

基于数据挖掘技术的图书馆改进方案

李体红, 丰树谦

(焦作大学 信息工程学院, 河南 焦作 454000)

摘要: 基于数据挖掘技术, 以 SQL 2000 数据库为基础, 设计了一种图书馆改进方案。详细描述了借阅记录、个性化服务、馆藏布置、图书分布和决策支持等管理和服务方面设计思路, 初步建立了图书馆综合化的管理服务系统。

关键词: 图书馆; 数据挖掘技术; 管理服务系统

中图分类号: TP392

文献标识码: A

文章编号: 1674-7720(2011)03-0004-02

Improved scheme of library based on data mining technology

Li Tihong, Feng Shuqian

(School of Information Engineering, Jiaozuo University, Jiaozuo 454000, China)

Abstract: Improved scheme of library was designed based on data mining technology and SQL 2000 database. Design idea of management and service was described, such as lending record, personalized service, library layout, books distribution and decision support, the combined platform of management and service was established.

Key words: library; data mining technology; platform of management and service

随着计算机技术和通信技术的发展, 高校图书馆的职能已从传统的读者服务和教育功能逐渐向读者提高、领导决策的职能转变, 在丰富的信息资源中为用户提供有效的信息和知识参考咨询, 完善图书馆的资源建设, 进行数字化的科学管理^[1-2]。现代图书馆正向自动化、数字化、信息化方向发展, 数据挖掘技术可在促进职能转变、建设图书馆资源、提高服务质量、拓宽服务范围等方面提供强有力的技术和服

务支持。本文针对焦作大学图书馆管理和建设工作的现状和不足, 采用数据挖掘技术, 提出一种图书馆改进方案设计, 为校图书馆的资源建设、读者服务提供有效保证, 并根据采集的信息和数据, 为教学和科研的发展提供依据。

1 数据挖掘技术概述

1.1 数据挖掘技术的定义

数据挖掘^[3-5]DM(Data Mining)是从大量数据中提取隐含在其中的事先不知道但又有潜在的有用信息和知识的过程, 而这些信息和知识是不完全的、有噪声的、模糊的、随机的。发现的知识主要用于信息管理、查询优化、决策支持、过程控制等, 也可以用于数据的自我维

护。狭义上的数据挖掘是数据库知识发现 KDD (Knowledge Discovery in Database)的一个关键步骤, 是提取数据和建立模型的重要环节。

1.2 数据挖掘的过程

数据挖掘大致可以分为四个步骤: (1) 资源发现, 即提取信息所需要的数据; (2) 信息的选择和预处理, 即从检索到的数据和信息中筛选和处理有用的信息; (3) 模型化, 即从单个或多个信息发现其中的规律, 建立普遍的模型; (4) 评价, 即对挖掘出的模型进行确认或者解释。

根据挖掘的类型不同, 数据挖掘的方法主要有以下四种^[6]:

(1) 关联分析。从大量不同属性数据中发现数据之间的关联性, 挖掘出数据间潜在的相互关系或者结构, 如超级市场的“购物篮”问题研究。

(2) 序列分析。从一定时间间隔内接连发生的事件中发现各数据之间的联系。把这些事件用一个序列表示, 序列中的各事件除具有时间属性外, 还具有统计上的概率性, 与关联分析的算法相似。

(3) 分类分析。把具有相同属性的样本看作同一个类别, 通过分析类别样本的特点, 得到决定样本属性的规

则和方法。如全国受教育程度分析中,可以依据文化程度来划分样本。

(4)聚类分析。与分类分析不同,聚类分析是把数据按照其相似性和差异性分成若干具有一定意义的子集,然后再分别描述不同的类别。

2 图书馆改进设计方案

焦作大学图书馆始建于1981年,经过近30年的发展,拥有了丰富的馆藏资源,可提供图书借阅、期刊查询、读者服务等基本功能,为全校教学和科研提供了强有力的支撑。针对我校图书馆的资源布置、信息检索和服务方面的缺点,采用数据挖掘技术,在现有图书馆资源和管理的的基础上,提出了图书馆改进方案。方案设计原理框图如图1所示。

