

HTML 向 XML 转换的研究*

黄晓斌

(中山大学信息管理系 广州 510275)

【摘要】 分析 HTML 和 XML 的特点,论述了 HTML 向 XML 转换的必要性,介绍了转换的有关原理、方法和工具。

【关键词】 HTML XML 转换 网络信息组织 【分类号】 TP393

A Study on the Conversion from HTML to XML

Huang Xiaobin

(Department of Information Management, Zhongshan University, Guangzhou 510275, China)

【Abstract】 This paper analyses the characteristics of HTML and XML, discusses the significances, means and the tool of conversion from HTML to XML.

【Keywords】 HTML XML Conversion Network information organization

1 HTML 与 XML 的主要特点比较

HTML(超文本标记语言)是用于创建网页和进行信息发布的通用语言,它以纯文本形式存储,以标签定义文档的组织,提供跨平台的文档共享。在 HTML 文档中,可以嵌入其它对象,如电子表格、视频、音频以及各种应用程序等内容,通过统一资源定位器能实现 Web 节点间的超文本链接。HTML 的文档标签主要分为八类:格式控制、字体控制、加载图片、定义超文本链接、特殊字符、颜色背景控制、表格、交互表格标签等。HTML 的结构主要有两种:一种是超文本结构,依据此结构,逻辑上相关联的结构信息在物理上链接起来,另一种是由 HTML 文本特点决定的文本组织结构,通过 HTML 语言用不同的方法将数据组织在文本中。总的来说,HTML 有如下特点:格式和语法比较简单,易懂易学,将数据加上一些控制标记即可,即使没有编程经验也可以轻易地使用 HTML 来设计网页;而且 HTML 所有的控制标记都是固定的,数目也有限,提供的功能与相关属性的设置也是固定的,容易记忆;规定比较灵活,如控制标记中的英文标记大小写没有区别,另外,对控制标记中要有相对应的结束标记也没有严格的要求。HTML 的简洁性更适合低成本信息发布的需求;HTML 作为 Web 中共同的信息描述方式通用性较强,可以实现不同平台的文档共享;创建也比较灵活,HTML 文档是纯文本文件,可利用各种各样的编辑工具进行创建。HTML 的主要缺点是:

(1) 表现过于简单。HTML 文件将数据和数据的表现集中在一起,形式较为单调,尽管它能表达脚本、表格等功能,但很难表达复杂的形式。

(2) 链路容易断,链宿地址改变后,链源不能自动纠正。

(3) 检索时所花的时间较长,检索到的内容针对性较差,返回的结果较多。

(4) 扩展性差,HTML 的标记集合是固定的,不允许用户自行定义他们自己的标识。由于网络技术发展得非常快,不断有新的数据格式的文档上网,这就要求要有一种比较灵活的标签机制才能满足网络信息不断发展的要求,但 HTML 不允许用户根据需要来创建新的标记,更无法表示许多特殊行业的数据。

(5) 缺少语义性,HTML 是一种标记技术,不能很好地揭示信息内容的本质,计算机无法知道各段文本的确切含义。HTML 在设计上是用来展示内容和手工浏览网页的,不适合用作网络信息资源的自动化组织管理。HTML 是一种显示描述语言,通过标记来定义文档内容以什么样的形式呈现,它仅仅描述了 Web 浏览器应该如何在页面上布置文字、图形等,并不揭示所标记信息的具体含义,HTML 只是一种表现技术。由于 HTML 是在 Web 技术发展的早期制定的,因此 HTML 许多功能受到一定的限制。

XML(eXtensible Markup Language)是由万维网协会(W3C)设计,专门为 Web 应用服务的 SGML 的一个重要分支。XML 是一种元标示语言,它提供描述结构化资料的格式。XML 由若干规则组成,可用于创建标记语言。XML 以其良好的数据存储格式、可扩展性、高度结构化、便于网络传输等优势将在许多领域应用,便于网页信息组织,不仅能满足不断增长的网络应用需求,而且还能够确保在与网络进行交互时,具有良好的可靠性与互操作性。概括起来,XML 的主要特点有:

