

# 集成 Web 数据的系统框架与实现方法

卢正鼎<sup>1</sup>, 张素智<sup>1,2</sup>

<sup>1</sup>(华中科技大学 计算机科学与技术学院, 湖北 武汉 430074)

<sup>2</sup>(郑州轻工业学院 计算机系, 河南 郑州 450002)

**摘要:** 基于我们正在研究和开发的项目 Panorama, 本文提出了一个基于 XML 和 CORBA 的 Web 数据集成系统框架。该框架可以集成来自 Web 上多个异构的数据源, 包括关系数据库、面向对象数据库、HTML 和 XML 文档及结构化文本文件。在系统实现中, 把 Web 看作是一个巨大的虚拟数据库, 以 CORBA 作为分布式对象模型, 以 XML 作为公共数据模型, 并且以 XML-QL 作为全局查询语言来完成 Web 上的数据查询和集成。文中还详细分析和描述了系统框架中的一些主要模块的实现方法, 具有很好的可操作性。

**关键词:** 数据集成; CORBA; XML; Web 数据; 异构数据源

中图分类号: TP311

文献标识码: A

文章编号: 1000-1220(2003)10-1759-05

## An Architecture of Integrated Web Data and Implementation

LU Zheng-ding<sup>1</sup>, ZHANG Su-zhi<sup>1,2</sup>

<sup>1</sup>(College of Computer science and Technology, Huazhong University of Science and technology, Wuhan 430074, China)

<sup>2</sup>(Department of Computer Science and Technology, Zhengzhou of Light Industry, Zhengzhou 450002, China)

**Abstract:** In this paper, we propose an architecture of Web data integration with XML and CORBA, which is based on our developing project, Panorama. The architecture can integrate the data from multiple data sources, including relational database, object-oriented database, HTML and XML documents, and structured text files. In the architecture, we view Web as a huge database, and take CORBA as the object model and XML as mediated data model, and use XML-QL to accomplish data query and integration on the Web. We also elaborate and analyze some implementing methods, such as integrator and wrappers corresponding to various data sources.

**Key words:** data integration; CORBA; XML; Web data; heterogeneous data source

### 1 引言

随着应用需求的不断增加和网络技术的发展,越来越多的用户希望能够同时访问和处理来自多个数据源的数据<sup>[3,4]</sup>。WWW 的迅速发展,更使其成为全球信息传递和共享的日益重要和最具潜力的资源。Web 上的资源不仅包括传统的数据库,如关系数据库和面向对象的数据库,它们都有很好的数据模型,而且还包括无结构和半结构的数据,如 HTML 和 XML 文档及文本数据。Web 数据的表现形式是不规则和多样的,要用传统的数据库技术来存储和管理 Web 上的所有数据是十分困难的。如何将 Web 上多个分布式异构数据源中的数据集成在一起为用户所用,是目前急需解决的问题<sup>[4,6]</sup>。

目前,XML 已成为新一代 Internet 上数据表示和交换事实上的标准<sup>[1,2]</sup>。像 HTML 一样,XML 也是一个标记语言,它具有许多新的特性,XML 是面向内容的,它具有更多的结构和丰富的语义,良好的可扩展性,简单易学和自描述等特点,适用于 Web 上数据的表示和交换。XML 中的文档数据类型定义(DTD)描述了文档内可以使用的元素类型及其元素之

间的关系。事实上,XML 为 Web 上的数据管理提供了新的数据模型,很多成熟的数据库技术将用于 Web 数据处理,把 Web 变成一个巨大的虚拟数据库。XML 使得将数据库技术和研究扩展到 Web 数据管理成为可能<sup>[1,3]</sup>。

为了给用户提供透明地访问 Web 上的数据,而不用考虑平台、位置和系统,在我们提出的 Web 数据集成结构中采用了 CORBA 技术。CORBA 是由对象管理组(OMG)提出的分布式对象标准,用来处理平台层的异构性和提供位置及实现的透明性<sup>[6,9,10]</sup>。利用 CORBA 技术,可以大大减少编写程序的复杂度,允许客户发现新加入的对象,并按即插即用的方式使用这些对象而不需要改变客户端的程序代码。

