

基于大数据挖掘与决策分析体系的高校图书馆个性化服务

文/张怡华

摘要

互联网科技在人们身边无处不在,大数据已经潜移默化的来到人们身边。图书馆作为现代社会的有价值的信息系统,通过大量的非结构化数据、半结构化数据可以帮助我们寻找隐藏在数据背后的世界。本文分析基于大数据挖掘与决策分析体系的高校图书馆个性化服务研究,希望能为以后研究提供参考。

【关键词】大数据 挖掘 决策 个性化服务

1 大数据时代图书馆面临的挑战和个性化需求

1.1 大数据时代服务平台的系统结构更加复杂和开放

1.1.1 服务架构和应用程序复杂度快速增长

目前的软件体系结构越来越难以应对和处理日益增长的软件复杂性,如何实现图书馆基础设施硬件个性化需求是我们孜孜不倦的追求。随着微服务架构的兴起,大规模整体式应用程序向着分布式微服务架构迁移,加大了系统结构的复杂性。

1.1.2 网络和大数据资源的开放性

随着互联网技术的飞速发展,可以利用网络进行传播,实现资源的最大共享。但是有利必有其弊,大数据的开放性涉及数据安全和隐私保护。相比于传统的数据资源形式,云计算、大数据、传感技术具有开放性、复杂性、多样性和海量性等特征。

1.1.3 以非结构化数据为主体

非结构化数据处理是大数据分析技术的难点和重点也逐渐得到客户对其价值的认可。过去那种有限内容和结构单一的数据库已经无法满足客户多样化需求,尤其是在图书馆个性化服务更是以非结构化数据为主体。

1.1.4 改变了传统信息安全交付模式

大数据时代的到来,给图书馆的服务带来了更大的挑战和机遇,一方面是大数据带来的个性化服务,另一方面是大数据改变了图书馆在传统IT情况下的信息安全交付模式。

1.2 图书馆大数据QOS(服务质量)保障与用户隐私保护冲突激

1.2.1 读者个性化服务中数据被过度分析

如果对用户数据存在过度的分析和使用,不仅仅不能帮助图书馆实现个性化服务,反而会导致分析的结果失真。更有甚者,引起用户隐私的泄露,带来被控诉的官司。

1.2.2 黑客会最大限度地收集图书馆和读者的

个人信息

大数据环境下经常出现图书馆用户隐私泄露,如何有效保护读者隐私权,已成为加强图书馆与读者之间信任感的重要武器。大数据时代,信息获取渠道增多,给图书馆用户的隐私保护带来极大的风险,进行数据分析和数据挖掘应该在不暴露用户个人隐私的前提下。

1.3 大数据挖掘与决策分析的图书馆个性化需求

1.3.1 辅助事前预测

我们应该利用此项技术来对自身的数据进行整理、组织、分析,为图书馆服务的模式、对未来发展趋势提供分析与预测。寻找用户的诉求,进而改进服务规模,提高服务质量,达到图书馆对大众的资源和服务的双向满足。

1.3.2 辅助事中感知

大数据的容量如此庞大,在海量数据中对有价值的数据进行甄别与筛选,再对其进行科学系统的挖掘,绝非易事,因此要提早作好各方面的准备。比如天津图书馆每个周六都有讲座活动,所以那一天来图书馆的读者比较多。中文报纸阅览室正对着报告厅,每一场报告结束,许多读者还会直接走进阅览室。有的老读者会直接向你反映这场报告会都讲了什么、感受如何,打动人心的地方在哪儿。这些读者的反馈是图书馆讲座服务的一面镜子,这是图书馆人在工作中所作的小数据集。

1.3.3 辅助事后反馈

在大数据中“沙里淘金”,需要事后实时反馈,事后做好相关文字记录,跟踪设备的记录。大数据还能图书馆构建全新的知识服务引擎提供必要的技术支撑。并且在技术应用与管理过程中还可以不断反馈信息

2 高校图书馆大数据挖掘及决策分析体系的架构和流程

2.1 高校图书馆的大数据挖掘及决策分析流程

2.1.1 多维度提取数据来源

从从各类业务和大数据存储交换平台内提取数据,或者从外部互联网(直接装载爬虫引擎)提取数据。

2.1.2 评估数据规模和样例分析

高校图书管理系统,用综合效益评价体系和评价方法,根据规模大小选择合适的分布式并行计算应用架构,也进行样例分析。

2.1.3 根据实际需求确定建模方式

读者用户监控和分析、分类、预测及辅助决策需要用数学建模方法,我们需要一种全新的面向Agent的需求建模方式。

2.1.4 根据需求确定输出方式和优化

确定了各评价指标的权重,运用TOPSIS方法对直接输出分析报告,确定一套科学的质量评价指标。

2.2 高校图书馆的大数据挖掘及决策分析的架构

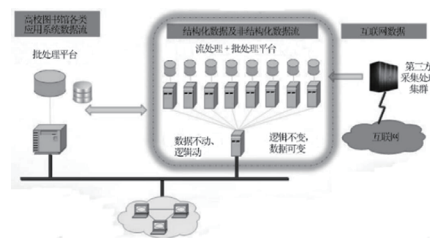


图 1

图书馆的大数据服务体系,其服务与运行模式的核心是客户,将图书的搜寻、查找、借阅、管理等流程加以优化,并有机的结合起来,利用云技术进行分析、预测和智能的辅助服务,建立具有特色的科学性很强的实用图书管理模块。大数据计算服务后,在大数据计算服务内进行一些数据计算和算法模型的训练。然后再将计算的结果导入ADS和云数据库中,其中导入的ADS支持BI系统,并且ADS能够多值列查询和毫秒级的实时响应,有利于生成BI报表;另一部分数据存入云数据库中,不仅降低了存储成本,同时也提高了数据的安全性。

3 大数据挖掘及决策分析体系在高校图书馆个性化服务中的应用

3.1 基于用户行为模型的个性化服务方案

3.1.1 用户行为本体数据库构建过程

如何构建合适的用户行为模型并基于海量的行为日志数据提供个性化服务,其本质是从异质的市场中寻找用户行为习惯及特点等诸多。构建大数据的存储、搜索、共享、分析和可视化等管理行为,科学构建CRM大数据库和管理信息系统。

3.1.2 用户显性兴趣和隐性兴趣本体提取过程

通过分析用户的浏览方式和内容,实时获取用户兴趣信息,设计一种利用用户日志库提取用户显性兴趣和隐性需求本体的个性化服务方案。高校图书馆大数据应用模式框架抓取工具,提出数字图书馆个性化用户兴趣领域本体的动态过程。

参考文献

- [1] 孙琳. 大数据时代图书馆服务体系创新研究[J]. 理论观察, 2013(04).
- [2] 胡莲香. 走向大数据知识服务: 大数据时代图书馆服务模式创新[J]. 农业图书情报学刊, 2014(02).

作者简介

张怡华(1978-),女,四川省广元市人。助理研究员,硕士学位。研究方向为计算机应用、数字图书馆技术与管理。

作者单位

成都信息工程大学 四川省成都市 610225