

面向对象的链接信息系统^{*}

Object Oriented Linked Information System

高 杨 李幼平

(四川绵阳919信箱517分箱 绵阳621900)

Abstract In 1989, T B-Lee submitted his famous report Information Management: A Proposal to CERN. In the report, novel information management architecture, Linked Information System, was coined. It is the milestone document indicating the shaping of Web idea. Linked Information System is non-linear information management architecture formed by circle nodes and arrow links. But the architecture still has two main limitations: the circle nodes are unstructured, and restricted personal cyberspace will make arrow links transfer array very complex. With the farther analyses on portal-helped WWW information management architecture, the authors present ideas of information order, high-order information. And the concept of structured circle nodes is also given, which complies with digital object in digital library technique. In fact, the more perfect information management architecture, object oriented linked information system, is developed in the paper. It is evolved from T B-Lee's Linked Information System.

Keywords Object Oriented, Information management, Cyberspace, Web, Digital library

1. Tim Berners-Lee 的报告

1989年3月至1990年5月, Tim Berners-Lee 完成了题为《信息管理: 一项提案》(Information Management: A Proposal)的报告, 提交给他所在的 CERN(欧洲核子研究中心)。这一报告是 Web 思想形成的标志, 也奠定了 T B-Lee 作为 Web 之父的崇高地位。报告中, T B-Lee 继承了 50 年代 Ted Nelson 提出的“超文本(Hypertext)”概念, 发展为“超媒体(Hypermedia)”, 并将他创造的信息管理结构命名为“链接信息系统(Linked Information System)”。

在 T B-Lee 的理念中, “链接信息系统”是指由圆圈节点(circle nodes)和箭头链接(arrow links)形象描述的非线状的信息管理结构。圆圈节点表示“实体(entities)”, 箭头链接则表示实体之间的“关系(relationship)”。链接信息系统不同于层次结构的树状信息系统、实体关系数据库系统、关键词系统(keywords system)等信息管理结构。T B-Lee 的工作, 是为了解决 CERN 信息管理遇到的诸多实际问题, 主要是研究人员众多(数千人)、人员流动(两年客座研究期限)造成的研究设备/成果共享、经验分享、项目管理等问题。T

B-Lee 期望在这一信息管理结构上, 能够形成一本内容和结构都在不断变化的书; 这部书, 还将随着组织或项目的变化而成长、演化。

提案中, T B-Lee 还卓有远见地预见到: “CERN 目前遇到了问题, 这些问题, 世界其余各处不久也将必须正视”。T B-Lee 的“链接信息系统”, 也就是 Web, 最终成为了“世界其余各处”的解决之道: Web 与 Internet 的结合, 就如同血液和血管的结合, 构成了网络社会的生命线。事实上, 对于 WWW 而言, Internet 就是 WWW 中的前两个 W: Internet 将 Web 信息管理的理念和架构推向了世界。而 Web 及其相关技术, 如 HTML、浏览器等, 又彻底改观了 Internet, 推动着 Internet “从小写的 i 走向大写的 I”(汪成为院士语)。

2. 链接信息系统的局限

10 年以后, 当绝大多数用户已经将 WWW 和 Internet 混为一谈, 而 Internet 的线路拥塞也为越来越多的人所诟病的时候, 我们不得不将审视的目光投注到 Internet 上, 并且将我们仔细的审视穿越 C/S (Client/Server) 或 B/S (Browser/Server) 网络通讯架构的浅层, 深入到 Web 最核心的信息管理结构中。我

^{*} 本项研究受中国工程物理研究院基金重点项目 97Z044 及四川银海科学基金资助。高 杨 博士, 研究方向为通信与信息系统。李幼平 工程院院士, 研究兴趣为核电子学、光电子学系统、通信与信息系统。

们要提出的问题是:10年前提出的“链接信息系统”结构,是否已暴露出不尽完善的地方?

