

联机目录： 用户和图书馆员需要什么



Copyright © 2009, OCLC Online Computer Library Center, Inc.
6565 Kilgour Place
Dublin, Ohio 43017-3395

ALL RIGHTS RESERVED. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying or otherwise, without prior written permission of the copyright holder.

The following are trademarks and/or service marks of OCLC: Connexion, FirstSearch, OCLC, the OCLC logo, WorldCat, WorldCat.org, WorldCat Resource Sharing and "The World's Libraries. Connected."

Third-party product, service, business and other proprietary names are trademarks and/or service marks of their respective owners.

Printed in the United States of America

ISBN: 1-55653-420-5

978-1-55653-420-1

联机目录:用户和图书馆员需要什么

主要作者

Karen Calhoun, Vice President, WorldCat and Metadata Services(副总裁,WorldCat 与元数据服务)

Joanne Cantrell, Marketing Analyst(市场分析师)

Peggy Gallagher, Market Analysis Manager(市场分析经理)

Janet Hawk, Director, Market Analysis and Sales Programs(主管,市场分析与销售策划)

图表、排版和编辑

Brad Gauder, Creative Services Writer(创意服务写作员)

Rick Limes, Art Director(美术指导)

Sam Smith, Art Director(美术指导)

合作作者

Diane Cellentani, Market Research Consultant to OCLC(OCLC 市场研究顾问)

目 录

中文版序言

摘要..... (1)

引言..... (2)

 信息查寻行为 (2)

 目录使用、目录用户和目录数据质量 (3)

 构建一个全新的数据库质量计划 (3)

研究方法 (5)

 终端用户的聚焦小组 (5)

 聚焦小组的研究规程 (6)

 终端用户弹出式问卷调查 (6)

 图书馆调查 (7)

数据质量:用户的需求 (9)

 主要研究结论 (9)

 详细分析 (9)

 概述 (9)

 发现 (11)

数据质量:图书馆员的需求 (17)

 主要研究结论 (17)

 详细分析 (17)

 概述 (17)

 期待的数据质量提高措施 (18)

 不同类型图书馆期待的提高数据质量措施 (19)

数据质量:图书馆员与终端用户 (30)

 主要研究结论 (30)

 详细分析 (30)

 概述 (30)

图书馆员和终端用户的需求比较	(31)
图书馆员对终端用户需求的预测	(34)
结论	(38)
图书馆员和终端用户	(38)
平衡终端用户和图书馆员的需求	(38)
可用性、用户为中心的设计及信息组织原则	(38)
元数据和内容	(39)
目录的传递、链接和更多的在线内容	(39)
发现、传递和丰富数据	(39)
主题词和主题信息	(40)
标准编号	(40)
何处获取数据	(40)
扩充数据	(40)
相关排序、分类和用户事务数据	(41)
建议	(41)
未来研究设想	(42)
译者后记	(44)

中文版序言

众所周知,OCLC 是全球最大的不以赢利为目的、始终坚持使用最先进的技术维护和管理电子资源系统并提供计算机图书馆服务的会员制合作和研究组织,其宗旨是为广大用户发展对全世界各种信息的应用以及降低获取信息的成本。目前全球有 112 个国家和地区的 71,000 多所图书馆都在使用 OCLC 的服务来查询、采集、出借、保存资料以及编目。

OCLC 一直注重学术研究与开发工作,与大学或研究机构合作开展各种研究活动,以实现其改善世界信息资源获取和共享的诺言。近些年来,OCLC 连续发表了意义深远而实用的学术报告,2003 年的《五年的信息格式趋势》(Five - Year Information Format Trends)、《图书馆:如何做比较》(Library: How They Stack Up)和《审视发展环境,认清前景趋势》(OCLC Environmental Scan: Pattern Recognition)、2004 年的《信息格式趋势:内容、而非传统的载体形式》(Information Format Trends: Content, Not Containers)、2005 年的《对图书馆和信息资源的认知》(Perceptions of Libraries and Information Resources)和《大学生对图书馆和信息资源的认知》(College Students' Perceptions of Libraries and Information Resources)、2007 年的《网络世界中的共享、隐私与信任》(Sharing, Privacy and Trust in our Networked World)以及 2008 年的《从意识到资助:美国图书馆经费政策研究》(Awareness to Funding: A study of library support in America)的报告。这些报告发表之后在许多国家的各级各类图书馆以及相关情报部门产生了较大的反响,其中的很多观点一直在不断被借鉴和引用。

2009 年 4 月,OCLC 再次发布题为《联机目录:用户和图书馆员需要什么》(Online Catalogs: What Users and Librarians Want)的研究报告。报告总结了 OCLC 从用户和图书馆员的不同角度对图书馆联机目录的质量究竟取决于什么的调查所得出的结论。

2008 年,OCLC 对终端用户和图书馆员这两大重要群体进行问卷调查。面向终端用户,项目组采用在 OCLC 免费的终端用户网站 WorldCat.org 设置弹出式调查问卷的方法;面向全球图书馆员,项目组采用网络调查方式。

《联机目录研究》报告展示了此次调研的结论:

- 对于终端用户,最重要的是元数据,这是他们判断该书目记录是否满足其需求的依据;
- 高端终端用户希望看到图书馆在线目录以便帮助他们确定合适的资料;
- 高端图书馆员则推荐使用图书馆在线目录,可更好地协助他们的工作。

调查结果显示:除了其他用途外,虽然图书馆在线目录往往被认为是一个信息发现工具,但对终端用户来说与文献传递相关的信息也是同样重要的。

报告论及了图书馆用户和图书馆员对数据质量的需求以及目录的使用、目录用户和目录数

据质量等问题。报告强调图书馆目录应结合搜索引擎目前流行的一些特征,使用户能够从中获得更多的内容,有效地增强目录的功能性。

许多图书馆联机目录的用户就是图书馆员,他们组成了使用目录的重要群体;而目录的终端用户(普通公民、学生以及科研教学人员)对目录数据的信息需求偏好和期望是不同的,利用目录数据完成自己工作的图书馆员也同样如此。

随着下一代图书馆联机目录的出现,OCLC 也开始制定如何开展下一代 WorldCat 数据质量计划。为此,OCLC 成立了一个专门研究小组,包括一些编目和元数据专家,通过大量调研来鉴别和比较目录的终端用户和图书馆员对数据质量期望的不同点,比较和分析不同类型的图书馆员对目录数据的质量期望有何不同。兼顾终端用户和图书馆员两方面的角度,建议目录数据质量应该关注的重点。

报告指出,经过调查研究得出结论:终端用户和图书馆员在目录质量优先的问题上存在较大的差异。终端用户和图书馆员使用目录之目的各不相同,使用目录数据目的之差异造成对编目数据质量的不同需求。

报告最后强调这个报告是写给那些为了寻求改进目录数据方法和在下一代图书馆目录的贡献、摄取、聚合、同步或链接多种资源中扮演角色的人们,并提出 8 点很好的建议和未来研究设想。

为了使中国的图书馆员更好地阅读这篇报告,经 OCLC 总部批准,中国科学院国家科学图书馆承担了翻译任务,张晓林馆长在百忙之中亲自审阅报告全文,《图书情报工作动态》编辑部在 2009 年第六、七期专刊刊登了本报告中文稿。对他们为促进中国图书馆事业进一步发展所作出的贡献表示衷心的感谢!

在获得中文稿翻译和出版的授权过程中,OCLC 副总裁王行仁先生给予了大力支持。

联机计算机图书馆中心(OCLC)北京代表处首席代表 丘东江

摘 要

终端用户在网络上的工作习惯和预期会影响他们是否选择使用图书馆联机目录。联机目录的界面有一定影响，目录中的数据质量也相当重要，都是用户评价相关产品和服务是否有用的一个主要决定因素——对终端用户如此，对图书馆员和图书馆工作人员也是如此。为了对构成目录数据“质量”的要素有一个全面的、有证可循的了解，OCLC 成立了一个研究小组，以推进如下工作：

- 鉴别和比较终端用户和图书馆员对目录数据质量的期望。
- 比较分析不同类型图书馆员对目录数据的质量期望。
- 从终端用户和图书馆员两个角度，建议目录数据质量应该关注的重点。

对于希望界定从终端用户和图书馆工作人员两个界面完善目录数据的需求的读者，本报告将有助于开拓思路。那些参与下一代图书馆目录和图书馆集成系统中对多源数据的提交、提取、同步和链接问题的研究者，也会从本报告中得到相关的启示和帮助。

报告的主要研究成果包括：

- 对于终端用户来说，获取所需内容与发现所需内容一样重要，甚至更加重要。
- 终端用户依靠并期望获得增强的内容揭示方式，包括文摘（或摘要）、目次等。
- 高级检索选项（支持按数据项检索）以及分面检索有助于终端用户优化检索，以及对大结果集的检索、浏览和管理。
- 终端用户和图书馆员对目录数据质量的关注重点存在重大差异。
- 终端用户和图书馆员都是有目的地使用目录和目录数据。终端用户一般是希望找到并获得所需要的信息；而图书馆员通常则是出于履行工作职责。图书馆员的工作角色会影响他们对数据质量偏好。
- 图书馆员希望提高数据质量，反映出他们对于目录中准确的、结构化数据的重要性的理解。

上述发现表明信息组织中存在的两个传统——其一来自于图书馆界，另一个来自于网络。图书馆员们对数据质量的看法仍受到传统信息组织原则的影响，而终端用户对数据质量的期望在很大程度上源于他们从网站上体验到的流行的信息组织形式。现在需要的是整合这两方面的优势，形成新的、扩展的定义来回答图书馆联机目录中的“质量”到底意味着什么。

在结论部分，本报告针对数据质量项目提出建议，在联机目录中寻求终端用户和图书馆员需求之间的平衡点，同时也提出了若干值得进一步研究的问题。

引言

“在图书馆数据库质量的决策过程中,长期存在并亟需解决的弊端之一是缺乏对用户需求和获益的科学研究,以及对数据库质量决策对用户所产生的影响方面的研究”^[1]。

图书馆专家们对目录数据的“质量”构成已经有了很深入的了解。图书馆员对目录质量的有记载的定义可以一直追溯至 1876 年 Charles Cutter 对图书馆目录的目标的陈述^[2];在此后的一百多年里,这些目标为图书馆员设计目录提供了指导。多亏 Cutter 和随后的理论家们的成果^[3],使得今天的图书馆目录建立在有规可循、一致的记录结构和标目结构的基础上,这有利于获得意外发现、提高已知条目的检索效率、拓展浏览方式。典型的图书馆目录一般都包含优良的元数据,也就是说,他们采取规范档控制、分类标准和内容编码标准来描述和组织相关资料。而上述这些均建立在 Cutter 提出的目录的实践的基础上,支持作者、题名、主题和不同版本的检索。

1989 年,由 OCLC 联机数据质量控制部门主持的一项研究证实,图书馆员们关于所在图书馆的目录质量的共同认识与他们对于 WorldCat 作为共享目录记录源的期望是一致的^[4]。当时的联机数据质量控制部门负责人 Carol Davis 发现,图书馆员对 WorldCat 关注的三个焦点质量问题是:重复记录(描述同一版本的有不止一项记录)、名称标目错误(非规范名称)和主题标目错误(非规范主题)。今天,OCLC 的 WorldCat 质量计划继续集中精力解决在数据库质量中存在的这三个重点问题。此外,《OCLC 书目格式及标准》(OCLC Bibliographic Formats and Standards)^[5]也证实了 Cutter 的经典原则在诸多领域仍然支撑着当今最佳的实践方式,诸如:标准化描述、统一的纪录编码、以及规范名称和规范主题。

很多作者已证实,Cutter 的目录设计理念是方便用户,而“书目记录功能需求”(FRBR)的设计理念也与其类似,FRBR 主要是基于用户搜寻、鉴别、选择和获取所需信息这一系列任务而建立的概念模型^[6]。但是,对相关文献资料的分析表明,无论是 Cutter、还是其后著名的理论家们、或者是 FRBR 和“资源描述与检索”(RDA)计划^[7]的设计者,都没有请信息用户对所构建的概念架构进行严格的检验。Fran Miksa 教授(同时也是图书馆员)对 Cutter 和早期图书馆员的旨在建立信息组织传统的工作进行了研究,发现“在图书馆界,科学的用户研究从 1920 年代才刚刚开始”^[8]。同样,在本章开头的引用的那段文字中,Janet Swan Hill 还告诫图书馆员,要清醒地理解如何以及做出哪种图书馆目录数据质量决策才能帮助用户找到和获得所需要的信息。

信息查寻行为

2003 年,《OCLC 环境扫描》报告(OCLC Environmental Scan)将自助性、满意度、无缝性等认

定为信息查寻者所期望的最终指标^[9]。这一报告指出,信息查寻者认为用法简单、获取方便和可用性强等因素,与信息质量和信息可信度一样重要。2005 年的《图书馆和信息资源的认知》(Perceptions of Libraries and Information Resources)从图书馆的角度更深入地探究了信息查寻者的行为和偏好,并揭示了他们搜寻信息时的一个明显现象:他们更多选择使用搜索引擎(84%),而不是图书馆网站(1%)^[10]。除上述对终端用户信息需求的研究案例之外,在交流、学习理论、社会学、心理学、用户研究、人机交互及其他领域还存在大量相关研究。社会科学研究者们已经开始探究信息查寻研究的众多范式^[11];在图书馆界最著名的研究恐怕是文献学家 George Zipf 的“最小省力原则(Principle of Least Effort)”。Marcia Bates 提交给国会图书馆(Library of Congress)的一份关于改善用户访问目录与门户的报告中 also 包含关于信息查寻的有用的文献综述^[12]。

Donald Case 在他的信息查寻著作中指出,大多数研究集中在信息资源(例如图书或者报纸)和系统(例如目录)方面,而忽视了信息用户的需求、动机和行为^[13]。换言之,大多数研究的重点在于信息对象和系统,而不是人。相反,可用性专家已经认识到根据使用情景设计系统的重要性——要对工作实践进行研究,并利用研究结果驱动信息系统的设计^[14]。Rochester 大学 River 校区图书馆在应用工作实践研究来解决图书馆相关研究问题方面走在前列^[15],其中一个例子就是对科研人员的研究工作进行研究,了解学者们是如何使用机构知识库的。

目录使用、目录用户和目录数据质量

用户习惯使用网络搜索引擎、网络书店和全文无缝链接。为了更好地满足此类用户的需要,最新的图书馆文献中包含了大量关于如何改变联机目录来更好满足用户需求的文章。目录数据质量控制的研究人员逐渐意识到,终端用户对搜索过程和结果的期望及其网上工作习惯,很大程度上决定了他是否会使用或者再次访问图书馆联机目录。美国国会图书馆的 John Byrum 在国际图联(IFLA)2005 年 8 月大会上发布的一篇论文中指出,图书馆目录应根据流行搜索引擎的特性,提供对更多内容的获取能力和提供显著增强的功能^[16]。谈到目录覆盖内的局限性以及侧重印本出版物这两点,Haverford 学院图书馆副馆长、书目和数字服务部负责人 Norm Medeiros 认为,“用户们更加期望和追求全文;而当他们查阅电子期刊、数据库和网站时,用户越来越忽略目录这种查询方式”^[17]。

