

超文本的标准化 Hypertext Standardization

刘连芳

Liu Lianfang

(广西计算中心广西软件新技术实验室 南宁市星湖路 32 号 530022)
(Guangxi New Software Technology Lab., Guangxi Computing Center,
32 Xinghu Road, Nanning, Guangxi, 530022
email: lfl@Sun.ibep.ac.cn)

唐培和 张伟刚 殷作勤 李春贵 韦爱忠
Tang Peihe Zhang Weigang Yin Zuoqin Li Chungui Wei Aizhong

(广西工学院 广西柳州东环路 545005)
(Industry College of Guangxi, Donghuan Road, Liuzhou, 545005)

摘要 综合介绍了超文本/超媒体技术的标准化的意义、内容、表示及管理超文本和多媒体文档的国际标准 HyTime。

关键词 超文本 超媒体 标准化 互换格式 HyTime

Abstract Since hypertext emerged, scholars have probed into the problems concerning hypertext standardization. This paper summarizes its significance content, and introduces briefly HyTime, an international standard for representing the structure of hypertext and multimedia documents.

Key words hypertext, hypermedia, standardization, interchange format, HyTime

1 前言

早在第一次国际超文本研讨会(1987)上就有人提出了超文本的标准化问题。

超媒体标准必须在超媒体完全有用之前就已经形成。但是,只有在完全懂得超媒体的使用方法之后才能形成有生命力的标准。只有通过具体应用才能弄清对超媒体的需求。但是,在还没有完全了解超媒体的功能需求之前,用户就需要使用功能强大的、灵活的超媒体系统。矛盾伴随着超媒体技术的发展。

2 为什么要标准化

每种计算机都具有自己的通讯与信息管理能力。大多数超媒体系统又往往依赖于各种计算机的特定环境。要打开信息共享的大门,以信息损失最少为条件,顺畅有效地把信息及控制从一个超媒体系统转移到另一个系统,就需要各种超媒体系统之间有某些各

方都遵守的协议,这就是标准。

如果超媒体系统是运行在分布式、异种机网络环境下,超媒体技术将不同人、不同地点、不同时间以不同方式建立的多媒体信息(图形、图像、视频、动画、音频等)组合起来,这时,标准化工作就更显得必要和合理。

此外,超媒体应用系统中的宝贵信息不应该仅仅在短时期内使用,还应该在几十年后甚至更长时间内为人们共享。那就要求不论软硬环境如何变化,若干年后超媒体应用信息都可读、可理解、可使用。异步标准化将在这方面起到举足轻重的作用。

3 标准(?)定义及组成

自超媒体产生到现在,人们对它的理解是多种多样的。例如把它理解为:

- 非线性文档
- 文件或节点间的跳转
- 文件或节点内的跳转
- 高亮显示与其定义相连的“热字”

- “热字”链接数据库
- 非线性小说或诗歌
- 工作进程注释链接器
- 多媒体链接器
- 快速访问与用户能力无关的知识

至今还没有超媒体的统一标准定义。不过，有一点是公认的，那就是它的本质是非线性，它强调用户自由。无论如何，我们可以把超文本/超媒体描述为存储、组织、浏览离散信息以及建立和表示信息之间关系的技术。

现在的超媒体系统，包括许多方法，都在不断变化。但在超媒体的组成上人们基本达成了共识，那就是超媒体是由链连接的节点，组成的网络信息系统。节点能容纳正文、音频、视频、动画、图形、图像、源程序、表格、目标程序及其它类型数据。

当然，还有很多问题没有获得统一的看法。例如，节点是否可以脱离链而独立存在？节点是否唯一地由每个特定应用的链集连接？是否允许链有属性？链只能连接一个节点吗？如果一个节点被删除了，链也消失吗？如果一条链正等待动态地连接满足特定要求的新节点时，是否还允许跟旧节点相连？

