

一种 Web 资源组织框架^{*}

A New Organization Framework of Web Resource

胡久乡 钟 瑜

(华中理工大学计算机科学与技术学院 武汉430074)

Abstract This paper analyzes characteristic and disadvantage of the organization of the Web resource, which is conflict with the expand of the application of multimedia in internet today. By enhancing the ability of HTML, we build a logical uniform and organized multimedia Resource Object. Based on such Resource Object, we provide a new organization framework of the Web resource, it can meet the exception of complex and cosmically multimedia application. It can help user quickly find the information what they want. It can effectively keep coherence of the information.

Keywords Web, HTML, Resource Object, Templet, Address method

1 引言

当前的 Web 技术主要采取的是 HTML (Hyper Text Mark-up Language) 文档的解决方法来对万维网上的信息进行组织表示发布、传输、定位。简单地说,即是对大量 HTML 文档进行组织,在 HTML 文档中嵌入包含有 URL (Universal Resource Locators) 的特殊标志,将 HTML 文档互相链接然后利用 HTTP (Hyper Text Transfer Protocol) 进行传输,其优点在于简单,维护方便,能够方便地构造出内容生动丰富的用户界面,但是 Web 技术所依赖的传输协议 HTTP 从本质上说是一个面向静态文档的协议,难以处理复杂的交互操作,并且最令人头痛的是 Web 上的资源往往缺乏有效的组织,其上信息的检索却相关,一致性的控制是十分麻烦的,对复杂和大规模应用往往难以胜任。针对这种情况,国内外的厂商和研究机构提出了一些解决的方法:其中 W3QS 提出类 SQL 语言查询 Web 并利用 UNIX 工具实现了原型系统^[1],但没有给出数据模型;WebLog^[2]提出类 Datalog 语言,着重于 Web 文档的内部结构;Mendelson 等人在文[3]及其后一系列论文中提出了 Web 查询语言 Web SQL 并实现了原型及应用举例,其数据模型类似对象模型,在基于 WebSQL 基础上给出了形式化的 Web 演算,此外还提出基于场地 (Locality) 的查询优化思想;RAW^[4]提出了一个 Web 关系代数及其操作。陈彦等提出了基于标记图的 Web 数据模型^[5],实际上,大多数学者和研究

人员将 Web 作为一种半结构化数据源,主要关注 Web 各个文档的链接关系,而不是从文档内容的内在逻辑关系去构建 Web 数据模型。

本文针对大量无组织零散的分布于网上各处的信息数据,提出一种在逻辑上有组织的统一的多媒体数据对象,它将有效地帮助用户检索到所需要的信息,适应复杂和大规模应用,有效地保证信息数据的一致性,同时也有利于网站的有效管理。另外,本文提出了分析该框架性能的基本参数模型。

2 Web 数据组织新框架模型

针对当前 Web 技术的特点与网络大型应用之间的矛盾,以及未来网上应用发展趋势和要求,我们结合了当前 Web 技术和数据库技术,扩充了 HTML 文档,提出了一种基于资源对象的 Web 资源组织框架模型。

2.1 框架系统的整体结构

当客户在客户端上的浏览器上发出对某一信息的要求,由客户端的对象客户端接收以 HTML 文端向 Web Server 发送。当 Web Server 接收到 HTML 后将其有关对象的 HTML 提交对象服务器,对象服务器对封装在 HTML 中的对象信息分析后,向数据库服务器提出相应请求,数据库服务器将依据请求,向对象服务器输出构建对象的原始的文本、图像、视频、音频信息数据,对象服务器将原始数据封装后在动态生成的 HTML 文档中嵌入多媒体信息对象(即资源对象),然后发送 Web Server,再由 Web 客户下载后,由对象客

^{*} 本研究由国家高性能计算基金部分资助,胡久乡 博士,副教授,主要研究方向:人工智能、网络技术、计算机可视化技术,钟瑜 硕士研究生,主要研究方向:网络技术。

户端将用户所需信息提交显示给用户。

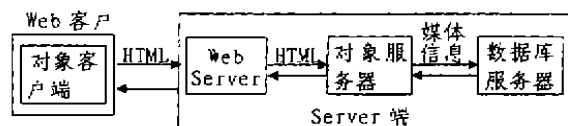


图1 框架系统的整体结构

2.2 多媒体信息对象(即资源对象)

在框架中我们看到资源对象是一个关键性的角色,正是它使得Web上存放的散乱信息变得高度有机统一,大大缩减了用户操作时间,克服了传统Web上交互的困难。其实资源对象是我们利用面向对象的技术对Web上的信息和相应操作进行有机组织得到的。

