

二十一世紀圖書館學叢書

任誌愈題

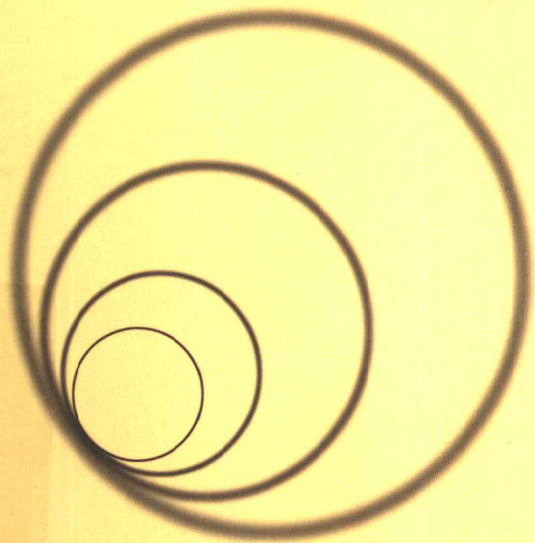
SERIES OF LIBRARY SCIENCE IN THE 21 CENTURY

丛书主编 丘东江

未来网络的基磐技术 — XML 的理论与应用

监修 孙承莹

著者 石塚英弘 李颖



华艺出版社

未来网络的基磐技术

——XML 的理论与应用

**BASIC TECHNOLOGY OF NETWORK IN THE FUTURE:
XML: THEORY AND APPLICATION**

监修 孙承鉴

著者 石塚英弘 李颖

华艺出版社

图书在版编目 (CIP) 数据

未来网络的基磐技术——XML 的理论与应用 / 石塚英弘 李颖著、孙承鉴监修
北京：华艺出版社，2002.1

ISBN 7-80142-368-2

I. 未... II. 石... III. 计算机网络-可扩充语言 IV. TP312
中国版本图书馆 CIP 数据核字 (2001) 第 063950 号

未来网络的基磐技术——XML 的理论与应用

华艺出版社

北京市海淀区北四环中路 229 号 10 层

邮编：100083 电话：82885151 转

北京市后沙峪印刷厂

850×1168 毫米 1/32 11.3 印张 168 千字

2002 年 1 月第 1 版 2002 年 1 月北京第 1 次印刷

定价：28.00 元

ISBN 7-80142-368-2 / Z · 186

《21世纪图书馆学丛书》

丛书题名

任继愈 国家图书馆馆长

顾 问

(以姓氏笔画为序)

马远良 上海图书馆馆长、研究员

马费成 武汉大学信息管理学院院长、教授、博士生导师

卢子博 南京图书馆原副馆长、研究馆员

卢克斯 (Claudia Lux) 博士 国际图联管理委员会委员、德国中
央暨州立图书馆馆长

孙承鉴 国家图书馆原副馆长、研究员

孙蓓欣 国家图书馆副馆长、研究馆员

吴建中博士 国际图联管理委员会委员、上海图书馆副馆长、研
究馆员

吴慰慈 北京大学信息管理系主任、教授、博士生导师

李华伟 (Hwa-Wei Lee) 博士 美国俄亥俄大学荣誉馆长、教授、
OCLC 杰出访问学者

李炳穆 (Byung Mock Rhee) 博士 韩国延世大学文献情报学科教授

孟广均 中国科学院文献情报中心研究员、博士生导师

张晓林博士 四川大学信息管理系主任、教授

倪波 南京大学信息管理系教授、博士生导师

倪晓健 首都图书馆馆长、教授

徐引箴 中国科学院文献情报中心主任、研究员、博士生导师

曾程双修 (Sally C. Tseng) 美国加州大学欧文图书馆期刊编目
主任、美国图书馆协会英美编目及元数据会议主任、全美华
人图书馆员协会执行理事长

程焕文 中山大学网络教育学院院长、图书馆馆长、教授

詹德优 武汉大学信息管理学院教授、博士生导师

潘寅生 甘肃省图书馆馆长、研究馆员

总序

21世纪是全球信息化的世纪,是知识经济占主导地位的世纪,是一个充满希望的世纪。它将对图书馆学情报学研究提出新的更高的要求。

丘东江同志主编的《21世纪图书馆学丛书》旨在顺应时代潮流,把握知识和网络时代图书馆学情报学发展的脉络,强调精品意识和求实创新的学术风格,全力推出理论与实践相结合的最新研究著述。