节点和链的这些问题会影响知识的表达。超媒体技术的目的在于管理信息和信息之间的关系，因此，它对知识表达的格式有一定的要求。最佳的表示格式取决于信息的类型和使用它的应用系统。未来的超媒体系统必须足够灵活，以支持各种表示格式。

4 需要哪些标准

讨论标准化，最重要的问题之一是确定超媒体哪些方面需要标准，哪些方面最好保持各自独立。这种确定应该是以一个系统的功能模型作为基础的。该模型应该具有足够的灵活性以适应意想不到的技术发展。下面是两种超媒体的特别情形。在第一种情况下，工作站上运行一个独立的超媒体应用；在第二种情况下，运行一个共享的超媒体应用，它可以通过宽带网与其它应用进行通讯，并且具有通向分布式数据库的信息通道。毫无疑问，在这两种情况下，涉及信息共享方面的标准是不同的。在第一种情况下，标准必须保证应用系统在许多目前的和将来的环境中的兼容性。在第二种情况下，标准必须保证作者、用户和机器之间的完全的信息共享。

标准化工作涉及三个方面。第一是独立的超媒体环境；第二是数据对象的表示；第三，也是许多专家认为至关重要的方面：人机接口。

支持超媒体标准的参考模型应该与许多其它应

用的参考模型共享大多数属性，它应该尽可能包含各种可能的因素，并且应该不依赖任何具体的硬件、软件及网络。

超媒体涉及不同类型的数据和它们之间的关系。由于复杂的连接结构和多种媒体类型，所以需要详尽明了的数据表示标准。丰富而灵活的标准对超媒体作者在交流和建立应用方面很有价值。超媒体应用可能是唯一地影响当前的以及发展中的数据对象的表示标准。

数据表示标准应该能满足并容纳已经存在的数据标准，还要易于拓展。

标准界面涉及的问题包括两方面：连接节点和浏览信息的操作。

5 同步及异步标准

标准可以看作是在同步和异步两维空间内起作用。同步标准使超媒体系统可以在任何相同时刻相互交换和共享信息。同步标准在某个范围内是可行的，并且是关键的第一步。而异步的延续性才是制定标准的重点。也就是说，同一个超媒体系统必须现在能用，若干年后仍旧能用。事实上，满足当前超媒体系统的任何超媒体数据库在将来都可能不复存在。同步标准给我们提供的是建立真正异步标准的经验和知识。如果幸运的话，同步标准将不间断地发展成异步标准。

对许多领域，同步比异步的延续性更重要。例如，1970年的一篇论述穿孔卡片新技术的论文，对今天的工程师而言没有多大吸引力。也许有一天科学历史学家想研究这种技术，但是，我们不能保留每一件东西。在这样的领域，必须丢弃过时的信息。

如果一个人的思想只在五或十年内有价值的话，他就不会太介意把这些思想放入同样生命短暂的一个超媒体系统里。生命短暂的系统对某些人（例如书商）而言是理想的，因为它们可以冲击旧教科书市场，使学生不得不经常买新的电子教科书。

然而，真正的出版意味着出版物将成为无限期的公开记录，而不是只公开几年。没有哪个学者能化费时间去建立不能使用30年以上时间的文档。例如，在中非或拉丁美洲工作的人类学家，他们有他们自己的研究问题，他们自己的结论可能很快会过时。但是他们还是描述了正在迅速变化的社会的人种。他们发表的人种论也许是我们对那些社会的最好的（也许是我们唯一的）记录。这些必须成为我们的信息基础设施的永久资料。我们正在不断地为我们的世界增加最基本的记录。这种记录必须永久保存。

标准如果不能满足科研对书库的严格要求，至多

它只能是个不完善的临时代用品。最坏的情况下,它将引起毁灭的流沙,使后人对超媒体不信任。

超媒体系统具有惊人的潜力,而且可能最终导致研究方法的革命。但是,在它具有异步延续性之前不可能具有这种影响。一个当前在 10 个系统上运行的具有同步延续性的数据库,如果 10 年后没有系统能运行它,那么它对于学者没多少用处。