2.2.1 资源对象的结构 谈到资源对象首先要提到对HTML文档的扩充,加入一个对象容器的概念,我们将所有同属于一个资源对象的所有Web页面放在一个对象容器,在通常情况下,容器中对外仅提供一个标签页。标签页上包含有对资源内容的简介,并且封装了资源对象浏览许可和进入资源对象操作。在获得资源对象浏览许可,并进入资源对象之前,对任何一个资源对象仅可见一个标签页。在进入资源对象后我们可以看到允许浏览的Web页,Web页之间的链接全部封装。对象之间可以互相包含,例如图2中A对象包含了B对象。简单说资源对象将包含以下成员和操作:

成员:

- 标签页:包含了对该资源对象的简介与封装,浏览许可和进入对象操作。
- 目录页:对该资源对象,资源使用操作的封装和进入相关页面操作。
- 资源页:提供资源页面,有时可能与目录页重合,合用。

操作:

- 浏览许可:对某些特定资源的访问进行身份检查,确保安全性。
- 进入对象:进入对象,提交目录页供用户使用资源对象。
- 资源使用:图片,文本显示,音频,视频播放,依资源对象不同。
- Down In:进入所包含的子对象操作。
- Up out:进入资源对象目录页。
- Jump out:进入相关的资源对象标签页。

资源对象之间除了上下包含之外,还可能有些别的关系,例如商品资源对象和厂家资源对象之间关系不使用包含关系,为了方便用户对资源收集,所以提供了Jump out操作,从整体上来说资源对象内部是有

机统一的,页面之间的链接全部封装,无法见到显示的URL链接。

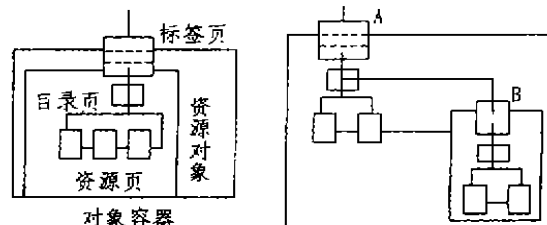


图2 资源对象

2.2.2 资源对象的浏览,使用设置 通常可以装一个index页面,将一批相关资源对象的标签页链接放入其中,或者可以将标签页放入一个网络搜索页面中,让用户通过对内容检索而得到相关资源对象的标签页。

2.2.3 资源对象与用户的交互 显然用户浏览使用资源对象其实就是一种交互,另外在对象服务器中存放有相关的模板供用户使用,由用户来创建资源对象,提供各种资源,并设定资源对象放置位置与使用许可。

2.3 对象服务器

对象服务器其作用在于完成对象的构造以及析构过程,负责将数据库中的原始的基本类型数据:文本、图像、声音、视频等信息构造成一个逻辑层次的资源对象,同时又负责将由用户以资源对象形式传回的信息析构成相应的基本类型数据存入目前通用的商用关系数据库中,其结构大致划分为:

- 对象模板:依据相关对象信息来指导,对象构造器,生成或析构对象。
- 对象构造器:依据对象板模板生成或析构资源对象。
- 智能Agent:依据对象资源请求状态,对一定资源对象采用缓存共享的方式。
- 对象缓冲区:对资源对象进行缓存。
- HTML模板:用来生成对象容器装入资源对象形成HTML页。

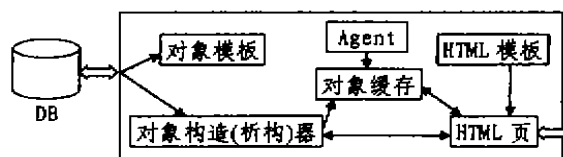


图3 对象服务器

显然对象服务器可以采用CORBA技术来实现,采用分布方式,并发操作将加快对象生成和交互使用

且对于巨型的复杂的应用来说该层服务器分布式技术合适且现实。

3 框架性能的分析

显然该框架组织对只有少量信息的 Web 服务或只需要少量信息的网上应用,没有什么益处,反而由于存取数据库,构建模型资源对象而大大降低反应时间。但是对于需要复杂多媒体信息的一些网上商场,电子商务,网上教学确能很好地提高效率,主要是由于对象的组织使人们在各个网页之间链接来、链接去的查找时间减少,基本上找到一个资源对象就可以找到一批相关的资源,从而提高工作效率。我们可知:

$$\text{总体时间 } T = t_{TP} + t_{BO} + t_{DB}。$$

t_{TP} 网络传输时间 $t_{TP} = V$, 总体数据量 V 网络传输率

显然 t_{TP} 与网络硬件操作系统的关系密切,我们能做的是减少重复传送,而其它一般没有办法。 t_{BO} 对象构建和析构时间 t_{BO} 数据库查询等操作时间。

