

XHTML——一种可扩展的超文本标记语言

XHTML—An Extensible Hyper Text Markup Language

16-18

王仲

TP312XH

TP393.4

(北京大学国家文字信息处理技术实验室 北京 100871)

Abstract A new specification used for Web browsing is presented in this paper. The characteristics and advantages of the specification are analyzed in detail in the view of technique. A typical example is demonstrated to reinforce the specification. Some opinion on the development of Web browsing in future is given at the end of the paper.

Keywords XHTML, HTML, XML, Web technique, Internet/Intranet

XHTML 最早叫做“HTML in XML”，是一种基于 XML 的超文本标记语言，也就是说，将以前用 SGML 定义的 HTML 改为用 XML 重新定义。从这个意义上讲，XHTML 实际上是 XML 的一种应用。表面上看，在 XHTML 中，其标记仍旧沿用以往的 HTML 的标记，各元素和属性的名称和用法也几乎不变，只不过由于改为用 XML 定义，就必须严格按照 XML 语言规范来处理了。提起 XHTML，就不能不先来看一下 XHTML 的发展历史。

1 XHTML 历史

XHTML 的前身是大名鼎鼎的 HTML，HTML 作为在 Internet/Intranet 网上进行数据浏览和数据交换的一种先进的文件格式规范，以浏览器为依托凭借着其丰富而强大的功能打破了传统的数据浏览方式，给人以耳目一新的感觉，也极大地推动了 Internet 的发展。但是，随着 HTML 应用的不断深入，其不足之处也逐渐暴露出来。首先，实际应用中的 HTML 相当不规范。虽然 W3C 早已制定出 HTML 的规范，但是出于商业动机，当今两大 Web 浏览器软件制作厂商—Microsoft 和 Netscape—为了抢占日益扩大的浏览器市场，纷纷在 HTML 中加入自己定义的非规范的标记，并在自己的浏览器中提供相应的特殊支持，旨在占领更大的浏览器市场份额，以达到左右 HTML 标准的制定并最终将竞争对手打垮的目的。这种不良竞争的结果，造成 HTML 标记的过度膨胀，大量与标准 HTML 不一致的标记充斥于 Internet 的大小角落，而且某些 HTML 标识符只能在 Netscape Navigator 上才能执行，而另一些只在 Internet Explorer 上才是有效的。这种情况给网页设计人员带来很大的麻烦。其次，随着技

术的飞速发展，新的 Web 浏览工具不断被开发出来，从小巧的掌上电脑到功能特异的能上网的手机再到最近被炒得沸沸扬扬的数字机顶盒，网上浏览再也不必局限于桌上电脑了。据权威人士预测，到 2002 年，大约 75% 的网页将会通过非桌上型电脑来浏览^[5]。但是，我们不仅又问，到底现在的 HTML 标准是不是真得适合在各种非桌上型电脑中进行浏览呢？很不幸，回答是否定的，现在的 HTML 不能解决这个难题。另外，HTML 只能利用现有的有限标记，而且大多数标记都没有实际的含义，这使得它越来越不适应人们不断增加的需求。在这种情况下，在各大知名厂商的压力下，W3C 组织于 1998 年底成立工作组，专门讨论 XHTML 标准的制定。好在此之前，W3C 已正式推出 XML1.0 规范，这无疑为 XHTML 规范的制定奠定了强大的基础。经过不断的讨论，该工作组于 1998 年 12 月 5 日提出第一个草案^[1]，并在随后次年的 2 月 24 日^[2]、3 月 4 日^[3]、5 月 5 日^[4]分别公布了阶段性草案。1999 年 8 月 24 日，第一个建议标准^[5]正式推出。不过，还应指出的是，这并不意味着就一定成为正式标准，事实上，XHTML1.0 的这份标准目前正处于激烈的讨论中，争议的焦点之一是名称空间(namespace)的解释问题，不过，现在总算有了一线希望，相信出路就在前方。

2 XHTML 的优点

XHTML 之所以能成为建议标准，就是因为它具有一系列优点：

可伸缩性 XHTML 可以根据用户的需要和浏览器的处理能力选用合适的 DTD，单是从这一点来看，就充分体现了一种时下比较流行的理念——“科技以人为本”。XHTML1.0 的标准中定义了三个 DTD，它

们分别是:严格式(Strict)http://www.w3.org/TR/xhtml/strict、过渡式(Transitional)http://www.w3.org/TR/xhtml1/transitional、分帧式(Frameset)http://www.w3.org/TR/xhtml1/frameset,每一种都有不同的应用。

