

# A Novel Digital Library Personalized Service Scheme

Liu Qing-Ao, Lei San-Ya

The First Aeronautic Institute of the Air Force, Xinyang, 464000  
123907828@qq.com

**Abstract:** In this paper, the characteristic of digital library personalized service and the main methods of data mining are introduced at first. By analyzing the requirement of personalized service and the use of data mining, a personalized service scheme using data mining is proposed. Based on computer network, this service scheme utilizes the database to produce service rule secondly. In this scheme, by mining coherent knowledge base, library historical data, borrowing data, reading log and feedbacking information of each readers, a service rule is produced providing personalized scheme for every readers and the produced service can be updated by information status. The service rule can give every reader some book list or journal list in which the readers may be interested. So when new books, journals or electronic magazines are updated by the digital library, according to the service rule this service scheme can select some of them to advice the readers to read or borrow.

**Keywords:** digital library; personalized service; data mining

## 一种新的数字图书馆个性化服务方案

刘庆敖, 雷三丫

空军第一航空学院, 河南 信阳, 464000  
123907828@qq.com

**【摘要】**介绍了数字化图书馆个性化服务的特征和数据挖掘的基本方法,提出了一种采用数据挖掘的数字化图书馆个性化服务方案。该方案利用计算机网络和数据库,采用数据挖掘的方法,通过对相关知识库、图书馆历史数据、读者借阅数据、读者日志、读者反馈信息的挖掘,生成服务规则,为每个读者提供个性化服务方案,并根据信息的状况进行服务更新。这种服务规则可以在数字化图书馆的图书馆、杂志或者电子资料更新时,选取一些推荐给读者,供他们阅读或者借阅。

**【关键词】**数字化图书馆;个性化服务;数据挖掘

### 1 引言

随着校园网信息化建设工作的逐步落实,图书馆的信息化建设也被提到议事日程上来,尤其是在一些高校,图书馆的信息化建设也是网络建设其中的一个重点。与传统的图书馆相比,除了信息量呈几何倍数增长外,经过被信息化改良过的图书馆还拥有了新的名字“数字图书馆”。数字化图书馆是在互联网的支撑下,以内容管理为核心,以海量信息处理、知识发现与加工交流为主要技术手段的智能知识服务基础平台,是运行在互联网上的、超大规模的、便于使用的、没有时空限制的知识中心。服务是数字图书馆工作的核心。目前数字图书馆信息服务模式经过了分散分布式的“资源/产品中心”“馆员中心”服务模式实践,正在

向集中式的“读者中心”服务模式深化发展。

数字化图书馆以文献的数字化为标志,图书馆通过网络向每个读者提供各种数字图书、数字化期刊等资料,实现学习的数字化。数字化图书馆以网络技术、数据库技术为基础。数字化图书馆以读者的自我学习为服务对象,而个性化服务是数字化图书馆所提供服务的最重要特征。

个性化服务是数字图书馆系统中的关键部分,信息服务的智能化,由单纯的信息提供转向信息生成。数字图书馆的个性化服务主要表现为两个层次:第一层次为按读者要求进行信息订制。读者根据自己的需要订制专门信息,其功能包括数字图书馆站内搜索,Internet搜索,时间、日期、重要事件的提示,并可帮

助读者建立个性化信息空间。第二层次则是数字图书馆挖掘读者兴趣模式，主动提供服务，使数字图书馆成为一个智能型、主动性的信息提供商。个性化服务可以提高信息获取速度和拓展服务形式。

## 2 数据挖掘

伴随着数据仓库的诞生，数据挖掘也得到迅速的发展。数据仓库的发展为数据挖掘技术开辟了新的战场，同时也提出了新的要求和挑战。数据挖掘就是按企业既定业务目标，对大量的企业数据进行探索和分析，揭示隐藏的未知的或验证已知的商业规律，且进一步将其模式化的数据处理方法，它的最大特点是能够建立预测模型，预测未来的情况。

数据挖掘与知识发现有着不可分割的联系，有的认为，数据挖掘与知识发现是等价的概念。人工智能领域习惯称知识发现，而数据库领域习惯称数据挖掘，也有的把知识发现当作发现知识的完整过程，而数据挖掘只是这个过程的一部分。知识发现的过程可以分为三个主要阶段：数据准备，数据挖掘和结果表达和理解。

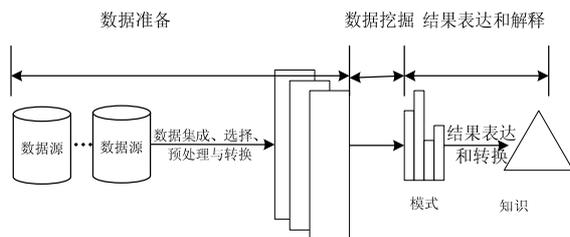


图 1 知识发现过程  
Fig.1 Process of knowledge finding

数据准备首先需要将多文件或多数据库运行环境中的数据合并处理，解决语义模糊性，处理数据中的遗漏和清洗脏数据等，这个过程称为数据集成；然后为知识发现的目标搜索和选择有关的数据，这包括不同模式数据的转换和数据的统一和汇总，进行数据选择，其目的是辨别出需要分析的数据集合，缩小处理范围，提高数据挖掘的质量；通过数据预处理对数据进行清理和充实；通过数据转换实现数据编码，数据库中字段的不同取值转换成数码形式将有利于搜索。知识发现的第二个阶段进行实际的挖掘操作，利用机器学习从数据库中发现有用的模式或知识。

