

三网融合区域云数字图书馆跨库检索服务研究

王晓明,张龙昌

(渤海大学信息科学与技术学院,辽宁锦州 121000)

摘要:为了促进区域云数字图书馆与三网融合的结合,使区域图书馆扩展信息资源,更好地为读者服务,针对目前区域云数字图书馆在三网融合环境下跨库检索的困难,对该环境下的跨库检索服务进行了研究。首先研究来自于三网的信息资源、异构终端并进一步研究跨库检索服务的需求,随后提出该环境下跨库检索服务的模型,再根据跨库检索服务的模型提出相对应的跨库检索服务的流程,最后对信息资源的管理与定制、信息资源的权限、信息资源的推送、跨媒体检索进行了研究。

关键词:区域图书馆;三网融合;云;跨库检索;服务

中图分类号:TP39

文献标识码:A

文章编号:1673-629X(2014)09-0245-04

doi:10.3969/j.issn.1673-629X.2014.09.058

Research on Cross-database Search Service of Regional Cloud Digital Library in Three Nets Fusion Environment

WANG Xiao-ming, ZHANG Long-chang

(College of Information Science and Technology, Bohai University, Jinzhou 121000, China)

Abstract: Research the cross-database search service of the regional cloud digital library in three nets fusion environment, to promote the combination of the regional cloud digital library and three nets fusion, making regional library expands information resources and serves readers, solving the retrieval problems. Firstly, research the information resources from three net and abnormal terminator, and further study the demand of cross-database retrieval service. Secondly, propose the cross-database search model, according to which the corresponding process of cross-database search service is put forward. Finally, research the management, the custom made, the authority and the delivery for the information resources and cross-media retrieval.

Key words: regional library; three nets fusion; cloud; cross-database search; service

0 引言

三网融合,在2010年政府工作报告中被注释为电信网、电视网、计算机网三网经过技术改造,逐步实现宽带通信网、数字电视网、下一代互联网三个网络之间的融合,提供基于语音、视频以及数据等各种媒体的应用和业务^[1-3]。三网融合图书馆的环境,为用户带来的信息资源种类更多、数量更多,能打破时间和空间的限制,体现“以读者为中心”的理念。

跨库检索,是信息检索的一种,首先通过统一的检索接口获得需求,然后根据数据源转化成对应的检索表达式,并发检索多个异构数据源,再集成、去重和排序检索结果,最后以统一的格式展现给用户。跨库检

索可以缩短检索的时间,增添便利性,提高检索的效率,目前国内这方面研究主要有:文献[4]利用信息技术整合数字资源、实现跨库检索;文献[5]采用PHP、MySQL和Apache的联合检索方法,利用CURL组件实现跨库检索;文献[6]基于OGSA,借助网格技术构建基本构架为Web Service的跨库检索平台;文献[7]利用Ajax通过应用异步实现检索,有效地缩短了检索结果的等待时间。然而以上研究均没有考虑到三网融合的环境,对云和区域考虑也较少,很难应用到三网融合环境下的区域云数字图书馆中;同时笔者以“三网”和“跨库检索”作为题名和篇名,于2013年10月19日分别在万方和China Academic Journals进行检索,获取

收稿日期:2013-10-31

修回日期:2014-02-11

网络出版时间:2014-07-17

基金项目:教育部人文社会科学研究青年基金项目(12YJC870030);辽宁省社会科学规划基金项目(L12CTQ008);辽宁省教育科学“十二五”规划基金项目(JG12DB149)

作者简介:王晓明(1979-),男,硕士,讲师,研究方向为云、三网融合、图形图像。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20140717.1235.059.html>

的论文数量均为 0 篇。为此文中首先研究三网融合区域云数字图书馆环境下跨库检索服务的需求,然后探索该环境下的跨库检索服务模型,其后进一步研究跨库检索的一次过程,最后对跨库检索的若干技术进行了探讨。

1 三网融合区域云数字图书馆跨库检索服务需求

1.1 信息资源研究

三网融合区域云数字图书馆环境下,信息资源来自于三网,其特征表现为:

(1)数量呈现海量。信息资源数量上的海量导致了跨库检索服务很难直接使用传统意义上的跨库检索,需要先缩小跨库检索服务的检索范围;

(2)模态异构化、多样化、复杂化。首先,信息资源种类很多,有文本、图像、视频、语音等;其次,各种信息资源之间存在很多差异,如文本往往是静态的而视频往往是动态的;再次,各种信息资源的格式往往不同,文本、图像、视频、语音等往往有不同的模态(格式);

(3)来源不同。信息资源可能来自区域内单独的图书馆,也可能来自多个成员馆,甚至可能来自电台、电视台等,读者对信息资源的使用权限可能会有差别;