CORBA 对象模型和 XML 数据模型都是用户和编程人员公认的工业标准而广为采用。在本文中,基于我们正在研究和开发的项目 Panorama,提出了一个基于 XML 和 CORBA 的 Web 数据集成系统框架。在这个框架中,把 Web 看作是一个巨大的数据库,采用 CORBA 作为对象模型,XML 作为数据模型,利用 XML-QL<sup>[3]</sup>作为查询语言来实现 Web 上数据的查询和集成。它可以集成包括关系数据库、面向对象数据库、

HTML 和 XML 文档以及结构化文件等多个数据源中的数据.

## 2 XML 数据管理的基础

### 2.1 XML 文档的分类

XML 文档分为两类<sup>[1]</sup>: 一是以数据为中心的文档, 这种文档在结构上是规则的, 在内容上是同构的, 它可以用作存储 Web 数据的一个容器. 另一个是以文档为中心的文档, 这种

```

<? xml version="1.0" standalone="" ?>
<! DOCTYPE bookstore SYSTEM "books.dtd">
<bookstore>
  <book year="2001">
    <title>XML 揭秘</title>
    <author>Michael Morrison </author>
    <publisher>清华大学出版社</publisher>
  </book>
  <book year="2000">
    <title>分布式数据库</title>
    <author>郑振楣</author>
    <author>于戈</author>
    <publisher>科学出版社</publisher>
  </book>
  ...
</bookstore>

```

图 1 以数据为中心的文档  
Fig. 1 Data-centered document

文档的结构是不规则的, 常用来在网页上发布描述性信息、产品性能介绍、E-mail 信息等. 图 1 和图 2 分别表示了这两种文档的例子.

```

<? xml version="1.0">
<paper>
  <title>集成 Web 数据的框架与实现</title>
  <description>
    <abstract>WWW...</abstract>
    <keyword>数据集成...</keyword>
    <para>1. 引言: 随着应用需求的不断增加和网络技术的发展, 越来越多的用户希望能够同时访问和...</para>
    <para>2. Web 数据管理的基础</para>
    .....
  <reference>
    <item><link url="http://www.w3.org">XM</item>
    <item><link url="http://www.omg.org">CORBA </item>
  </reference>
</paper>

```

图 2 以文档为中心的文档  
Fig. 2 Document-centered document

我们把以数据为中心的 XML 文档看作是一个数据库, 称为 XML 数据库<sup>[2]</sup>, 其内部的结构化数据称为 XML 数据. XML 数据库的模式由它所遵循的 DTD (或 XML Schema) 来定义. 图 1 中的 DTD 定义在外部文件 books.dtd 中, 内容为:

```

<! ELEMENT bookstore (book)+>
<! ELEMENT book (author+, title, publisher)>
<! ATTLIST book year CDATA #REQUIRED>
<! ELEMENT title (#PCDATA)>

```

```

<! ELEMENT author (#PCDATA)>
<! ELEMENT publisher (#PCDATA)>

```

### 2.2 XML 数据模型

XML 数据与半结构化数据类似, 可以说 XML 数据是 Web 上的半结构化数据. 文献中已经提出了许多描述半结构化数据模型<sup>[5,11,23]</sup>, 大多数都是采用基于带标签图的方法作为半结构化数据模型, 最典型的的就是斯坦福大学提出的 OEM 模型. XML 数据模型也是一个图/树模型<sup>[3]</sup>, 该模型分为有序数据模型和无序数据模型两种.

定义 1: 一个无序的 XML 数据模型是:

- 一个图(或者是树)G, 图中的每个节点由唯一的对象标识符(OID)来表示;
- 图的边用元素标签来标记;
- 图的节点由属性-值对的集合来标记;
- 图的叶节点由一个字符串来标记;
- 图有一个明确的树根节点.

定义 2: 一个有序的 XML 数据模型也是一个 XML 图, 但是图中的每个节点有一个总体序号. 对于从 XML 文档构造出的图, 节点的自然序号就是它们在文档中的顺序号.

