

数字图书馆智能化服务及其关键技术探析

程风刚

(盐城师范学院图书馆,江苏 盐城 224002)

[摘要]智能化服务是一种全新的信息服务模式,它主要借助于网络并通过计算机模拟或实现类似于人的智能行为,为用户提供信息服务。同传统图书馆服务相比,数字图书馆智能化服务具有明显的优势,而先进的技术则是推动其发展的主要动力,本文着重对智能化服务所涉及的关键技术进行了分析。

[关键词]数字图书馆;智能化服务;关键技术

[中图分类号]G250.76

[文献标志码]A

[文章编号]1005-6041(2010)03-0016-03

随着现代信息技术的迅速发展,传统图书馆逐步向数字图书馆转变。为了不断满足读者需求,图书馆应不断引进并整合智能代理、数据挖掘、专家系统、机器学习等先进技术,加快人工智能和信息服务的融合,积极为用户提供智能化服务。本文试从智能化服务的概念入手,着重对智能化服务所涉及的关键技术进行分析。

1 智能化服务的概念及特征

智能化服务是指在网络环境下通过计算机模拟或实现类似于人的智能行为,为不同用户提供信息服务。这是一种全新的信息服务模式,服务系统能迅速、准确地理解用户的意图,自主或交互地执行各种模拟任务,为不同用户提供自动化、网络化、数字化服务,真正体现以用户为中心的服务理念。

在智能化服务过程中,智能是整个服务的灵魂,是区别于传统服务、纯网络服务的重要标志。先进的智能化服务应具备以下几个特征:能主动分析、预测用户需求及期望目标,把相关问题描述出来并提供给用户;强调服务与被服务之间的一一对应,强调量体裁衣式的个性化服务;集成了专家系统、机器学习、人机接口等功能,能自我学习和自我调整,知识库更新较快。

智能化服务系统不是一个简单的系列操作过程,它是由智能断定、获取、精练、开发创新、提供信息和信息应用效益评估等系统组成的完整的有机统一体。^[1]

2 智能化服务关键技术分析

技术是推动数字图书馆发展的主要动力,数字图书馆的智能化服务正是在计算机、网络通信、人工智能等新技术的基础上发展起来的。目前,智能化服务主要涉及智能代理、数据挖掘、资源整合、智能检索、信息过滤、智能推送、信息安全等关键技术。

2.1 智能代理技术

智能代理(Intelligent Agent)常被称为“会思维的软件”。它具有以下主要特征:1)智能性。具有与执行任务相关的知识,能够进行相关的推理、判断或计算,比较准确地理解用户需求,提供针对性服务;2)代理性。能引导并代替用户完成某些任务;3)自主性。在没有外界干预的情况下,能自主决策,自主完成任务;4)通讯能力。每个智能代理都有标准的接口,采用统一的通信语言进行信息的交流,以便获取信息,充实和更新知识库;5)推理能力。能利用当前知识和经验,根据环境变化,以一种类似于人的思维方式推理和规划。这是智能代理区别于其他软件的关键所在;6)自适应性。为了维护自治和推理能力,智能代理必须能评估外部环境的当前状态并将其融入下一步行动的有关决策之中;7)可信赖性。智能代理必须准确可信地代表用户,这是智能代理能否被用户接受的关键。^[2]

智能代理作为一种新型的人工智能技术,其快速发展无疑为数字图书馆开展智能化服务提供了强大的技术支持。但由于智能代理技术还不完善,图书馆界对智能代理的研究和利用目前还主要集中在信息导航、信息检索等方面。

2.2 数据挖掘技术

数据挖掘(Data Mining)也称数据库的知识发现(Knowledge Discover Database),就是利用信息技术和信息分析手段从大量的数据中提取有用信息和知识的过程。数据挖掘所得到的结果,一定是潜在的、事先未知的。通过数据挖掘,可以从各种数据库、数据仓库或其他信息存储中提取有用的知识、规律或高层信息,并从不同角度加以观察或浏览。数据挖掘的主要功能包括发现、预测、关联分析、聚类、概念描述、偏差分析等。^[3]

数据挖掘在数据组织与分析、知识发现及信息深层挖掘等方面具有强大优势,图书馆可以运用关

联规则分析、聚类分析、决策树、时间序列分析等数据挖掘方法,找出数据库中蕴藏的对于服务有用的规则、描述、预测,为图书馆改进服务措施、提高服务质量提供决策支持。如利用关联规则分析从图书馆借阅记录中发现规律,并利用这些规律改进图书推荐服务;利用时间序列分析方法对图书馆流通数据库或者 Web 日志进行挖掘,找出图书流通规律,以便科学安排流通部门的服务工作;利用分类模式挖掘方法对图书馆用户管理数据库中的历史数据进行挖掘,分析不同用户群体的特征,细分群体的偏向爱好,开展个性化服务。