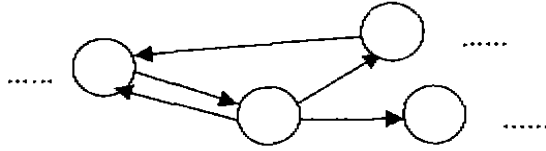


图1 链接信息系统示意图

图1是链接信息系统的示意图。这是一个非线状的信息资源管理结构。沿用 T B-Lee 的表示法,图中圆圈节点表示信息实体,即内容;箭头链接表示实体之间的内容联系,当箭头链接的表示法与 Internet 上 URL (Uniform Resources Locator) 资源标识法相结合时,在浏览器 HTTP(HyperText Transmit Protocol) 协议功能的支持下,就能够在分布式网络环境中实现信息资源的无缝链接。

链接信息系统的缺陷同样也可以从圆圈节点和箭头链接两方面来考察。

图1中,圆圈节点表示的是0阶信息,即内容本身。我们在这里给出信息“阶”的概念:如果定义某信息实体为0阶信息,那么由信息实体衍生的各级信息则为相应的高阶信息。“高阶信息”的概念与“元数据/信息(Metadata/Meta-information)”概念是吻合的,因为“元”在英文中以词头“meta”表示,其含义为“关于xx的xx”,所以“元数据/信息”的含义就是“关于数据的数据”或“关于信息的信息”,正好对应为1阶信息。依次类推,“关于元信息的信息”则为2阶信息。

为了更好地理解信息“阶”的概念,我们以科技论

文的通用结构为例,进行简要分析。

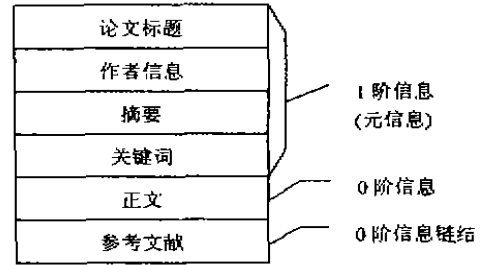


图2 例:科技论文中的信息“阶”分析

一篇科技论文可以成为图1所示链接信息系统结构中的一个“圆圈节点”。但是,对于科技论文的结构化特征,图1中的结构却不能加以利用。因此,对链接信息系统的第一个改进,是引入“高阶信息”,将简单的“圆圈节点”改进为“结构化的圆圈节点”,如图3所示。

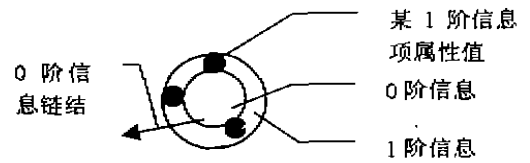


图3 结构化圆圈节点

“结构化节点”反过来正说明了 W3C(World Wide Web Consortium) 在 Meta, XML(eXtensible Markup Language) 方面的研究为何具有特别重要的意义: Meta 的研究将形成高阶信息的通用、标准数据格式, XML 将为高阶信息提供统一、规范的容器(container)。

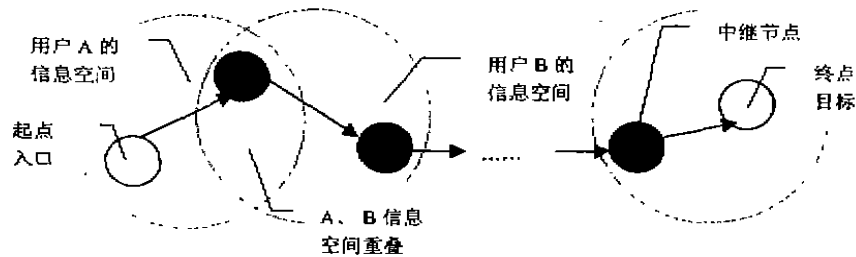


图4 个人信息空间有限性引发的链接序列复杂性

再考察图1链接信息系统中的“箭头链接”,正是“箭头链接”的自由存在,才使得链接信息系统成为一类非线状的信息管理结构,而与其它信息管理结构(如:树状的层次结构,2维表的关系式数据库结构)区别开来。然而,“箭头链接”的建立是由人工直接指定的,且一旦定义后,链接关系就随之固化下来。当链接信息系统成为 WWW 这一全球浩瀚资源库赖以生存