例如, 图 1 中的 XML 数据文档的有序数据模型可以表示如下, 如图 3 所示.

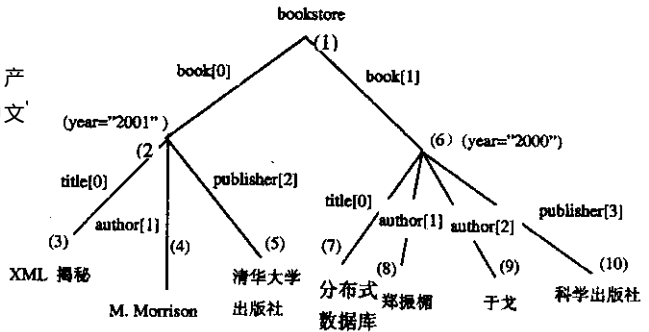


图 3 一个有序 XML 数据模型的例子

Fig. 3 An example for ordered XML data model

XML 图便于与不规则的数据进行映射, 它适合于描述异构的和动态变化的 Web 数据. XML 数据库的模式由 DTD 定义, 并且 DTD 决定了 XML 文档的结构.

### 2.3 XML 数据的查询与转换

XML 数据与关系数据和面向对象数据不同, 所以传统的查询语言(SQL 和 OQL)都不能直接适用于 XML 数据<sup>[8,17]</sup>. 然而, XML 数据模型与半结构化数据模型很相似, 并且一些处于研究阶段的查询语言已被开发出来, 如 XQL, XSL, XML-GL, LOREL 和 XML-QL<sup>[3]</sup>. 由 AT&T 实验室开发的 XML-QL 是一种可以对 XML 数据进行查询的语言, 并且利用 XML-QL 的查询方式可以实现 XML 数据的抽取、转换和集成.

XML-QL 的结构是 Where-In-Construct 的形式, 类似于 SQL 的 Select-From-Where 结构, 但它们的含义和使用方法不同. XML-QL 结构中的 Where 子句包括两部分<sup>[13]</sup>: 模式和条件表达式, 查询出的结果要满足两个条件:

(1) 数据项的类型和值必须与指定的模式相匹配;

(2) 数据项的值要满足条件表达式. 在我们所提出系统框架中, 采用了 XML-QL 作为数据查询和数据集成的语言.

Web 上的数据不仅包括传统数据库中的数据, 而且还包括 XML 数据、HTML 文档和结构化文件中的数据. 由于在系统结构中采用了 XML 数据模型作为公共数据模型(CDM), 这就产生了其他模型的数据向 XML 数据的转换<sup>[83]</sup>, 包括:

- (1) 关系数据向 XML 数据的转换;
- (2) 对象数据向 XML 数据的转换;
- (3) HTML 文档向 XML 数据的转换;
- (4) 结构化文本文件数据向 XML 数据的转换.

在我们提出的系统框架中, 这些转换都是通过不同的包装器(Wrapper)来实现的.

## 2.4 XML 数据的存储和索引

因为把 XML 数据看作是一个数据库, 所以就应当考虑 XML 数据存储的问题. XML 数据存储方法分为四种<sup>[13,14]</sup>: 关系数据库、面向对象数据库、直接存成 XML 文本文件和专门设计的 XML 数据存储策略. 我们在设计中采用了将 XML 数据直接存储在 XML 文本文件的方法, 以便于用 XML-QL 进行转换和集成. XML 数据模型一般采用树/图结构数据索引<sup>[23]</sup>. 伯克利大学的 Xset 系统和斯坦福大学的 DataGuide 系统都为 XML 数据和半结构数据建立了索引.

## 3 Web 数据集成的体系结构

### 3.1 在设计中需要考虑的因素

Web 数据集成系统的目标就是从 Web 上多个异构数据源中查询所需的数据, 并把这些查到的数据合并在一起提交给用户<sup>[13]</sup>. 在 Web 数据集成过程中需要注意一些问题, 包括:

(1) 中间模式的说明和生成, 包括中间模式的选择以及各局部模式与中间模式的转换, 如选择 XML 作为中间模式.