同时,许多图书馆联机目录的用户就是图书馆员和图书馆工作人员,他们构成了使用目录的重要群体。因此,如同终端用户(例如普通公民、学生、科研教育人员)具有对目录数据的信息需求、偏好和期望一样,利用目录数据完成自己工作的图书馆员也同样如此。

构建一个全新的数据库质量计划

当下一代图书馆联机目录开始出现,当 OCLC 开始考虑如何开展下一代 WorldCat 数据质量

计划时,建立起对目录数据“质量”的有据可循的、以用户为中心的理解,对于使图书馆联机目录成为不同群体(包括终端用户和图书馆员)的重要的、经常访问的信息源,是至关重要的。为此,OCLC 成立了一个专门研究小组,包括编目和元数据专家 Karen Calhoun、一位独立的来自市场研究机构的专业人士、以及 OCLC 市场分析小组的成员们,致力于:

- 鉴别和比较目录终端用户和图书馆员之间对数据质量期望的异同点。
- 比较分析不同类型图书馆员对目录数据的质量期望的异同点。
- 兼顾终端用户和图书馆员两个角度,提出目录数据质量应关注的重点的建议。

虽然研究小组的许多成果是与 WorldCat 数据质量计划密切相关,但是很多研究成果都可以推广应用于目前的图书馆联机目录和图书馆集成系统(ILSs)的完善。本报告主要介绍这些可推广的研究结果。对于那些希望确定用户需求、以便从终端用户或图书馆工作人员角度来完善目录数据的读者,本报告会提供有用的概念。对下一代图书馆目录和图书馆集成系统中利用多源数据进行提交、提取、集成、链接感兴趣的读者,也能得到相关启示和帮助。

参考文献:

- [1] Hill, Janet Swan, "Is it worth it? Management decisions related to database quality," *Cataloging and Classification Quarterly* 46 (1) (2008): 5-26.
- [2] Miksa, Francis L., Charles Ammi Cutter: Nineteenth-Century Systematizer of Libraries, (dissertation, University of Chicago, 1974) p. 373, <http://www.ischool.utexas.edu/~miksa/dissertation.html>. Cutter's 1876 objectives of the catalog are quoted in Miksa's comprehensive study of Cutter's contributions to librarianship:
To enable a person to find a book of which the author, the title or the subject is known.
To show what the library has by a given author, on a given subject or in a given kind of literature.
To assist in the choice of a book as to its edition (bibliographically) or as to its character (literary or topical).
- [3] A comprehensive review of past and present cataloging theory may be found in *The Future of Cataloging: Insights from the Lubetzky Symposium*, April 18, 1998, University of California, Los Angeles, edited by Tschera Harkness Connell and Robert L. Maxwell, Chicago: American Library Association, 2000.
- [4] Davis, Carol C., "Results of a survey on record quality in the OCLC database," *Technical Services Quarterly* 7 (2) (1989): 43-53.
- [5] OCLC Bibliographic Formats and Standards, 4th edition, <http://www.oclc.org/bibformats/en/>.
- [6] IFLA Study Group on the Functional Requirements for Bibliographic Records, Functional requirements for bibliographic records: final report, UBCIM publications (München: K. G. Saur, 1998) new ser., v. 19.
- [7] Joint Steering Committee for the Development of RDA, Full Draft of RDA, 2008, <http://www.collectionscanada.gc.ca/jsc/rda-fulldraft.html>.
- [8] Miksa, Fran, "Information organization and the mysterious information user" (lecture delivered at the School of Information, University of Texas - Austin), forthcoming in *Libraries and the Cultural Record* 4 (4) (Fall 2009).
- [9] De Rosa, Cathy, Lorcan Dempsey and Alane Wilson, *The 2003 OCLC Environmental Scan: Pattern Recognition: A Report to the OCLC Membership* (Dublin, OH: OCLC Online, 2003).
- [10] De Rosa, Cathy, *Perceptions of Libraries and Information Resources: A Report to the OCLC Membership* (Dublin, OH: OCLC,

2005).

- [11] Case, Donald O. , Looking for information: a survey of research on information seeking, needs, and behavior (San Diego CA: Academic Press, 2002) p. 140. The Principle of Least Effort asserts that information seekers will use information sources that are the easiest and most convenient, even when better or more authoritative sources are available, but not as easy or convenient to use or find.
- [12] Bates, Marcia J. , " Improving user access to library catalog and portal information: final report," prepared for the Library of Congress, 2003. <http://www.loc.gov/catdir/bibcontrol/2.3BatesReport6-03.doc.pdf>. The section on information – seeking behavior is on pages 4 – 6.
- [13] Case, p. 6.
- [14] See for example Holtzblatt, Karen, and Hugh Beyer, " Contextual design: using customer work models to drive systems design," in Conference companion on human factors in computing systems(New York NY: ACM Press, 1996) p. 373 – 374.
- [15] Foster, Nancy Fried, and Susan Gibbons, " Understanding faculty to improve content recruitment for institutional repositories," D – Lib Magazine 11 (1) (2005) , <http://www.dlib.org/dlib/january05/foster/01foster.html>.
- [16] Byrum, John D. , " Recommendations for urgently needed improvement of OPAC and the role of the National Bibliographic Agency in achieving it" (paper prepared for the 71st IFLA General Conference and Council, Oslo, Norway, August 14 – 18, 2005) , <http://www.ifla.org/IV/ifla71/papers/124e-Byrum.pdf>.
- [17] Medeiros, Norm, " Driving with eyes closed: the perils of traditional catalogs and cataloging in the Internet age," Library Computing 18 (4) (1999) : 300 – 305.

研究方法

OCLC 研究小组采用了以下三种方式来确定终端用户和图书馆员对目录数据质量的要求：(1) 聚焦小组 (focus groups)，(2) 在 WorldCat.org 网站上的弹出式调查表，(3) 互联网上进行调查。OCLC 面向终端用户的网站 WorldCat.org，帮助研究小组观察聚焦小组和弹出问卷回答者对新近设计的联机图书馆目录的数据元素的反应。第三种调查方式的对象是图书馆员，他们可能分别从用户界面和工作人员界面访问 WorldCat，这包括来自 WorldCat.org、WorldCat on First Search、Connexion 编目系统、可访问 OCLC 目录的 Z39.50 协议端口（主要由编目员使用）和 WorldCat 资源共享服务系统 WorldCat Resource Sharing（通常由负责馆际互借的员工使用）。

研究图书馆员对书目数据质量的要求时，需要把 WorldCat 的所有界面都考虑进去，因为本研究的主要兴趣是目录数据质量，而不是某个特定界面；关于特定界面的兴趣退居次位。最终目标是得到一个有据可循的、能代表不同群体用户需求的、以用户为中心的关于数据质量需求的全面认识。在那些 WorldCat 和图书馆目录可对照的方面，把研究方法的焦点集中于数据质量而不是界面，有助于将研究 WorldCat 所获得的结果用于更为一般的图书馆联机目录和集成自动化体系。

终端用户的聚焦小组

2008 年 5 月在俄亥俄州哥伦布市，OCLC 委任 Bule Bear 公司组织了对根据用户特点选定的三个聚焦小组的调查。这三个小组分别是：18—24 岁的大学生组、25—59 岁的偶尔检索型用户组、以及教师和研究生组（本报告称为学者组）。Blue Bear 公司的调查员主持了每个小组的讨论。这三个小组中的参与者情况如下：

大学生组	偶尔检索型用户组	学者组
<div>参与者信息</div> <ul style="list-style-type: none">• 8 位参与者；• 年龄 18—24 岁；• 7 位大学生；1 名刚获得学士学位；• 6 位使用过图书馆网站；• 6 位目前拥有图书馆证；	<div>参与者信息</div> <ul style="list-style-type: none">• 8 位参与者；• 年龄 25—59 岁；• 普通互联网用户；• 职业：篮球教练、演讲家、教师、流行音乐主持人、商业人士和作家；• 5 位使用过图书馆网站；• 5 位目前拥有图书馆证；	<div>参与者信息</div> <ul style="list-style-type: none">• 8 位参与者；• 年龄 25—59 岁；• 3 位研究生和 5 位教员；• 研究生专业分布：计算机科学、语言学、学校咨询与教育学；• 教员领域分布：德语研究、政治学、通信、经济金融和社会人类学；• 都出版过著作或者发表了被引用 10 次以上的论文；

网络使用情况	网络使用情况	网络使用情况
<ul style="list-style-type: none">• 在线时间(每周):6 位参与者花费 6—10 小时、2 位花费 1—5 小时(不计 e-mail);• 所有人都用过电子商务网站;被频繁提到的网站包括: textbooks.com Amazon, Yahoo!, Bestbuy.com and eBay;	<ul style="list-style-type: none">• 在线时间(每周):6 位参与者花费 10 小时以上(不计 e-mail)、2 位花费 6—10 小时;• 所有人都用过电子商务网站;被频繁提到的网站包括: Amazon, eBay, Barnes & Noble and half.com;	<ul style="list-style-type: none">• 在线时间(每周):半数参与者花费 10 小时以上(不计 e-mail)、3 位花费 6—10 小时、1 位花费 1—5 小时;• 所有人都用过电子商务网站;Amazon and Barnes & Noble 被所有人提及;

聚焦小组的研究规程

参与访谈者都通过笔记本电脑联网。调查者通过专门的软件和网络连接,能截获联网过程的网络计量参数、评论、页面标记,以便及时回顾或在会后分析。本报告中将会展示截图和一些评论。

用户使用中的评论通过下图所示的评价符号来记录。绿色的加号代表正面评价,被调查者喜欢他们看到的;红色标签代表负面评价,被调查者不喜欢他们所看到的。



聚焦小组成员先参加结构化的网络活动,接下来进行讨论。这些活动包括:

- 任务 1:根据用户自己设计的题目利用 WorldCat.org 进行在线检索。
- 任务 2:通过 WorldCat.org 在线检索《枪炮,病菌和钢铁》(Guns, Germs, and Steel)。
- 对任务 1,2,聚焦小组讨论自己的体验,包括系统浏览 WorldCat.org 检索过程。
- 任务 3:根据他们(大学生、偶尔检索者、学者)自己的兴趣或专业,利用 WorldCat.org 在线检索专门的资料。
- 对任务 3,聚焦小组讨论自己的体验。
- 对整体体验进行最后讨论。

终端用户弹出式问卷调查

OCLC 委托 ForSee Results 公司在世界范围内实施这个在线弹出式问卷调查,调查对象是 OCLC WorldCat 网站的终端用户。ForeSee 在 2008 年 5 月 12 日至 2008 年 7 月 9 日期间,回收到 11151 个回复。如果按照显示给用户的弹出问卷次数,调查回复率是 4%。

当用户点击 WorldCat.org 上的 Detailed Record 时,在线问卷调查就弹出。这个调查问卷含有 25 个封闭式问题和 1 个开放式问题。其中一个是要被访者描述自己(“Which best describe you?”),它的答案能将调查者划分到相应的群体内:学生、老师/教授、其他一般检索者以及图书馆员和图书馆工作人员。为了集中于终端用户的体验,来自图书馆员的被调查者未列入终端用户数据中。在总数为 11151 个回复中,68% 来自于终端用户。

终端用户是这样描述自己的:

年龄	
小于 18 岁	5%

19—30 岁	24%
31—40 岁	17%
41—50 岁	20%
51—60 岁	20%
大于 61 岁	13%
不愿意透露年龄	1%

终端用户总数	7583
学生	28%
老师或教授	22%
工商业专业人士	19%
其他	16%
退休者	9%
医疗卫生专业人士	6%

学生类别	
研究生/博士研究生	54%
大学生/专科学生	30%
高中	10%
初中	3%
其他	2%
小学	1%

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查)

图书馆调查

OCLC 委托市场研究公司 Marketing Backup 来实施基于互联网的调查,对象是通过不同界面使用联机书目数据库的图书馆员。

OCLC 给世界上来自于大学图书馆、公共图书馆、专业图书馆的共计 3522 个 OCLC 成员邮寄了邀请函。为了收集所有使用 WorldCat 数据的用户的观点,研究组将邮件分为三个组成部分:馆长、编目主任以及公共服务主任。OCLC 请接收者将邀请函转寄给所有 WorldCat 数据的图书馆用户,无论他们从事馆藏建设、采选、编目、馆际互借还是参考咨询。

从 2008 年 9 月 10 日至 2008 年 11 月 30 日,共计有 1397 位图书馆员回复了这个调查。因为一个机构有多人被邀请参加调查,所以无法计算回复率。

回复者的工作职责有如下几类:馆长、公共服务/参考服务、编目、采访、馆际互借。多数人选择多个职责(58%);42%的人认为他们在图书馆中只担任一种角色。在整个报告中,图书馆员(回复者)是指所有的图书馆员和图书馆工作人员。

图书馆类型	
学术	52%
公共	25%
专业	18%
其他	5%
工作职责(可多选)	
编目	64%
公共服务/参考咨询	44%

馆藏建设	31%
采选资料	28%
馆际互借	24%
馆长	21%
地区	
北美	74%
非北美	26%

来源:联机目录:用户和图书馆员需要什么, OCLC, 2009(图书馆调查)

本报告中揭示了不同类别被调查者间的若干显著差异。统计意义上的显著差异是指回复意见中那些并非由随机因素引起的差异。

数据质量：用户的需求

内容概览

用户想要得到什么？

- 在线内容的直接链接——无论是文本还是多媒体格式；
- 评价性内容，如总结/摘要、目录、摘录；
- 相关的搜索结果；
- 资源可获得性信息——资源是否可以得到、如何得到；
- 带有高级选项和引导性选项的简单关键词检索；

主要研究结论

获取所需内容的要求与发现这些内容的要求同等重要，甚至更加重要。那些能够支持从发现过渡到获取的恰当、准确、可靠的数据元素是非常重要的。

用户希望实现从发现到传递的无缝链接；用户希望能即刻知道资源是否可以获取，如果可以又如何获取。对于网络资源，无论是文本还是多媒体资源，用户希望能有更多的直接链接或者其它更便捷的网络内容访问方式。