的信息管理结构基础时,人为、凝固链接的矛盾就暴露出来了。

首先,个人能够指定的链接只能发生在个人(或个体)所知的信息空间中,无法直接与整个信息空间(cyberspace)建立周全的链接。基于个人信息空间的重叠,虽然可能通过许多“中继内容节点”进行链接传递,最终实现信息节点的无限链接,波及整个信息空间;然

而,一组知识序列所对应的链接路径,往往是极其复杂而曲折的。图4揭示了这一矛盾。

简而言之,受个人信息空间有限性的客观制约,T B-Lee 链接信息系统结构虽能通过个人信息空间内在的、重叠式蜂窝结构,提供目标内容节点的“可获得性(accessibility)”,但要付出链接序列复杂性的代价:最终用户只能在无序的信息空间中,通过漫游式的访问、遍历长度未知的链接序列才可能达到真正所需的目标内容节点。这一结构,无法确保用户顺利、便捷地抵达目标。

链接序列的复杂性,以及目标内容节点的若隐若现,是导致 Internet 带宽拥塞的原因之一;反之,Internet 带宽拥塞又消磨着信息搜索者遍历繁复链接序列的勇气和耐心。

3. 引入门户网站的 WWW 信息管理结构

网络门户(Portal)站点的出现,部分弥补了上述缺陷。图5表述了引入网络门户网站之后的 WWW 信息管理结构。图5中的结构有这样一些特点:

·作为用户(0阶信息的搜索者)、WWW 内容提供商(0阶信息提供者)之外的第三方,门户网站提供浩瀚

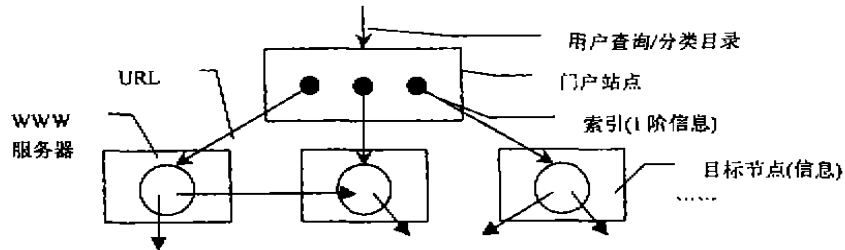


图5 门户网站 WWW 信息管理结构

4. 改进:面向对象的链接信息系统

受门户网站 WWW 信息管理结构的启发,我们进一步考察改进系统结构的途径与方案。图5蕴藏的主要问题是:0阶信息与1(高)阶信息分离。如果0阶信息能够与1阶信息密切结合为一体,那么即便继续沿用门户网站这一第三方解决方案,也能够大大提高信息查全率与查准率。

4.1 “结构化圆圈节点”与“数字对象”

改进后的链接信息系统结构如图6所示,图中圆圈节点与箭头链接的含义与 T B-Lee 的定义一致,但圆圈节点增加的外圈部分表示按照一定的元数据规范从0阶信息中提取的1阶信息。外环中的图案小圈,表示一个元数据名-值对(name-value pair)。这样,T B-Lee 结构中简单的圆圈节点就演化成了较复杂的“结构化圆圈节点”。

WWW 信息资源的索引,即1阶信息,在门户网站上,用户可用自由词查询,或者沿门户提供的分类目录层层深入(又回到了“层次结构”),直至获得所需信息的 URL 地址;再通过一次网络通讯连接,就能得到目标内容;

