1 说明了静态 HTML 页的访问(箭头 1-4),装入 Java Applets (箭头 5-8),使用 CGI 稿本生成

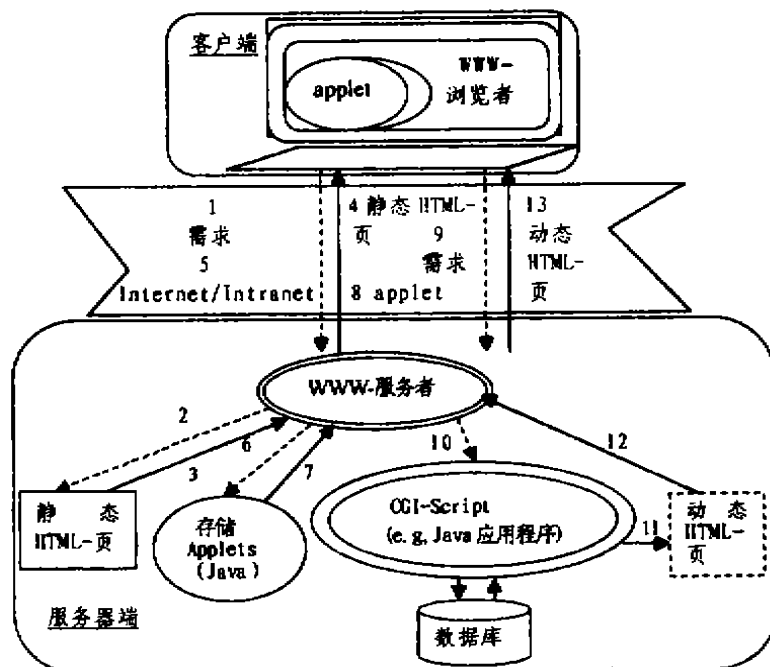


图 1 使用静态 HTML 网页,Java Applets 和 CGI Script

HTML 页(箭头 9-13)。

2.1.1 服务器端方法 网上访问数据库最古老大概也是最广泛应用的方法是基于公共网关接口(CGI)的。CGI 是一个标准,它允许 Web 服务器访问任意的外部应用,称做 CGI 稿本。所有的 Web 服务器都支持 CGI。

一个 CGI 稿本在服务器端被执行,这是由 Web 服务器根据客户端的请求启动的。CGI 稿本接受由用户提供的参数及隐含变量,生成动态 HTML 页(例如:通过 CGI 稿本恢复数据库中的数据),并且传到 Web 服务器。Web 服务器再把生成的 HTML 页传给 WWW 客户。

CGI 稿本可由任何语言实现,但是,为了确保服务器端应用的方便应该使用广泛的支持技术。根据 Web-CDS 的建议可能达到这一点的是用 Java 实现 CGI 稿本。

克服了 CGI 方法的一些不足(例如:对每一数据库的访问都要连接与释放连接),许多 CGI 工具是有

效的。其中之一就是 Web,它支持带有嵌入稿本的 HTML 模板,由 TCL(工具命令语言)实现的稿本允许变量和控制结构,Web 格式化 HTML 部分并根据 HTML 模板扩充 TCL 象 OraTcl 遵循同样的方法产生执行的 Tcl Script。

2.1.2 可执行内容的方法 用 Java 写的程序允许与数据库系统的更详细的交互,根本上是因为其克服了无国界 HTTP 的局限。因此 Java 与数据库的结合非常重要。在各种方法中最重要的就是 Java 数据库连接(JDBC),这是为 Java 访问相关数据库所制定的开放标准,JDBC 是由一些厂商支持的访问数据库的访问程序接口。

另外,对象 Java 和活动的 X 控件有如 JavaScript 的稿本方法。这些稿本由浏览器解释和执行,这样增进了网上的交互性和通讯,例如对输入的一致性检查。

2.1.3 中间层 使用 CORBA 在网上访问数据库主要有两种方法。其一是使用 CGI 稿本作为 CORBA 客户,这些 CORBA 客户转而通过 CORBA 服务器

对数据库系统访问。其二使用 IIOP 在作为 CORBA 客户的 WWW 客户与 CORBA 服务器之间直接通讯。此外对数据库系统来说任意深入的系统和应用同样可由 CORBA 使之一致。

2.2 各种方法的比较

比较网上访问数据库的各种技术时,所用标准应尽可能独立于特定的技术因素。对不同数据库特征的支持包括支持会话管理、事务和功能所支持的范围。一般的特征包括安全性、性能和同用户友好性。软件成本包括培训、开发和维护的费用。开放性指的是对不同类环境的支持、可移植性并支持不同厂商的产品。