(4)均以二进制的形态存在。

1.2 终端研究

三网融合区域云数字图书馆环境下的服务,既要面对图书馆内部,又要面向图书馆外部,其终端有如下特征:

(1)多样化。能够在三网使用的终端有电话、电视、台式计算机、笔记本电脑等多种,其类型不同;

(2)所支持的格式和内容不同。不同的终端设备能够支持的信息资源格式和内容不同,这导致部分信息资源无法被一些终端设备直接使用;

(3)能够一次性接收的数据量不同。起因于不同终端所能够存储的数据量不同;

(4)能够支持的计算复杂度不同;

(5)屏幕性能不同。

1.3 跨库检索服务需求

在研究信息资源和终端的基础上,文中认为:

(1)跨库检索服务需要考虑到检索到信息资源后的推送过程,只有能够获得和使用检索到的信息资源跨库检索服务才有意义,那么既要考虑到信息资源,包括信息资源的模态、内容等,又要考虑到终端,避免所推送的信息资源无法被终端直接使用;

(2)跨库检索服务需要与个性化相结合,以便实现“以读者为中心”的图书馆服务理念,既可以根据用

户的个性缩小检索的范围,又可以根据不同的终端对信息资源设计对应的获得和使用方案;

(3)跨库检索服务需要与权限相结合,读者可能只拥有部分信息资源的使用权限,也可能拥有全部信息资源的使用权限,读者可以被允许检索自己没有权限的数据库,也可以不被赋予此权力,但不能允许读者使用自己没有权限的信息资源;跨库检索服务与用户权限相结合,可以在一定程度上既保障区域云数字图书馆的正常经营又可以保留各成员馆的独立性,同时也体现了“以读者为中心”的服务理念;

(4)跨库检索服务需要与跨媒体检索相结合,来自于三网的信息资源模态多样化、异构化,但各种模态都需要检索;

(5)跨库检索服务需要与语义检索相结合,但语义检索的内容仅限于信息资源的简要介绍(如论文的摘要),既可以借助用户个性化缩小检索的范围,又可以缩短跨库检索的总时间;

(6)跨库检索服务需要考虑信息资源实际存在的位置,以便减少网络流量。

2 跨库检索服务模型

根据信息资源、终端、需求的研究,提出对应的跨库检索服务模型,整个服务成分布式资源库层、跨库检索服务层、用户信息服务层三个层次,如图 1 所示。

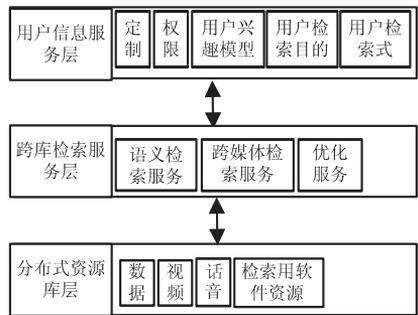


图 1 三网融合区域云数字图书馆跨库检索服务模型

用户信息服务层,主要用途是获得需求,其使用统一的界面根据数据源转换成检索表达式,主要包括定制、权限和用户兴趣模型三个组成部分,其中:

(1)定制既包含对用户所需要的信息资源的定制,同时又包含用户使用各种终端时所接收的信息资源的定制策略;

(2)权限针对信息资源的使用,可以限定跨库检索的范围;

(3)用户兴趣模型用于为后续的语义检索服务和跨媒体检索服务提供基础,从读者共性和个性的角度建立相关的模型,通过模型与检索服务结合的方式缩小后续检索的范围,提高检索的检准率,缩短跨库检索所需要的时间。

跨库检索服务层,主要包括语义检索服务、跨媒体检索服务和优化服务,在云中,具有并发的能力,能够同时检索多个数据库的信息资源,其中:

(1)语义检索服务优势在于有效地支持对知识和信息的检索,可以实现语义的推理与匹配,在实际应用中中文中推荐采用语义网检索的实例检索;

(2)跨媒体检索服务的优势在于可以在与人的高层认知一致的前提下跨越媒体检索,虽然对部分视频或图像的检索不是几个关键字就可以检索到的,但是可以通过交互处理检索要求和内容、通过用户兴趣模型、定制和权限进行补充,然后再检索;

(3)优化服务主要包括去重、排序等方面的优化,包括根据距离远近、网络流量等方面的优化服务。

分布式资源库层,是跨库检索服务的基础,包括两个方面的内容:

(1)数据、视频、语音这些来自于三网的信息资源,其可以而且最好包括以文本模态存在的关于内容的简要介绍;