(1) 自描述性 XML 通常包含一个文档类型声明,因而 XML 文档是自描述的。XML 文档中的数据可以被任何能够对 XML 数据进行解析的应用所提取、分析和处理,并以所需格式显示。XML 表示数据的方式独立于应用系统,并且这些数据能重用。XML 中的标签是没有预先定义的,使用者要自定义所需的标志。XML 是能够进行自解释的语言,使用文档类型定义 DTD(Document Type Definition)来规定

收稿日期:2002-07-12

* 本文研究受到广州市社会科学界联合会的资助

• 18 •

数据,用XSL(eXtensible Style Sheet Language)来描述这些文档如何显示的机制。由于XML能够标记更多的信息,所以能使用户很容易地找到需要的信息。

(2)可扩展性 XML的主要特性体现在可扩展性。可扩展性是HTML的弱点,HTML被定义成一套固定的标签集,由于网页需要表达的内容越来越丰富而复杂,HTML规定的标签远不能满足页面设计的需要。为了解决HTML目前的限制并且考虑到未来发展的需要,要求页面标签具有可扩展性,能够创建新的标签。XML的标签是由DTD定义的,因而易于扩展,不像HTML的格式一成不变。

(3)开放性 XML作为一种可扩展性标记语言,其自描述性适用于不同应用间的数据交换。XML最大的优点是对数据描述和传送能力,因此具备很强的开放性。为了使基于XML的业务数据交换成为可能,就必须实现数据库的XML数据存取,并且将XML数据同应用程序集成,进而使它同现有的业务规则相结合。支持XML的数据库可以开发基于XML的动态应用如动态信息发布、动态数据交换等。XML提供描述不同类型数据的标准格式,提供了对多语种的支持,具有国际通用性。

(4)结构良好 XML建立在基本嵌套结构的基础之上,文档组织良好,数据高度结构化,结构性强,文件结构嵌套可以比较复杂,能表示面向对象的等级层次。XML文档是一种树型结构,可把某类中所有文档的共同属性,如标题、作者、段落、图片等抽象提取出来,定义成格式良好的DTD,把每个属性的内容放到XML的DTD相对应的节点中去,最后把XML文档存储到数据库中。这种方法可以解决半结构化和非结构化文档的统一存储、管理和浏览这一系列问题,而HTML不支持深层的结构描述。

(5)互操作性强 XML可以在多种平台上使用,可以用多种工具进行解释。XML支持用于字符编码的许多主要标准,交互性好。用户与应用进行交互时,使用XML可以非常方便地在本地客户端进行排序、过滤和其它的数据操作,不需要与服务器进行交互,减轻了服务器的负担。

(6)语义性强 HTML文档只是包括格式和结构的标记,而XML可以自行设计有意义的标记便于异构系统之间的数据交换和信息检索,实现机器与机器之间的信息交换。XML文档组织良好,数据高度结构化,标记对人和机器都可读。HTML依靠文档内容搜索,产生大量的无意义结果。利用XML可以很容易地按照XML定义的标签分类,HTML文档只是包括格式和结构的标记。而XML可以自行设计有意义的标记,便于异构系统之间的数据交换和信息检索。

(7)内容和表现相分离 XML提供了一种结构化的数据表示方式,使得用户界面与结构化数据分离。同一数据可用不同的形式展现给不同的使用者。HTML只是描述数据的外观,而XML描述数据本身。能够在结构化的数据中嵌套程序化的描述以表明如何显示数据,同样的数据集允许指定不同的显示方式,使数据更合理地表现出来,并且用多个查看方法,而不必向网页服务器发出另外的请求,减少了服务器的数据交换量和浏览器的响应时间。