(2) Web 数据源的完备性, 包括对不同数据源的数据完备性进行评估, 分析不同数据源中的数据是否有重叠, 设计中要过滤掉重复的数据.

(3) 各数据源查询数据能力的不同要求考虑对局部数据访问和抽取数据能力的限制, 尽可能把局部数据源能胜任的工作放到各数据源来完成, 以减少网络数据的传输.

(4) 查询分解与优化. 设计中存在全局查询到各数据源的查询分解和转换, 同时还要考虑查询优化和形成查询计划.

(5) 包装器的设计与实现. 包装器(Wrapper)程序实现各不同数据源到中间模式之间的模式转换和查询翻译. 采用 XML 数据作为中间模式有利于包装器的设计和实现.

### 3.2 体系结构

根据以上讨论和分析, 我们提出一个集成 Web 数据的系统框架, 如图 4 所示. 在这个系统框架中, 我们采用 XML 数据模型作为中间模式, 用 XML-QL 作为查询语言来完成对 Web 上不同数据源的查询、模式转换和数据集成. 主要程序模块包括: 应用程序和可视化用户接口, XML-QL 查询处理器, XML 数据集成器, 元数据字典和各不同数据源的包装器.

• 应用程序和可视化用户接口用于按全局模式(或集成模式)提交用户请求, 并且以界面友好的形式将查询结果返回给用户. 在应用程序中的数据处理是通过 XML-QL 来完成的, 包括全局的查询, 数据统计和数据集成等, 主要是因为从集成器输出的数据是 XML 数据形式.

• XML-QL 处理器包括两部分: XML-QL 分析器和 XML-QL 查询分解器. XML-QL 分析器用于分析 XML-QL 的语法和测定语义的有效性, 而 XML-QL 查询分解器则是通过与系统元数据字典合作将全局查询分解成与不同数据源相对应的多个子查询.

• 元数据字典用于存储多个数据源的元数据和全局(或集成)模式, 以及局部数据源和全局数据模式之间的映射关系.

• 包装器: 不同的数据源有不同的包装器, 它主要完成从全局子查询(XML-QL)到局部数据查询的翻译(如 SQL 或 OQL), 以及从局部查询结果到 XML 数据的转换, 如从关系数据到 XML 数据的转换.

• 集成器(包括过滤器)用于将多个来自于局部数据源包装器的中间查询结果(XML 数据)按全局(或集成)模式合并在一起, 形成一个虚拟数据库, 用户应用程序直接对这个虚拟数据库进行操作. 过滤器是集成器中的子模块, 专门用于过滤来自不同数据源的数据, 将重复的数据过滤掉.

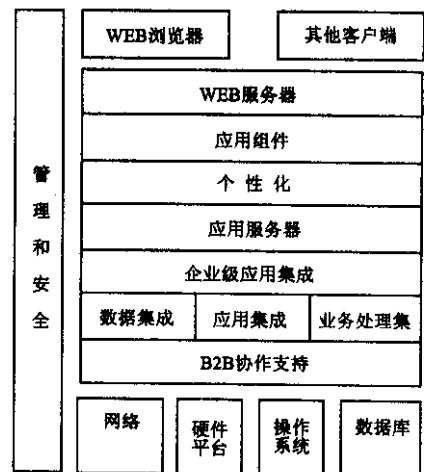


图 4 Web 数据集成的系统框架

Fig. 4 Architecture of integrated system for web data

系统框架中的主要程序模块都是按 CORBA 对象模型来实现的, 分布在 WWW 上的多个异构数据源都作为 CORBA 对象进行注册, 并通过 ORB 和 IIOP 进行通信. CORBA 为系统框架提供了可靠的对象模型和基础, 用户仅面对应用程序和可视化用户界面发送用户查询请求和接收查询返回的结果, 通过 XML-QL 处理器、集成器、不同数据源的包装器与元数据字典的配合, 实现 Web 上数据集成的透明性和一致性. XML-QL 处理器和集成器合在一起也叫做中间器(Mediator). 通过中间器, 用户发出查询请求, 就可以透明地得到集成的查询结果.