除作者和标题之外的其它与发现相关的信息（如总结、摘录和目录）是联接用户发现 - 传递体验不同阶段的重要内容。

在检索过程中，用户期待和重视评价性信息，希望这些信息会有助于发现、最终有助于获得资源。

图书馆需要为用户提供更容易的方法来判断资源是否满足其需求；对于非数字资源，由于往往需要到图书馆去借阅，图书馆需要提供数据来帮助用户判断是否值得花时间去获取。

搜索结果必须具有相关性，而且相关性也必须是显而易见的。

当用户检索时，他们对自己应获得结果的类型是有所预期的。用户希望图书馆目录能够基于这种期望返回相关结果。在后台处理过程中，图书馆目录需要提供相关的数据元素和排序算法，让用户明白返回这种搜索结果的原因及其顺序。

“关键词检索为王”，但高级搜索选项（支持字段检索）以及分面检索有助于用户对检索的优化和对大结果集的导航、浏览与管理。

用户希望能进行类似于谷歌检索那样的简单搜索，而且结果能够与他们的期望一致。与此同时，用户也注重高级检索和引导性检索功能，这样可以不必浏览大量的检索结果就可以找到所需的资源。用户将检索结果的分面细化视为控制难以管理的大量检索结果列表的有效方法。

详细分析

概述

本研究通过两种方法收集用户的相关数据：聚焦小组访谈以及在线弹出式调查。参与聚焦小组访谈的用户分为三种类型：年龄在 18 - 24 岁之间的大学生、年龄在 25 - 59 岁之间的偶尔检索者、教师和研究生（本报告中称为学者）。当用户利用 WorldCat.org 检索时会随机弹出问卷

进行在线调查。

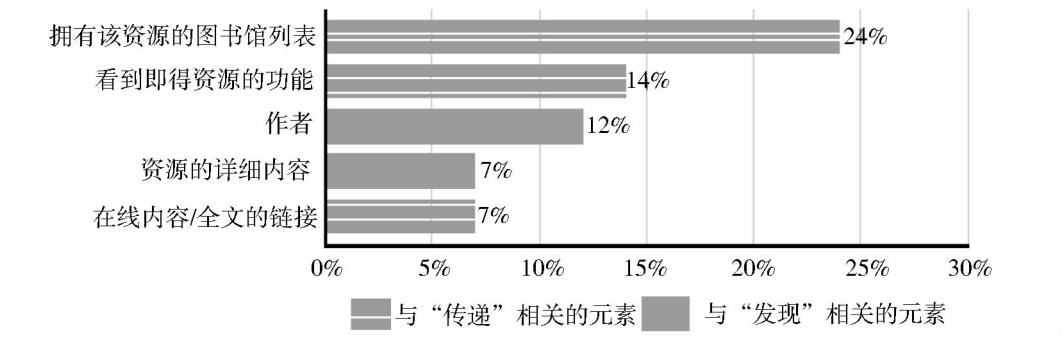
最重要的数据元素

当问及在确定自己所需资源的过程中目录记录中什么信息最重要时,弹出式调查的对象对数据元素的选择不一,既有支持用户传递的数据,也有一些支持用户发现的数据。下图中与“传递”相关的元素用带有水平条纹的矩形柱表示,与“发现”相关的元素用实心矩形柱表示。

最重要的数据元素^①

在确定您所需要的资源时,什么样的信息是最重要的?

基数:弹出式调查中的调查对象



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查)

排名前五位的选项中,有三项是支持传递的数据元素:24%的调查对象表示他们想看到拥有所需资源的图书馆列表;14%的调查对象表示想知道哪些是即得资源;7%的调查对象希望能有指向在线内容和全文的链接。

排名前五位的选项中,有两项是与发现相关的数据元素:作者(12%)和资源的详细内容(7%)^②。

当特别问及在寻找自己所需资源过程中,哪种数据元素最有用时,聚焦小组调查对象主要提及与发现相关的数据元素,如图书及封面图片的预览功能、利用不同标准细化检索的功能以及总结/摘要。然而在讨论过程中,他们也提到了与传递相关的数据元素。其中一位访谈对象表示,我们正生活在一个电子资源即时获取的“即买即得”世界。这也正是用户期待于图书馆的:成为将用户从描述内容的元数据连接到内容本身的桥梁。

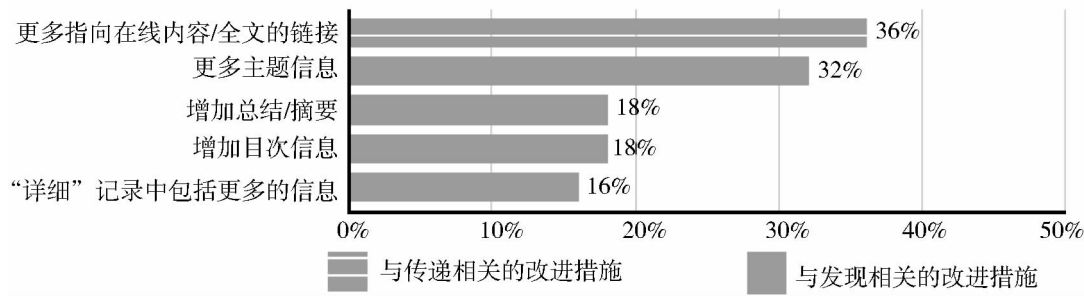
所期望的数据质量改进措施

当问及如何改进目录会有助于用户寻找所需资源时,弹出问卷回答者中有三分之一以上回答说,更多的指向在线内容/全文的链接是他们最想看到的改进措施。正如前面关于最重要数据元素的回答那样,这个关于改进措施的问题得到相同的发现:对于用户来说,与发现相关的信息元素和与传递相关的信息元素是同样重要的。

① 注:为了更好地分析其它数据元素,标题(由于是最为普遍的选项)未被包括在图表之中。
② 在进行弹出式调查时,WorldCat.org 的“details”标签包括基本的书目信息,如果有目录和总结/摘要,这些信息也被包括在其中。

排名前五位的用户期望的数据质量改进措施

哪些变化将最有助于您找到所需的资源？



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查)

从用户经常选择的其它改进措施(例如从“更多主题信息”到“详细记录中包括更多的信息^①”)来看,被调查者表现出一种期望,想得到标准目录中未包含的数据元素。用户所期望的数据元素可能是结构化的或非结构化的,但它们都能帮助用户对结果集中资源的有用性进行评估,从而判断哪些资源值得花费时间去获取。

发现

本文接下来的两部分将会提供更多关于用户发现体验和传递体验的观察资料。本部分将首先讨论发现部分,重点将讨论不同用户类型间的差异——参加聚焦小组访谈的偶尔检索型用户、大学生型用户以及学者型用户以及弹出式问卷调查对象。

检索

聚焦小组的对象讨论了他们的检索经历,包括检索结果的相关性以及检索过程中所用到的辅助工具,如高级检索功能和分面检索功能。

检索结果的相关性。研究结果表明,用户的喜好和期望越来越多地受其利用搜索引擎(如谷歌)和网络书店(如亚马逊)经历的影响。当用户利用图书馆目录进行检索时,他们希望自己的搜索能准确地检索到他们想要的资源,他们想得到相关的结果。

对聚焦小组的研究发现,当用户搜索已知内容时,他们的检索策略经常是成功的,能返回相关的结果。这样,他们能很容易地找到自己想要的内容。一位访谈对象说她想找的内容恰恰就是第一条检索结果。进行已知资源检索时,一位学者型访谈对象说,“这是我的首选”。另一位访谈对象兴奋地解释了他是如何利用检索不仅得到了想要的内容而且还能发现新的内容:“我检索一个熟悉的资料,结果发现了一个比在亚马逊中得到的还要全面的列表”。

当检索已知内容时,聚焦小组成员通常都能成功地找到这些资源;采用的检索策略也比较简单,就是输入题名或作者。但恰恰是当用户对某主题进行一般性检索时(也就是检索未知资源时),当检索返回了与它所需内容不相关的结果时,他们表现出了不满。或许是因为用户不知道如何构建一个合适的主题检索策略。研究结果表明,用户希望仅输入少量单词,图书馆目录就能返回他们想要的内容;他们希望图书馆目录仅通过检索框中的几个术语就能“知道”他们

① 在进行弹出式调查时,WorldCat.org 的“details”标签包括基本的书目信息,如果有目录和总结/摘要,这些信息也将会包括在其中。

想要查找的内容。此外,如果所使用的单词在不同背景下有不同的含义,用户仍然希望通过检索能准确地返回他们想要的内容。如上所述,用户利用搜索引擎以及在检索中利用自然语言的经历好像已经影响到了他们对图书馆的期望。正如一位调查对象所要求的那样,“应该让它(图书馆目录检索)变得像谷歌图书搜索那样简单”。

聚焦小组的成员还提到,他们期望图书馆目录能根据流行检索词的通用含义,在检索算法中采用权重方式进行区分,以保证图书馆目录能找到用户所需要的内容。当这种权重区分很弱时,用户感觉图书馆目录未能满足他们的期望。一位用户谈到他曾对“running”进行检索,返回的结果中包括了公司经营(running a business)和一些书名中包括“running”的小说。实际上,他想检索的是关于跑步这项运动和锻炼的图书或其它资料,所以目录给出的结果让他非常吃惊。他觉得图书馆目录应该理解“running”这个词的一般用法,而且通过权重使其侧重运动或锻炼。

改善检索结果的相关性是一个非常有趣的数据质量问题,编目员所提供、获取、管理或挖掘的不同类型的元数据无法全部解决该问题。但是显而易见的是,用户的期望可能(至少是应该)会促进图书馆目录检索相关性功能的进一步发展。

高级检索。支持高级检索功能需要分别对书目数据元素进行标引,如标题、作者、主题、格式、出版日期等等。高级检索功能的有效运行还需要利用受控词汇加以保证,如海明威(Ernest Hemingway)的所有作品通过 Hemingway 的名字查询都可以检索出来。

聚焦小组成员在搜索已知资源时会利用 WorldCat.org 的高级检索功能。在聚焦小组访谈中发现,单一检索框好像有助于用户进行一般性检索(即未知资源)。但是,当用户查找一个已知资源时,检索的精确能力可能是检索成功的保障。高级检索功能可以使用户在检索结果中更快地找到所需内容,因为它所提供的结果更加精确。

Advanced Search

Enter search terms in at least one of the fields below

Keyword:
Return items with these words appearing anywhere

Title:
Return items with these words in the title

Author:
Return items by author

Subject:
Return items in this subject area

Guns, gems and steel
e.g. The Old Man and the Sea DVD

Always nice to be able to specify - or include all info

Fishing
e.g. Fishing

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(聚焦小组)

聚焦小组访谈中还发现,一部分学者型用户最初使用了除基本的关键词检索之外的更多的功能。主持人与他们讨论了对检索策略构建的期望,用户希望将高级检索功能作为一种可选项应放在显著位置,最好还有相应的使用说明。访谈对象能够理解高级检索的基本逻辑,但好像未能理解其全部功能。

优化检索。假设已经对数据进行了适当标引,那么这些已经存在于目录数据库中的字段

(例如作者、出版时间和类型等)可以被抽取和提供出来(通常放在网页左边栏目)进行分面导航,这也被称为分面检索或者分面浏览。分面检索可以对在检索框中输入检索词的直接检索方法进行补充,并可消除其中的不精确性。分面可以使用户在大量的、难以控制的检索结果集中不断地细化他们的选择。图书馆员已经在下一代目录界面中积极采用分面检索,因为利用图书馆编目标准和惯例所创建的数据库中已经包括了这些数据元素。研究表明,分面是对图书馆员创建、维护的受控术语和规范标目的另一种有效应用。

聚焦小组成员认为,WorldCat.org 中的分面浏览是一种“非常全面的检索结果优化方法”。查看比较短的检索结果列表是可行的,但关键词检索经常会返回数百个结果,这就使得分面浏览成为一个快速细化选择的有效工具。一位大学生表示,“当你得到数百个结果时,你就需要去优化检索,但得到 15 个结果时就有没必要了”。



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(聚焦小组)

聚焦小组的一些成员不太熟悉分面浏览,因此不能很快地理解该功能。而其他用户一看到分面浏览选项就明白如何使用分面浏览以及它应该返回什么样的结果。这一点在大学生和偶尔检索者中表现得特别明显。此外,他们还表现出使用分面浏览的强烈倾向。

与发现相关的信息元素

聚焦小组成员也讨论了从检索结果列表中判断所需内容时,哪些信息元素是最为有价值的。对这个内容的调查,弹出问卷调查结果与聚焦小组的调查结果是一致的。两组受访者都依赖并希望能有增强型内容,包括总结/摘要和文献的目录。

研究结果还表明,除了增强型内容外,对于部分用户来说社会性信息也非常重要。特别是对于部分用户来说,只要述评文章作者和评分者是可信赖的,这些评述和评分信息也是很有用的帮助发现所需内容的信息元素。

总结/摘要。三类聚焦小组都选择了总结信息作为改善目录的一种途径,而且弹出问卷回答者也在所提建议中经常提到它们。大部分聚焦小组成员认为,检索时在检索结果列表中和单个详细记录显示页面中,总结信息都应该列于比较显著的位置。一位被调查者提出:“我希望在结果页面中能够列出一小段(一行)关于图书的短评,就像谷歌为你提供一小段关于被链接网站的内容那样”。

研究表明,在搜索未知资源时总结信息是最有用的信息。聚焦小组成员以及弹出问卷回答者都对快速确认所需内容表现出兴趣,以便判断这些内容是否值得花时间去看看。当进一步问及寻找所需内容时什么信息最为重要,一位学者评价道:“老实说,没什么特别的东西,我想应该是总结信息,那样我就可以决定是否去获得它”。一位被调查者表达出同样的观点:“请包括一些关于内容的描述,这样浏览者在出发前就能知道是否值得花时间去图书馆找这本书了”。

其它与发现相关的研究结果

在与发现相关的最重要数据元素的调查中,聚焦小组与弹出问卷调查的结果相互印证,包括资源的详细信息、总结/摘要以及目次信息。因为聚焦小组访谈是交互式的,所以可用于挖掘封面、摘录、述评、评价以及用户提供的列表在辅助发现以及资源选择上的作用。

封面。三类聚焦小组成员都对显示图书封面表示欢迎。正面评价出现在所有聚焦小组访谈中,如“我喜欢看图书的照片”、“我喜欢比较突出的封面图像”以及“喜欢看到图书像什么样子”。大学生们对看到图书封面图片表现出特别突出的兴趣。访谈对象认为在检索结果列表和单个记录详细显示中给出封面都有积极的价值。

摘录。三类聚焦小组成员都希望能对检索结果抽取内容摘录,并希望能够在在线查看这些内容摘录。他们并未将这种愿望局限于图书,也希望能够对诸如音频和视频文件这样的载体内容进行摘录。



来源:联机目录:用户和图书馆员需要什么, OCLC, 2009(聚焦小组)