2.3 数字资源整合技术

随着图书馆资源数量的增加以及类型的丰富,将会出现越来越多异质、异构的数字资源。图书馆需要整合这些资源,建立相关联的数字资源体系,为用户提供迅速、方便、简洁的智能化服务。

数字资源整合就是对各种自主的、异构的乃至离散式分布的数字资源采用各种整合技术实现透明无缝集成,以便用户在统一的界面下获取和利用各种资源。目前,在国外 SFX 作为一种比较先进的数字资源整合技术被提出并开始应用。运用 SFX,可以把不同来源和不同通信协议的数字信息完全融合起来,使不同类型、格式的数字资源实现无缝链接。国内对异构资源整合技术的研究稍晚,主要有 CALIS 资源统一检索平台、汇文一站式文献检索系统、同方 USP 统一检索平台等。MetaLib 内嵌的 SFX 是目前数字资源整合最优良的解决方案。MetaLib 中的 SFX 扩展服务实现了各种资源之间的透明链接和导航,并且可以与图书馆现有的全文传递、馆际互借、参考咨询等服务无缝集成。它同时整合了技术、服务、资源等功能,这是 MetaLib 区别于其他资源整合系统的显著优势。哈佛大学、麻省理工学院以及国内的清华大学、复旦大学等高校的图书馆都已应用了这一技术。此外,南京师范大学、北京理工大学等图书馆也实现了 MetaLib 内嵌 SFX 的整合检索。^[4]

2.4 智能检索技术

智能检索是信息检索和人工智能研究的一个交叉领域。它将信息检索从目前基于关键词的层面提高到基于知识(或概念)层面,对知识有一定的理解和处理能力,能够正确运用分词技术、同义词技术、概念搜索、短词识别以及机器翻译等技术。因而,智能检索更具有信息服务的智能化、人性化特征,能为用户提供更方便、准确地搜索服务。

随着人工智能技术广泛应用于信息检索领域,现已产生多种智能检索系统模型。归纳起来,主要有 2 大类:一类是基于机器学习的大范围信息检索

系统,一类是智能化的用户个性化信息检索系统。这些检索系统需要依赖许多先进的技术来实现它的各项功能,所涉及的关键技术主要包括:1) 信息收集和存储技术,分人工和自动 2 种方式。其中自动方式是由“网络 Robot”来完成的;2) 信息预处理技术,包含信息格式支持、转换和信息过滤;3) 信息索引技术,建立索引涉及的信息语词切分和语词语法分析,进行词性标注及相关自然语言处理,建立检索项索引及检索结果处理技术。其中检索结果处理技术是关键技术,其核心是依据计算结果与查询词的相关程度来排序。^[5]

由于智能检索所涉及的相关技术不是很成熟,现有的智能检索系统仍然存在一些缺陷。大部分的概念检索,只是单纯对输入的关键词进行概念扩展,检索结果并不能达到预期的效果。目前,国内外许多公司和研究机构已投入大量资金和人力用于智能检索的研制和开发,智能检索系统将变得越来越“聪明”。

2.5 信息过滤技术

信息过滤技术是指根据用户兴趣帮助用户过滤掉那些他不感兴趣的信息。常用的信息过滤方法有 2 种:一是基于内容的过滤方法,主要是通过资源与用户描述的比较来推荐资源。可以利用基于关键词的方法将信息流与用户描述进行匹配计算,具有简单、有效等优点,其缺点是难以区分资源的特征与形式,只能发现那些与用户曾经感兴趣的相似资源,而不能为用户发现新的感兴趣的资源;二是基于协作的过滤方法,主要是根据用户的相似性来推荐信息资源。这种方式能够发现新的、用户可能感兴趣的资源,其缺点是在系统使用之初,由于系统资源还未获得足够多的评价,因而很难利用这些评价来发现资源。随着用户和资源的逐渐增长,这种方法的性能将会降低。^[6]以上 2 种过滤方式各有利弊,应综合利用这 2 种方法的优点,构建混合型的信息过滤模式,以提高信息过滤的性能。

信息过滤技术可以过滤掉与用户无关的信息,能够满足不同用户的信息需求。随着数字图书馆的迅速发展,信息过滤技术必将在智能化服务中发挥越来越重要的作用。

2.6 智能推拉技术

推拉技术是一种新的网上信息服务技术,从“信息源”与“用户”的关系来看,可分为信息推送与信息拉取 2 种技术。信息推送技术是由“信息源”主动将信息发送给“用户”,如新书通报和定题服务等。这种技术时效性好,应用面广,但针对性差,难以满足用户的个性化需求。信息拉取技术是由“用户”主动从“信息源”中提取所需信息,如信息查询。