$$t_{BO} = \begin{cases} t_{模板} + t_{建立对象} + t_{封装HTML} \\ t_{模板} + t_{析构对象} + t_{解封装HTML} \end{cases}$$

$$t_{BO} = t_{建物理连接} + t_{查询} + t_{输出}。$$

由于缓冲区的建立,因此

$$T = t_{TP} + t_{建缓冲区} + t_{传输缓冲区}。$$

由于缓冲区是牺牲相应内存和硬盘空间,故不会太大

$$t_{建缓冲区} + t_{传输} \ll t_{BO} + t_{DB}。$$

所以如果能提高缓冲池命中率,效率必然会大大提高。由于在某些情况下对象请求共用一数据库物理连接,故 t_{DB} 又可为

$$t_{DB} = t_{建物理连接} / n + t_{查询} + t_{输出}。$$

一般情况下建立物理连接起动一个新进程是比较耗时的,复用技术可使物理连接时间由 n 个请求分摊,但我们又必须注意在某一微观时刻只能有一个请求占有物

理通道,故增大 n 必然使 t 轮询时间增大(即各个请求平均等待时间)所以必须对 n 取一个合理的值范围。

必须注意到以上都是在没有进入临界状态的情况,因为我们知道任何系统的资源总是有限的,随着客户连入增大,进程不断开设,系统可用资源将减少,到最后进入临界状态,即系统已无可用资源,当有一个新客户到达时,系统为了替其建立进程和分配资源,必然导出其它进程到硬盘,从而引起系统抖动,不断导出,性能急剧下降,所以必须通过实验测出系统临界状态,确定用户上限。

结论 本文针对大量无组织零散的分布于网上各处的信息数据,提出一种在逻辑上有组织的统一的多媒体数据对象,它有效地帮助用户检索到所需要的信息,适应复杂和大规模应用,有效地保证信息数据的一致性,同时也有利于网站的有效管理。另外,本文提出了分析该框架性能的基本参数模型。该框架结构对 Web 资源进行了有效的组织,建立了基于面向对象观点的资源对象模型,比较适用于复杂的多媒体信息数字网上应用。

参考文献

- 1 Konopnicki D, Shmueli O. W3QS: a query system for the World Wide Web. In: Proc the 21st Very Large Data Base. Zurich Switzerland, 1995. 54~65
- 2 Lakshmana L V S, Sadri F, Subramanian L N. A declarative language for querying and restructuring the Web. In: Proc of RIDE'96. New Orleans, 1996
- 3 Mendelzon A O, Mihaila G A, Milo T. Querying the World Wide Web. In: Proc of PDIS'96. 1996. 80~91
- 4 Fiebig T, Weiss J, Moerkotte G. RAW: a relational algebra for the Web. In: Proc of Workshop on Management of Semistructured Data. Tucson, Arizona, 1997
- 5 陈彦, 徐宏炳, 王能斌. 基于标记图的 Web 数据模型. 计算机学报, 1999, 3: 306~312

(上接第24页)

复率达到90%所需的平均时延大约为240ms,而采用基于链路的恢复机制时,恢复率达到90%所需的时延大约为440ms。这主要是因为基于“树”的恢复机制可能利用一条更短的路径完成恢复任务。当节点失败时,无论采用何种机制,恢复性能都将大幅度降低,但基于“树”的恢复机制的性能还是明显优于基于链路的恢复机制。

结论 BVP 算法和泛洪式恢复算法满足了不同业务和用户对网络生存性的不同要求,在实际网络中应用很广泛,是 ATM 网络自愈技术的基础,很多以后的改进算法都是在此基础上进行的。当然这两种传统的自愈算法并不是十全十美的,本文所讨论的 ATM 多播树就是一个例子,当采用这两种算法时,恢复率并

不是很高,且时延大,我们针对多播树的结构特点,对传统的自愈算法进行改进,以适应多播树中失效信息的恢复,改进后的算法明显地提高了 ATM 多播树的性能。

参考文献

- 1 Kawamura, Sato K, Tokizawa I. Self-rerouting ATM networks base on virtual path concept[J]. IEEE J SAC, 1994, 12(1): 120~127
- 2 Anderson J. Fast Restoration of ATM Network[J]. IEEE J SAC, 1994, 12(1): 128~138
- 3 马丁·德·普瑞克(程时端译). 异步传递方式. 邮电出版社, 1995
- 4 Tanenbaum A S. Computer Networks 清华大学出版社(影印版), 1996
- 5 刘夏兵, 等. 基于 ATM 构造现有网络技术. 数字通信, 2000(1): 42~44