总而言之,超媒体技术需要同步标准,也需要异步标准,而且异步标准更重要。

6 数据互换格式建议

现在的各种类型超媒体系统可以归结为两大类。

第一类超媒体系统管理的数据有:文档、文档内的锚、锚之间的链以及与文档、锚、链相关的标准属性,象名称、建立的时间、文档的作者、锚的作者、链的作者等等。当前的大多数超媒体系统至少是第一类系统。

第二类超媒体系统除了管理第一类系统的所有数据之外,还管理如下信息:用户定义的对象和类型、用户定义的性质和关键字、文档、锚、链、性质的版本历史。

所有超媒体系统都需要以某种形式存储系统中的链数据。必须设计一种可以移植的互换格式以便支持相互链接的超媒体数据可以在不同系统之间交流。

布朗大学提出一种互换格式。^[1]这种互换格式包括了第一类超媒体系统中基本的链相关信息。任何理解这种互换格式的应用系统都可以捕抓现有超媒体系统的链信息。这种互换格式连同转换和传输文档数据的方法一起,可以使超媒体系统之间完全地共享信息。

这种互换格式对在相似的第一类超媒体系统之间传输数据是很有用的,对把第一类系统信息传送到第二类系统,或者从第二类系统传入第一类系统也是有用的。从第一类到第二类,需要补充第二类所需的额外信息。从第二类到第一类,则可以忽略第二类的额外信息。

互换格式并不涉及文档本身的格式。以下略为详尽地介绍这种格式。

6.1 基本对象

大多数超媒体系统涉及的信息基本上是一样的,所用的对象名可能不同。第一类超媒体系统管理文档、锚、链和系统属性。数据对象存在数据库中。在互换格式中,每种数据对象对应于一个独立的数据文件。

6.1.1 文档

18

文档用于存放应用系统的信息。它由两部分组成,一部分是应用系统的实际内容,另一部分是浏览应用系统所必须的信息。实际内容可以是正文、图形、视频、音频等。

6.1.2 锚

锚是文档中的位置。链与锚相连。锚是应用系统特有的,它是浏览应用系统的辅助标记。

6.1.3 链

链用于连接锚。链是有向的,从源锚连向目标锚。

6.1.4 系统属性

属性是由系统预先定义的,它与文档、锚、链有关。对第一类超媒体系统,属性由名字、作者、建立时间组成。

6.1.5 用户自定义属性

用户自定义属性也与文档、锚、链有关。它们可以增强超媒体信息处理和检索的灵活性。

6.2 数据文件结构

与互换格式有关的文件包括 5 个信息文件和若干文档文件。5 个信息文件分别记录有关链相关对象信息。每个文档对应一个文档文件。

6.2.1 文档信息文件

文档信息文件包含涉及所有超媒体文档的通用信息。这些信息帮助应用系统获取文档的物理位置与文档相关的用户自定义访问权限,检索文档的作者和对文档作最后一次修改的相关信息。文件中的唯一的文档标识符使应用系统可以访问锚文件中的锚信息。

6.2.2 锚文件

锚文件中存放了有关所有文档中所有锚的信息。通过这些信息应用系统可以知道锚的位置、作者、最后一次修改情况等等。唯一的锚标识符可以用于访问链文件中的链信息。

6.2.3 链文件

链文件包含了有关所有锚之间链的信息。这些信息用于为用户导航。链文件还包含有关链作者和最后一次修改链的情况。链名和唯一链标识符用于与其它文件保持一致以及适应未来的功能扩展。

6.2.4 属性定义文件

顾名思义属性定义文件包含定义系统中使用的属性的信息。系统预定义属性,象名字、作者、建立时间等,不在这个文件中定义。

6.2.5 属性文件

属性文件记录哪个对象具有哪些属性以及这些属性的值。

6.2.6 文档文件

每个文档文件的格式是由它的内容以及使用它的应用系统的需求决定的。系统之间对这些信息的交流不属于互换格式涉及的范围。然而,上述 5 个链相关文件的某些字段,间接地依赖于文档文件的存在、系统属性及文档内容。这些在系统实施中 useful。