《21世纪图书馆学丛书》由国内外图书馆学专家学者合力而成,并特邀著名人士担任顾问;国家图书馆馆长任继愈为该丛书题写书名;国际图联执委、OCLC 总裁分别为部分著作题词、作序;国际图联秘书长和 UBCIM 计划主席和国外著名专家授予翻译权。

《21世纪图书馆学丛书》以探讨图书馆学情报学最新理论与实践为主要内容,其中包括:《知识管理的理论与实践》、《国际图联(IFLA)与中国图书馆事业》(上、下)、《元数据导论》、《未来网络的基磐技术——XML 的理论与应用》、《Z39.50 的原理、应用与技术》、《OCLC——全球在线计算机图书馆中心》、《虚拟图书馆的昨天、今天和明天》、《国际标准书目著录(ISBD)》(上、下)、和《国外图书馆学重要著作选译》。

我希望这套丛书的问世,对我国图书馆事业的发展,提高图书馆学研究水平有所裨益。

吴慰慈

2001年7月30日

前 言

现代生活离不开因特网，而在网络上存在的实体，均为数字化资源。因此有关数字资源的记述标准堪称网络的基本技术。目前为止，由于 HTML 的简洁性，该技术被广泛地应用。随着网络的普及，人们通过网络进行的活动不断丰富，如，电子商务、数字图书馆及其它较为复杂的网上信息系统。为实现各种各样的网上需求，现有的 HTML 技术已显得力不从心。由此，产生了可扩展的数字资源记述语言，即 XML 标准。

自 1998 年 2 月，万维网联盟 W3C 首次推出 XML 标准以来，该标准就受到举世瞩目。XML 的应用扩展到网络的各个领域，其势有增无减。为使基于 XML 标准的数字资源更好地适用于社会经济结构，与 XML 标准有关的新标准化课题也在持续发展。目前，议论的焦点从作为网络信息交换的表记标准 XML 的研讨，开始向利用 XML 文件如何进行信息交换的标准行为（即解决方案）转移。事实上，象企业间数据无国境共享的标准化活动、电子商务信息的国际标准（ebXML）等正成为热点。也就是说，当今世界对 XML 的期待，已大大超出了单纯的文献数字化标准的范围，它涉及到网络上全部行为，如电子商务、数字图书馆、远程教育和医疗等等。由于利用了 XML 技术，使得网上的信息系统的构筑简单易行。XML 文件的再

利用性和交换性使得信息系统的开发的成本不断降低。XML正起着21世纪高度信息化社会的信息交换/信息共享的基磐作用。

随着全球化的进展，网络的关联技术被看成世界一体化的标志。XML作为下一代因特网的标准格式，在各国不断地被研讨和应用，基于XML的解决方案层出不穷。欧美及日本已进入适用化阶段。

本书是我于1999年9月赴日本筑波参加在图书情报大学召开的数字图书馆国际研究会期间，有幸认识了本书的作者中国留学生李颖博士生及其导师石塚英弘教授。他们对XML（可扩展置标语言）理论及实践方面有很深的造诣和丰富的经验。为了在我国推广普及XML技术，加速网上信息资源的组织和共享，我协助他们撰写了“未来网络的基磐技术”这部书。本书通过大量的实例力求对XML理论进行通俗易懂的解说。同时，也列举了XML的应用。由于XML技术在不断的发展，本书的最后，对其作了展望。因篇幅关系，我们不可能囊括所有的XML话题，有兴趣的读者可通过附录所提供的线索，进一步查阅相关资料。

孙承鉴

2001年4月于北京

寄 言

XML の誕生

今、HTML に代わる次世代の Web 用言語として XML (eXtensible Markup Language) が注目されている。XML の適用範囲は広く、インターネット上で情報を送り、操作し、利用する分野であればどの分野でも使うことができる。適用が注目されている分野の例を挙げれば、たとえば、デジタル図書館、電子出版、電子商取引などがある。