相关主题以及主题分类。大学生聚焦小组的用户喜欢将“相关主题”作为内容有用性的判断方法。在 WorldCat.org 单一记录显示页面中,“相关主题”标签排在受控主题术语前。当作为分面类别显示在左框时,聚焦小组许多成员发现它们有助于检索结果的优化和浏览。

社会性信息。聚焦小组成员对社会性信息(例如创建评论、共享列表功能等)的反应不一。大学生聚焦小组成员更欢迎用户提供的信息,而且很明白自己区分权威与非权威评论的能力。

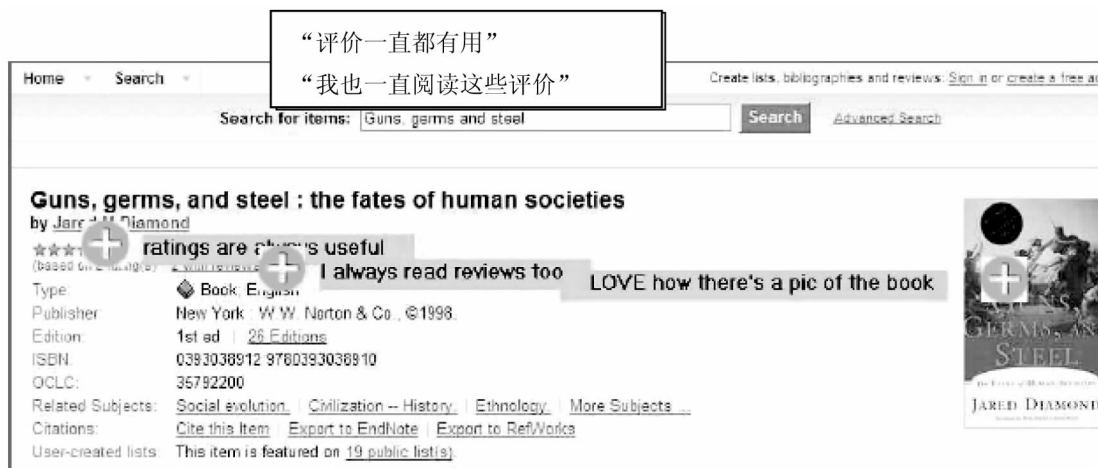
整体来讲,编辑评价要比用户提供的信息更加有价值,尤其是当聚焦小组成员的信息需求

是针对学术或专业目的时,这种观点更突出。

当不是进行学术检索时,用户提供的评价至少在一定程度上还是会吸引所有三类聚焦小组的部分成员。学者对专业评价更有兴趣,而对其它用户的观点表现出的兴趣不大,除非该用户是“专家”。

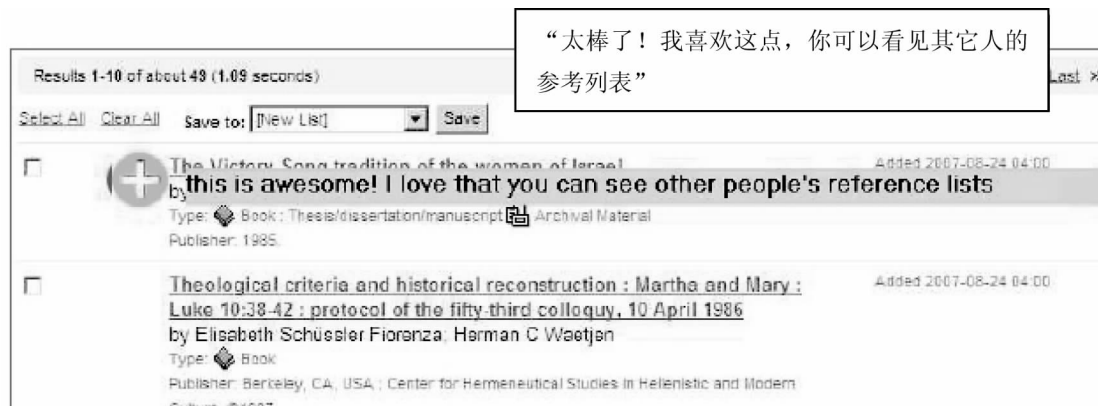
尽管部分偶尔检索型用户感觉评价信息是“枯燥无聊的”,但也有访谈对象表示,因为他的职业(篮球教练),他喜欢阅读评价;“我们彼此学习”。

大学生聚焦小组成员表示,当判断需要什么样的资源时,评价将是有用的。尽管他们也愿意了解谁写了这个评论,但并不是所有的人都认为这些评论必须是“专业”的。他们觉得每个人都有不同的品味和观点,而且这些观点都有价值。虽然不是所有的人都认为专业评价和编辑评价是必须的,但都承认这些评价是有用的。



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(聚焦小组)

大学生聚焦小组成员认同用户创建列表的价值,它们可以帮助用户利用 WordCat.org 来跟踪和共享资源。用户可以利用这些列表引用相关资源,可以与朋友和同事共享这些列表。这些访谈对象认为,能够了解别人出于什么目的正在利用什么资源是非常有益的。同时他们还认为,了解谁在利用资源也会增加该资源的可信度。



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(聚焦小组)

偶尔检索者以及学者们并没有感觉到用户创建列表的有用性。一位偶尔检索者将用户创建列表描述为“一群读了某本书的随机用户群把它放到列表某一位置…在进行检索时一点用都没有…”。这些访谈对象希望能知道更多的关于这些列表作者的信息。

传递

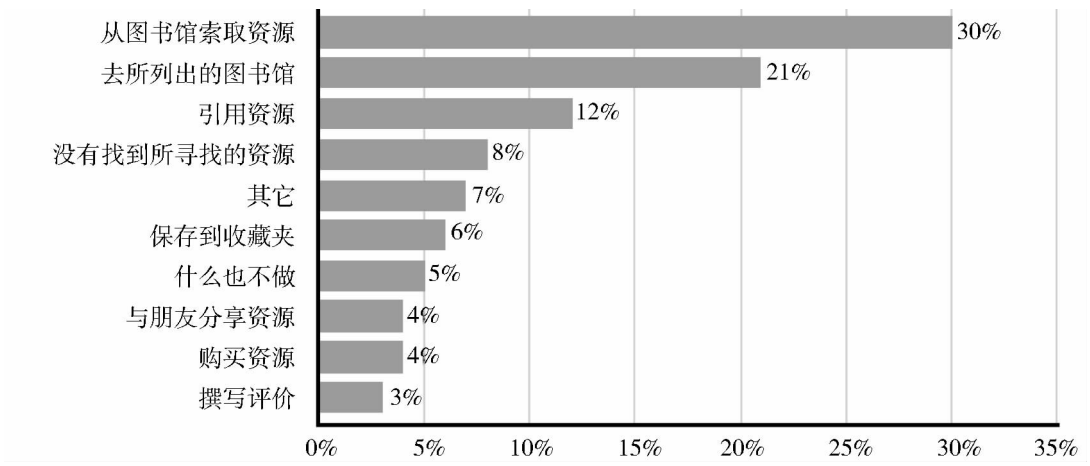
发现是重要的,但传递的重要性不亚于发现,如果不是更重要的。本研究发现,从发现到传递的无缝、便捷的连接对用户是非常重要的。这点或许看起来显而易见,但我们仍需记住:对大部分用户来说,如果不能得到所需内容,发现本身就只是浪费时间。

弹出式问卷调查结果也证实了传递作为检索目标的重要性。在询问用户如何处理从WorldCat 检索到的信息时,大部分弹出问卷回答者都选择了与传递相关的选项;30%的调查对象将向图书馆索取资源,21%的调查对象将访问列出的图书馆。

如何处理信息

基于检索到的信息您将做什么？

基数:弹出式调查的调查对象



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查)

资源所有权与可获得性。研究结果表明,用户想知道资源是否可获得,最好是在对检索结果排序时。对于聚焦小组的部分成员来说,与传递相关的信息不仅是有用,而且是极其重要的。一些偶尔检索者甚至建议,如果资源不可获得,它就不应该出现在检索结果列表中。正如一位访谈对象所言,“如果你都无法获得这个资源,为什么把它列进结果中? …”。一位大学生组成员希望设立一种“优选图书馆”账号,并据此显示检索结果。

这些研究成果表明,我们仍需投入更多的努力去研究哪些数据元素该如何改进,以使用户对资料的传递体验尽可能是积极的,无论资料本身是印刷型、授权使用的或是数字资料。

访问在线内容。研究结果特别强调,用户对从图书馆目录直接链接到和方便获得在线内容(无论是在线阅读还是下载)的需求。传递,实际上是即时传递,是用户体验不同可下载媒体后对图书馆的期望:“可下载媒体是最伟大的概念,但2008年必须解决它1当你购买了它,就应即时得到它。电子书和MP3的下载是一个必须的条件”,一位偶尔检索型者总结信息即时满足需求时这么说。

从数据质量的角度来看,从图书馆目录便捷访问在线内容的需求可能需要增加链接元数据管理以及第三方数据互操作方面的投入。

数据质量:图书馆员的需求

内容概览

图书馆员的需求是什么?

- 合并重复记录;
- 解决排印错误;
- 升级简洁的记录;
- 增加评价内容,比如目录、摘要、或者封面等。

主要研究结论

与用户一样,图书馆员都是有目的地查找目录和目录数据。用户一般寻找并获得所需信息;图书馆员一般是因工作需要而利用目录数据。图书馆员的数据质量偏好主要取决于他们的工作任务。

图书馆员的工作任务一般分为两大类型:涉及到与终端用户直接接触的工作任务;主要支持图书馆业务的工作任务,比如为图书馆选择和采购新资料。图书馆员的工作类型对其关于目录数据质量的选择偏好有很大的影响。

重复的记录(同一版本的多条记录)阻碍了图书馆员的工作。

不论是在北美或在其他地方,无论是何种类型的图书馆,无论图书馆馆员的工作类型是什么,合并 WorldCat 中重复的记录都是重中之重。一位图书馆员的回应代表了普遍的情绪:“我想消除所有重复的记录!我花了很多时间试图确定各个记录有什么不同,然后选择我所需要的。如果没有那些重复的记录,我的工作将会轻松很多。”对于图书馆员来说,合并、删除或更好地管理重复的记录似乎是提高效益和效率的当务之急。

图书馆员认为用目次数据丰富编目记录是一个重点。

从一长串改进数据质量潜在措施中,被调查者一致把“将目次添加到书目记录中”为提高数据质量的第二个重点(在合并重复记录之后)。各种类型的图书馆员对此持一致意见。

除去合并重复记录和添加目次信息外,其它关于数据质量工作的优先性选择因图书馆员工作角色、图书馆类型以及地区的不同而存在很大的差异。

在图书馆界,对提高目录数据质量的期望和偏好似乎取决于多种因素。北美内部和北美外部,不同类型图书馆,不同角色图书馆员间,对于第三个提高数据质量的重点没有形成共识。例如,北美外图书馆选择最多的第三项重点是为非英语材料添加更多记录,而这个选择在北美内部图书馆的提高数据质量选项中甚至没有进入前十。在学术和公共图书馆中,升级简洁的数据记录位列第四,但是在专业图书馆的选择中,这项内容也没有进入前十。

详细分析

概述

研究小组通过网络调查方式从 1397 名图书馆员中收集到数据,其中 74% 来自北美,26% 来自其他国家。网络调查的对象是通过用户或工作人员界面访问 WorldCat 的图书馆员,包括通

过 WorldCat. org、WorldCat FirstSearch、WorldCat Z39. 50 协议端口(主要用在 WorldCat 上编目)和 WorldCat Resource Sharing(一般是馆际互借使用)等界面。

接受调查的图书馆员包括图书馆馆长、采访、编目、馆藏开发、资源共享和参考咨询等类别,他们可以选择一种以上的工作角色。本章分析对比了不同类型的被调查者,揭示了不同类别被调查者间的若干显著差异。统计意义上的显著差异是指回复意见中那些并非由随机因素引起的差异。

对图书馆的研究是为了评估哪些数据元素被图书馆员认为在从 WorldCat 发现资料最有帮助,哪些提高数据质量的举措是图书馆员所需的。

本研究集中于图书馆员对 WorldCat 数据质量的偏好,而不是某个具体 WorldCat 界面。本报告介绍了关于提高 WorldCat 数据质量的发现。在那些 WorldCat 和图书馆目录都存在的数据质量问题上(例如重复记录),本研究的结果适用于一般的图书馆联机目录和集成系统的数据质量要求。

期待的数据质量提高措施

整体图书馆受访者期待的数据质量提高措施

为了编制一份全面的数据质量改进的清单,反应尽可能多的观点,调查设计者研究了许多目录数据的研究结果,提出了 18 个改进措施选项,要求被调查者选择一个最希望的措施,另外再从 18 个选项清单中选择希望的其他措施。下面图表给出了所有被调查者的选择排序。从下图可见,位列前三的提高数据质量措施是:合并重复记录(52%),在详细书目记录中添加目次信息(40%),在详细书目记录中添加摘要(28%)。

名列前茅的措施被选择的原因

调查提示被调查者说明选择最希望的措施的原因。建议给详细书目记录添加目次信息的 113 个被调查者表示,这样可以为图书馆员和用户在确认内容时提供足够的详细信息(53%),或者能够加强检索(24%)。有 8 位被调查者说明了选择为详细书目记录添加摘要的原因,超过三分之二(68%)认为摘要可以帮助图书馆员和用户确定所需内容。

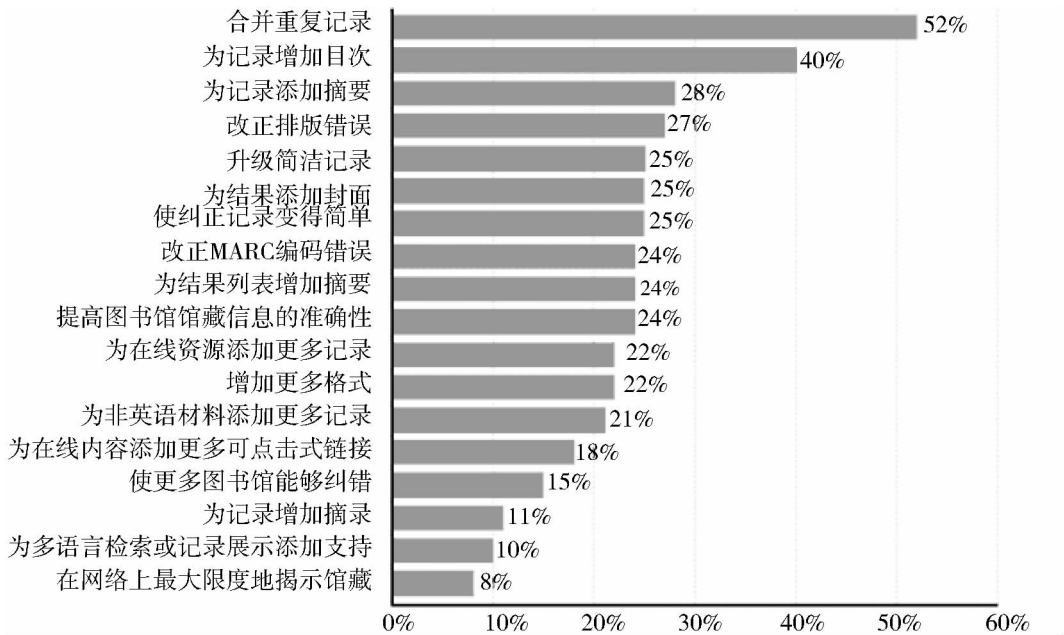
不同类型被调查者所期待的提高数据质量措施

所有被调查者的选择排序受到被调查图书馆构成情况的影响,其中一半多来自高校图书馆,近四分之三来自北美图书馆,近三分之二的工作职责与编目相关。按照不同类型被调查者来分析的提高数据质量措施选择,揭示了不同类型图书馆、不同类型工作角色以及不同地区的一些有趣差异。