## 4 程序实现方法

### 4.1 集成器的程序实现

由于集成器的输入(即多个中间查询结果)都是 XML 数据,所以可以使用 XML-QL 来实现集成器的程序功能.假设在 Web 上有两个不同数据源的中间结果存放在 www.a.b.c/bookstore.xml 和 www.x.y.z/books.xml 中.这两个 XML 数据文档相似,但所遵循的 DTD 不一定完全相同,现假设 DTD 中包含下面共同的部分:

```
<! ELEMENT book(author+, title, publisher, ...) >
...
<! ELEMENT title (#PCDATA) >
<! ELEMENT author (#PCDATA)>
```

如果用户仅想查询两个数据源中所有书籍的标题和作者,用 XML-QL 实现的数据集成程序如图 5 所示.

```
WHERE
  (book) $b(</>IN{"www.a.b.c/bookstore.xml",www.
                                     x.y.z/books.xml"})
  (title) $t(</>
</>IN $b
CONSTRUCT
  (result)
  (title) $t(</>
  WHERE(author) $a(</>IN $b
  CONSTRUCT (author) $a(</>
  (</>
```

图 5 一个数据集成的例子

Fig. 5 An example for data integration

### 4.2 元数据字典的设计

元数据管理对多个异构数据源的数据集成是十分重要的<sup>[4,24]</sup>,因为不同的数据源都有它自己的接口和属性,当处理查询和选择查询结果时必须知道局部数据源的接口和属性.在上述结构中,元数据字典用于对系统元数据的管理,它是通

```
Interface MetadataDictionary{
  Boolean QL-Available(in QuerySourceType qsType);
  Boolean Registry(in QuerySourceMeta metadata);
  Boolean Unregistry(in QuerySourceName qsName);
  ...
}
```

图 6 元数据字典的接口设计

Fig. 6 Interface design for metadata dictionary

过使用 XML 中的 DTD 进行建模和设计的,以便于与中间器和包装器进行合作.元数据接口设计如图 6 所示.所有参与的数据源必须在元数据字典中注册,元数据字典按局部数据源的类型按层次进行管理.

### 4.3 包装器的设计

包装器主要完成从全局子查询到局部数据查询的翻译和从局部查询结果到中间数据(XML 数据)的模式转换.不同类型的数据源应该有不同包装器,所以它们的实现方法也不相同.

4.3.1 XML 数据源 因为 XML 数据源可以直接用 XML-QL 进行查询,并且返回的结果也是 XML 数据形式,所以这

种数据源的包装器只需要将 XML 数据源查询的结果转换成全局模式,可以利用 XML-QL 实现模式的转换.

4.3.2 HTML 数据源 Web 上存在大量以 HTML 形式存放的数据,但由于其特点的限制,除搜索引擎外还没有更好的方法用来查询 HTML 文档,提取 HTML 文档中的数据更加困难.在我们的设计中,首先将 HTML 文档全部转换成 XML (或 XHTML)文档,然后对生成的 XML 数据文档进行查询,并按 XML 数据的形式将返回查询结果给集成器.

目前有两种方法可以将 HTML 文档转换成 XML (或 XHTML)数据文档,一是手工方法,另一种是自动转换工具.采用手工转换时要注意它们之间的区别和对应的转换规则<sup>[18]</sup>.HTML Tidy 是由 HP 公司开发的一个自动转换工具.还可以将两种方法结合起来使用,先用自动转换工具进行转换,再用手工作法进行微调.

4.3.3 结构化文本文件 许多企业数据和遗留系统中的数据都存在结构化文本文件中.这种数据源的包装器可以通过编写程序从数据文件中读出数据,然后根据定义在元数据字典的数据模式要求,将数据写入到 XML 数据文档中.这个包装器程序由 C++(或 JAVA)来实现.当查询这种数据源时,首先用该包装器将结构化文本文件转换成 XML 数据文档,然后用 XML-QL 对生成的 XML 数据文档进行查询.最后,查询的结果通过 ORB 和 Internet 传送到集成器.

