

●刘友华

开发数字图书馆需考虑的几个问题

摘要 数字图书馆开发者须熟悉系统内容和用户需求;要通过维护系统品质和易访问性来保证系统的持久可用;使用导航模块等工具技术,确保大多数用户能简易省时地使用系统;保证系统的开放性是为了接纳潜在的用户群体,并使系统容易适应变化的运行环境和技术更新。参考文献 15。

关键词 数字图书馆 系统开发 质量度量 系统维护

分类号 G250.76

ABSTRACT The author thinks that digital library developers should know system contents and user requirements, ensure the long-term system usability through the system quality and accessibility, facilitate the users by providing navigation models and ensure the openness of the system. 15 refs.

KEY WORDS Digital library. System development. Quality measurement. System maintenance.

CLASS NUMBER G250.76

数字图书馆的服务对象不仅是个体用户,也必须包括集团或团体用户。目前许多数字图书馆基于已有的传统形式图书馆进行开发,建成后可能同时提供电子和传统形式的信息资源,也可能仅提供电子格式的内容,当然也有的考虑直接建成数字图书馆。数字图书馆的信息内容要求无论在哪一台计算机上,使用何种浏览器,以何种数据格式,都应能持久地可靠地被接入和访问,并能最大程度地满足用户需求和方便用户的使用。从数字图书馆系统开发生命周期的角度看,这些目标也是软件工程所要求的质量度量因素需考虑的主要方面。本文将从尽可能广阔的视角着眼,提出适用于绝大多数甚至是所有数字图书馆开发的框架性指导原则。尽管在某种意义上这些原则显而易见,或是基于他人的经验,并被实践活动所支持着,但开发人员由于往往会受时间、精力等多方面压力的影响而偏离甚至背弃这些原则。坚持把握好它们,既有利于耗费大量精力的数字图书馆开发设计人员,也有利于期望系统长期顺利运行并持续盈利的投资者。

1 开发有用的数字图书馆系统

1.1 了解系统中的内容

数字图书馆内容是最令人感兴趣和最有价值的部分。其创建者需要管理和决定相关内容,诸如恰当地选择数字图书馆所包含的对象、对象存储格式,以及对不同对象赋予用户不同权限等。

在开发早期,决定元数据的种类和数量很重要,对已被标记索引的元数据,是正在运行的系统内容集

合中不可分割的主要部分。理论上说,无关内容、关系的存在是难以避免的,但若实际所占比例过大并且处理不当,则会加大检索利用信息的难度,甚至会使用户反感。系统中内容标识的深浅详细程度也影响其可用性,不同的处理所耗费的精力和资金差别很大,且针对的用户范围也有不同。

1.2 研究用户需求和用户行为信息

一定的用户群体通常会有大量相同的信息需求和信息检索的行为特征。数字图书馆系统的开发,需要研究用户需求和行为,并对潜在的用户作出预测,这可能成为改善前边提出的了解并确定系统内容的依据,并在最大程度上使用户方便快捷地检索到准确的信息内容。数字图书馆用户培训也必须建立在对用户需求和行为研究的基础上,就像“现代汽车”的设计无论怎样完善和人性化,人们也要经过学习才能安全驾驶——仅仅通过开发设计上单方面的改进,无法实现交通运输效率的根本性提高^[1]。

2 维护系统的长期可用性

2.1 及时跟踪数字图书馆的内容

除了对开发数字图书馆系统过程的高质量要求,在系统运行过程中也要维持提供高质量的服务状态。系统运行中必然会出现各种问题,比如软硬件出错、陈旧,以及数据不一致或内容需要纠正更新。比如视频、音频资料,需要对它们进行周期性评估,确保它们适合发展着的格式和标准,不会因格式过于陈旧而不能被用户用现有的工具观看或收听。

2.2 完全支持 Web 访问

尽管严格说来,支持 Web 访问并不是数字图书馆必要的属性。然而,它是一种可供用户访问各种网络资源的普及方式,现在的大多数图书馆系统都被设计为可通过 Web 访问。好的 Web 页面设计应充分考虑不同网络 and 用户之间的差异因素,不同 Web 导航方式的选择等等。而且有些差异又是多层次的,如用户使用的浏览器在显示信息的方式上可能会有所不同。开发者应该创建一个尽可能简洁的界面,许多在客户端能够灵活控制的属性应当避免在服务器端进行额外控制,这样利于不同类型的网络 and 用户接入,也利于接纳潜在用户^[2]。