这种技术针对性强,用户能够有目的地查询自己所需的信息,但时效性差,需要用户掌握一定的检索技术。

为了适应信息服务需要,信息推送和拉取2种技术应相互结合、优势互补,并在此基础上融入人工智能、机器学习、知识发现等技术,从而形成智能信息推拉技术。这项技术是当前 Internet/Extranet/Intranet、数据库系统及其他信息系统为用户提供信息服务的一个发展方向。智能推拉技术的引入,可以提高网络及数据库的智能水平,从而从根本上解决“推送”和“拉取”技术应用过程中所遇到的难题。^[7]

智能推拉技术是实现信息检索、发布和传播的重要手段。数字图书馆利用智能推拉技术,一方面,可以从所“推送——拉取”的信息中发现其内在规律,提取用户最关心、最感兴趣的有用信息,以便更有针对性、更及时地向用户推送实用信息;另一方面,用户也可以更快、更准确地从资源数据库中获取到所需信息,大大提高服务质量和效果。

2.7 信息安全技术

图书馆在提供智能化服务的同时,也面临着严重的信息安全问题。如果服务系统一旦受到黑客攻击,感染病毒,将导致大量数据的丢失或破坏,其损失将是灾难性的。因此,图书馆必须增强信息安全意识,不断加强和完善信息安全防护措施。

在网络环境下,信息安全防范体系是分层次的,不同层次有着不同的信息安全问题。根据信息传输和应用网络的情况,信息安全技术应分为物理层安全、系统层安全、网络层安全、应用层安全和管理层安全5个层次。^[8]但是,技术不是万能的,图书馆应逐步建立融技术、管理、人才、制度等为一体的多层次信息安全防护体系,才能有效地维护网络信息安全。

3 结 语

近年来,上述诸多关键技术 in 数字图书馆服务

中得到广泛应用,大大提升了数字图书馆服务质量和水平,进一步拓展了信息服务空间。目前,数字图书馆所提供的智能检索、智能推送、智能咨询、智能辅导、智能决策和智能 Web 站点等服务形式,都离不开先进技术的支撑。因此,在数字图书馆建设过程中,要自觉地以科学方法为指导,密切关注相关技术的研究动态和发展状况,不断探索、引进和应用智能化技术,促进智能化服务的快速发展。同时,也要在实践的基础上,不断补充、完善、拓展相关技术,推进智能技术与信息服务的融合,使智能化服务更好地满足读者需求。

[参考文献]

- [1] 张安珍. 论网络环境下的智能信息服务[J]. 情报理论与实践, 2004(6): 659—662.
- [2] 王惠仙. 智能代理及其在数字图书馆中的应用[D]. 中山大学, 2001: 3—6.
- [3] 朱立红. 高校图书馆的数据挖掘技术应用与用户研究[J]. 图书馆杂志, 2008(6): 39—40.
- [4] 毛秀梅, 刘青华, 王春梅. 高校图书馆数字化服务技术应用分析及优化策略探讨[J]. 图书馆建设, 2007(5): 81.
- [5] 马 静. 网上信息资源及其检索技术智能化研究[J]. 图书情报工作, 2001(1): 57—58.
- [6] 焦玉英, 王 娜. 信息过滤技术在数字图书馆的应用[J]. 中国图书馆学报, 2006(3): 46—48.
- [7] 刘志敏, 董 慧. 推拉技术在数字图书馆中的应用[J]. 中国图书馆学报, 2003(2): 60—63.
- [8] 俞承杭. 信息安全技术[M]. 北京: 科学技术出版社, 2005: 28.

[收稿时间] 2009-12-24

[作者简介] 程凤刚(1973—), 男, 本科, 馆员, 主要从事数字图书馆服务研究, 已发表论文十余篇。

[说明] 本文系盐城师范学院人文社会科学研究项目(项目编号: 07YCKW032)成果之一。

(上接第12页)

6月5日

广西壮族自治区新闻出版局下文通报,《图书馆界》在全区关于都市类报纸和社会科学类期刊编校质量检查中成绩优异。《图书馆界》的差错率仅为万分之0.66,在全区100多种期刊中名列第9。

6月26日

广西图书馆学会秘书长秦小燕出席了广西壮族自治区社科联举办的“2009年第一次社科专家活动日”,并提交了书面建言稿。

7月3—6日

“川、吉、苏、桂、冀”五省(区)图书馆学会第十一届学术研讨会在河北省秦皇岛市召开。来自四川、吉林、江苏、广西和河北五省(区)各系统图书馆的108名代表出席了会议。我学会由理事长徐欣禄带队,组团13人参加会议。

7月6—9日

广西图书馆学会理事长徐欣禄、副理事长陈大广赴北京参加由中国图书馆学会举办的“中国图书馆学会第八次会员代表大会”。会上11个先进学会,397名优秀会员和42名优秀学会工作者受到表彰。广西图书馆学会获得教育与培训单项先进奖;副理事长陈大广被评为优秀会员;秘书长秦小燕被评为优秀学会工作者。

(下转第35页)