なぜ HTML ではなく、XML なのか？ その理由は、HTML の限界を超えるものとして XML が考案されたからである。Web の本家であり、HTML の規格を策定した W3C (World Wide Web Consortium) 自体が HTML の限界を認識し、その限界を打破するものとして XML の規格を策定した。XML は Web 用言語としての HTML の長所を継承した。また、HTML の欠点を解決するために機能を根本から拡張した。

HTML では決められたタグしか使えないが、XML では必要に応じて自由にタグを設定し、それを使って情報を表現できる。送信側はタグを使ってデータの内容を知らせることができるし、受信側はタグを使ってデ

一夕内容に合わせたデータ処理が可能である。一方、HTML の場合は受信側は表示のみでデータ処理ができなかった。XML はこれらの欠点を解消し、多くの分野で従来に無かった新しいシステムの構築を可能にした。これが多くの分野で XML が注目されている理由である。

この本では、XML の基本を分かりやすく解説し、その応用を述べ、将来を展望する。応用では特にデジタル図書館と電子商取引について解説する。インターネット上の Web (World Wide Web) は、それを見た人に新しい巨大な共用情報資源の存在を実感させた。事実、アメリカではインターネット上の図書館システムをデジタル図書館と呼び、その研究開発を進めている。HTML で始まったデジタル図書館は XML の出現によって更に発展すると期待される。また、電子商取引は現在多くの期待を集めている分野である。

この本が一人でも多くの人に XML の長所を知らせ、XML を適用した新しいシステムの開発に寄与できれば幸いである。

石塚英弘 2001年3月 日本、筑波研究学園都市にて

XML 的诞生

(译文)

当今,作为 HTML 替代物的下一代 Web(Would Wide Web)用语言 XML(eXtensible Markup Language)正日益受到瞩目。XML 的适用范围非常广泛,诸如因特网上信息传递、操作等,凡有信息利用的领域,就可应用 XML。而应用 XML 的领域,象数字图书馆、电子出版、电子商务等正倍受青睐。

为何是 XML,而非 HTML?因为 XML 超越了 HTML 所持有的限制。Web 的发祥机构 W3C(World Wide Web Consortium)自身意识到 HTML 的界限,为打破该界限,策划了 XML 标准。XML 继承了 Web 用语言 HTML 长处,同时,为解决 HTML 的问题,从根本上对其机能进行了扩张。

HTML 只能使用预先规定的标签,XML 却可依需要自由设定,并利用这些标签来描述信息。信息发送侧通过标签传达所送数据内容,接受侧则通过标签来处理与标签吻合的数据内容。对于 HTML,接受侧对数据进行的唯一的处理是在浏览器上显示它,而不能对其做任何更多的处理。利用 XML,在许多学科领域均可构筑前所未有的新系统。这也是 XML 在众多的领域受到注目的缘由。

本书对 XML 的基础进行了简明易懂的解说,并阐述了其相关应用,同时也对 XML 的未来进行了展望。作为 XML 的应用,对数字图书馆及电子商务做了特别的阐述。

让浏览 Web 的人实感其上新型的、巨大的共享信息资源的存在。在美国，因特网上的图书馆系统被称为数字图书馆，其研究开发不断被推进。以 HTML 开始的数字图书馆会因 XML 的出现而被期待得到更进一步的发展。与此同时，电子商务也正成为众目期待的领域。

本书寄期待于将 XML 的长处传达与众多的中国友人，哪怕只是一人，如能运用 XML 开发新系统，将是我的荣幸。

石塚英弘 2000 年 3 月于日本筑波研究学园都市

目 录

第一章 置标语言的变迁

- 1.1 从 SGML、HTML、到 XML.....1
- 1.2 SGML、HTML、XML 三者的关系.....5

第二章 置标语言的定义与 SGML

- 2.1 置标语言(Markup Language)的定义.....9
- 2.2 标准通用置标语言(SGML: Standard Generalized Markup Language).....10
 - 2.2.1 SGML 概念的由来.....10
 - 2.2.2 SGML 产生的背景—从纸张到 SGML.....11
 - 2.2.3 SGML 的发展历程.....17
 - 2.2.4 SGML 文档的构成.....19
- 2.3 SGML 的应用.....23
 - 2.3.1 美国及加拿大.....23
 - 2.3.2 欧洲.....25
 - 2.3.3 亚洲.....26
- 2.4 SGML 的利弊.....27