2.3 维持系统的持久性

Jeff Rothenberg 指出,“‘21 专家组’的研究给出了令人印象深刻的结论:目前根本无法对数字信息存储的效果作出任何保证”^[3]。虽然存储问题在档案和图书馆方面的考虑由来已久,但也仅是在最近才有较多的研究机构对它产生兴趣。“图书馆研究组织”(Research Library Group)在一项对其成员机构的调查结果中,将“技术淘汰作为对数字信息集合的最大威胁”^[4]。研究者已经提出一些建议来处理数字对象的存储,以应对技术的更新淘汰。经常提及的一种方式“移植策略”,即将数据从一种文件格式转换为另一种格式,或从一种软件环境转换到另一种环境,也可能从一种物理媒体转移到另一种媒体,它要求数字图书馆的维护人员必须持续不断地用新数据格式或新媒体来表现现有的资料集合以跟上新技术。相关研究人员提出的另一个建议是“仿真策略”,即要求仿真整个软件系统,以运行在将来未知的系统上,透过仿真软件访问原有数据^[4]。

数字对象被永久存储于物理媒体并不等于具有持久的可用性。数字图书馆联盟(Digital Library Federation)指出了数字对象的存储和持久性之间的一个重要区别:存储涉及一个对象的技术寿命和质量,持久性是一个更广泛的概念,包含了存储,但也指对象是否仍然在将来以任何形式存在。持久性暗含了对对象维护的概念,同时也意味着对其可访问性的持续保持^[5],否则会缩短数字图书馆的平均寿命^[6],降低数字图书馆的可访问性。

3 创建易用的系统

3.1 尽可能实现操作自动化

自动化工具能使数字图书馆系统简便易用,提高

用户利用资源的工作效率。以下拉列表、报表生成器、自动文摘、信息提取、翻译服务、自动语音识别、资料统计与科学建模计算、图像再现处理等各种形式出现的导航模块工具,不仅让关注内容的专业人员不必耗费大量精力于工具调配及其检索过程,甚至还可以利用这些工具帮助他们完成少部分专业处理工作,如自动抽取符合要求的信息条目,并按要求进行排序,生成诸如 HTML 页面、分类报表等结果报告。

3.2 数字图书馆系统各模块应无缝整合

开发数字图书馆系统应该考虑模块化设计,模块之间尽可能地相互独立,如对数据库模块的更新不应影响用户对导航模块工具的使用,反之亦然。Profiles in Science 的开发者通过增加诊断服务模块来提供专业人员访问初步建立的主数据库,允许他们在系统正式投入使用之前发现并改正其中暗含的错误和矛盾,而这一模块在系统正式建成后可被很容易删除^[7]。

尽管模块之间的独立性很重要,其间的无缝整合同样不可忽视。无论何种功能、以何种方式实现,模块间的衔接都应该对用户透明,给用户一体化的印象。Digimorph DL 提供来自生物学样品的 CAT 扫描数据,而它所提供的工具可以使用许多方法巧妙处理这些数据,以分离骨骼结构和器官,以及显示可以从不同角度观察的旋转 3 维图像^[8]。Perseus 系统提供了接入收藏有伦敦历史地图和图片系统的接口,还提供能让用户在历史城市街道上进行虚拟 3 维穿行的工具^[9]。ARION 系统向用户提供海洋学数据并提供处理这些数据的算法,帮助用户通过列出每一步所需要用到的数据和算法来建立一个总体的过程,然后系统从数字图书馆获取所需的数据和算法来执行过程^[10]。CHLT 系统正在开发一种综合的阅读环境来帮助学生和人文学者阅读、理解,并且解释古典原文^[11]。

3.3 界面能引导用户贯穿整个复杂的任务

简洁而强大的导航模块工具可用来实现复杂的功能,但用户使用数字图书馆依然是一个存有大量交互性的过程,必须通过简明易懂的界面向用户反映运行情况,并允许用户对其进行控制。Perseus 系统其数据结构和辅助工具建设相当成功,经常被他人参照仿效,但它的界面却不够理想,有碍充分使用其丰富的信息资源。