4.3.4 关系数据库和面向对象数据库 传统的关系数据库和面向对象数据库是 Web 上存储数据的两个最主要的方式.目前,大部分的商用数据库管理系统(DBMS)都提供了对 XML 的支持<sup>[1]</sup>,如 Oracle 8i,DB2 和 SQL Server 等.也就是说,这些 DBMS 在局部可以使用 SQL 或 OQL 进行数据查询,而查询结果以 XML 数据形式返回.为了加快商用数据库包装器的设计,我们尽可能使用数据库商家提供的转换功能,这是因为 DBMS 商家更了解该数据库的数据模型和程序接口.包装器接收 XML-QL 查询并通过翻译器将该查询翻译成局部查询(SQL 或 OQL),而查询结果则以 XML 数据的形式直接传送到集成器.翻译器是包装器中的一个子类,目前只实现了从 XML-QL 到 SQL 有限的翻译.

## 5 总结

将 Web 作为一个巨大的数据库进行数据集成是一项具有十分重要意义的课题.要实现这个目标,还有许多问题要解决.目前单一的信息搜索方法已不适应新一代 Web 的应用,而更多的是采用数据库的方法来管理 Web 数据.XML 将变成 Web 上数据表示和交换的标准和通用的语言.Web 数据源要求结构化的组织和访问,包括数据模型、数据查询和集成.Web 数据集成是一个具有挑战性的项目,包括数据模型的转换、查询分解和查询翻译、查询优化和索引等.在本文中提出了一个基于 XML 和 CORBA 的 Web 数据集成框架,并详细描述和分析了实现的主要方法.该结构采用 CORBA 作为对象模型,XML 作为数据模型,XML-QL 作为查询语言对 Web 上异构的多数据源中的数据进行集成.该结构具有以下优点:

(1) 采用了 CORBA 技术, CORBA 是一个开放的系统模型, 它支持分布式对象环境中各软构件之间的通信, 以及数据源的动态定位和集成. CORBA 也是基于构件和模块化的设计方法, 它使将来替换某些构件更加容易. 另外, CORBA 提供的 IIOP (Internet Inter-ORB Protocol) 协议使得 Web 上的各对象更容易通信和互操作.

(2) 采用 XML 数据模型作为公用数据模型 (CDM), XML-QL 作为查询和集成语言. 采用 XML 数据模型既可与 Web 上的数据特点相一致, 又解决了选择公共数据模型的问题. 作为查询和集成语言, XML-QL 与 SQL 或 OQL 相比, 虽然不太成熟, 但可以减少大量查询之间的转换.

(3) 元数据字典使用 XML 语法中的 DTD 进行设计, 与公共数据模型相一致, 这有利于它与集成器、XML-QL 处理器以及各包装器之间的合作.

(4) 不同的数据源设计不同的包装器, 而不是采用统一的包装器方法<sup>[4]</sup>, 以减小系统设计的复杂性.

(5) 尽可能利用已有的技术. 特别对于传统的商用数据库, 充分利用 DBMS 所提供的转换功能, 简化包装器的设计.

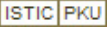
该结构简单易行, 但仍存在一些问题需要解决, 如查询的翻译、优化和索引. 在 Panorama 项目中, 这些工作将不断地得到改进和完善, 以使该系统更加稳定和实用.