3 实例分析

3.1 使用 CGI 和 HTTP:WebChiq1

为方便用户能通过浏览器直接访问 Web 上的数据库,我们开发了基于中文数据库查询语言 Chiq1 的 CGI 接口程序 WebChiq1。通过 WebChiq1,用户可以在 HTML 文本中直接嵌入 Chiq1 语句,完成对数据库的查询操作。

本质上 WebChiq1 是一组 CGI(公共网关接口)程序,在因特网服务器软件 Windows httpd 支持下运行。CGI 是为了使 Web 服务器不仅能访问 HTML 页面信息,而且能透明访问其它信息源(如 DBMS)而设置的标准接口。基于这种方法,用户只要在客户端的 HTML 文本中把适当的 Chiq1 语句嵌入<chiq1>与</chiq1>标记之间,服务器端就会由 Web 服务器启用有关的 CGI 程序 WebChiq1,将查询提交给指定的数据库服务器,从数据库找出数据,取代 HTML 文本中的<chiq1>与</chiq1>标记,从而使目标文本能显示数据库数据。WebChiq1 是 VC 开发的,对数据库的访问是基于 ODBC 接口,因此具有很好的可移植性。

WebChiq1 可支持静态和动态两种形式的 Chiq1 语句。静态 Chiq1 是将定义好的 Chiq1 语句写入 HTML 文本中,从而固定了数据在最终网页上显示的位置和格式;而动态 Chiq1 是将带有动态变量的 Chiq1 语句写入 HTML 文本中,根据用户输入的变量赋值,而获取相应的数据。

3.2 使用 Java 和 JDBC:Nchiq1 界面

通常,由 C 语言实现的应用程序可以通过 ODBC 与数据库联系,从而访问数据库,类似地,与 Java 语言相对应的 JDBC(Java DataBase Connectivity)就可以实现 Java 语言与数据库的连接。开发人员无须了解低层的处理,直接采用 JDBC 提供的类以及方法就可以实现与数据库的连接。给用户提供一个友好的图形界

面就要采用 AWT 包(Abstract Window Toolkit),采用现成的各种组件,对这些组件稍加改动即可使用,大大方便了开发人员。有了 Java 及其辅助技术 JDBC API、JDBC drivers 以及 AWT,就可以建立一个连接和访问数据库的图形界面。这样,也就可以实现中文自然语言对数据库的网上查询。

与纯粹的 HTML 比较,Java 在开发图形用户接口方面需要实质性的努力,另一方面 Java 提供了如多窗口和用户界面等舒适的方面。如果不用数据压缩,WWW 服务器和 WWW 浏览器之间的通讯次数会减少而通讯量会增加。关于在 WWW 环境下访问数据库,与 HTTP 相比,Java 允许内部的客户服务器通讯,这样很容易通过访问多表的网络支持数据交互。

3.3 创建一个联合构架:WWW-UIS

在一个联合构架中,各独立构件建立一个松散的整体系统。联合构架扩充了联邦和多数据库系统(FDBMS),FDBMS 可在公共的结构下从不同的数据库管理系统获得信息,在公共的结构中每个参与的 DBMS 仍保持自身独立性。

图 2 表示一个联合构架的例子,作为 FDBMSs 的扩展,不仅数据库((x)DBS),而且地理信息系统(GIS)、专家系统(XPS)以及深入的应用程序加入了联合,CORBA 用于使中间件一致。而且,正如从动态数据库中所知,WWW-UIS 根据 ECA 规则合并了探知事件。

横向来说,我们区分用户层和服务层的服务。系统层的服务提供如访问数据库和 HTML 页的准备等基本功能。用户层服务就是把几个系统层的服务合并为使用者可用的较高层次的服务。Java Applets 使用 IIOP 与其系统端通讯。

纵向来说,我们要区分信息和数据的服务,信息服务基于元数据,即数据的数据,它们帮助用户找到相关数据源,如数据服务和报告。数据服务可访问各种不同的数据。信息服务和数据服务都传送为生成 HTML 页作为输入的结果。

总结 本文提供了从网络上访问数据库的不同方法。由于这些方法在一些标准上不同,所以一般不能只推荐一种技术,但是,若只要求访问一个数据库的有限功能,那么传统的 CGI 和 HTTP 方法就足够了。为了增强服务器端的可移植性,要将 Java 和 JDBC 结合起来,与此对比,如果要求较细的功能和较高的交互性,如并发更新代表不同表单的几个数据库表,在客户端和服务端使用 Java 是比较好的方法。