3.4 提供强大的数据库检索功能及数据关联结构

对数字图书馆工具的设置需要超越简单的单个数据库中的词汇表控制和原文检索,要能在相关多个

数据库或数字图书馆中进行复杂检索。比如,检索某对象时还关心与其相关的对象及属性;检索复合对象;检索基于某类对象属性的大体情况等。这些除需复杂的检索算法,还涉及不同系统协同工作能力和分布式数据库的接口处理问题。尽管在这一领域也有许多相关研究,但事实上仍然没有找到普遍适用的操作方法。比如,如何识别不同系统中同一种属性的不同表现形式,以及将这些不同属性类型的检索结果联结、统一排序;如何比较和处理初步检索结果;如何发现来自不同系统的检索结果的冲突和矛盾;如何将检索自不同系统不同类型的数据结合起来创建统一的报告等等。就像当初提出以引擎驱动汽车时,它被建立在马车的外形之下,人们称之为“不用马的马车”,只有经过长期发展,才充分利用了许多新技术并具有现代意义的汽车成型^[12]。许多关于数字图书馆的开发实践活动应该使得数图系统不断突破“不用马的马车”的瓶颈。

数字图书馆开发者还应尽可能地将丰富的数据关系通过链接反映出来,支持用户通过不同数据和对象之间的链接进行查阅、解释。链接不仅存在于同一个数字图书馆之内,不同数字图书馆之间也应当在相互关联的信息之间建立链接,共享资源,以提高信息资源的利用率。各种链接复杂交错,需要对它们分类整理,提高清晰性和可维护性等^[13]。

3.5 允许用户拥有自己的信息空间

数字图书馆不但应该允许用户访问,而且要能够让用户建立自己的个人或团体信息空间。用户可以根据需要将自己信息空间中的内容与数字图书馆管理员或其他用户共享,能找到他们在数字图书馆各种活动的历史记录,也能作为他们研究成果的存储媒介和交流媒介等。

4 数字图书馆必须是开放的系统

4.1 充分考虑开发过程中所涉及的人员

来自不同背景和提供不同专业意见的用户对建立数字图书馆都能作出各自不同的贡献。计算机专家关注在技术上实现各种服务功能的可行性以及局限性,图书馆学专家长期以来一直是信息资源的管理者,也是信息服务的提供者,懂得不同用户的信息需要。开发数字图书馆的团队必须协同合作,在最大程度上将各自理解的用户需求融合在一起。

4.2 开放的接口

开放的接口包括进入数字图书馆系统的接口和

进一步访问数字图书馆信息的接口。Christine Borgman 将信息访问定义为“对计算机网络和对其中所包含的可用内容的可连通性,这一技术必须是可用的,用户具有必需的技能 and 知识,而内容本身也应处于有用的和可用的格式状态”^[14]。在任何情况下,为了可持续访问和使用,公开共享的非专有工具应当成为首选,而要避免采用一些需要通过特殊要求才能获得的工具作为访问数字图书馆及其内容的入口。

4.3 采用和坚持标准

数字图书馆创建过程中,采用标准格式开发的应用程序更容易升级、实现互操作和移植^[15]。开发过程中,资源密集和劳动密集的资源部分尤其需重视和坚持标准的采用,这些内容往往是系统中最有价值的部分,同时也是系统更新、移植或转换时最需要保持可用性的部分。

4.4 预期未来可能的变化

数字图书馆开发者应充分考虑日益变化的技术问题,甚至对技术发展趋势作出一定预期。假如数字图书馆的有关文档是以某特定的字处理程序创建的,那么必须依赖这种字处理程序才能解读文档的格式及其内容,这种依赖性会导致当这种字处理程序发生技术淘汰或替换时,原先的文档将无法或很难再被获取,致使当初的收藏价值消失殆尽。因此应有必要的处理方法,能在原来预期和计划的情形发生改变时及时响应,以确保数字资源持久的可访问性。

5 结语

本文运用现代软件工程学的观点探讨了数字图书馆系统开发生命周期中应当关注的质量方面的问题。美国近期建立的 Profiles in Science 系统就是从 90 年代早期的一个数字图书馆移植过来的,在当时将图书馆内容提供在互联网上服务的想法还不很成熟甚至没有成型,两个系统的软硬件环境也完全不同,但移植工程却非常成功,这正是因为系统开发设计的思路是科学的。建设内容丰富并有价值的数字图书馆的同时,实质上也就是在向现有的和预期的用户作出长期提供有效信息服务的重要承诺的过程。数字图书馆开发者应当关注本文所讨论的诸多方面并结合所涉及的实践活动,使得数字图书馆有价值的内容能被正确处理,以尽可能高的质量展现出来,也使得数字图书馆所有的内容具有持久性,保证系统的易用性和开放性。