## References:

- Bertino E, Catania B. Integrating XML and databases[J]. IEEE Internet Computing, July-August, 2001. 84~88.
- Salminen A, Tompa F W. System desiderata for XML databases [C]. Proceedings of the 27<sup>th</sup> VLDB Conference, Roma, Italy, 2001.
- Deutsch A, Fernandez M, Florescu D et al. A query language for XML[J]. Computer Networks. 1999, 31:1155~1169.
- Chang Y S, Ho M H, Yuan S M. A unified interface for integrating information retrieval[J]. Computer Standards & Interfaces. 2001, 23:325~340.
- Garcia-Molina H, Hammer J, Ireland K et al. Integrating and accessing heterogeneous information sources in TSIMMIS [C]. Proceedings of the AAAI Symposium on Information Gathering, Stanford, California, USA. 1995. 61~64.
- Bouguettaya A, Benatallah B, Ouzzani M., Hendra L. WEBFINDIT: an architecture and system for querying Web databases[J]. IEEE Internet Computing, July-August, 1999. 30~41.
- Levy A Y, Rajaraman A, Ordillc J J. Query heterogeneous information sources using source description [C]. Proceeding of the 22<sup>th</sup> VLDB(1996). 251~262.
- Fernaadez M, Tan W C, Suci D. SilkRoute: trading between relations and XML [J]. Computer Networks. 2000, 33:723~745.
- Dogac A, Dengi C, Kilic E et al. A multidatabase system implementation on CORBA[J]. IEEE 1996. 2~11.
- D' Ambrogio A, Iazeolla G. A CORBA-Based approach to design gateways for multidatabase systems[J]. IEEE 1997, 49~51.
- Abiteboul S, Buneman P, Suci D. Data on the Web from relations to semistructured data and XML [M]. San Mateo, CA: Morgan Kauffmann Publishers, 2000.
- Liu L, Pu C, Han W. An XML-enabled data extraction toolkit for web sources[J]. Information System. 2001, 2:6563~583.
- Meng Xiao-feng. Survey of Web data management[J]. Computer research and development. Apr. 2001, Vol38. 385~395.
- Zhang Z. Q, Zhou L Z, Feng J H. Introducing Web database[J]. Computer Science. 2001, 10(28):1~6.
- Azavant F, Sahaguet A. W4F user Manual [EB/OL]. 2000. <http://db.cis.upenn.edu/w4f/>.
- Cluet S, Delobel C, Simeon J et al. Your mediators need data conversion[J]. Acm Sigmod Int'l Conf on Management of Data, Seattle, Washington, 1998.
- Chamberlin D, Robie J, Florescu D. Quilt: An XML query language for heterogeneous data sources [C]. In: Proc of WebDB' 2000, Dallas, Texas, 2000. 18~19.
- Morrison M, Boumphrey F, Brownell D. XML unleashed [M]. Sams Publishers. 2000. 260~350.
- Florescu D, Levy A, Mendelzon A. database techniques for the World Wide Web: A Survey [J]. SIGMOD Record, 1998, 27(3).
- Shanmugasundara J et al. Relational databases for query XML documents: Limitation and opportunities [C]. In: Proc Of the 25<sup>th</sup> VLDB. Edinburgh, Scotland, 1999, 302~314.
- Papakonstantinou Y, Molina H G, Widom J. Object exchange across heterogeneous information sources [J]. In: Proc of Int'l Conf on Data Engineering (ICDE). Taipei, 1995, 251~260.
- Abiteboul S, Widom J, Lahiri T. A Unified Approach for Querying structured data and XML [EB/OL]. <http://www.w3.org/TandS/QL/QL98>. 1998.
- McHugh J, Widom J. Query optimization for XML [C]. In: Proc of the 25<sup>th</sup> VLDB. Edinburgh, Scotland, 1999, 315~326.
- Chantal R, Pierre S J, Dan V. Semantic integration of XML heterogeneous data sources [J]. IEEE, 2001, 199~208.
- Dongwon L, Chu W W. CPI: Constraints-Preserving Inlining algorithm for mapping XML DTD to relational schema [J]. Data & Knowledge Engineering 2001, 39:3~25.

## 附中文参考文献:

- 孟小峰. Web 数据管理研究综述 [J]. 计算机研究与发展, 2001, 38, 385~395.
- 张志强, 周立柱, 冯建华. Web 数据库技术简述 [J]. 计算机科学, 2001, 28, 1~6.

作者: 卢正鼎, 张素智  
作者单位: 卢正鼎(华中科技大学, 计算机科学与技术学院, 湖北, 武汉, 430074), 张素智(华中科技大学, 计算机科学与技术学院, 湖北, 武汉, 430074; 郑州轻工业学院, 计算机系, 河南, 郑州, 450002)  
刊名: 小型微型计算机系统   
英文刊名: MINI-MICRO SYSTEMS  
年, 卷(期): 2003, 24(10)  
被引用次数: 7次