可扩展性 正如 XHTML 字面上所体现的那样,XHTML 具有一定的扩展性,这要归功于名称空间(namespace)的引入。有了名称空间,用户可以在 XHTML 文档中任意添加自己需要的标记,两个比较典型的例子是:MathML 和 SVG,MathML 和 SVG 作为 XML 的一种具体应用,可以灵活地嵌入 XHTML 文件中,将使得 XHTML 增色不少。现在,前者已经成为 W3C 正式推荐标准,而后者也即将成为建议推荐标准。

可重用性 XHTML 的推出原本是作为 XML 的过渡标准,因为 XML 尚处于一个不断完善阶段,还有大量的工作需要去做,因此短时间内将很难推出一个完全支持 XML 的浏览器,即使声称支持 XML 的 Internet Explorer 5.0 目前也只是提供部分的支持,而在这过渡期内,运用 XHTML 完全可以达到既定目标。另一方面,XHTML 从根本上来讲,也是一种规范的 XML,即使将来 XML 彻底替代了 XHTML,XHTML 文档也不会为 XML 浏览器或其它基于 XML 的工具所不容。

3 XHTML 的规则

在讨论 XHTML 规则之前,我们先来看一个典型示例。

```
<? xml version="1.0" encoding="GB2312"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML1.0Frameset//EN"
"http://www.w3.org/TR/xhtml1/DTD/strict.dtd">
<html xmlns="http://www.w3.org/TR/html">
<head>
<title>一个典型的 XHTML 示例</title>
<meta http-equiv="Content-Type" content="text/html; charset="GB2312"? />
<script type="text/JavaScript">
<![CDATA[
function count(n){
if(n<=0)
return"True";
else
return"False";
}
]}>
</script>
</head>
<body>
<p>一个典型的 XHTML 示例</p>
<hr />
<form action="mailto:wzb@bd748.pku.edu.cn" method="post">
<input type="button" name="Submit" value="submit" />
</form>
</body>
</html>
```

前面提到,XHTML 实际上是 XML 的一种应用,

因此,某些在 HTML4.0 中规范的标记必须加以修改才能符合 XML 的约定。具体体现在以下几个方面。

文档必须是规则的 文档的规则性是 XML 中的一个全新概念,其基本要求是:元素必须有结束标记或以特殊的结束标记,并且标记间必须严格嵌套。下面是两个对照例子:

错误: <p>This is an emphasizedparagraph.</p>

正确: <p>This is an emphasizedparagraph.</p>

元素和属性名必须小写 XHTML 要求所有的元素和属性都必须以小写的形式出现,这是因为 XML 是大小写敏感的。比如:<tr>和<TR>就是两个不同的标记。

结束标记不可省略 在 HTML 中,有些标记不加结束标记都是无关紧要的,浏览器会自动处理这种忽略。但是,这种情况在 XHTML 中是不允许的。下面是两个对照例子:

错误: <p>This is a paragraph.<p>here is another paragraph.

正确: <p>This is a paragraph.</p><p>here is another paragraph.</p>

属性必须用引号括起来 所有的属性必须用引号括起来,即使那些是数值的属性。下面是两个对照例子:

错误: <table rows=3>

正确: <table rows="3">

属性不可最小化 在 HTML 中,属性允许最小化,比如有些属性: selected、checked、compact 在值为 TRUE 时可直接放入标记中,但这在 XHTML 中要求必须将属性表达书写完整。下面是两个对照例子:

错误: <option selected>

正确: <option selected="selected">

特殊的空元素 在 XHTML 中,定义为空的元素可以有两种表达方式,即或者用结束标记(象 HTML 中那样)或者在起始标记后">"前加"/"。下面是两个对照例子:

错误:
<hr>

正确:
</br> <hr/></hr>
</br>

属性值中的空格处理 在 XHTML 中,解析器会将属性值中的前置和后置的空格过滤掉,将中间所有的空格(包括换行符)缩为一个空格,除非设置 xml:space="preserved"。

Script 和 Style 元素置于 CDATA 中 在 XHTML 中,Script 和 Style 元素被定义为 #PCDATA,因此为避免 Script 或 Style 中的 "<" 和 ">" 被解析

器误认,必须将 Script 和 Style 置于 CDATA 中,下面给出示例:

```

<script>
  ! [CDATA]
  ...unescaped script content...
</script>

```

使用正确的元素和名称空间 XHTML 标准中规定,文件的根元素标记必须是<html>,而且要用 xmlns 这个特别的属性来注明 XHTML 专属的名称空间,如:

```
<html xmlns=http://www.w3.org/TR/html">
```

严格声明 head 和 title 在 HTML 中<head>和<title>可以省略,但在 XHTML 中,一律得按规则加在适当的位置,另外,<title>必须是<head>中第一个出现的元素。

输出类型 经过验证的 XHTML 文件是一个真正的 XML 文件,因此理论上讲可以为任何兼容的 XML 浏览器或软件所理解,当然也可以输出成 HTML 文件,在 HTML 浏览器中显示,要输出成 HTML 类型,必须在 meta http-equiv = "Content-Type" 的类型设为 "text/html"; 同样,要输出成 XML 类型,必须在 meta http-equiv = "Content-Type" 的类型设为 "text/xml".