(下转第 81 页)

年代末至 60 年代, 60 年代以后, 成为政府、学术界和产业界共同关注的热点问题。从我国信息法律和信
息政策的研究轨迹来看, 早期的研究一般集中在科技
信息领域。进入 90 年代以后, 随着现代信息技术、信
息活动、国际信息交流和因特网的发展, 出现了诸如
网络安全、信息伦理、电子犯罪、信息隐私、知识产
权、国际信息合作等许多新问题。学术界对信息法律
的研究进一步加强, 并取得了许多成果。可以看出, 信
息法学是我国近年来信息学和法学等众多领域研究
的重点内容之一, 信息法学研究预测也是信息法律
预测的重要内容。

由于信息技术的不断发展, 人类从事信息生产实
践活动的不断增多, 信息关系会越来越复杂, 而且处
于产生和消亡的变动中。因此信息法律预测工作要
经常进行, 且要随着预测理论和方法的发展, 不断采
用新的预测方法和技术, 这样才能更好地促进信息法
制建设的发展。

参考文献

- 1 王志荣. 信息法概论. 北京: 中国法制出版社, 2003

- 2 马海群. 信息法学. 北京: 科学出版社, 2002
- 3 倪建民, 沈志坤, 公丕祥. 法律预测的理论与方法. 北京: 法律出版社, 1988
- 4 蒋坡. 国际信息政策法律比较. 北京: 法律出版社, 2001
- 5 贾文中, 黄瑞华. 试论信息法的体系. 情报理论与实践, 1997(1)
- 6 查先进. 信息化浪潮下国家信息法的部门法地位. 中国图书馆学报, 2003(5)
- 7 吴宏亮, 颜小云. 论我国信息立法的几个基本问题. 中国图书馆学报, 1999(2)
- 8 马费成等. 我国信息法规建设现状评价与对策研究. 情报学报, 2004(2)
- 9 Nick Moors. Information policy and strategic development, a framework for the analysis of policy objectives. Aslib Proceedings 1993(11~12)

相丽玲 山西大学管理学院教授, 硕士生导师。通信地址: 太原市。邮编 030006。

屈宝强 南开大学国际商学院图书馆系博士研究生。通信地址: 天津市。邮编 300071。

(来稿时间: 2004-06-22)

(上接第 60 页)

参考文献

- 1, 7, 12 Dagobert Soergel. A Framework for Digital Library Research. D-Lib Magazine. December 2002. <http://www.dlib.org/dlib/december02/soergel/12soergel.html>
- 2 Alexa T. McCray, Marie E. Gallagher. Principles for Digital Library Development. Communications of the ACM, May 2001; 49-54
- 3 Rothenberg, J. Avoiding Technological Quicksand: Finding a Viable Technical Foundation for Digital Preservation. Rep. To Council on Library and Information Resources, Jan. 1999
- 4 Kenney, A. and Rieger, O. Moving Theory into Practice: Digital Imaging for Libraries and Archives. Research Libraries Group, Mountain View, CA, 2000
- 5 Waters, D. The Digital Library Federation: Program Agenda. A program of the Council on Library and Information Resources, June 1998
- 6 Kahle, B. Preserving the Internet. Sci. Am., Mar. 1997
- 8 The Digimorph Digital Library was established by Tim Rowe at the University of Texas at Austin. <http://digimorph.org/aboutdigimorph.phtml>

- 9 Gregory Crane. The Perseus Digital Library. <http://www.perseus.tufts.edu/PR/vr.ann.html>
- 10 ARION, Forth. http://dlforum.external.forth.gr:8080/papers/ARION-paper_v52.pdf
- 11 Jeffrey Rydberg-Cox. CHLT (Cultural Heritage Language Technologies). University of Missouri, Kansas City. <http://www.chlt.org>
- 13 Kopak, Richard W. Functional link typing in hypertext. ACM Computing Surveys, Dec. 1999; 31-34
- 14 Borgman, C. From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World. MIT Press, Cambridge, MA, 2000
- 15 Strand E., Mehta R., and Jairam R. Applications thrive on open systems standards. Standard View 2, 3 Sept. 1994; 148-154

刘友华 南京大学信息管理系副教授, 博士生。通信地址: 南京市。邮编 210093。(来稿时间: 2004-12-09)