(8)可粒状更新 XML不需要每次在改变数据后重组整个数据结构,减轻了Web服务器的工作量,增强了服务器的性能。由于用来描述数据结构的标记符的重复性而能够进行极好的压缩,提高传输大型数据的速度,XML也允许加进其他数据,所加入的信息能够进入存在的页面,不需要浏览器重新发一个新的页面。

总的来说,HTML和XML都是用来进行网络信息组织和传播的,都是用文本形式编写存储的,并且都是基于结构化

信息的国际标准。但HTML只说明数据看起来是什么样的,而XML则说明数据是什么意思。利用XML可以建立自己的标记,这些标记可以更准确地描述用户所要的东西,而HTML不能用来定义新的应用,这是XML与HTML的最大区别。XML突破了HTML的一些局限性,具有广阔的应用前景。

2 转换的必要性

目前大多数网页仍然主要是由HTML组成,HTML的固有缺点使原来的网络信息组织模式难以符合新的发展要求。由于XML的相关技术发展不断成熟,越来越多的网站逐渐用XML进行设计。在这个过程中既要新的内容以XML的方式存储和传播,同时也要考虑到兼容原来的数据,因而需要把现有的HTML网页转换成更灵活处理和应用的XML数据形式。同时,要把以往积累下来的HTML文档在新的环境下继续发挥作用,向XML转换无疑是一个解决办法。HTML向XML转换还有利于对网页信息进行集成、提取、检索、过滤或挖掘分析。

(1) 便于网络环境下异构系统的信息集成

在网络环境下,由于存在着异构的操作系统平台和数据库,导致信息交流和共享的困难。异构系统之间数据交换的办法之一是采用统一的信息交换格式。XML因其具有自定义性及其可扩展性等优势,便于表达各种类型的数据,可作为异构数据库之间的中间件,解决数据的统一接口问题,便于不同数据库之间的信息交换和共享。XML可以用来构造数据集成的中间层,将源数据转换成集成数据,简化集成系统的查询转换,为用户提供涉及多数据源的统一的查询机制,用统一的方式使用来自不同数据源的各种各样的数据,屏蔽各数据源在结构、运行环境上的差异。

(2) 便于网络信息的组织管理

随着网上信息的内容及应用快速增加,实现网络信息管理自动化是非常必要的。通过HTML进行用户交互操作则往往由于服务提供者改变了网站布局、增加了新的内容或改变了URL,使得浏览器出错,而更改需要花费大量的时间。XML则能通过浏览器提供综合服务,采用XML进行数据的描述和接口格式的定义,并进行目录等信息的交换和自动更新,以实现目录的自动分类处理。由于信息资源增长的速度,HTML在描述非结构化数据上缺乏足够的力量,而任何通过人工组织网络信息资源的做法都是不可行的。元数据是为了解决提高网络信息资源有序组织和整理的效率之间的矛盾而产生的,但元数据难以对所有的网络信息资源进行客观著录,而利用具有语义的XML标记数据,可以较大地增加它们在不同网络系统间交换的灵活性和被理解的程度,从而较好地解决网络信息组织管理问题。

(3) 便于网络信息的检索和过滤

由于XML标记明确地表达了它的涵义,搜索引擎就可以按关键词和内容之间的关系对特定的信息进行准确定位,从而能根据用户所提供的关键词,清楚地知道用户表达的语义而返回正确的结果。XML的语义结构化可将其作为结构化数据的交换标准,将改善信息检索的效果。HTML无法知道一篇文章的题名、著者、文摘、结论,而XML可以提供有关这些结构的信息。HTML要完成一个查询过程通常是比较复杂的,但有了XML就可以自动完成一系列的查询过程。XML使