4 XHTML 及 Web 浏览语言展望

为了很好地发挥 XHTML 的优点,满足各种不同的需要,还有许多工作要做。目前,W3C 正在紧锣密鼓地进行 XHTML 相关标准的制定工作。有关的标准有:HTML 的模块化、子集与扩展性、文档配置文件(Document Profiles)。届时,随着 XHTML1.1 规范的推出,XHTML 文档的编辑者和浏览器能按实际需要去套用不同组合的 DTD。那么,今后 Web 浏览将全采用什么语言呢?在这里,我们不得不提到 XML。XML 是一种目前在业内广为流行的新兴标记语言,且大有成为 Web 浏览技术上的明日之星的趋势,但是不是说 XML 马上就可成为取代 HTML 的后继者呢?不是的,

XML 的确有许多优势(详细论述参见参考文献中提到的文章),但它仍存在一些短时间内无法克服的缺点。XML 还有很长的路要走,而 XHTML 做为对 HTML 的扩展,其优势使其成为 HTML 的理想继承人,现在已经有针对 XHTML 应用的支持,MathML 和 SVG 就是很好的例证,因此,我们预测在不久的将来,XHTML 将会成为浏览器中的核心页面语言,直至 XML 技术的真正成熟之日才会逐渐退出历史舞台,关于 XHTML 和 Web 浏览技术,我们将在今后的工作中进行相应的跟踪研究。

参考文献

- 1 W3C 的第一个 XHTML 草案. <http://www.w3.org/TR/1998/WD-html-in-xml-19981205>
- 2 W3C 的第二个 XHTML 草案. <http://www.w3.org/TR/1999/WD-html-in-xml-19990224>
- 3 W3C 的第三个 XHTML 草案. <http://www.w3.org/TR/1999/WD-html-in-xml-19990304>
- 4 W3C 的第四个 XHTML 草案. <http://www.w3.org/TR/1999/xhtml1-19990505>
- 5 W3C 的第一个 XHTML 建议标准. <http://www.w3.org/TR/1999/PR-xhtml1-19990824>
- 6 Hu Lao. Two-Tigers Workshop. No-crap XML, 1999. 10-14
- 7 Bray T, et al. Extensible Markup Language (XML) 1.0 Specification. 1998. <http://www.w3.org/TR/REC-xml>
- 8 DOM (Core) Level 1 Specification Recommendation. <http://www.w3c.org/TR/REC-DOM-Level-1>
- 9 CSS1 (Cascading Style Sheet 1). <http://www.w3c.org/TR/css1.html>
- 10 CSS2 (Cascading Style Sheet 2) <http://www.w3c.org/TR/css2.html>
- 11 Bray T Using XML to Build the Annotated XML Specification. <http://www.xml.com/xml/xmlannotation.html>
- 12 Murata Madoto (Fuj Xerox Information Systems). DTD Transformation by Patterns and Contextual Conditions In: SGML/XML'97 Conference Proceedings. 1997. 153~169
- 13 Jelliffe R. Using XSL as a Validation Language. http://www.open-oasis.org/xml/xslAsValidator_19990124.html

(上接第46页)

参考文献

- 1 Goma H. Software Design Methods for the Design of Large-Scale Real-Time System. The Journal of Systems and Software, 1994, 25: 127~146
- 2 Selic B, Gullekson G, Ward P. Real-time Object-Oriented Modeling. John Wiley & Sons, Inc., 1994
- 3 邵维忠,梅宏.统一建模语言 UML 述评.计算机研究与发展, 1999(4)

- 4 Nissanke N. Realtime Systems, Prentice Hall, 1997
- 5 Gheorghe S, McGee J. Using Object-oriented modeling to design complex real-time embedded systems. Real-Time Engineering, 1995, 1(4): 11~25
- 6 Jacobson I. Formalizing use-case modeling. JOOP, 1995, 8(3): 10~14
- 7 Jacobson I, et al. Object-Oriented Software Engineering, A Use Case Driven Approach. ACM Press, Addison-Wesley, 1992
- 8 Olsen A, et al. Systems Engineering Using SDL-92. North Holland, 1994