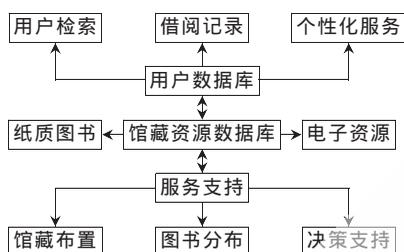


图1 方案设计原理图

系统以SQL 2000数据库为基础,分为用户数据库和馆藏资源数据库,并根据两个数据库的信息提供服务支持。用户数据库主要用来为图书馆馆藏资源的使用者提供有用户检索、借阅记录和个性化等主要功能;馆藏资源数据库用于搜集和提供图书馆的资源,包括纸质图书、电子资源等;服务支持根据系统搜集的信息,提高图书馆在馆舍布置、图书分布及决策服务等管理方面的服务,促进图书馆的科学管理。在图1所示的方案设计中,数据库挖掘技术主要应用于借阅记录、个性化服务和服

2.1 借阅记录

读者借阅记录存储于用户数据库中,每条记录包含读者的标识码和借阅图书的标识码,如表1所示。

表1 借阅记录用户数据库

读者标识码	图书标识码
s070401	A0105124, B0205652, ...
s070420	A0301821, C0402513
s070514	D0301025
...	...

采用数据挖掘技术挖掘读者信息和对应的图书信息时,为提高挖掘效率,采用数据表形式存放挖掘记录,并根据读者的标识码和图书的分类及存放地址的表链接,如表2和表3所示。

借阅记录采用关联分析方法,通过唯一的读者标识码在数据库内部自动把三者连接起来,显示完整的图书借阅信息供管理人员、读者查询。同时,数据库自动记录

表2 读者信息表

标识码	姓名	类别	年级	院系	日期
S070401	甲	A1	2	A	08/05/12
S070514	乙	B1	3	B	09/10/23
...

表3 图书信息表

标识码	借阅者	书名	类目	册数	地址
S070401	甲	A1	文学	2	馆1
S070514	乙	B2	机械	1	馆3
...

图书的借阅频率、图书类别的使用频率和图书的交叉借阅频率,并存储为管理人员提供的参考信息。

2.2 个性化服务

采用数据挖掘技术,根据数据库记录的图书借阅频率和图书类别使用频率,数据库在读者查询相关图书时,自动把频率较高的图书置前,并根据读者的兴趣爱好向读者推荐相关图书。当读者所需要的图书已被全部借阅或者图书馆没有收藏时,自动向读者推荐3~5种的相关替代图书或电子资源。

2.3 服务支持

服务支持主要包括图书馆馆藏布置、图书分布和决策服务。

馆藏布置根据数据库统计的图书类别使用频率和图书交叉借阅频率来布置不同类别的图书库在图书馆的位置,使关联度较高的图书库紧邻。如借阅了文学类图书的读者同时借阅艺术类图书的频率较高,则可将文学图书库和艺术图书库相邻。同时,对于同一类目的图书,可根据使用率的大小在图书库内设置阅览桌的数量,使图书馆利用更加合理。

图书分布根据同一类目的图书使用率,把借阅率高的图书放在图书馆的前面,方便读者借阅的同时也提高了图书馆的利用率。

决策支持为图书馆的科学管理提供可靠依据。根据图书借阅频率和图书类别使用频率,图书馆管理人员在购买新的图书时可优先考虑频率较高的相关图书和图书类目,使资金利用效果最大化。

基于数据挖掘技术的图书馆比传统图书馆在图书管理、服务等方面具有更大的优势。通过关联分析方法,使图书借阅记录成为图书馆管理和服务的平台,为个性化服务、馆藏布置、图书分布、决策服务等提供了可靠的参考,使图书馆成为一个综合化的管理和服务平台。本方案在建立高效的数据挖掘规则和服务体系方面仍需进一步研究。

参考文献

- [1] 司徒浩臻. 数据挖掘技术在图书馆信息服务中的应用[J]. 现代图书情报技术, 2005, 10(129): 15-19.

- [2] 董云鹏.数据挖掘技术在图书馆中的应用[J].现代情报, 2006, 11(11): 131-132.
- [3] 陈文伟,黄金才,赵新昱.数据挖掘技术[M].北京:北京工业大学出版社,2002: 189-200.
- [4] 孟晓明.浅谈数据挖掘技术[J].计算机应用与软件, 2004, 24(8): 34-36.
- [5] 唐笑林.数据挖掘技术的研究与应用[J].华东理工大学学报(自然科学版), 2008, 34(2): 290-295.
- [6] 梁协雄,雷汝焕,曹长修.现代数据挖掘技术研究进展

[J].重庆大学学报,2004,27(3): 21-27.

(收稿日期:2010-10-15)

作者简介:

李体红,女,1965年生,副教授,主要研究方向:软件工程、操作系统。

丰树谦,男,1965年生,副教授,硕士,主要研究方向:软件工程、数据库工程。