得代理信息检索变得更加个性化。利用XML还可以对网络不良信息进行发现和定位,从而采取有效的过滤措施,净化网络环境。

(4) 便于实现网络数据挖掘和知识发现

从HTML文档中提取信息往往比较困难,而XML具有语义性,便于进行信息提取分析。XML能够使不同来源的数据很容易地结合在一起,使检索多个的不兼容的数据库成为可能,从而为解决网络的数据挖掘带来新的途径。XML的扩展性和灵活性允许描述不同种类应用软件中的数据,从而能搜集的网页中的数据记录。同时,由于XML的数据是自我描述的,其标签具有语义性,数据不需要有内部描述就能被交换和处理。因此,XML可看作是一种半结构化数据的模型,可以容易地将XML的文档描述与关系数据库中的属性对应起来,进行信息查询与抽取。随着XML作为在交换数据的一种标准方式的出现,网络的数据挖掘和知识发现将会变得容易实现。

3 转换的基本原理

要实现HTML数据向XML转换,关键是给出HTML的内容数据及其关系的一种组织方式,找出这种方式在XML模式中相应的表达规则,建立HTML标识到XML模式的一种映射,从而实现从HTML内容到XML结构的转换。由于HTML中既有用来表现特定意义的内容数据,又有用来呈现数据格式的数据,要把HTML自动有效地转换成XML数据,主要问题是实现内容数据和格式信息的分离。HTML文件中放置内容数据的有关项目主要有:页面的标头(head)、段落(p)、图像(img)、表单(form)、表格(table)以及多页面(frame)等。HTML文件中放置内容数据主要为文本、表格和表单等。表格的语法形式是<table>...</table>, <tr>用来定义表行, <th>用来定义表头, <td>用来定义表元。表头中原子性内容数据都在<th>和</th>之间,根据这些规则可以界定表格中的内容数据。基于对HTML的结构和语法分析,文献^[3]定义了一套标记规则用来提取HTML文档中的含有特定的内容数据,利用这些规则可在HTML的网页上对内容和呈现内容的格式实现进行分离,按照内容数据间本来的关系组织这些数据,标记规则也是XML模式直接构件依据。在标记规则作用下,HTML的源文件的标记分两个层次展开:一是根据页面不同部分的语义的组织关系,进行区域逻辑上的划分,二是在区域内对各种不同类型内容数据进行划分。因此,标记分为区域标记和逻辑标记,在形式上用HTML中的注释符号加上标记信息来表示。标记信息表达出不同类型内容的界定范围,其语法具有一定的层次性,区域标记是标记可变内容所在区域,分起始和结束标记;内容标记是标记特定区域内某一内容块的标记,分为文本、文本链、图像、图像链、表单、表格和多页面标记等。区域标记序号对HTML源文件中界定区域从头到尾用自然数标记,所有内容标记的内容块也编上记号,用区域标记序号和内容标记的序号界定区域内某一特定的内容块。自动转换系统一般由配置、转换和维护三个模块组成,配置模块的主要功能是按照标记规则对HTML的源文件进行区域和内容块的划分,并加上标记;转换模块基于标记规则,通过把带标记的HTML源文件向

XML的映射,转化为XML模式所规定的XML文档;维护模块基于对象接口模型操作XML文档。

4 转换的主要方法

4.1 HTML文档直接向XML的转换

XML是一种比较严格的标记语言,它要求文档中所有的语句都完全符合XML规范,只要XML文档有一点差错,XML应用程序就不能正确处理。因此要将HTML转换XML,首先第一部是给HTML加上XML声明,告诉客户端处理程序这是一个XML文档及有关的资源在什么地方;其次是检查HTML文档有没有语法错误,如果有的话应该严格按HTML的设计规范进行更正;再次是定义实体,与HTML不同的是,XML中所有的实体必须先定义后使用,如HTML文档用到的一些图形文件就是实体,在XML中必须先定义才能使用。