最近,对如 WWW,Java,CORBA 等公开的 Inter-

net 技术的集成提出了许多异质的技术问题。这些技术为向异质语义的挑战提供了一个良好的基础,例如

用仲裁技术把不同知识库统一为信息服务的一部分(见图 2)。

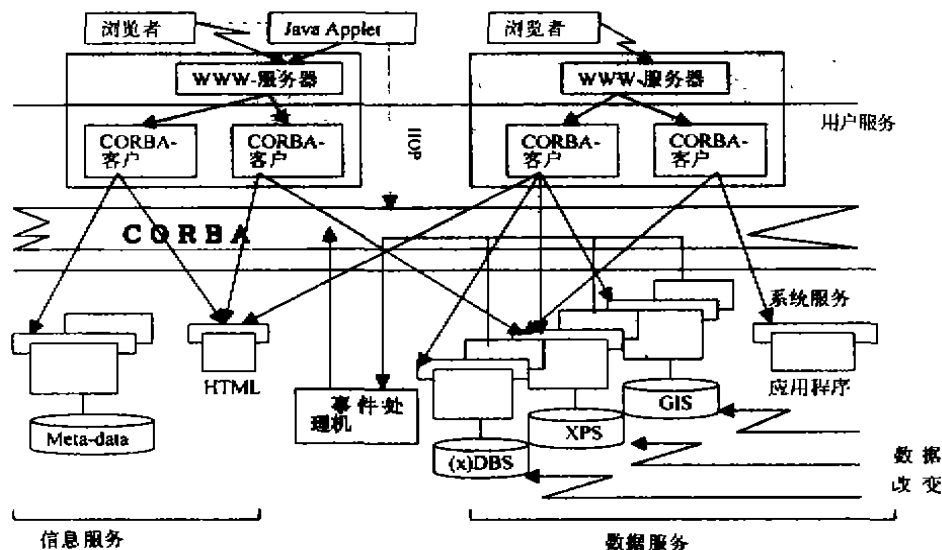


图 2 WWW、CORBA 和 Java 协作的联合体系结构

(上接第 59 页)

Bennett 的结果相比较,他采用 BP 算法训练了 10 个隐含节点数为 6 的神经网络,平均训练样本的正确率为 98.3%,测试样本为 94.9%^[13]。

参考文献

- 1 杜方、王小同. 前向神经网络学习算法: 学习速率矩阵法. 西北工业大学学报, 1994, 11(4): 403~406
- 2 Fahlman S E, Lebiere C. The Cascade-correlation learning architecture. In: Touretzky D S, ed. Advances in Neural Information Processing Systems 2. San Mateo, CA: Morgan Kaufmann, 1990, 524~532
- 3 Reed R. Pruning algorithm-A survey. IEEE Trans. on Neural Networks, 1993, 4(5): 740~747
- 4 Zurada J M. Lambda Learning rule for feedforward neural networks. IEEE ICNN, 1993, 1808~1811
- 5 Moraga C H J. The influence of the sigmoid parameters on the speed of backpropagation learning. In: Mira J, ed. From Natural to Artificial Neural Computation. Berlin: Springer, 1995, 195~201
- 6 向国全、董道珍. BP 模型中的激励函数和改进的网络训练法. 计算机研究与发展, 1997, 34(2): 113~117
- 7 Chen C-T, Chang W-D. A feedforward neural network with function shape autotuning. Neural Networks, 1996, 9(4): 627~641
- 8 吴佑寿、赵明生、丁晓青. 一种激励函数可调的新人工神经网络及应用. 中国科学(E 辑), 1997, 27(1): 55~60
- 9 Yamada T, Yabuta T. Neural network controller using autotuning method for nonlinear function. IEEE Trans. on Neural Networks, 1992, 3(3): 595~601
- 10 Ash T. Dynamic node creation. Connection Sci., 1989, 1(4): 365~375
- 11 Baum E B, Lang K L. Constructing hidden units using examples and queries. In: Lippman P, Moody J E, Touretzky D S, eds. Advances in Neural Information Processing Systems 3. San Mateo, CA: Morgan Kaufmann, 1990, 904~910
- 12 Chen Q C. Generation-shrinking algorithm for learning arbitrary classification. Neural Networks, 1994, 7(9): 1477~1489
- 13 Bennett K P, Mangasarin O L. Neural network training via linear programming. In: Pardalos P M, ed. Advances in Optimization and Parallel Computing. Amsterdam, Elsevier Science, 1992, 56~67