·WWW 上,Web 页,即 HTML 文件,是0阶信息的主要形态。由于 HTML 只能控制页面在浏览器中的显示效果,不提供内容的语义结构,因而门户网站需要利用搜索引擎技术,主动扫描 WWW 信息资源,并主要是从 HTML 文档头(<head></head>元素之间的部分)获得索引数据;

·依赖门户网站提供的 WWW 资源索引机制,用户可以避免遍历图4中描绘的复杂链接序列,直奔主题。门户网站的 WWW 索引结构引入了一类新的链接,在图中以另一类箭头形状区分,这类新链接同样以 URL 标记,但它独立于 Web 页,由门户网站第三方建立并维护,且由文档的1阶信息指向文档本身(0阶信息)。两类箭头链接,可以用“外引内链”来概括。

·在网络通讯互连基础上,0阶信息与1阶信息(元信息、索引)分离存放,0阶信息存放在通常的 WWW 服务器上,1阶信息存放在门户网站的服务器上。

比照数字图书馆(D-Lib)技术中“数字对象(Digital Object)”概念的定义,我们认为,“结构化圆圈节点”正是“数字对象”的一种图形化表述方法。

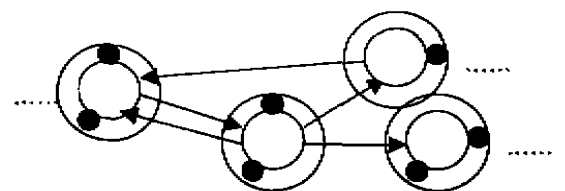


图6 改进后的链接信息系统

在 D-Lib 技术中,数字对象是本地或互连的数字仓储(Repositories)中表示信息的基本逻辑单位,它们存储在仓储中,用句柄(Handles)来标识。存储在数字对象中的信息称为内容(content);内容被分成数据

(data)和关于数据的信息(属性(properties)或元数据(metadata))。

4.2 动态聚合能力

完整的数字对象定义、特别是依赖于数字对象中的高阶信息标引,使结构化圆圈节点具备了动态聚合的能力。搜索引擎、聚类软件可以充分利用这种能力,动态、高效、准确地按照用户查询要求,即时组织信息资源并提交。图7给出了面向对象的链接信息系统响应用户查询请求、动态聚合后的效果。

类似于门户站点的 WWW 信息管理结构,图7中也存在两类链接:依赖1阶信息建立的动态链接,以及0阶信息内部固化的链接。在分布式网络环境中,面向对象链接信息系统的好处在于,能够将内容节点的结构化数据直接复制到搜索引擎,避免搜索引擎盲目索引,从而提高了索引的水准。在本地海量信息资源的应用环境,只有结构化的内容,才可能在智能软件帮助下,

聚合形成品质良好的“我的读物”。比较而言,简单的圆圈节点虽然也可在全文智能检索软件帮助下,形成“我的读物”,但其品质完全依赖软件的智能水平,而软件智能必将面临模糊语义识别等方面的巨大困难。

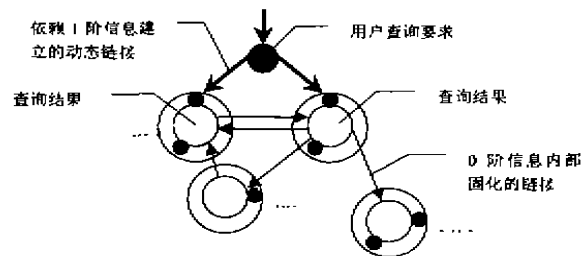


图7 面向对象链接信息系统的动态自组织

(下转第24页)

(上接第41页)