最后一个阶段根据最终读者的决策目的对提取的

信息进行分析，把最有价值的信息区分出来，并且通过决策支持工具提交给决策者。这一步骤的任务不仅是把结果表达出来，还要对信息进行过滤处理。如果不能另决策者满意，需要重复以上数据挖掘的过程。

## 3 提出的数字化图书馆个性化服务方案

根据数字图书馆个性化服务要求和结构，提出了基于数据挖掘的个性化服务方案。该方案通过对相关知识库、图书馆历史数据、读者借阅数据、读者日志、读者反馈信息的挖掘，生成服务规则，为每个读者定制个性化服务方案，并根据信息的更新状况进行服务更新。其服务方案如下：

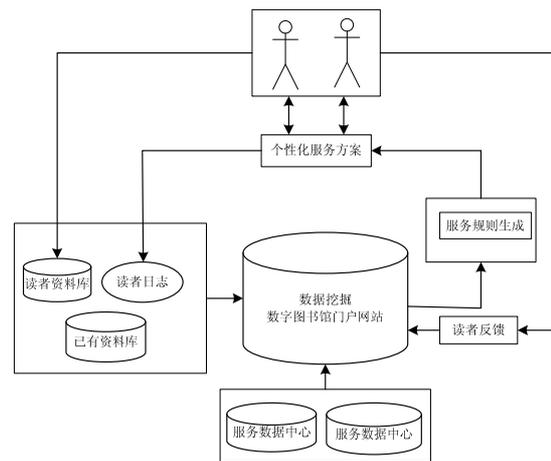


图 2. 提出的个性化服务方案  
Figure 2. Proposed Personalized Service Project

### (1) 个性化服务方案

个性化服务方案是为读者提供的个性化服务的集成，个性化服务方案是根据读者定制、读者反馈和相关数据挖掘来生成的。每一项服务中可以根据不同的考虑指定不同的选择参数。读者个性化服务方案是一个及时调整的动态方案，是个性化的体现点。如根据每个读者的阅读习惯、阅读知识范围，定期向每个读者推荐图书、杂志等相关资料，便于读者学习。

### (2) 数字图书馆门户网站

数字图书馆门户网站是囊括了所有读者个性化方案的整体，同时提供公共的图书查询等通用性服务，用以对外广播和发布如图书更新、杂志更新等信息。

### (3) 读者资料库

读者资料库包含读者一些基本信息，这些信息是通过一定加工整理而生成的，是准确的、规则的信息，通常是最基本的读者资料，其来源主要是读者的注册提交。读者资料库的准确性是数据挖掘实现的前提，因此，应对读者提供的资料进行鉴别。

#### (4) 读者日志

读者日志是纪录读者的动态交互信息形成的信息资料,包括读者数字图书馆门户和利用个性化服务方案的日志信息。读者日志是一个不规则的信息资料库,要实现数据挖掘必须先经过信息的抽取,形成可以用于数据挖掘的规则信息库。

#### (5) 已有的知识库

已有的知识库是已经存在的知识库,包括读者阅读习惯、读者来访规律等信息。主要是用来指定个性化服务规则的参考和验证。

#### (6) 服务数据中心

业务数据中心是实际业务发生的数据,也包括企业内部经过信息抽取或者规则化后的管理内容和数据。

#### (7) 读者反馈

读者反馈是读者针对个性化服务方案的反馈,例如对个性化服务方案的评判,读者选择个性化服务参数的特点等。读者反馈信息表达读者对数字化图书馆的需求,是读者学习内在需求的外在表现,对于这些需求,可以通过数据挖掘的方法生成指导读者学习的方案。

#### (8) 数据挖掘

数据挖掘是针对读者反馈、业务数据中心、已有的知识库:读者资料库及读者日志等信息充分利用相关的数据挖掘方法来生成读者的个性化服务方案规则,这些规则包括读者的兴趣所在,读者可能的潜在需求,读者关注不同服务的权重、读者从未关注的信息等。

#### (9) 读者规则生成

读者规则生成是制定读者个性户服务的依据。

在个性化服务系统中,数据挖掘技术应用于挖掘基于时间序列的数据、基于关联规则的数据等,数据挖掘方法有自动聚集、决策树、神经网络等。挖掘基于时间序列的数据是从时间角度来分析数据的变化发展规律,发现数据之间潜在的规则和联系。

## 4 结束语

个性化服务是数字化图书馆的重要特征,本文介绍了数字化图书馆个性化服务的特征和数据挖掘的基本方法,提出了一种采用数据挖掘的数字化图书馆个性化服务方案。该方案通过对相关知识库、图书馆历史数据、读者借阅数据、读者日志、读者反馈信息的挖掘,生成服务规则,为每个读者订购个性化服务方案,并根据信息的更新状况进行服务更新。

## References (参考文献)

- [1] C. L. Borgman, The invisible library: Paradox of the global information infrastructure[J]. *Library Trends*, 51 (4),2003, 652-653.
- [2] M. A. Keller, V. Reich, and A. C. Herkovic. What is a library anymore, anyway?[J] *First Monday*, 2003, May 5,8-13.
- [3] G. McCalla. The Ecological Approach to the Design of E-Learning Environments: Purpose-based Capture and Use of the Information about Learners[J]. *Journal of Interactive Media in Education*, 2004,7-11.
- [4] C. L. Borgman. Digital libraries and the continuum of scholarly communication [J]. *Journal of Documentation*, Vol.56 (4),2000, 412-430.
- [5] Tzu-Fang Sheu , Nen-Fu Huang , Hsiao-Ping Lee, Hierarchical multi-pattern matching algorithm for network content inspection[J], *Information Sciences: an International Journal*, v.178 n.14, p.2880-2898, July, 2008.