期待的数据质量提高措施

你会推荐以下哪些数据质量提高措施？

基数:被调查的图书馆员



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

不同类型图书馆期待的提高数据质量措施

此分析揭示了不同类型图书馆被调查者在选择上的若干显著差异。下面的表格列出了来自高校图书馆,公共图书馆和专业图书馆的被调查选择的前十项提高质量的措施。

期待的数据质量提高措施

你将推荐以下哪些数据质量提高措施

根据图书馆的类型

提高措施	总数		高校图书馆		公共图书馆		专业图书馆	
	排名	百分比	排名	百分比	排名	百分比	排名	百分比
合并重复目录	1	52%	1	57%	1	48%	1	56%
为记录增加目次	2	40%	2	46%	5	27%	2	44%
为记录添加摘要	3	28%	3	29%	6	27%	3	28%
改正排印错误	4	27%	5	28%	3	29%	4	27%
升级简洁记录	5	25%	4	29%	4	29%	14	17%
为结果添加封面	6	25%	11	23%	2	35%	12	18%
使纠正记录变得简单	7	25%	6	26%	11	25%	6	24%
改正 MARC 编码错误	8	24%	7	26%	9	25%	8	21%
为结果列表增加摘要	9	24%	9	23%	7	27%	10	19%
提高图书馆馆藏信息的准确性	10	24%	13	22%	8	25%	5	25%
为在线资源添加更多记录	11	22%	8	26%	14	14%	7	24%
增加更多格式	12	22%	10	23%	10	25%	11	19%
为在线内容添加更多可点击式链接	14	18%	14	20%	13	15%	9	21%

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

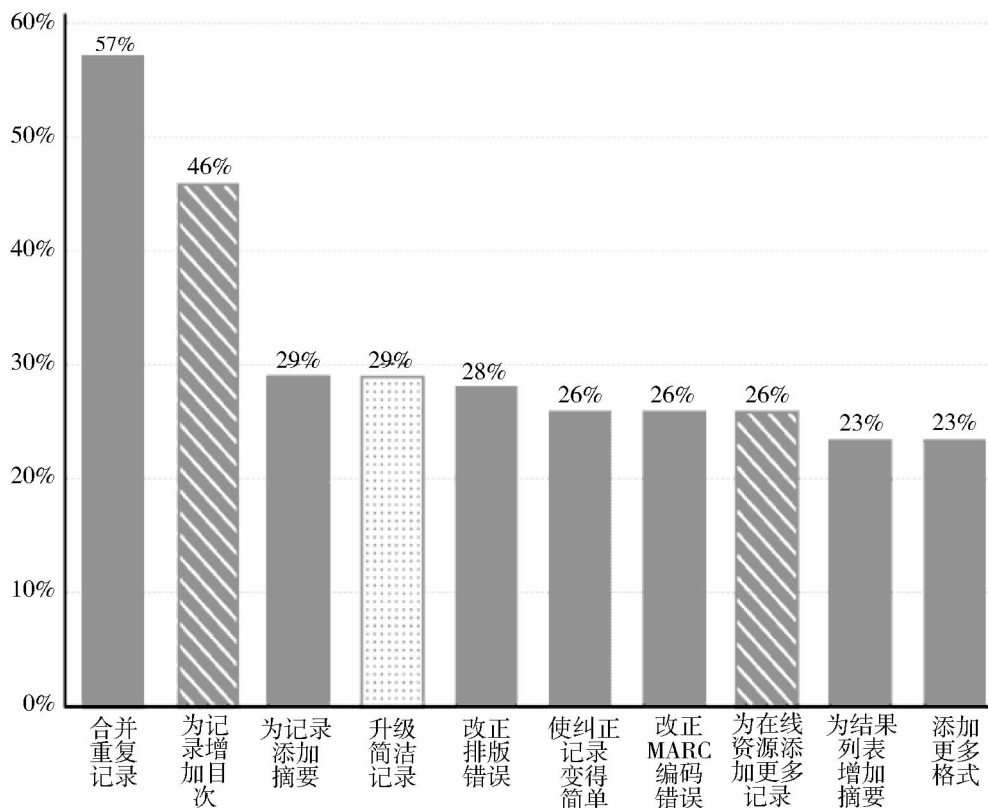
在下面的图表中,显著差异存在形式用栏目形式来表示。例如,条纹栏表示选择此项措施的高校图书馆员和专业图书馆员显著多于公共图书馆员,虚线栏表示选择此项措施的高校图书馆员和公共图书馆员显著多于专业图书馆员,格子栏表示选择此项措施的公共图书馆员显著多于高校图书馆员和专业图书馆员。

高校图书馆被调查者。与专业图书馆被调查者相比,高校图书馆和公共图书馆的被调查者明显更愿意优先选择升级简洁记录(注意虚线栏)。

与公共图书馆被调查者相比,高校图书馆和专业图书馆的被调查者明显更倾向于选择为记录增加目次信息(注意条纹栏)。更多的高校图书馆和专业图书馆的被调查者明显更倾向于把为在线资源添加更多记录列入前十项改进措施内——而此项措施在公共图书馆被调查者的选择中没有出现在前十(注意条纹栏)。虽然在任意一组中,为非英语资料添加更多记录都没有进入前十,但很显然的,与专业图书馆的被调查者相比,更多的高校图书馆和公共图书馆的被调查者选择了它。

排名前十位的数据质量提高措施
你将选择以下哪些数据质量提高措施

基数:高校图书馆被调查者



来源:联机目录:用户和图书馆员需要什么, OCLC, 2009 (图书馆调查)



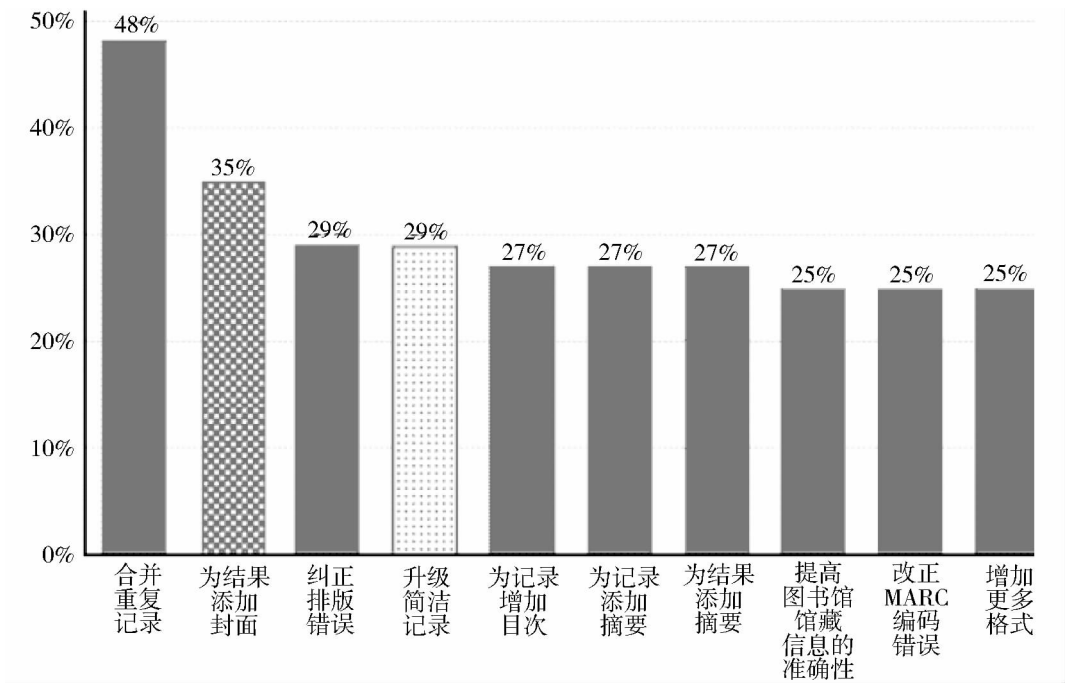
专业图书馆和高校图书馆被调查者显著多于公共图书馆被调查者





公共图书馆和高校图书馆被调查者显著多于专业图书馆被调查者

公共图书馆受访者。不同于其它受访组,公共图书馆被调查者选择的第二项提高数据质量措施是为结果添加封面(35%),而不是为记录增加目次信息。与高校图书馆的 23% 和专业图书馆的 18% (都没有列入前十)相比,很显然地,更多的公共图书馆被调查者选择了它(注意格子栏)。如前所述,与专业图书馆被调查者相比,公共图书馆和高校图书馆被调查者明显更倾向于优先选择升级第三级记录或其它简洁记录(注意虚线栏)。

排名前十位的的数据质量提高措施
你将选择以下哪些数据质量提高措施
基数:公共图书馆被调查者

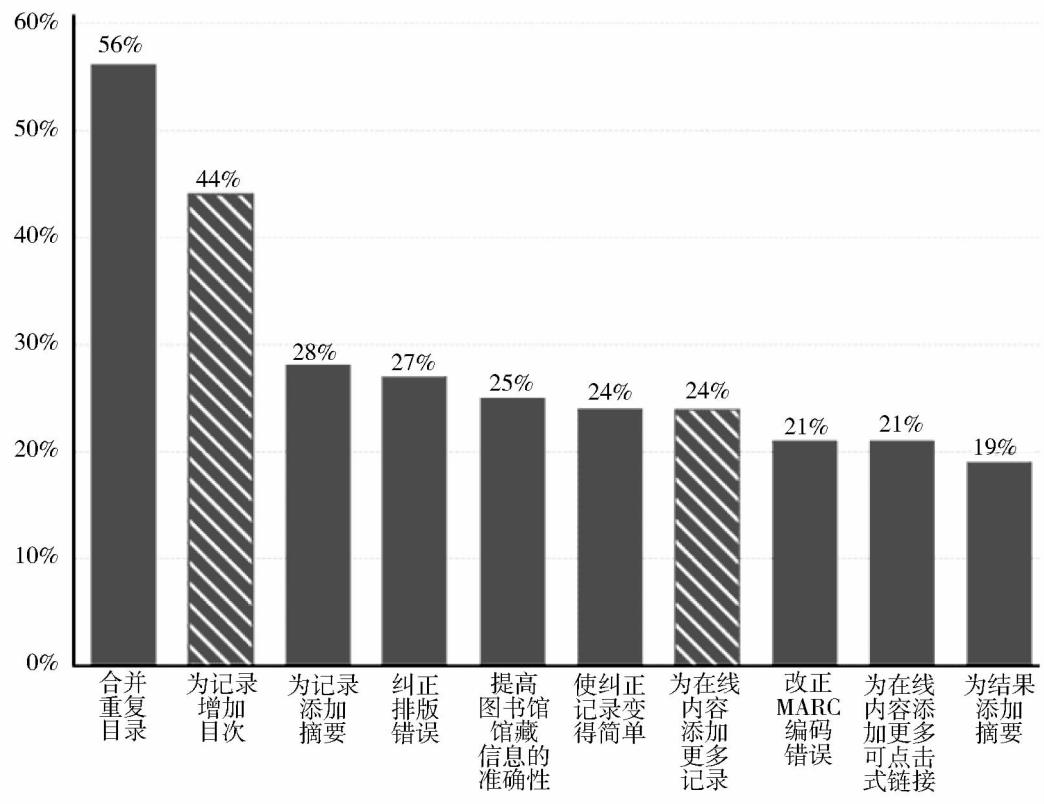


来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

-  公共图书馆被调查者的选择显著多于高校图书馆和专业图书馆被调查者
-  公共图书馆和高校图书馆被调查者的选择显著多于专业图书馆被调查者

专业图书馆被调查者。如前所述,与公共图书馆的被调查者相比,专业图书馆和高校图书馆被调查者更倾向于优先选择为记录增加目次和为在线内容添加更多记录(条状栏)。

排名前十位的数据质量提高措施
你将选择以下哪些数据质量提高措施
基数:专业图书馆被调查者



来源:联机目录:用户和图书馆员需要什么, OCLC, 2009(图书馆调查)