6.3 系统实施

下面从数据格式化和字段定义的角度介绍互换格式。

6.3.1 数据格式化

为了使数据互换尽可能直接了当,就必须让要交换的数据的格式简单。

数据值都以一般的 ASCII 格式保存,使之易读、易修改、易移植。

文件中每个数据记录以回车换行符分隔。记录中每个数据字段以标记符 Tab 定界。为避免冲突,在文档和路径名中不允许用标记字符 Tab。

数据值或是字符串,或是数字。字符串长度任意,数值以 4 字节表示。十进制 ASCII 数字对应于 32 位无符号长字。

表示时间的所有数值以 Unix 的 GMT 格式存放。GMT 格式以 1970 年 1 月 1 日午夜以来的“嘀嗒”数来表示时间。一年大约 3 150 万“嘀嗒”。

预定义的系统属性值通过实用程序 Export 取自操作系统。这些属性包括建立的时间、修改的时间、作者、修改者和名称。

因为某些应用系统可能需要一些没在互换格式中特别标识的数据,为这个目的,专门设置了一些字段。为给予最大限度的灵活性,这些字段里的数据以字符串格式任意地存贮。如果要在其它应用系统中使用这些数据,就需要进行专门的转换。

6.3.2 地点标识

每个记录的第一个字段有一个地点专用 ID。这个值由两部分组成。一个是每个地点(或机器)的唯一编号(以互换格式表示)。第二个是数据转移涉及的数据库的编号。这个编号对每个地点而言是唯一的。地点 ID 和对象本身唯一的 ID 组合在一起就保证了在不同地点之间进行数据交流时能永久地保持对象标识的一致性。

7 信息表示与管理标准

在讨论信息表示与管理标准之前,让我们先了解一下一般超媒体的功能结构。

当前的多数超媒体系统的功能都可以用三类操作来描述:

- 聚合 (aggregation) 操作: 把节点和链聚合成

更复杂的对象

- 导航 (navigation) 操作: 允许用户在超媒体中浏览

- 表现 (representation) 操作: 显示或打印超媒体信息

在把数据从一个超媒体系统转移到另一个系统时,有些信息应该被保留。以上三类操作都使用或建立这些信息。这一点最容易从聚合操作中看出。在聚合操作下,对超媒体的永久结构进行改变。然而,聚合对象的结构和语义,一个系统与一个系统不同。有些系统提供多种聚合对象。从这样的系统转移信息除了要涉及到包含在聚合对象内的信息外,还要描述可能存在的各种聚合对象。

把用户从一个节点带到另一个节点的导航操作符通常比较短命,对超媒体没有持续的影响。然而,某些系统提供的导游机制应该在转换中予以保留。由于这些信息在一段重要时期内对许多用户可能是有用的。要求用户在浏览某些特定节点前必须要看一个特殊节点的限制属于这类范畴。

表现类操作符也使用在交换时应保留的信息。这些操作符或许比前两类更依赖系统。节点、链和聚合对象可能具有一些属性,这些属性说明如何显示这些对象。比如,正文节点可能有这样的属性,它指出正文是按 ASCII 码显示,还是通过 PostScript 解释器显示。

假定这三种操作符涉及的都是有价值的信息,问题是以什么格式编码适于数据交换。如果只有少量聚合操作,并且结构和语义都是已知的,每一种都带有属性,用于说明如何执行导航以及如何表现信息,那么就可以为每一种聚合操作公布标准编码。然而,这是不可能的,因为现有的系统没有足够的共同性。不同系统支持不同类型和不同层次的聚合,不同系统可能支持,也可能不支持节点同时存在于不同的非嵌套的聚合体中。