第三章 超文本置标语言 HTML

3.1 HTML 的定义.....	29
3.2 HTML 与 World Wide Web.....	32
3.3 HTML 局限性与 XML 的诞生.....	37

第四章 XML（可扩展的置标语言）的概要

4.1 XML 产生的背景.....	46
4.2 XML 与 SGML.....	51
4.3 XML 与 HTML.....	53

第五章 XML 的逻辑构造和物理构造

5.1 XML 的思维方法.....	55
5.1.1 XML 的思维方法.....	55
5.1.2 利用 DTD 定义文件的构造.....	60
5.1.2.1 文件要素的置标记和 DTD.....	61
5.1.2.2 属性和 DTD.....	63
5.1.3 XML 与 Web 的对应.....	63
5.2 XML 文件整体逻辑构造.....	65
5.2.1 XML 宣言.....	66
5.2.2 DTD.....	67
5.2.3 名字空间.....	77

- 5.2.4 层次构造 XML 的实例·····80
 - 5.2.4.1 合法的 XML 文档·····81
 - 5.2.4.2 结构完整的 XML 文档·····84
- 5.3 XML 的物理构造 (XML 的 Source 文件) ·····85

第六章 XML 的文档

- 6.1 XML 文档的种类·····87
 - 6.1.1 有效 XML 文档·····87
 - 6.1.2 格式完整的 XML 文档·····89
- 6.2 XML 宣言的文法解析/如何进行 XML 宣言·····91
 - 6.2.1 XML 版本宣言·····91
 - 6.2.2 XML 文字编码宣言(encoding declaration)·····92
- 6.3 DTD 的文法·····93
 - 6.3.1 何谓 DTD·····93
 - 6.3.2 要素型宣言·····95
 - 6.3.2.1 “父子”关系的表现·····96
 - 6.3.2.2 内容模式的指定·····98
 - 6.3.2.3 要素内容·····99
 - 6.3.2.3.1 出现顺序的表现·····99
 - 6.3.2.3.2 出现次数的表现·····103
 - 6.3.2.3.3 出现顺序、出现次数组合来表现
要素内容·····107

6.3.2.4	混合内容	111
6.3.2.5	空要素	113
6.3.2.6	任意要素 (ANY)	114
6.3.3	属性一览宣言	115
6.3.3.1	属性一览宣言的记述	116
6.3.3.2	依数据类型定义属性值的侯补	117
6.3.3.3	依列举型定义属性值的侯补	125
6.3.3.4	侯补值的定义	125
6.3.4	实体宣言	129
6.3.4.1	XML 中实体的定义与实体的参照	129
6.3.4.2	实体的宣言	134
6.3.4.3	解析对象实体与解析对象外实体	138
6.3.5	记法宣言	138
6.3.5.1	实体宣言中的记法名的指定	139
6.3.5.2	作为属性的记法	141
6.3.5.3	处理命令目标的记法宣言	142
6.3.6	DTD 的指定方法	142
6.3.6.1	DOCTYPE 宣言	142
6.3.6.2	DTD 外部子集的指定	143
6.3.6.3	DTD 内部子集的记述	144
6.3.6.4	外部子集和内部子集的并用	145
6.3.7	DTD 中的注释语句	146

6.4 置标 XML 文档(XML 的实现值).....	146
6.4.1 XML 的实现值的概要.....	146
6.4.2 起始和结束标签.....	146
6.4.3 属性的指定.....	152
6.4.4 对文件加标签.....	153
6.4.5 表示空要素的标签.....	155
6.4.6 文字的置换.....	156
6.4.7 XML 文档可用的文字.....	160
6.4.8 XML 中文字的大写与小写的区别.....	161
6.4.9 注释.....	163
6.4.10 CDATA section.....	163
6.4.11 处理命令.....	168

第七章 XLink 与 Xpointer

7.1 HTML 的链接功能.....	170
7.2 XLink 的产生及其扩张链接功能.....	172
7.2.1 复数链接.....	173
7.2.2 双向的链接.....	180
7.2.3 各种处理的指定.....	182
7.2.3.1 表示的指定.....	182
7.2.3.2 含义属性-role 和 title.....	186
7.3 链接的指定方法(链接的文法).....	186