

高校数字图书馆个性化服务模型探究

孙士新,李海燕

(亳州职业技术学院 信息工程系,安徽 亳州 236800)

摘 要:信息资源的爆炸式增长,数字图书馆的传统模式已经不再适应具有较高信息素养的读者的需求。通过对传统高校数字图书馆个性化服务系统中的用户模型、图书使用模型、推送方式的研究,分析现有图书馆个性化服务系统在用户需求模型、用户使用模型、推送方式等方面的不足,提出新的图书馆个性化服务模型,并列挙新的个性化服务模型的优势所在。

关键词:数字图书馆;信息资源;服务模型

中图分类号:G251

文献标志码:A

文章编号:1672-8580(2011)01-0131-03

随着信息技术的发展,图书馆信息化水平的提高,个性化服务在高校数字图书馆中得到了充分的利用,但仍然存在诸多的不足之处。从传统高校数字图书馆个性化信息服务系统现状出发,分析现有图书馆个性化服务系统存在的问题,探究高校数字图书馆个性化信息服务系统模型。

一、传统高校数字图书馆个性化信息服务系统

高校数字图书馆个性化信息服务系统主要经历用户建模、图书建模、信息个性化推送三个环节。用户建模是收集和修改与用户习惯、兴趣、需求相关的信息的过程,其结果将产生一个表示用户特有背景知识或兴趣、需求的用户模型。图书建模是获取和维护与图书利用率、利用时间、图书需求相关的信息过程,其结果将产生一个表示图书特有背景、利用率和图书需求的图书模型^[1]。信息个性化推送阶段将以上两个模型为依据,运用各种推送技术寻找出与其匹配的资源,将这些资源推送给各自感兴趣的用户。

1. 高校数字图书馆个性化信息服务中的用户模型

传统高校数字图书馆个性化信息服务系统中用户访问的信息一般都是非结构化的信息,用户浏览网站一段时间后,系统利用明获取方式或暗获取方式提取用户兴趣。明获取是根据用户注册信息、留言、打分的方式进行;暗获取是系统通过分析用户的访问日志、内容、连接等提取用户兴趣。在提取用户兴趣的基础上,采用各种方法为用户建立兴趣模型。

高校数字图书馆个性化信息服务系统的用户兴趣

模型主要采用主题表示法、加权关键词向量表示法、关键词表示法等方法。随着用户访问次数的增多,用户兴趣的连续变化,用户行为也在不同程度的改变,所以,其用户模型是一个不断被建立和维护的模型,系统定时地搜索用户的反馈信息和对信息推送的评价情况,及时更新用户兴趣文件,以保证推送信息的即时性。

2. 高校数字图书馆个性化信息服务中的图书使用模型

传统高校数字图书馆个性化信息服务系统只记录图书的访问量、读者对新书的需求,系统根据读者输入关键词去检索相关书籍信息。图书管理人员可以通过系统统计随时了解图书的访问情况,系统获取用户兴趣的方法有明获取和暗获取两种。明获取是用户提交想看的书籍清单、对搜索书籍的评价等方式来获取;暗获取是系统记录书籍的访问次数、访问时间等提取图书使用情况。根据提取的图书信息,图书管理员可以随时了解图书的闲置信息。

高校数字图书馆个性化信息服务系统采用关联规则挖掘算法来提取图书使用情况。传统数字图书馆个性化服务系统并没有真正建立图书使用模型,一般只记录了图书访问量高低。通过对国内几所高校图书馆电话调查发现,大多数图书馆并没有对访问率很低的书籍进行重新推送,甚至把这些书籍直接淘汰。在称呼上,清华大学称访问量极低的图书为呆滞图书,北京大学称存放访问量极低的书籍的书库为密集书库。

3. 高校数字图书馆个性化信息服务中的推送方式

收稿日期:2010-04-15

作者简介:孙士新(1980-),男,安徽亳州人,主要研究方向为职业教育技术、数据挖掘。

信息推送(Information push)是通过一定的技术标准或协议,在 Internet 上通过定期传送用户需要的信息来减少信息过载的一项新技术。信息推送服务是运用退送技术来实现的一种个性化主动信息服务方式^[2]。信息推送技术也是第三代浏览器的核心技术,它能够根据用户兴趣需求,主动的将用户需求的信息分类推送到用户设备中,从而改变了系统被动提供信息的局面,有效地改变着人们获取信息的方式。常用的推送方式有邮件推送、手机短信推送、用户专用网页推送、频道式推送等形式。