## 参考文献(27条)

1. Dogac A; Dengi C; Kilic E A multidatabase system implement on CORBA 1996
2. Fernaadez M; Tan W C; Suci D SilkRoute: trading between relations and XML[外文期刊] 2000(1-6)
3. Morrison M; Boumphrey F; Brownell D XML unleashed 2000
4. Chamberlin D; Robie J; Florescu D Quilt: An XML query language for heterogeneous data sources 2000
5. Cluet S; Delobel C; Simeon J Your mediators need data conversion 1998
6. Levy A Y; Rajaraman A; Ordillc J J Query heterogeneous information sources using source description 1996
7. Bouguettaya A; Benatallah B; Ouzzani M; Hendra L WEBFINDIT: an architecture and system for querying Web databases[外文期刊] 1999(4)
8. Garcia-Molina H; Hammer J; Ireland K Integrating and accessing heterogeneous information sources in TSIMMIS 1995
9. Chang Y S; Ho M H; Yuan S M A unified interface for integrating information retrieval[外文期刊] 2001(4)
10. Deutsch A; Fernandez M; Florescu D A query language for XML[外文期刊] 1999
11. Salminen A; Tompa F W System desiderata for XML databases 2001
12. 张志强; 周立柱; 冯建华 Web数据库技术简述[期刊论文]-计算机科学 2001(10)
13. 孟小峰 Web数据管理研究综述[期刊论文]-计算机研究与发展 2001(04)
14. Dongwon L; Chu W W CPI: Constraints-Preserving Inlining algorithm for mapping XML DTD to relational schema 2001
15. Chantal R; Pierre S J; Dan V Semantic integration of XML heterogeneous data sources 2001
16. McHugh J; Widom J Query optimization for XML 1999
17. Abiteboul S; Widom J; Lahiri T A Unified Approach for Querying structured data and XML 1998
18. Papakonstantinou Y; Molina H G; Widom J Object exchange across heterogeneous information sources [外文会议] 1995
19. Shanmugasundara J Rational databases for query XML documents: Limitation and opportunities 1999
20. FLORESCU D; Levy A; Mendelzon A database techniques for the World Wide Web: A Survey[外文期刊] 1998(03)
21. AZAVANT F; Sahaguet A W4F user Manual 2000

22. [Zhang Z Q;Zhou L Z;Feng J H](#) [Introducing Web database](#)[期刊论文]-[Computer Science](#) 2001(10)
23. [Meng Xiao-feng](#) [Survey of Web data management](#) 2001
24. [Liu L;Pu C;Han W](#) [An XML-enabled data extraction toolkit for web sources](#)[外文期刊] 2001
25. [Abiteboul S;Buneman P;Suciu D](#) [Data on the Web from relations to semistructured data and XML](#) 2000
26. [DAmbrogio A;Iazeolla G](#) [A CORBA-Based approach to design gateways for multidatabase systems](#) 1997
27. [Bertino E;Catania B](#) [Integrating XML and databases](#) 2001

#### 引证文献(7条)

1. [胡迎松, 宁海霞](#) [一种新型的Web挖掘数据采集模型](#)[期刊论文]-[计算机工程与科学](#) 2007(2)
2. [李颖富, 杨宏, 苑津莎](#) [基于Web服务/XML的配电网信息发布集成研究](#)[期刊论文]-[电力系统通信](#) 2006(3)
3. [李徐](#) [基于Web服务及XML的数据交换系统的研究与实现](#)[学位论文]硕士 2006
4. [张志强](#) [上海网交会网管系统设计与实现](#)[学位论文]硕士 2006
5. [张素智, 李宏伟, 卢正鼎](#) [WEB数据集成系统中元数据的管理](#)[期刊论文]-[计算机工程与应用](#) 2005(21)
6. [陈丽娟, 陶莉](#) [基于Web服务的XML数据集成方案研究](#)[期刊论文]-[福建电脑](#) 2005(8)
7. [李颖富](#) [基于Web服务的配电网设备信息发布系统研究](#)[学位论文]硕士 2005

本文链接: [http://d.g.wanfangdata.com.cn/Periodical\\_xxwxjsjxt200310007.aspx](http://d.g.wanfangdata.com.cn/Periodical_xxwxjsjxt200310007.aspx)