■专业图书馆和高校图书馆被调查者的选择显著多于公共图书馆被调查者

不同工作岗位期待的提高数据质量措施

下页表格列出了六种图书馆员角色中至少一种角色类别选择的位列前十的提高数据质量措施。这六种馆员角色是:编目、采访、图书馆馆长、参考咨询、馆藏建设和资源共享。

不论哪种工作岗位,合并重复目录都是首要选择,为记录增加目次位于第二选择。除了参考咨询和资源共享工作人员外,为在线资源添加更多记录在其它组中都进入前十位。

为非英语材料添加更多记录在编目人员的选择中进入前十,但在其它任何组中都没有进入前十。

馆长组是唯一没有将“纠正或丰富记录使工作变轻松”这一项列入前十的一个组。

此表后面列出了每个角色提出的前十个提高数据质量措施;不同工作角色前十项选择的显著差异将在后面给出。

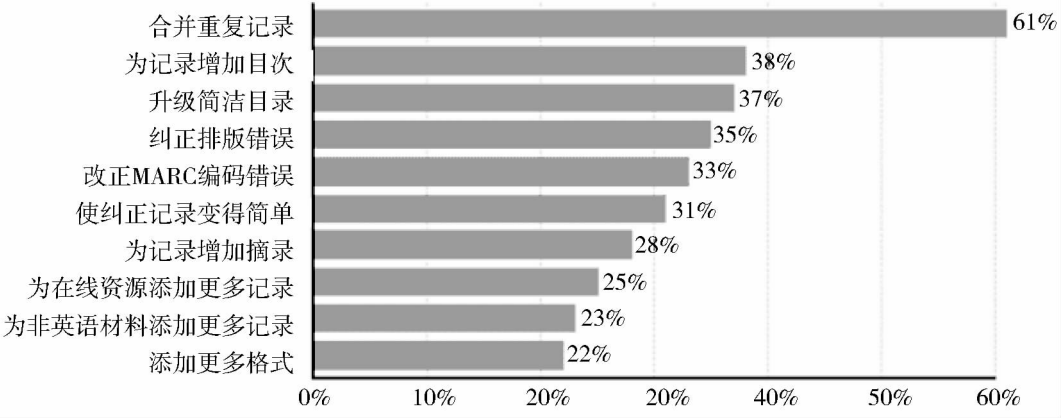
期望的数据质量提高措施
你将选择以下哪些数据质量提高措施
基数:工作岗位

提高内容	总数		编目		采访		图书馆馆长		参考咨询		馆藏建设		资源共享	
	排名	%	排名	%	排名	%	排名	%	排名	%	排名	%	排名	%
合并重复目录	1	52%	1	61%	1	55%	1	51%	1	52%	1	55%	1	49%
为记录增加目次	2	40%	2	38%	2	40%	2	47%	2	45%	2	46%	2	42%
为记录添加摘要	3	28%	7	28%	7	27%	5	30%	3	30%	3	31%	4	27%
改正排版错误	4	27%	4	35%	3	31%	8	24%	6	26%	4	29%	9	21%
升级简洁记录	5	25%	3	37%	4	30%	14	18%	13	19%	12	23%	14	16%
为结果添加封面	6	25%	11	22%	8	26%	4	30%	7	25%	7	26%	8	23%
使纠正记录变得简单	7	25%	6	31%	5	29%	12	21%	9	22%	8	25%	10	20%
改正 MARC 编码错误	8	24%	5	33%	6	29%	11	21%	10	21%	9	25%	11	19%
为结果列表增加摘要	9	24%	14	19%	12	21%	3	31%	4	29%	6	27%	5	25%
提高图书馆馆藏信息的准确性	10	24%	12	20%	9	26%	10	22%	5	29%	5	28%	3	38%
为在线资源添加更多记录	11	22%	8	25%	10	26%	9	23%	12	19%	10	25%	12	19%
增加更多格式	12	22%	10	22%	11	22%	7	26%	8	23%	11	24%	7	23%
为非英文材料添加更多记录	13	21%	9	23%	13	20%	13	21%	14	15%	14	16%	13	17%
为在线内容添加更多可点击式链接	14	18%	15	14%	14	17%	6	28%	11	21%	13	22%	6	23%

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

以下两页的图表给出了每种工作岗位所选择的前十个提高质量措施,这些工作岗位包括:编目、采访、图书馆馆长、参考咨询、馆藏建设和资源共享。

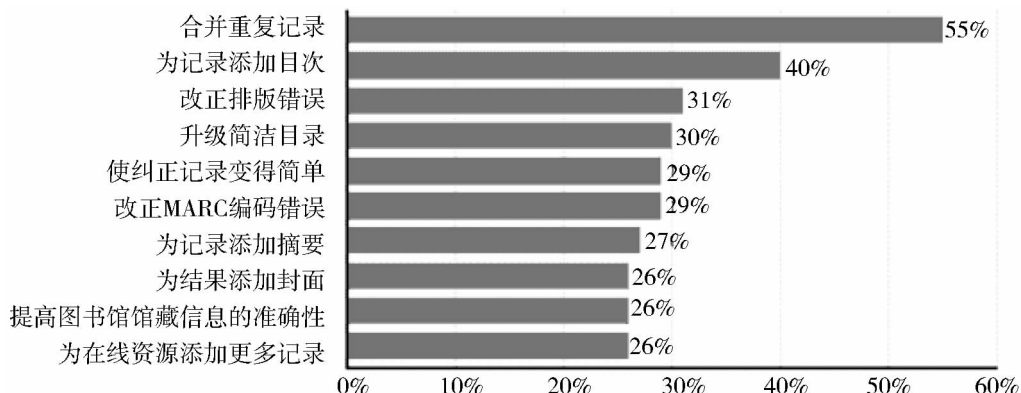
编目——排名前十位的质量提高措施
你将选择以下哪些数据质量提高措施



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

采购——排名前十位的质量提高措施

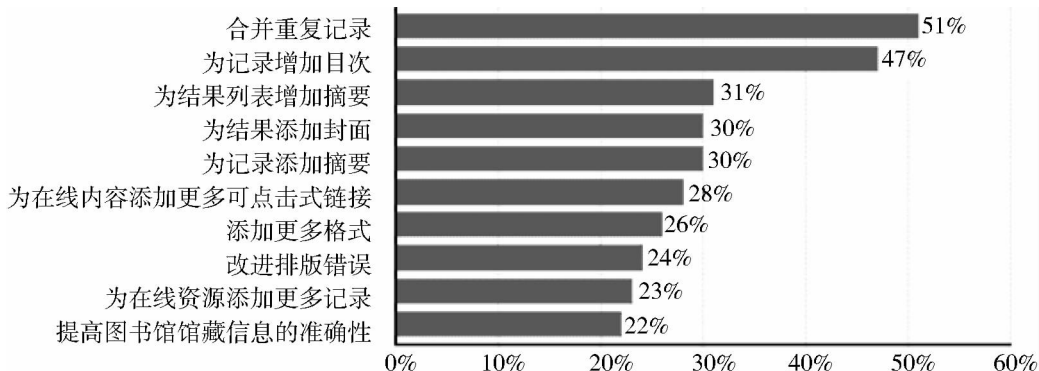
你将选择以下哪些数据质量提高措施



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

图书馆馆长——排名前十位的质量提高措施

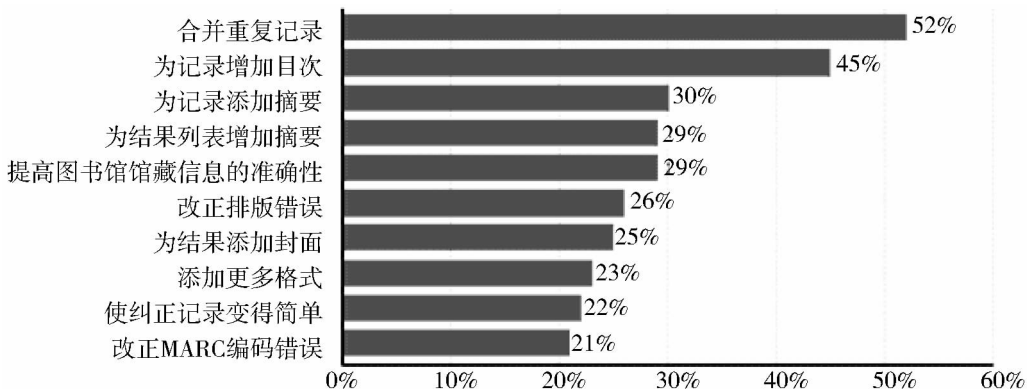
你将选择以下哪些数据质量提高措施



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

参考咨询——排名前十位的质量提高措施

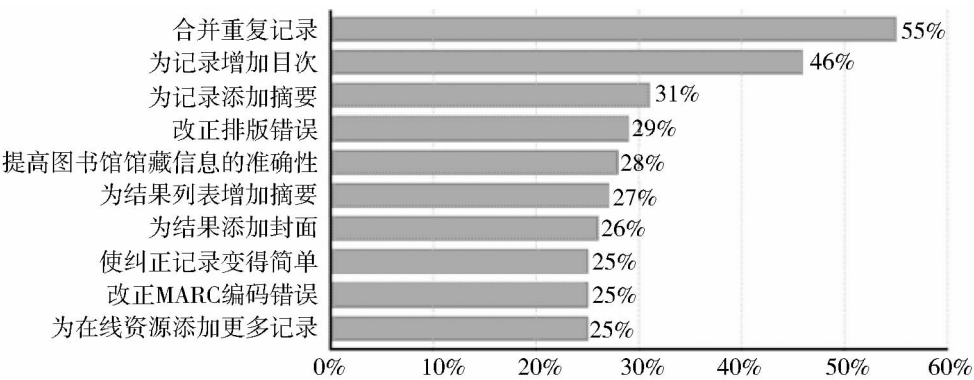
你将选择以下哪些数据质量提高措施



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

馆藏建设——排名前十位的质量提高措施

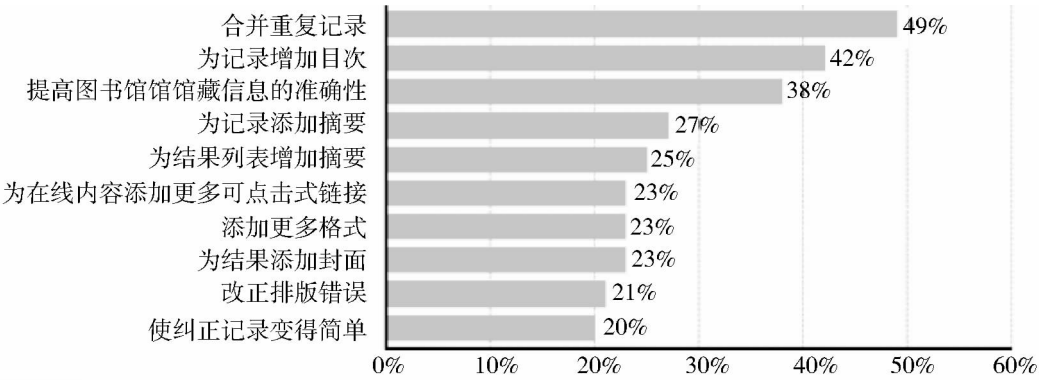
你将选择以下哪些数据质量提高措施



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

资源共享——排名前十位的质量提高措施

你将选择以下哪些数据质量提高措施



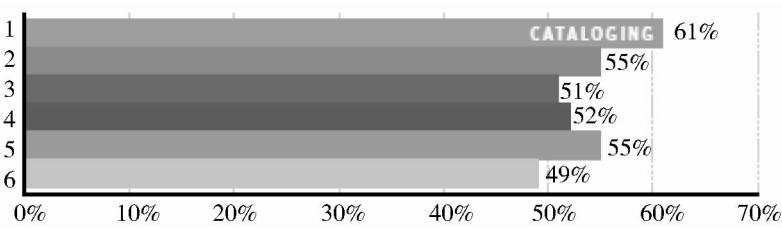
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

工作职责为编目的被调查者

很显然,与图书馆馆长、参考咨询人员或者资源共享人员相比,更多的编目岗位的被调查者将合并重复记录作为首要选择。

合并重复记录

你将选择以下哪些数据质量提高措施



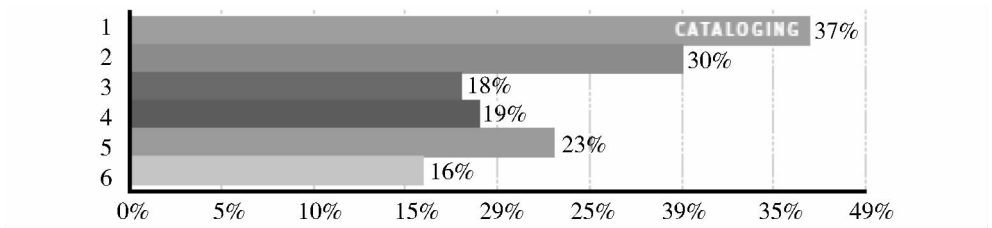
1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

类似地,与图书馆馆长、参考咨询、馆藏发展和资源共享人员相比,编目人员更加倾向于选择升级第三级记录或者其他简洁记录。

升级简洁记录

你将选择以下哪些数据质量提高措施



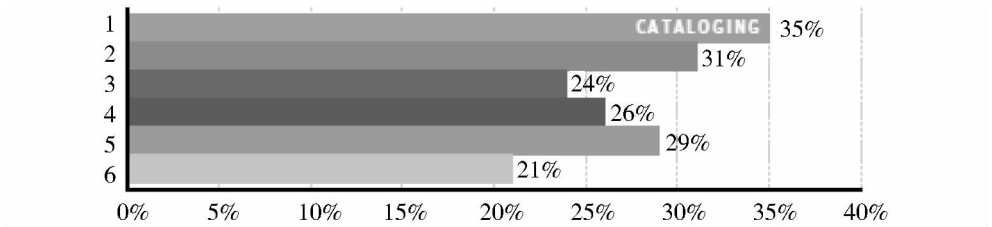
1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

与资源共享、参考咨询人员和图书馆馆长相比,编目人员更倾向于选择改正排版错误。

改正排版错误

你将选择以下哪些数据质量提高措施



1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

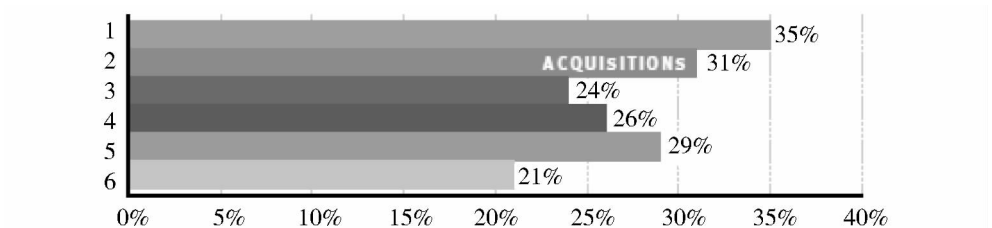
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

工作岗位为采访的被调查者

与工作岗位为资源共享的被调查者相比,采访人员更倾向于选择改正排版错误。

改正排版错误

你将选择以下哪些数据质量提高措施



1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

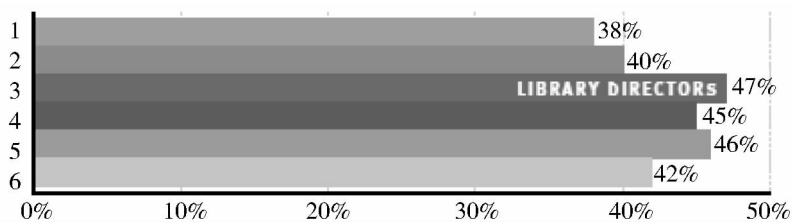
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