由于HTML网页数量庞大,进行人工转换是十分困难的。在HTML文档向XML文档的自动转换过程中,要求HTML文档结构良好,即所有元素的首尾标记必须配对,所有元素的嵌套层次结构必须正确,所有的属性值都以双引号(“”)的形式出现,所有自说明的元素要以</>结束。一般在转换过程中采用将HTML文档分析加工为DOM树的形式,在此基础上再作进一步的转换。转换过程主要解决的是HTML文档及其集合要表达的模式信息,可以通过程序自动进行抽取,并利用手工对HTML文本进一步加工,把遗漏的模式信息加以补充,从而形成一个完整的数据模式。根据抽取出来的模式,确定各对象的属性名和对象间的语义关系;指定HTML文档中待转换的信息的标记,并明确这些被转换的HTML文档信息与XML的文档标记的对应关系;根据这些关系,扫描HTML文档并输出相应的XML结果。

4.2 利用XHTML进行转换

XHTML采用了一种模块化方案,将现有的HTML模块化。它可以分成两个部分:

- ① 核心部分,包括所有访问设备必须支持的基本标记集;
- ② 基于核心的各种扩展模块。

XHTML文档遵循XML格式,因此可以用标准的XML工具浏览、编辑和检查;同时XHTML又继承了HTML的特点,适合现有的HTML的应用工具,而且还有更好的表现;因此可以说XHTML是现有的HTML通往结构化XML的桥梁。

要把HTML向XML迁移,可利用XHTML对HTML文档进行改装,XHTML结合了HTML和XML的优点,由于它和HTML比较接近,所以可以很容易地利用HTML进行改装和简化,形成新的XHTML文档,实现HTML向XML的过渡。在转换过程中应该注意几个问题:首先,XHTML对大小写标记敏感,因此,在HTML中用来增强文档可读性的有些技巧在XHTML中就不适用了,如在HTML中定义元素属性时使用大写字母,而具体的数值可使用小写,这样可读性

会强一些,但在 XHTML 中定义元素的属性必须使用小写。其次,XHTML 严格要求元素必须以标记开始和标记结束。在 HTML 中不是很严格地按照在每一段的开始使用<p>而在结尾使用</p>,但在 XHTML 中所有非空的元素都要求被关闭。再次,所有的 XHTML 属性都需要使用引号来表示,例如 HTML 中

4.3 利用智能代理进行转换

智能代理通常是指模拟人类行为,能够根据所感知的环境自主运行和提供相应服务的程序。在网络范畴内可以定义为在网络环境下代理用户或其他程序,以主动的方式完成有关操作的软件实体。智能代理能创建与实现目的有关的计划,定时和交互地执行,对网络环境变化做出反应。智能代理具有以代理性、自主性、智能性、交互性、机动性、适应性等特点,能够在用户不需要干预的情况下自行获取、分析和处理数据;这些技术特性在网络信息组织管理中得到充分应用。网络智能代理能够识别网页特征,可以自动检索有关内容,然后转换成相应的 XML 格式。文献^[4]介绍了美国的虚拟信息中心利用 VIPAR 系统把网上的 HTML 新闻报纸信息转换为 XML。通过检索智能代理在网上的新闻信息资源中查找有关的 RDF,包括元数据、路径、文章标识和结构规则等,然后根据文章的结构用 XML 标签表示文章的标题、作者、日期、位置、地点等元素,把非结构的 HTML 自动转换为包括许多计算机能够识别处理信息的结构化 XML,并在此基础上能够通过 XML 标签进行内容检索。

5 常见的转换工具

目前,许多部门在考虑开发能够将 HTML 自动转换为 XML 的应用软件。根据网络信息资源的检索,现有如下一些常见的转换工具:

(1) HTML Tidy (<http://www.w3.org/People/Raggett/tidy/>)

HTML Tidy 可以运行在多种平台上,用来检查并清除 HTML 文件的标记错误,并能重新格式化 HTML 文件,使之有更好的可读性。HTML Tidy 现已成为一个能把 HTML 向 XHTML 转换的多功能工具。

(2) HTML-Kit (<http://www.chami.com/html-kit/>)

