

关于建立文物考古数据库的几个问题

张鹏程 (陕西省考古研究院)

数据库技术是一种高效、简便的管理复杂日常事务的方法,目前普遍运用于金融、企业、情报、交通等行业,它能方便地管理大量烦杂的数据,快速直观地监视和分析数据的变化。

在考古工作中,资料的整理、统计、分析以及各类文物的管理是一项烦杂的工作,需要耗费大量的人力、物力,即便这样仍然不能迅速、全面掌握资料的情况。如果将数据库技术引入到考古工作中,不仅能大大简便资料整理工作,而且能够把整个考古工地的规划、管理、进度纳入到一个系统事务中,将工作的准备、实行、整理、研究变成一个紧密衔接的流程。

下面将在考古工作中实现数字化管理需要的几项准备作一简要介绍。

一、硬件环境的构建

大型数据库的硬件要求比较高,而运行中小型数据库对计算机的需求则很低。在 PC 机中,考虑执行和数据输入速度的因素,最低需配置主频率在 300M 以上的奔腾级中央处理器、64M 内存、4G 以上硬盘,考虑到数据库系统的扩展、执行效率等建议配置主频率在 1G 以上的奔腾级中央处理器、256M 以上内存、40G 以上硬盘。出于数据安全性的考虑,应当使用磁盘阵列技术(RAID),使用 2 或 4 块硬盘建立数据安全系统,或设计良好的磁盘备份计划,保证数据的安全。苹果机也可以作为一种选择,但苹果机存在跨平台移植困难和数据库软件相对 PC 平台少的问题,不适合作为数据库构建、开发的硬件基础。

二、软件环境的构建

目前的操作系统常见的有 Microsoft 的 WIN95、WIN98、WINME、WINNT、WIN2000、WINXP、WINSERVER2003、WINDOWS VISTA 以及 LINUX、UNIX、苹果机。在这些系统中 Microsoft 的 WINDOWS 系列使用广泛,软件和硬件支持良好,选择范围较宽,安装简便,用户界面简单易用且为大多数人所熟悉,但其软件本身太过臃肿,严重浪费硬件资源,系统昂贵。此外若作为开放的数据库服务中心安全性也不够高。LINUX 是免费操作系统,安全性好,开发工具免费使用不需额外购买,但软硬件支持不足,日常维护复杂,操作不方便目前多为专业人员操作或作为服务器使用。UNIX 系统和 LINUX 相似,目前多为服务器,且其本身分化为多个不同的操作系统,只有少量专业人员使用。

综合考虑操作系统需选用 Microsoft 的 WINDOWS 系列或根据情况选用 LINUX。WINDOWS 系列中 WIN95、WINNT 已在淘汰之列,而 WINME 不提供 DOS 操作,故障率高也不能考虑。WINSERVER2003 是 WINXP 的服务器版本,价格昂贵。如果硬件配置高则不建议使用 WIN98,因其对内存的管理不好,故障多,对目前大量的移动设备支持较差,系统崩溃的风险高。WINDOWS VISTA 在目前阶段其安全性和兼容性尚需完善。WIN2000、WINXP 是目前比较好的选择,但这两个操作系统对硬件的需求较高,对系统的管理也相对复杂。

三、数据库软件的选择

数据库有多种划分方法,当前流行的数据库多为关系数据库管理系统,可分为三类。

A. 桌面型数据库管理系统。以 PC 机、微型机系统为运行环境,主要支持一般办公需要,操作方便,如 dBase、FoxBASE、FoxPro、Office 系列的 Access 等。

B. 主流数据库管理系统。强调系统的功能强大、完备,具有非常庞大的数据存储和管理能力,提供很全面的数据保护和恢复功能,适合全局性、关键性的数据管理工作。如:Oracle (甲骨文)、Power Builder 等。

C. 介于以上两者间的 Microsoft SQL Server。

目前经常见到的数据库系统主要有 Microsoft SQL Server、Power Builder、FoxPro、Office 系列的 Access。1. Power Builder 作为一个使用非常广泛,影响巨大的数据库产品在很多的关键性应用中发挥重要作用,在大型专业数据库中首屈一指,它容易使用,且提供图形化设计界面,对 WINDOWS 有良好的支持,可以方便地管理图形对象,具有极高的开发效率和质量,对多种数据库技术有较好的支持,但需要比较专业的人员使用,因它的窗口和内容没有汉化。如构建大型关键数据库,从数据库扩展的角度来看 Power Builder 是首选。2. FoxPro 是很广泛使用的中小型数据库,它比使用相对专业数据库要简便得多,内容汉化也很好,提供图形化设计界面,但仍然需要学习其编程方法后方能使用。3. Office 系列的 Access 集成在 Office 环境中,这样它可以很方便的和 Excel、Word 交换信息,并和系统结合紧密,操作简单,使用方便,能熟练使用 Excel、Word 的操作者在简单学习后,就可以使用它来编制数据库了。但它对其他数据库不提供支持,也不具有很好的数据保护功能,无法独立运行(必须安装 Access),最大能支持 100 万条数据。4. Microsoft SQL Server 多用于网络环境。

对于不要求数据库系统强大的扩展功能和很大的数据存储、管理,不要求数据移植,而要求简单易用的使用者来说,FoxPro 和 Office 系列的

Access 是很好的选择,这两者中 Office 系列的 Access 是更为简便的数据库系统。

数据库技术在考古工作中实现所需的硬件条件比较低,使用 Office 系列的 Access 技术要求也很低,而 100 万条数据的管理量即足够使用,投入很低,但会很大程度的提高工作效率。