工作岗位为馆长的被调查者

尽管各种工作职责的被调查者都将为记录添加目次信息列入前十位,与编目者相比,更多

的图书馆馆长选择了这一项。

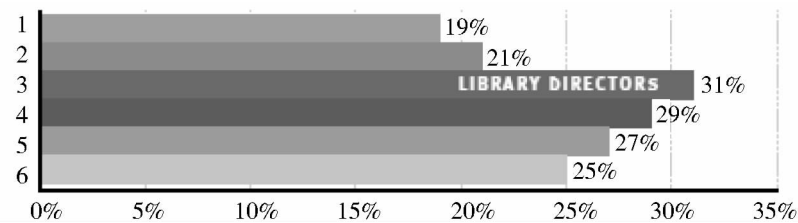
为记录添加目次
你将选择以下哪些数据质量提高措施



1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

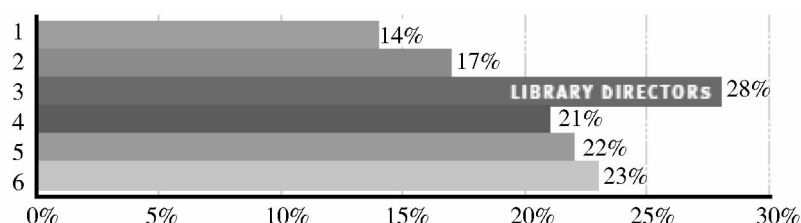
与负责采访和编目的被调查者相比,图书馆馆长更倾向于优先选择为结果列表上的条目添加摘要和增加更多的在线内容链接。

为结果条目添加摘要
你将选择以下哪些数据质量提高措施



1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

增加更多的在线内容有效链接
你将选择以下哪些数据质量提高措施



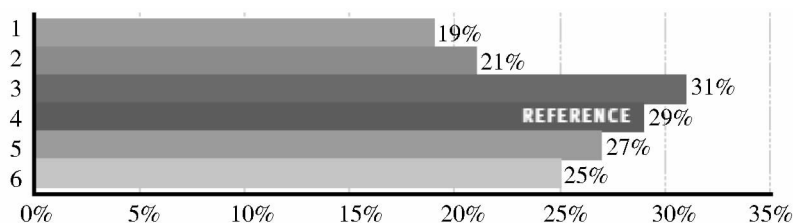
1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

工作岗位为参考咨询的被调查者

与编目被调查者相比,参考咨询馆员更倾向于优先选择为结果列表添加摘要和提高图书馆馆藏信息的准确性和实时性。

为结果列表添加摘要

你将选择以下哪些数据质量提高措施

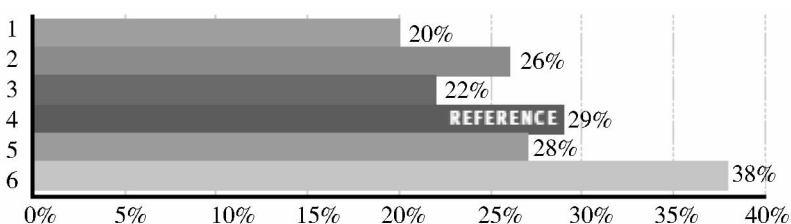


1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

来源:联机目录:用户和图书馆员需要什么, OCLC, 2009 (图书馆调查)

提高图书馆馆藏信息的准确性

你将选择以下哪些数据质量提高措施



1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

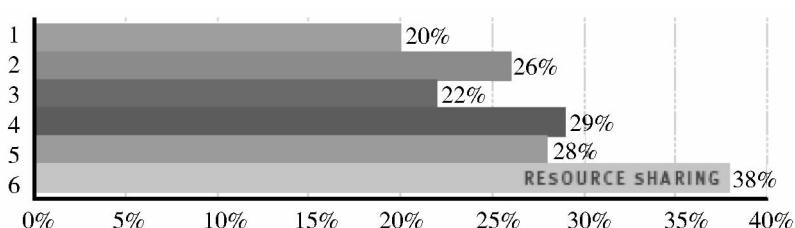
来源:联机目录:用户和图书馆员需要什么, OCLC, 2009 (图书馆调查)

工作岗位为资源共享的被调查者

与其它任何组的被调查者相比,负责资源共享的被调查者更加倾向于选择提高图书馆馆藏信息准确性和实时性。

提高馆藏信息的准确性

你将选择以下哪些数据质量提高措施



1 编目 2 采访 3 图书馆馆长 4 参考咨询 5 馆藏发展 6 资源共享

来源:联机目录:用户和图书馆员需要什么, OCLC, 2009 (图书馆调查)

不同地区被调查者期望的提高数据质量措施

本部分揭示了北美内外部不同区域图书馆之间的若干差异。北美外部图书馆的被调查者更加倾向于选择为非英文材料和在线材料添加更多记录这两项,而在北美内部,这两项都没有进入前十。另一项进入北美外部受访者选择列入前十位而没有被北美内部列入前十位的是添

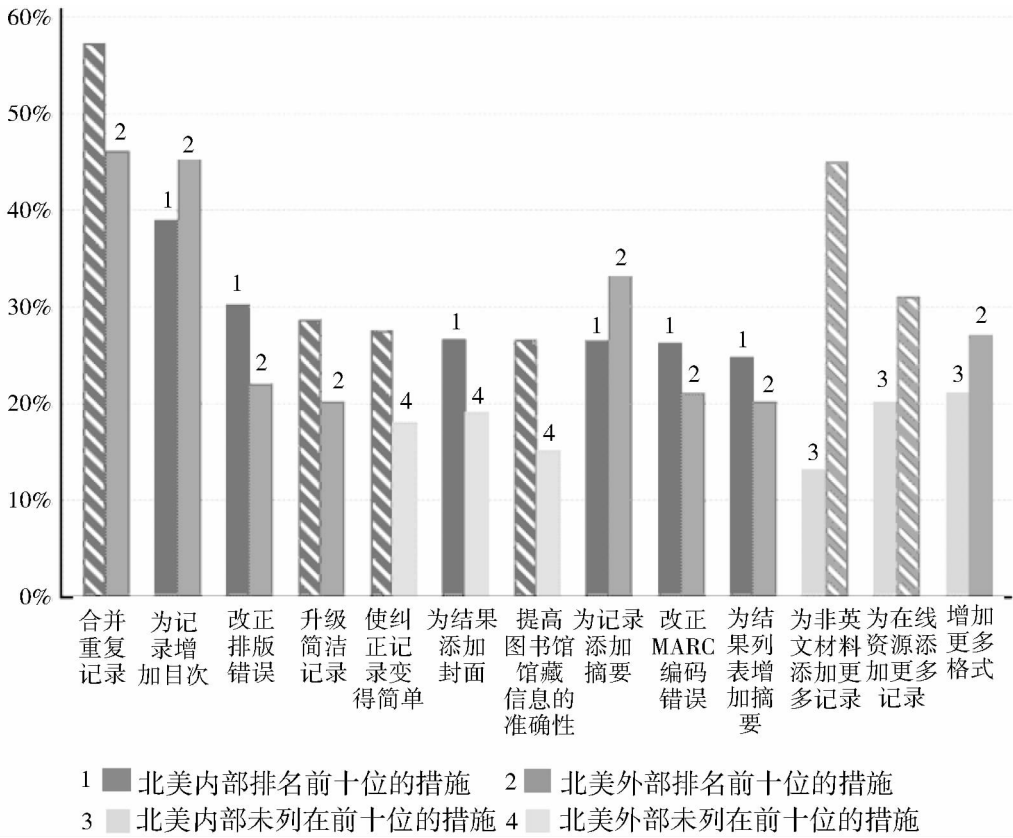
加更多格式(例如文章、图片、图表、音频、视频等)。

尽管无论北美内外的被调查者都选择合并重复记录为首要选择,但很显然地,与北美外部相比,更多的北美被调查者选择了此项。同样地,更多的北美被调查者选择了升级第三级记录和其它简洁记录、使图书馆纠正和丰富记录变得简单、提高图书馆馆藏信息的准确性和实时性。使图书馆纠正和丰富记录变得简单、为结果列表上的条目增加封面、提高图书馆馆藏信息的准确性和实时性位列北美被调查者选择的前十位,但在北美外部被调查者的选择中却没有进入前十位。

排名前十位的提高数据质量措施

你将选择以下哪些数据质量提高措施

基数:北美内部和外部的图书馆被调查者



条纹栏指存在明显不同

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

数据质量:图书馆员与终端用户

主要研究结论

- 终端用户和图书馆员都高度重视与印刷型或其他非数字资料的传递相关的数据元素,包括了解资料保存在哪里,哪些是可以立即获得的。

- 终端用户最重视通过“建立更多在线内容的链接”来改进目录质量,而图书馆员最重视整合重复记录。

图书馆员把“更多在线内容的链接”排在他们改进措施列表中的倒数第三位。然而,当被要求描述终端用户的偏好时,许多直接为终端用户服务的图书馆员都能准确预测到终端用户对从目录链接到在线内容的高度期待。

尽管图书馆员的数据质量优先次序明显不同于终端用户,但许多直接为终端用户提供服务的图书馆员非常了解终端用户关注的质量问题,例如丰富的内容、更多的关于非印刷载体资料的记录(例如音频和视频)及从目录中可获得的示例文本/摘录。

- 图书馆员选择的重要目录数据元素往往与已知资料检索相关,例如 ISBN 就是图书馆员确认所需资料时优先选择的数据元素。

- 终端用户对于发现所需资料的重要目录数据元素的选择,表明他们重视并且期待那些评价性的信息(如摘要),以帮助他们发现并选择所需资料。

- 增加目次和摘要信息对于终端用户和图书馆员同样重要。

通常情况下,标准的目录记录不包含此类数据元素。

- 除目录和摘要信息之外,终端用户和图书馆员对于目录数据质量的需求往往不同。

这种不同似乎主要是受利用目录的目的影响,比如:终端用户要发现并获取所需信息,图书馆员要完成一项工作任务(例如订购新书)。

- 图书馆员对数据质量改进措施的选择反映了他们对目录中精确、结构化数据的重要性的重视。

终端用户从这种结构化数据中受益,例如当他们通过分面类别优化检索时;然而,终端用户往往对相应的数据质量要求不清楚,而正是这些数据质量要求才能支持那些准确和一致的检索功能的实现。这种现象是图书馆员和终端用户之间在数据质量改进措施选择上存在不同的另一个动因。

详细分析

概述

本章第一部分在两个层面上比较终端用户弹出问卷调查结果和图书馆员调查结果:

- 确定所需内容的最重要的目录数据元素。
- 推荐的目录数据质量提高措施。

本报告中,终端用户的调查结果是基于对 WorldCat. org 的使用反馈。对于“对确定所需内容来说最重要的目录数据元素”,研究团队抽取了那些使用 WorldCat. org 的图书馆员的反馈信

息作为比较。

关于“期望的目录数据质量改进措施”,调查结果显示的是终端用户对 WorldCat. org 的反馈和从任一 OCLC 接口使用 WorldCat 数据的图书馆员的反馈。

在那些 WorldCat 和单个图书馆联机目录都具有的同样的数据质量问题上,本研究结果可推广到单个图书馆的联机目录需求。

如上所述,基于检索者利用数据的目的,本研究只评价其数据质量方面的需求,而不是交互界面的需求;对于终端用户来说,目的是发现和获得内容来满足信息需求,而对图书馆员来说,目的是完成工作任务。

本章的第二部分分析了图书馆被调查者关于对终端用户帮助最大的数据质量改进措施的预测,还比较了图书馆员的预测和终端用户实际需求之间的不同。

图书馆员和终端用户的需求比较

最重要的数据元素

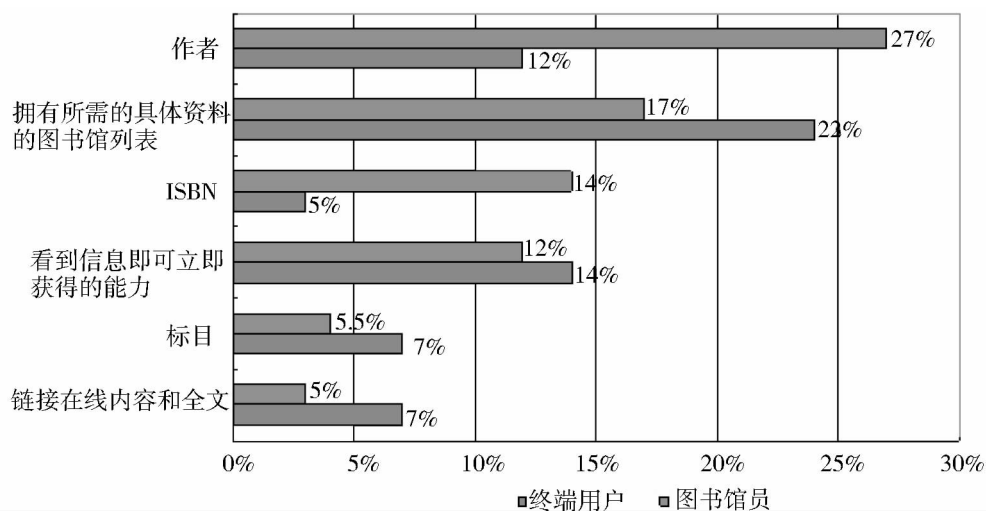
图书馆员和终端用户被问及目录记录中哪些信息对确定其所需内容是最重要的。下表详细展示了图书馆员和终端用户的选择。

比较图书馆员和终端用户的选择,两者均认为“作者、拥有该资料的馆藏单位列表及资料可获得性”是重要信息。值得注意的是,“ISBN”在图书馆员中被认为是第三位重要的数据元素,而仅有 3% 的终端用户认为它重要。

最重要的信息

在帮助您确认所需资料时,什么信息是最重要的?(图书馆员调查)

在帮助您确认所需资料时,什么信息是最重要的?(终端用户弹出式问卷调查)



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查与终端用户弹出式问卷调查)

图书馆员调查中提示被调查者提出,在一切条件都具备的理想情况下,他们将如何改进目录数据库。一个来自公共图书馆的从事编目和馆际互借的被调查者写下一段话,解释了为什么图书馆员认为 ISBN 重要的原因:“如果书目记录当前的版本有 ISBN 号或出版者号,我的工作

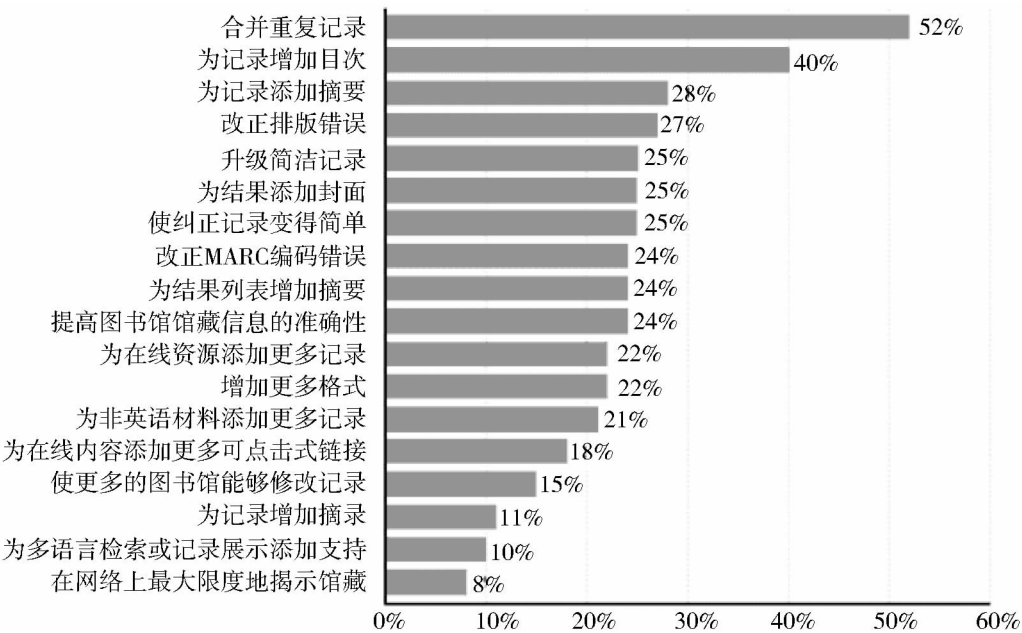
将容易得多。”其他对 ISBN 元素进行评论的人则希望,所有的记录都有准确的 ISBN 并且提高旧记录 ISBN 的精确性。