HTML-Kit 是一个免费的程序,能在许多平台上运行。它不仅能帮助进行 HTML 的编辑、格式化输出、合法性检查、预览和发布,而且还能在图形界面上进行 HTML 向 XHTML 的转换。在其用户界面上,有一个窗口显示源文件,另一个窗口显示转换后的标记语言结果,还有一个窗口显示出现的错误以及对 XHTML 进行改进的建议。

(3) Xspltit (<http://www.percussion.com/XMLzone/technology.htm>)

Xspltit 是由 Percussion Software 公司推出的新工具。Xspltit 让 Web 开发者能把 HTML 文档转换成相应的 XSL 样式表。Xspltit 能创建一个包含格式定义的 XML 的 DTD(文件类型定义)文件,还能使用静态的内容来创建一个 XML 样本文档。

(4) 网上 XML 的转换工具 (<http://www.html2XML.com/Html2XMLIntro.asp>)

这是一个网上的转换工具,可以对网络 HTML 格式资源的 URL 或上传一个 HTML 文档进行转换为 XML,输出 XML 文档格式,通过 Internet Explorer 5.0 (或更高版本)内嵌的样式表进行显示,还可以下载 XML2html.xsl 工具把 XML 文档转换回 HTML 文档。

(5) HTML2XML (<http://www.infoncall.com>)

HTML2XML 是因科技术(上海)有限公司提供的一套用来将 HTML 文档自动转换为 XML 文档的工具。目前主要针对以表格数据为核心的 HTML 格式文件。该工具功能主要有:提供基于 XML 来表达如何从 HTML 网页获取复杂结构;HTML 到 XML 声明性文档的映射,可以根据相应的解析模板自动产生 XML;提供可视化工具使得开发更加迅速和便捷。HTML2XML 包括两个工具:HTML2XML 模板生成器和 HTML2XML 解析引擎。通过 HTML2XML 模板生成器的图形用户界面和 HTML2XML 解析引擎,可以将 HTML 文件中的<Table>标记中的数据根据指导性文件,转换成 XML 格式数据,供其他应用程序进行进一步处理。HTML2XML 模板生成器提供方便的用户界面,HTML 网页内容编辑人员选定所需要的 HTML 内容后,以可视化的图形方式用鼠标进行拖拉操作便可完成对 HTML 内容的获取,用户不必了解所编辑 HTML 文件的源代码,当保存编辑结果后,即可生成该类 HTML 文件的解析模板和解析结果。HTML2XML 解析引擎支持 Service 和 API 两种用户界面。Service 界面不需要用户有较深的编程经验,API 界面为开发人员提供更灵活的编程接口。模板使用人员在开发具体应用时,通过解析引擎装载不同模板解析得到相应的结果,解析结果返回 XML 格式的字符串和保存为指定文件,以供进一步处理。若模板发生错误或未指定模板,解析引擎则按无模板的方式进行处理。此时解析引擎解析所有 Table 中的数据到 XML 文件中。HTML2XML 既可以将 HTML 转换成的 XHTML,也可以将其转换为独立于应用的 XML 通用格式,然后通过 XSL 进行发布。

参考文献:

- [1] 吴敏. XML 的研究现状及展望. 微型电脑应用, 2001, 17(4): 5-9
- [2] 孟小峰. Web 数据管理研究综述. 计算机研究与发展, 2001, 38(4): 385-3
- [3] 李青山. 一种基于内容的 HTML 到 XML 转换策略. 计算机工程与应用, 2001, (9): 37-32
- [4] 陈维义. HTML 如何向 XML 迁移. <http://www.people.com.cn/GB/channel5/29/20000802>
- [5] 董慧等. 网络信息资源开发与利用. 武汉: 武汉大学出版社, 2001
- [6] Thomas E. Potok. An ontology-based HTML to XML conversion using intelligent agents. <http://www.computer.org/proceedings/hicss/1435/volume4/14350120babs.htm>

作者: [黄晓斌](#)
作者单位: [中山大学信息管理系, 广州, 510275](#)
刊名: [现代图书情报技术](#) PKU CSSCI
英文刊名: [NEW TECHNOLOGY OF LIBRARY AND INFORMATION SERVICE](#)
年, 卷(期): 2003(1)
被引用次数: 9次