目前存在的高校数字图书馆个性化服务系统主要有两种:一是基于 RSS 技术的图书馆个性化服务系统^[2-3],采用此系统的国内高校主要有:清华大学、复旦大学等;二是 MyLibrary^[4],此系统最早在中科院国际科学数字图书馆、天津泰达、浙江大学等地使用,目前比较完善使用的有浙江大学、郑州大学、中南民族大学、武汉大学、清华灵捷图书馆等。

二、传统数字图书馆个性化服务系统的不足

传统高校数字图书馆个性化信息服务系统在一定程度上改善了信息推送方式,但仍然存在不足,其中突出的问题有以下 3 点。

1. 构建用户需求模型困难

用户在访问中搜索的信息都是非结构化的信息,系统提取用户需求时需要花费大量的精力在信息的预处理上。网站信息的无序性和非结构化给个性化服务带来了困难。

2. 图书使用模型有待建立

现有高校数字图书馆个性化信息服务系统只对图书访问量进行统计,并没有利用访问量与时间、读者群的关系建立图书个性化模型。图书管理员只统计了某图书的访问次数,而没有建立图书访问量与时间的对应序列,没有挖掘图书访问量随时间的变化趋势。

3. 信息推送动态性和时效性欠佳

传统的信息推送方式采用邮件方式、频道推送和用户专用信息网页。这些方式缺乏动态性和时效性,不能即时、无缝地把信息推送给用户。

基于上述问题,本文设计了一个高校数字图书馆个性化信息服务系统,该系统将用户个性化服务、图书使用模型、RSS 技术结合,能更好地为用户提供个性化服务、展示图书资源个性化、及时把最新信息主动推送给用户,促进图书资源整合,真正实现高校数字图书馆信息推送的个性化。

三、高校数字图书馆个性化服务系统模型

高校数字图书馆个性化信息服务系统模型主要包括用户兴趣模型、图书使用模型、个性化推荐和信息推

送模块,系统结构如图 1 所示。

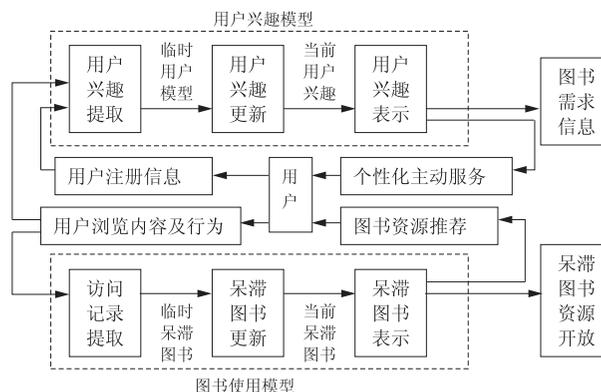


图 1 高校数字图书馆个性化信息服务系统模型

本系统的应用背景是高校数字图书馆个性化服务,面对的用户是学生、教师、科研工作人员。系统还提供了友好的搜索界面,用户可以选择分类或输入相关的关键词搜索图书信息。系统收集和处理了两个方面的数据,即用户有意识提交的数据和用户无意识提交的数据;用户有意识提交的数据有用户注册信息、用户对网页的评价等;用户无意识提交的数据有用户搜索信息时输入的关键词、用户访问网页的行为等。系统通过这两个方面的数据获取用户兴趣。依据用户兴趣模型,采用个性化推荐技术在系统图书资源库中查找与用户兴趣模型匹配的图书资源。RSS 信息推送主要通过资源解析、元数据抽取、信息封装三个过程,最后把 RSS 文件向用户推送。用户可以通过浏览器随时查看自己感兴趣的信息资源。

1. 用户兴趣模型

用户兴趣模型是按照用户查阅信息源——图书信息,采用相关方法提取用户兴趣特征,最终建立用户兴趣建模的过程。系统资源库中的信息分两类:基于 RSS 格式的信息源和非 RSS 格式信息源。资源库中 RSS 格式的信息源很少,多数为非 RSS 格式的信息源。RSS 格式的信息是由实际信息的元数据如:标题、摘要、连接等组成的具有标准 RSS 格式 XML 文档;非 RSS 格式信息是一般的无结构的 Web 页面信息。利用 Web 挖掘技术,在 Web 日志中提取隐含的有用信息,比如 Web 用户访问模式、用户感兴趣的相关页面等。系统可以把信息源分以下两种类别来处理:

(1) 非 RSS 格式的非结构化信息

每个图书页面都有一个页面描述,首先通过对页面描述进行文本分析得到各个页面包含的关键词和其对应的权值,然后通过计算相似度合并内容相似的页面,最后针对不同页面上出现的相同关键词,而且权值不同的情况,合并相同关键词的权值。依据以上提取的