(2)跨库检索需要使用的其他软硬件资源,如服务器、跨库检索所使用的程序等。

3 跨库检索服务流程

在研究跨库检索服务的基础上,提出跨库检索一次过程,如图2所示。

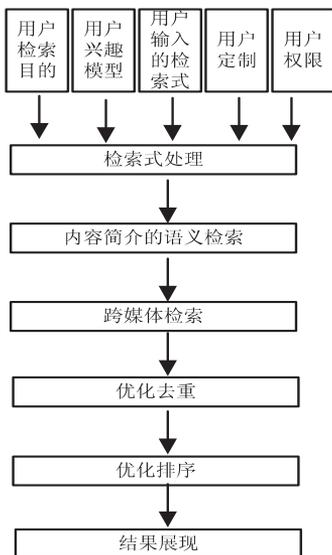


图2 三网融合区域云数字图书馆跨库检索一次检索过程图

首先根据用户兴趣模型、用户检索目的、用户定制信息、用户权限和检索式进行检索式的处理,其目的是找到满足用户需求和应用目的的检索式,为后续的跨库检索打下基础。在此过程中

(1)用户兴趣模型用于应用人群及其兴趣点的定位,其是可选项,在用户进行跨库检索服务的时候可以

由用户选择是否是自己所需的,如果是自己所需的则将其应用人群及兴趣模型作为参考之一;否则不以用户兴趣模型为依据。用户兴趣模型可以获取用户兴趣点,可以通过用户的年龄、性别、知识层次等背景确定用户所属人群,然后再逐步伴随用户的使用获得用户的个性进而调整用户兴趣模型。应用人群及兴趣点的定位不仅可以用于实现检索式处理时的个性化而且可以用于计算跨库检索服务获得的结果与用户需求相符合的程度,以便更有效地排序和去重,并为其他服务打好基础。

(2)用户检索目的对应服务目的,可以利用检索界面提供常用的服务目的由用户选择,需要与终端、信息资源的类型结合起来使用,可以用于缩小进一步检索的范围。

(3)检索式的处理其实质是以用户所输入的检索式为基础,根据用户检索目的和兴趣点加入用户的个性化信息、结合用户权限和定制,调整用户输入的检索式以形成符合用户需求的检索式,其处理不一定一步到位,也可以通过人机交互逐步完善。

在获得完善的检索式之后,并发检索满足定制和权限的数据库和信息资源,先通过语义检索缩小检索的范围,再通过跨媒体检索进一步提高检准率,然后再优化去重、优化排序,最后展现。信息资源的去重和排序在实现的过程中应该至少包括以默认值去重和排序、用户个性化去重和排序两种不同的方式,默认值可以由系统根据运行环境的特点设置;用户个性化信息则可以根据用户的目的结合终端和信息资源设置,可以根据信息资源存在的位置与终端之间的距离按照位置就近原则去重和排序,可以根据信息资源内容与用户需求的匹配程度去重和排序,可以根据信息资源的模态去重和排序等。优化的去重与排序可以提高后续推送服务的质量、速度,实现便捷性和灵活性。

4 关键技术

4.1 跨媒体检索

跨媒体检索主要解决不同模态的多媒体对象之间的相关性匹配问题,近年国内对跨媒体检索的研究主要有文献[8-11]。其中文献[8]对跨媒体的识别和表达、数据处理的机制等方面做出了综述;文献[9]提出基于综合性推理模型适用于多媒体的语义挖掘和跨媒体的检索技术,效果稳定而且准确;文献[10]分析不同模态内容特征间的相关性,利用子空间映射、先验知识修正拓扑结构,度量跨媒体的相关性;文献[11]学习底层特征的潜在关系,在降维得到的子空间修正聚类质量,借鉴了反馈样本。在国外,可以利用的技术和成果主要有:跨媒体数据间相关性的识别,文献[12]

利用流形对复杂多媒体文档模型进行描述,协调不同层次的数据,得到这些数据最优的向量描述;不同类型媒体数据间的关联挖掘,文献[13]提出了模型并实现了学习算法,从多个层面和侧面论证了正负相关反馈的异构特性;借助机器学习分析跨媒体数据间的相似性,如文献[14-15]。

4.2 信息资源的管理与定制

文中认为三网融合区域云数字图书馆可以以区域图书馆联盟的形态出现,也可能以单个图书馆的形态存在,信息资源可以按照分层次分类别的方式进行管理。

(1)如果图书馆以图书馆联盟的形态出现,则按照信息资源所属馆分别管理,用户定制自己感兴趣的成员馆的信息资源,可以只定制一个或者几个成员馆的信息资源,获取的信息资源也是由对应的成员馆获得再推送;

(2)按照信息资源所属的学科分类别,分别为用户提供定制,交叉学科的信息资源可以在多个学科的定制中都出现,以便于在海量的信息资源中快速地检索;

(3)按照信息资源所涉及的知识难易程度分层次,分别供用户定制,便于按照用户知识层次检索和获得信息资源;