参考文献(6条)

1. 吴敏 [XML的研究现状及展望](#)[期刊论文]-[微型电脑应用](#) 2001(04)
2. 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)
3. 李青山 [一种基于内容的HTML到XML转换策略](#)[期刊论文]-[计算机工程与应用](#) 2001(09)
4. 陈维义 [HTML如何向XML迁移](#)
5. 董慧 [网络信息资源开发与利用](#) 2001
6. Thomas E Potok [An ontology-based HTML to XML conversion using intelligent agents](#)

本文读者也读过(10条)

1. 张文斌. 陈恩红. 王进 [一种基于多叉树的HTML到XML的转换方法](#)[期刊论文]-[小型微型计算机系统](#)2003, 24(4)
2. 黄伟. 刘娟. HUANG Wei. LIU Juan [一种基于DOM树的HTML转换为XML的方法](#)[期刊论文]-[电脑知识与技术 \(学术交流\)](#) 2006(7)
3. 曹风华. CAO Feng-hua [XSLT在XML向HTML转换中的作用](#)[期刊论文]-[现代计算机 \(专业版\)](#) 2010(3)
4. 方睿. 韩斌. 陈灵 [WordML文档转换器研究](#)[期刊论文]-[计算机应用](#)2006, 26(z1)
5. 李雪竹. LI Xue-zhu [一种基于XML的Web数据抽取的实现](#)[期刊论文]-[科学技术与工程](#)2008, 8(9)
6. 贾长云. 程永上. JIA Chang-yun. CHENG Yong-shang [HTML表格向XML的智能转换](#)[期刊论文]-[计算机工程](#) 2009, 35(14)
7. 陈艳梅. 张斌 [HTML到XML转换技术的研究与实现](#)[期刊论文]-[现代图书情报技术](#)2003(5)
8. 李昕. 李丽萍. 常革新. LI Xin. LI Li-ping. CHAN Gge-xin [基于XML的文档的动态产生](#)[期刊论文]-[辽宁工程技术大学学报](#)2006, 25(1)
9. 王富强. 王默玉. WANG Fu-qiang. WANG Mo-yu [数据岛在HTML的嵌入显示](#)[期刊论文]-[电脑与信息技术](#)2005, 13(1)
10. 张燕. 贾焰. 黄晓斌. 周斌. 顾剑. ZHANG Yan. JIA Yan. HUANG Xiao-bin. ZHOU Bin. GU Jian [一种基于本体的网格服务匹配方法](#)[期刊论文]-[计算机科学](#)2008, 35(6)

引证文献(9条)

1. 霍江 [HTML文档的预处理](#)[期刊论文]-[电脑知识与技术 \(学术交流\)](#) 2007(21)
2. 黄晓斌 [网络主题图的原理及应用](#)[期刊论文]-[情报理论与实践](#) 2004(3)
3. 涂颖哲 [图书馆网站的标准化问题研究](#)[期刊论文]-[科技信息](#) 2008(19)
4. 陈勇 [基于XML的地图网络发布技术的研究](#)[学位论文]硕士 2004
5. 刘江宏. 刘金瑄 [一种基于Web的HTML到XML数据转换方法](#)[期刊论文]-[计算机与数字工程](#) 2009(1)
6. 张弘. 张新平. 吴祖立 [也谈网上审稿](#)[期刊论文]-[宁波大学学报\(教育科学版\)](#) 2004(4)
7. 李姝 [基于XML的Web数据挖掘研究](#)[学位论文]硕士 2007
8. 吴治宗 [基于XML的文档处理技术的研究与实现](#)[学位论文]硕士 2006
9. 王娜 [XML在信息检索系统中的应用及评价研究](#)[学位论文]硕士 2005

本文链接: http://d.g.wanfangdata.com.cn/Periodical_xdtsqbjs200301007.aspx