关键词和权值而建立用户模型。

(2)RSS 格式的信息

RSS 可以将网站当作一系列频道(channels)的集合,每个频道又包括一系列资源(Items),所以通过对频道和资源的描述可以实现对作为资源集合的网站描述^[9]。系统在注册信息模块为用户提供了兴趣关键词输入的入口,用户也可以通过搜索关键词栏目提交自己感兴趣的关键词,系统会为每个用户输入的关键词设置一个权值。系统通过对用户注册信息、定制信息等建立用户兴趣模型。

2.图书使用模型

图书使用模型是借助用户浏览图书的内容和行为获取图书被访问情况,统计每本图书的访问量,把访问量较低的图书提取出来更新呆滞图书库,然后对这些呆滞图书进行重新推送和作为呆滞资源对外开放。系统经过一段时间的运行,记录了图书的访问量、访问时间、读者群等信息。通过对各图书的访问量、访问时间的挖掘,发现了各图书的访问量变化与一年中各时间段的关系,把每个时间段的闲置书籍提取出来,形成图书馆闲置书籍与时间的对应关系序列。图书馆可以在不同的时间段有不同的闲置书籍对外开放或重新推送给相关的读者,以形成图书馆图书信息个性化。

3.个性化推荐

个性化推荐模块的主要功能是资源的组织和相似度计算。按照用户兴趣模型组织数据,借助现有推荐技术计算结果集。个性化推荐模块按照用户兴趣模型在信息库中为用户检索图书信息,把检索到的信息页面进行分析而提取关键词,按照用户兴趣描述检索页面,形成检索结果集。把检索的结果与用户兴趣相比较,计算相似度,把相似度高于预定值的页面即优化结果集推荐给用户。

4.信息推送模块

信息推送是对优化后的结果数据抽取元数据,包

括即标题(title)、链接(link)、摘要(escription)等元素,借助设计的对象类,将用户所需信息转化为 RSS 格式的信息,即 RSS Feed,且根据资源的变化而更新 RSS Feed。用户可以借助 RSS 阅读器,将推送的 RSS Feed 添加到阅读器中,新的信息会源源不断地被推送给用户面,从而节省了用户获取信息的精力和时间。

四、数字图书馆个性化服务系统模型的优势

高校图书馆个性化信息服务系统模型较传统的个性化系统区别在于:在用户建模和信息推送模块中糅合了 RSS 技术,提出了图书使用模型。采用结构化的 RSS 标准简化了用户建模时的预处理工作(即收集用户兴趣)。采用 RSS 包的解析工具,系统可以很容易的获取该 Web 页面的 Title、URL、Description 等信息,经过对这些信息的分析、处理获取用户兴趣。推送模块把得到的用户兴趣信息采用 RSS 标准封装而且推送给用户,使用户无需打开网页便能够查阅到自己感兴趣的信息。提出了图书使用模型,系统提供了时间与图书访问量的变化趋势,为各高校图书资源整合提供决策支持,为提高呆滞图书利用率起促进作用。把 RSS 技术糅合到图书馆个性化服务系统中为读者提供了更及时、更有效的个性化信息服务。

参考文献:

- [1] 欧阳烽.Web 数据挖掘与高校数字图书馆个性化服务[J].数字图书馆技术论坛,2008,(1):103-107.
- [2] 高文,刘峰,黄铁军.数字图书馆原理与技术实现[M].北京:清华大学出版社,2000:10-158.
- [3] 周志峰.基于 RSS 的高校图书馆重点学科信息导航系统研究[J].现代情报,2008,(11):67-72.
- [4] 汪琴.国内外数字图书馆个性化服务研究及展望[J].数字图书馆论坛,2006,(3):43-47.
- [5] 董倩.利用数据挖掘技术创建数字图书馆个性化服务系统[J].科技情报开发与经济,2007,(12):13-14.

责任编辑:梁雁

Research of the Individual Service Model in College Library

SUN Shi-xin, LI Hai-yan

(Bozhou Vocational and Technical College, Bozhou 236800, China)

Abstract: The information resources are increasing like explosion. The traditional model of the digital library hasn't adapted to the need of the readers who have more and more information attainment. Through the research of the user's model of college digital library's individual service system, the book usage model, the pushing style, the analysis of the lack of the user's need model in the existing library individual service system, the user's usage model, it puts forward to the new library individual service model.

Key words: digital Library; information resources; service model