四、数据库的设计

数据库的设计取决于实际工作的需求。对于一般的文物藏品管理主要需要解决文物编号、类别、名称、级别、保存状况、存放情况、既往研究等项目的编目管理,可以通过建立一个键的核心数据库,如编号、类别、名称、存放情况构成的一个基本数据库来索引全部文物藏品,而级别、保存状况、既往研究等项目可以通过建立分类数据库来实现,通过建立关系数据库的方式来使用核心数据库统纳全部藏品。这样可以保持核心数据库的稳定性,同时实现了分类数据库的可扩展性,最终通过建立查询数据库的方式,按照不同的需求将数据库的内容推送至操作者的面前。建立这样的一个数据库是比较简便和易于维护的,但是这样的数据库仅适合能够自行建立数据库并进行维护的操作者,如果要推广至普通的用户则面临易用性的障碍,要解决这一问题必须设计窗体。

窗体的设计比较复杂,它不单是数据库的呈现问题,它的图形化界面涉及到操作的易用程度、用户感受、使用习惯、美观度等问题,直接影响着用户对数据库的接受程度,功能强大但操作复杂,或者操作简便但功能单一都面临淘汰。怎样在两者之间寻找平衡是设计的关键,同样因窗体设计的复杂程度影响着可自由设计数据库的人员数量。但是这些问题只是一些技术上的困难,如果要建立更为复杂的数据库,那么实现需求,难度将远远超过技术困难。

数据库设计中最为困难的是如何处理文物管理工作中所面对的复杂多样的文物自身状态。如年代的问题,文物的年代主要有如下一些表示法:1. 使用放射性方法测定的年代,如 C^{14} 测定得出的时间段;2. 考古研究给出的文化或期别,如半坡四期;3. 历史纪年时期如唐开元或汉代;

4. 准确年代，如 1840 年。这些年代表示使得我们仅仅能够对其进行记录、索引而不能对其进行便捷的机器排序或统计。这种困难具体表现在：1. 半坡四期是仰韶文化半坡类型四期还是仰韶时代半坡文化四期类型；2. 开元通宝的纪年是以其铸造年代记还是使用年代，而开元通宝本身又不是开元年间所发行。3. 汉代涵盖了西汉、东汉，而时间跨度达到数百年。研究的不足、文物本身所承载信息的复杂等等都造成了数据库深层次利用难以实现。除文物自身承载信息的复杂多样外，人们对文物的判读也造成了千差万别的描述和命名方式。如陶器，其颜色关系陶器烧制的火候、工艺等问题，而陶器的颜色因不同的质地，不同的火候、不同地域等因素本身就千变万化，而判读者有自身的喜好、文化修养、以至一时的情绪都造成了描述的差别，如此判读的颜色信息就失去了进行检索所需要的准确性。而陶器的器型、纹饰的复杂程度再加上读者的个人差异就更使得此类数据库的建立更加困难，所以建立此类数据库就目前而言仅适用于单一使用者。

文物自身承载信息的复杂程度、研究的不断深入、文物的判读者这几个方面的限制是目前数据库设计中所面临的主要困难，其不论对于大型数据库的设计还是小型数据库的设计都同样造成了很大困扰，而解决的方法就目前而言可能还遥遥无期。对小型数据库的而言技术力量的薄弱使得窗体的设计和功能的实现也成为了瓶颈。那么这些问题目前有没有一种可行的、有效的解决办法，笔者认为还是可以找到的。

五、对文物考古数据库发展的展望

目前我们的数据库建设仅仅能够有效地满足人们对藏品管理的需要，如果需要对藏品进行出于一定研究目的的分类、索引、统计、甚至比较，还是较难实现的。但是如果设定一套规则，在一定程度上规范对藏品的描述，还是可以在有限的范围内实现以上目标的。如我们对可能输入的年代进行阶段划分，将细致的年代描述都

归纳至较大的年代描述中，如将“西汉中期”归入“西汉”，将“半坡四期”归入“仰韶晚期”。对于器物的描述问题，色彩则直接放弃细微的描述，直接归入红色、黄色、黑色、灰色等几个简单的类别，对器型和纹饰的描述也简便为仅描述主要的特征，唇、沿等更能体现陶器特点的特征适当抛弃。这样做的好处是我们可以尽可能大的建立一个通用的系统，并提供给尽可能多的人使用，缺点就是失去了很多必要的特征，这个缺憾可以通过提供线图或者照片来弥补。

目前计算机技术在飞跃发展，关于进行多特征图像检索的研究已经有所开展，尤为重要的是目前一些大型的商业公司已经开始着手研制可用于互联网的图形图像、声音视频的检索技术，在 Google 公司的免费产品 Picasa 中运用了可定制关键字、唯一识别编码、搜索重复图片及不同色彩等功能，这一技术的进一步发展必然会诞生可以通过复合特征进行准确图像搜索及比较的技术。但是这一技术未来使用的困难主要在于计算机硬件和网络速度的限制，我们目前所配备的硬件设备将不能实现新技术的有效利用。

如果可以在尽可能大的范围内建立这样一个通用型的数据库，不论对于田野发掘、室内整理还是文物管理都能提供一个基础平台。这个平台应该建立几个不同的版本，分别适用于发掘、整理和文物管理等不同工作，而这个平台还应该是开放的，提供简便的工具允许不同的使用者根据需要定制自己特需的内容，同时数据库还需要提供一种与文字处理、电子表格等软件便捷的进行互换数据的工具。

这个大型的数据库必须为未来的技术留下足够的发展空间，技术的发展速度惊人，如果在最初的设计中没有考虑到这一问题，那么可能在这一大型数据库建立之时，新的技术已经对这一数据库形成了挑战。

数据库是建立于需求之上，进而试图解决复杂机械劳动的利器，同时它还是技术发展链条中的一员，必须将其嵌入整个技术发展的潮流中去，做到与时俱进。

(责任编辑 宋远茹)