(4)用户所使用的终端环境可以允许定制,以便充分发挥三网融合环境的优势,实现以用户为中心的服务。

4.3 信息资源的权限

权限需要细分,对于部分图书馆和部分用户而言,可以获得的信息资源可能不是全部,而是部分,那么就可以只提供对应于用户权限的信息资源,可以在初步检索的过程中就检查权限,只检索符合用户权限的信息资源,可以减少检索的信息资源数量、提高速度和安全性并保障了信息资源使用的合法性;或者在排序的过程中检查权限,这样虽然增添了系统的负担,但是一定程度上可以起到推广的作用。文中推荐前者。

4.4 信息资源的推送

信息资源的推送,是将用户通过跨库检索后获得的信息资源通过处理再推送给用户,某种程度上讲也意味着“结果的优化”,这种处理实质就是在计算复杂度、存储空间、格式、网络能力和屏幕性能几个方面加工、转换和配置信息资源,使最终获得的信息资源能够被各种异构终端直接使用,详见文献[16]。此外,文中认为根据用户的定制,在信息资源与终端之间建立映射关系,根据映射关系将满足用户权限的信息资源加工、转换和配置成终端能直接使用的、最符合用户需求和定制的信息资源再推送给用户。

5 结束语

基于三网融合区域云数字图书馆的环境,文中首先研究了该环境下的信息资源和终端,再提出该环境下跨库检索服务的需求和模型,然后提出该环境下跨库检索的流程,最后对部分关键技术进行了研究和介绍,但对该环境下跨库检索之上的服务研究较少,在以后的研究中将逐步完善。

参考文献:

- [1] 袁超伟,张金波,姚建波. 三网融合的现状与发展[J]. 北京邮电大学学报, 2010, 33(6): 1-8.
- [2] 徐 恪,林 嵩,吴建平. 基于用户效用的三网融合演化模型与分析[J]. 计算机学报, 2013, 36(5): 903-914.
- [3] 刘 斌,寇 爽. 三网融合下的图书馆信息服务[J]. 图书馆理论与实践, 2013(2): 48-50.
- [4] 屈冠军. 基于实现跨库检索的数字资源整合技术探讨[J]. 图书馆, 2009(6): 86-88.
- [5] 何 芸,张 倩,赵霞琦. 跨库检索在高校教学参考系统中的应用与实现[J]. 情报理论与实践, 2010(8): 96-99.
- [6] 翟姗姗,陈卓群,夏立新. 基于 OGSAs 的数字图书馆跨库检索平台构建[J]. 情报科学, 2011, 29(2): 292-297.
- [7] 周 强,许雁冬,李 宇. 应用 Ajax 提高跨库检索系统 SRU 接口的检索速度[J]. 计算机技术与发展, 2010, 20(10): 203-206.
- [8] 吴 飞,庄越挺. 互联网跨媒体分析与检索:理论与算法[J]. 计算机辅助设计与图形学学报, 2010, 22(1): 1-9.
- [9] 杨 易,郭同强,庄越挺,等. 基于综合推理的多媒体语义挖掘和跨媒体检索[J]. 计算机辅助设计与图形学学报, 2009, 21(9): 1307-1314.
- [10] 张 鸿,吴 飞,庄越挺,等. 一种基于内容相关性的跨媒体检索方法[J]. 计算机学报, 2008, 31(5): 820-826.
- [11] 张 鸿,吴 飞,庄越挺. 基于特征子空间学习的跨媒体检索方法[J]. 模式识别与人工智能, 2008, 21(6): 739-745.
- [12] Yang Y, Zhuang Y T, Wu F, et al. Harmonizing hierarchical manifolds for multimedia document semantics understanding and cross-media retrieval[J]. IEEE Transactions on Multimedia, 2008, 10(3): 437-446.
- [13] Zhuang Yueting, Yang Yi, Wu Fei. Mining semantic correlation of heterogeneous multimedia data for cross-media retrieval[J]. IEEE Transactions on Multimedia, 2008, 10(2): 221-229.
- [14] Liu J, Wang B, Li M J, et al. Dual cross-media relevance model for image annotation[C]//Proceedings of the ACM international conference on multimedia. Augsburg: [s. n.], 2007: 605-614.
- [15] Sankar K, Jawahar C V. Probabilistic reverse annotation for large scale image retrieval[C]//Proceedings of computer vision and pattern recognition. Minneapolis: [s. n.], 2007: 1-6.
- [16] 王晓明. 三网融合环境区域云数字图书馆构建研究[J]. 计算机技术与发展, 2014, 24(5): 227-230.

三网融合区域云数字图书馆跨库检索服务研究

作者: 王小明, 张龙昌, WANG Xiao-ming, ZHANG Long-chang
作者单位: 渤海大学 信息科学与技术学院, 辽宁 锦州, 121000
刊名: 计算机技术与发展 
英文刊名: Computer Technology and Development
年, 卷(期): 2014(9)

本文链接: http://d.wanfangdata.com.cn/Periodical_wjfz201409058.aspx