参考文献

- 1 Rover E M, Toh C-K. A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks. *IEEE Personal Communications*, 1999, 6(2): 46~55
- 2 Jubin J, Tornow J D. The DARPA Packet Radio Network Protocols. In: *Proc. of IEEE*, 1987, 75(1): 21~32
- 3 Perkins C E, Bhagwat P. Highly Dynamic Destination-Sequenced Distance-Vector Routing (DSDV) for Mobile Computers. *Comp. Comm. Review*, 1994(Oct.): 234~244
- 4 Murthy S, Garcia-Luna-Aceves J J. An Efficient Routing Protocol for Wireless Networks. *ACM Mobile Networks and App. J., Special Issue on Routing in Mobile Communication Networks*, 1996(Oct.), 183~197
- 5 Chen Tsu-Wei, Gerla M. Global State Routing: A New Routing Scheme for Ad-hoc Wireless Networks. In: *Proc. IEEE ICC' 98*
- 6 Iwata A, et al. Scalable Routing Strategy for Ad Hoc Wireless Networks. In: *Proc. IEEE ICC' 98*
- 7 Joa-Ng M, Lu I-T. A Peer-to-peer zone-based Two-Level Link State Routing for Mobile Ad Hoc Networks. *IEEE Journal on Selected Areas in Communications, Special Issue on Ad Hoc Networks*, 1999(Aug.) 1415~1425
- 8 Chiang C-C. Routing in Clustered Multihop, Mobile Wireless Networks with Fading Channel. In: *Proc. IEEE SICON' 97-1997*, 197~211
- 9 Jiang Mingliang, Li Jinyang, Tay Y C. Cluster Based Routing Protocol, IETF Draft, August 1999
- 10 Perkins C E, et al. Ad Hoc On-demand Distance Vector Routing, IETF Draft, October 1999
- 11 Johnson D B, Maltz D A. The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks, IETF Draft October, 1999
- 12 Johnson D B, Maltz D A. Dynamic Source Routing in Ad Hoc Network Mobile Computing, T. Imielinski and H. Korth, Eds., Lulwer, 1996. 152~181
- 13 Park V D, Corson M S. A Highly Adaptive Distributed Routing Algorithms for Mobile Wireless Networks. In: *Proc. INFOCOM' 97*, Apr. 1997
- 14 Toh C-K. A Novel Distributed Routing Protocol to Support Ad hoc Mobile Computing. In: *Proc. 1996 IEEE 15th Annual Int'l. Phoenix Conf. Comp. And Comm.*, Mar. 1999. 480~486
- 15 Toh C-K. Long-lived Ad Hoc Routing Based on the Concept of Associativity. IETF Draft, May 1999
- 16 Dube R, et al. Signal Stability Based Adaptive Routing for Ad Hoc Mobile Network. *IEEE Personal Communications*, 1997(Feb.): 36~45
- 17 石坚, 邹玲, 董天临, 等. 遗传算法在组播路由选择中的应用. *电子学报*, 2000, 5(128): 88~89
- 18 石坚, 董天临, 石冰心. 网络路由算法的研究. *计算机工程与应用*, 2000, 9
- 19 Guerin R, Orda A. QoS-based Routing in Networks with Inaccurate Information. *Theory and Algorithms. Infocom' 97, Japan*, April 1997
- 20 Lorenz D H, Orda A. QoS Routing in Networks with Uncertain Parameters. *Infocom' 98*, March 1998

表7

编号	method	Pre	Run
1	Read	Φ	True
2	Read	Φ	True
3	Write	{1,2}	False
4	Read	{3}	False

表8

编号	Method	Pre	Run
2	Read	Φ	True
3	Write	{2}	False
4	Read	{3}	False

(8)方法 Read2运行结束后,Write3投入运行,方法请求队列变为表9。

(9)方法 Write3运行结束后,Read4投入运行,方法请求队列变为表10。

表9

编号	method	Pre	Run
3	Write	Φ	True
4	Read	{3}	False

表10

编号	Method	Pre	Run
4	Read	Φ	True

由于达到 objRW 对象的方法请求是动态的,在任何时刻都可能新的方法请求到来,读者自己可以试验一下新的方法请求到来时对方法请求队列的影响。

(上接第33页)