由于聚合的扩展性,就需要有一种处理超文本交换的双层法。第一步是建立描述语法。用这种语法去描述超文本中出现的聚合结构,包括聚合体是否以及如何相互关联,是顺序的、嵌套的还是别的什么。在语法描述中还应该包括与特定的聚合体相关的导航和表现属性。这样,就可以为每一种不同的聚合操作符准备编码,也可以相应地对超文本的文档编码。然后,就可以对这些聚合描述和编码超文本本身进行转移。到达另一个系统后,接收系统就需要决定如何在本系统中实现语法描述中出现的聚合类,其中包括导航和表现属性。一旦完成这步工作,接收系统就能相

对容易地将被编码的超文本文档放入自己的系统。

把聚合类描述和被编码的超文本文档组合在一起的思想,对国际标准文档类型描述语言 SGML (Standard Generalized Mark-up Language 标准通用置标语言)的用户来说似乎相当耳熟。它的确与文档处理的其它领域的现有技术十分相像,但是有一个重要的区别。聚合结构及其导航、表示属性可以用语法表达。用同样的方法,SGML 文档类型定义描述文档的元素和它们的属性。语法描述和被编码的超文本可以依次地转移到其它系统,就象 SGML DTD 可以在文档置标之前传送一样。

事实上,由于多种原因,SGML 已经适合于超文本的数据交换。SGML 允许定义聚合元素,这些聚合元素可以以复杂方式相互关联,包括顺序、嵌套和覆盖。SGML 元素可以有各种属性,适合于导航和表现信息的需要。然而,现行超文本系统中的聚合结构的变种无法为交换数据设计文档类型定义的标准集。象在传统出版系统里一样,超文本没有公共文档类型集。这就导致了在超文本交换过程中的一个不必要的步骤:处理新到达的超文本之前,接收系统需要决定如何实现一个给定的聚合类。对具有不同聚合结构的系统而言,这可是一个不小的任务。

上面概括了在超文本交换中应保留的长期有用信息的特性。核心思想是在转移超文本之前先用语法描述超文本中使用的聚合类及相关导航、表现属性。SGML 恰恰提供将描述和文档转移组合起来所需要的功能,所以 SGML 可以作为超文本交换的工具。

HyTime (Hypermedia/Time-based Structuring Language) 1992 年成为表示超文本和多媒体文档结构的国际标准。HyTime 是对 SGML 的语义扩充集。HyTime 适用于在当前乃至未来表示和管理信息对象。这些信息对象组成许多不同的复杂文档,而文档

可能分布在许多计算机系统里,而且也可以由多个主人、作者、编辑和用户建立和使用。HyTime 使超媒体和基于时间的文档享有同样的系统独立性和普遍的智能。这些都是 SGML 为静态的、独立的、基本的基于文本的文档提供的。HyTime 保证使处在异构和永不停止变化的技术环境里的许多种不同文档和它们所表示的信息能长时期、灵活地重用。

8 结束语

为了最大限度地共享信息,不同的超媒体系统之间,同一个分布式网络超媒体系统的各部分之间,当前的超媒体系统与将来的系统之间都需要信息和控制的转移。要顺利地完成这些工作,所有的超媒体系统都应该遵循某些共同的约定,即标准。

超媒体的标准化大致涉及模型、数据表示及接口三方面。如果其它技术(如多媒体)的某些标准(如文档描述,图像、视频压缩)也适用于超媒体,则超媒体标准可以借用它们。

标准化将赋予超媒体技术更强的生命力。

参考文献

- 1 Victor A. Riley. An Interchange Format for Hypertext Systems: the Intermedia Model. Hypertext Standardization Workshop, 1990.
- 2 Gregory Crane. Standards for a Hypermedia Database: Diachronic vs. Synchronic Concerns. Hypertext Standardization Workshop, 1990.
- 3 Nicholas C K., Rosenberg L H.. Interchange Standards for Hypertext. Hypertext Standardization Workshop, 1990.
- 4 DER-THANQ CHEN. Standardizing "HyperVocabulary": A Proposal. Educational Multimedia and Hypermedia, 1994, 129~134.