4.3 面向对象链接信息系统的困难

面向对象链接信息系统的难点集中体现在两个方面:节点结构化的规模及其技术手段。节点结构化规模过小,意味着系统检索入口少,动态聚合只能在少数几个维度展开;反之规模过大,又意味着标引繁复,甚至会出现高阶信息数据量远大于0阶信息数据量的情形。节点结构化有两种可能的技术手段:人工标引或机器标引。人工标引可能会受到作者习惯性惰性的抵制;机器标引又将遭遇全文智能检索软件所面临的技术困难。机器标引还会带来另一个问题:既然创作与标引分离,那么是否有必要保持0阶信息与高阶信息一体化的节点结构呢?我们是否只需将管理 WWW 这一浩瀚信息资源库的期望寄托在搜索引擎智能的全面提升之上呢?

• 24 •

小结 本文提出的同步控制模型包括互斥控制与条件同步控制。为自动实现互斥控制部分,本文扩充了预编译系统和运行系统,不仅大大减少了继承异常的发生,而且降低了开发人员进行并发面向对象程序设计的难度。条件同步控制采用了基于对象状态的继承异常解决方案,实现控制代码跟方法体的分离,解决了部分代码的复用,进一步降低了继承异常的发生。

随着 Internet 网迅速普及,特别是对 NOW 研究不断深入,我们可以把同步控制扩展到 NOW 环境,将通过同步控制的方法请求分布到另外的空闲工作站上运行,以提高对象内的并行性。另外可以改进依赖关系的判定算法,剔除一些假依赖关系,进一步提高对象内的并行性。

参考文献

- 1 Matsuoka S, Yonezawa A. Analysis of inheritance anomaly in object-oriented concurrent language. In: Agha G, Wegner P, Yonezawa A, eds. Research Directions in Concurrent Object-Oriented Programming. Cambridge, Mass: MIT press, 1993. 107~150
- 2 America P Pool-T: a parallel object-oriented language. In: Yonezawa A, Tokoro M, eds. Object-Oriented Concurrent Programming. Cambridge, Massachusetts: MIT Press, 1987. 199~220
- 3 Andler S. Predicate path expressions. In: Proceedings of 6 POPL, ACM, 1979
- 4 Ben-Arm. Principle of concurrent and distributed programming. Prentice-Hal International, Hemel Hempstead, 1991
- 5 陈家俊,赵建华,郑国梁. C++的一种并发扩充方案. 软件学报, 1998, 8: 586~591
- 6 Cornell G, Horstmann C S. Core Java. SunSoft Press, 1996
- 7 Yonezawa A, et al. Modelling and programming in an object-oriented concurrent language ABCL/1. In: Yonezawa A, Todoru M, eds. Object-Oriented Concurrent Programming. Cambridge, Massachusetts: MIT Press, 1987. 55~89

这些问题的实质,仍然是信息管理成本与效益之间的矛盾冲突。成本与效益折衷会倾向于什么样的信息管理结构呢?我们无法断言,只能等待未来技术与商业的发展作出抉择。

参考文献

- 1 Tim Berners-Lee. Information Management: A Proposal 1989. 3, 1990. 5. Available at: <http://www.w3c.org>.
- 2 李幼平. 构思“第五传媒”. 北京. 计算机世界周报, vol. 49-D1, 1998. 12. 21
- 3 高杨,李幼平. 内容主动服务技术. 北京. 计算机世界周报, vol. 50-D7, 1998. 12. 28
- 4 王云,赵武文,高杨. 内容表示语言与统一内容定位. 北京. 计算机世界周报, vol. 50-D9, 1998. 12. 28
- 5 唐世谓,杨冬青,王军. 互连信息空间 InterSpace. 北京. 计算机世界周报, C5版, 2000. 2. 21