期望的数据质量提高措施

图书馆员调查和终端用户弹出问卷调查的结果表明,与目录打交道的图书馆员和那些利用目录发现和获取信息的终端用户在数据质量提高措施方面有着很大的不同。下表从高到低列出了图书馆员所期望的数据质量提高措施(可以选择多项)。

整合重复记录(52%),为详细书目记录增加目次(40%)和摘要(28%),是图书馆员首选的三个数据质量改进需求。

期望的数据质量提高措施
以下提高措施,您推荐哪个?
基数:被调查的图书馆员



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

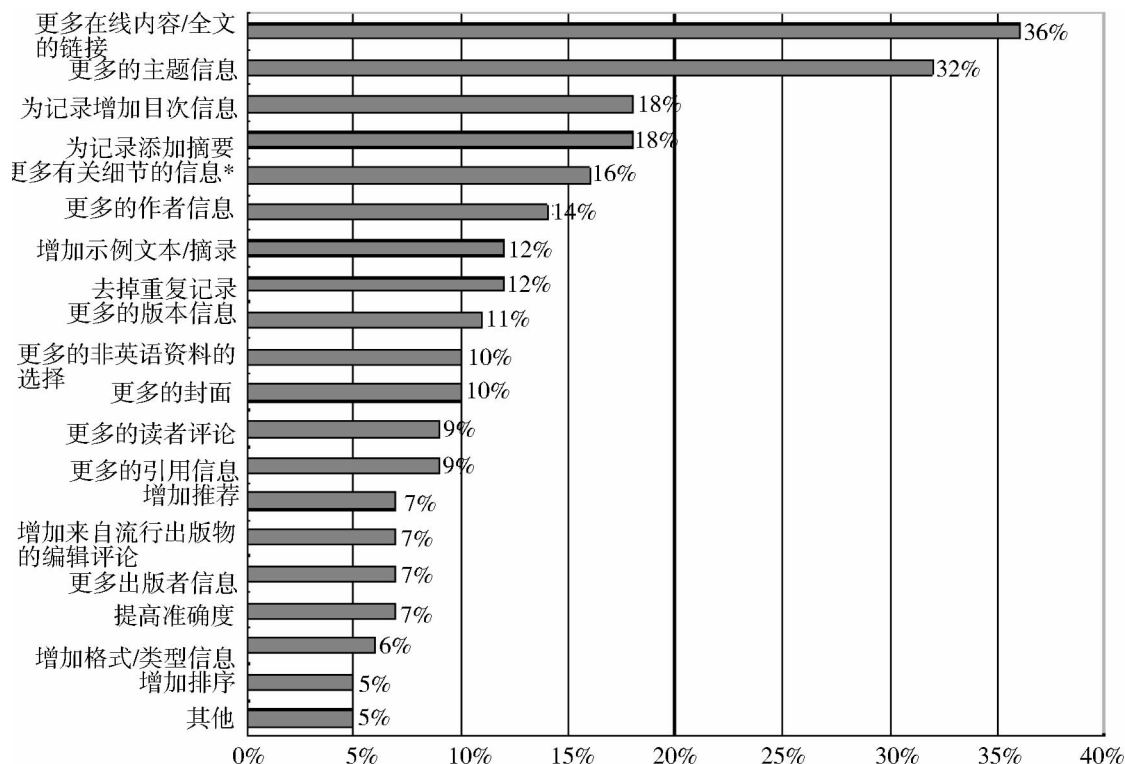
下面的表格从高到低列出终端用户弹出问卷调查所选的数据质量提高措施。该调查中的选择措施列表与图书馆员调查中的措施列表肯定不同,但还是有一些可比较的选择项。

终端用户首选的前两项是“更多的在线内容/全文链接”和“更多的主题信息”,并列第三的是“增加目次”和“增加摘要”。从“更多的主题信息”到“增加示例文本/摘要”,终端用户似乎在要求那些标准的目录描述所不包含的数据元素。终端用户把与纠正数据错误有关的选项(删除重复记录和提高标目准确度)分别排在第 8 和第 17 位。

期望的数据质量提高措施

在您鉴定所需资料时,哪些提高措施对您来说帮助最大?

基数:接受弹出式问卷调查的终端用户



来源:联机目录:用户和图书馆员需要什么, OCLC, 2009 (终端用户弹出式问卷调查)

* 在弹出式问卷调查时, WorldCat.org 上的“Details”包括基本书目信息以及目次、摘要等更为丰富的详细信息。

两类被调查者的期望措施列表中有多个选项不同,但存在足够的重复项使得两者可以进行比较。图书馆员和终端用户对增加目次和摘要的重要性的认识是一致的;除此之外,双方存在多种分歧。我们将两类调查列表中重复的选项制成了表格,并且将这些选项按排列位置分为高、中、低三个部分,结果如下表所示。在这些选项当中,图书馆员和终端用户对“更多在线内容链接”的看法明显不同;终端用户将这一改进措施列为首选高等级部分,而图书馆员将其列入低等级。对于“增加示例文本/摘录”,终端用户也明显比图书馆员更加热衷。

图书馆员和终端用户的数据质量提高措施排序等级的比较

以下提高措施,您推荐哪个?(图书馆员调查)

在您确定所需内容时,哪些提高措施对您帮助最大?(终端用户弹出式问卷调查)

	相关度排序		
	●图书馆被调查者	○终端用户被调查者	
可比较的改进项目选择*	高级	中级	低级
合并重复记录	●	○	
更多在线内容的链接	○		●
为记录增加目次	○●		
为记录增加摘要	○●		
增加封面	●	○	
增加更多格式		●	○
增加更多的非英语文献记录		○	●
为记录增加内容摘录	○		●

来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查与终端用户弹出式问卷调查)

高级检索、优化检索与分面检索

前文已提到,要支持高级检索、浏览检索结果和对检索结果的分面导航等功能,涉及书目数据元素的字段索引和为确保目录检索结果的一致性和可靠性而建立和管理受控标题表。

相对于终端用户来说,图书馆员更加注重修正目录数据错误(比如整合重复记录、修正错字和 MARC 编码错误),因为这类错误降低了目录数据对高级检索和分面检索的有效支持。本研究中的终端用户认为在有些情况下高级检索和分面检索是有帮助的,尤其是当面对大型检索集合时。但是终端用户往往不清楚支持这类功能所需要的数据结构和操作。图书馆员知道如何才能支持更高级的检索功能,因此,相对终端用户来说,图书馆员、尤其是编目工作者对数据库纠错等赋予更高的优先级是有道理的。

图书馆员对终端用户需求的预测

那些直接为用户提供服务的图书馆员被要求预测终端用户会推荐的数据质量提高措施(用“图书馆员对终端用户的预测”在图表中标识出来)。本节将对终端用户观点的预测结果和终端用户的实际观点进行比较。

调查要求直接为用户提供服务的图书馆员从下表所示的 18 个措施中选择对终端用户来说最有帮助的措施。将近一半的图书馆员(48%)认为是“为检索结果列表增加更多的封面”,其次是“为记录增加目次”(44%),以及“为检索结果列表增加摘要”(43%)。

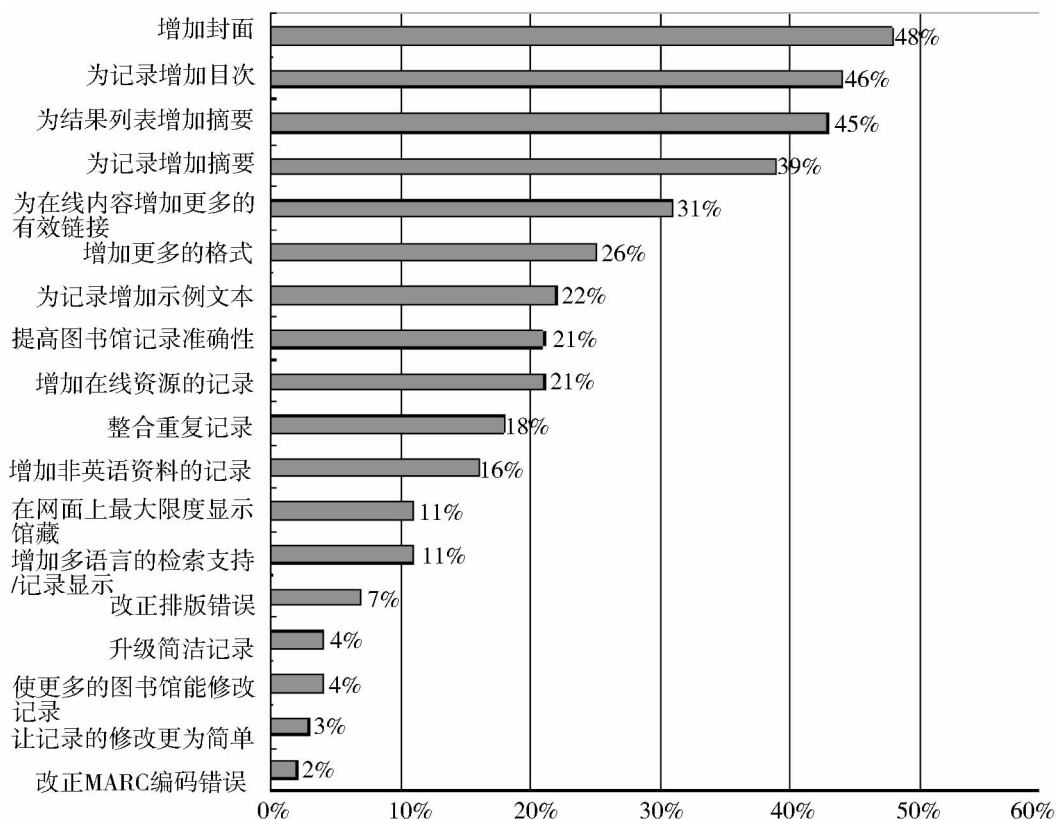
按图书馆类型对调查结果的分析表明,相对于高校图书馆和专业图书馆来说,公共图书馆员明显更倾向于“为检索结果列表增加更多的封面”对用户最有帮助(公共图书馆员有 69%,高校图书馆员为 45%,专业图书馆员为 31%)。

那些直接为用户服务的图书馆员中有四分之一以上认为,对用户来说有帮助的措施是“为详细书目记录增加摘要”(39%),“为在线内容增加更多的有效链接”(31%),以及“增加更多的格式”(25%)。

图书馆员对终端用户观点的预测:所期望的数据质量提高措施

以下提高措施中,您认为哪个对您的用户来说最有帮助?

基数:直接为用户服务的图书馆员



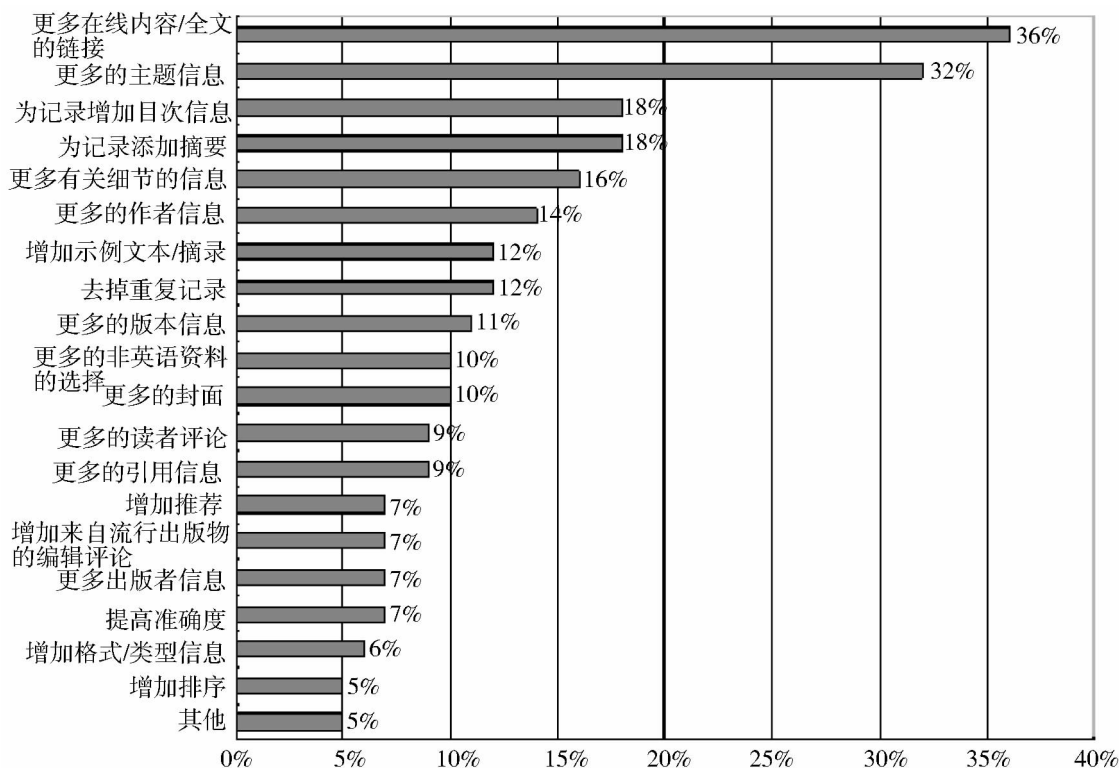
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(图书馆调查)

可见,图书馆员们意识到了终端用户关注诸如摘要一类的扩展内容,也知道他们期望对在线内容/全文提供更多的有效链接,优先增加示例文本/摘录。以上选项在终端用户的提高数据质量措施列表中也得到反映,如下所示。

期望的数据质量提高措施

在确定您所需内容时,哪些提高措施对您帮助最大?

基数:参与弹出式问卷调查的终端用户



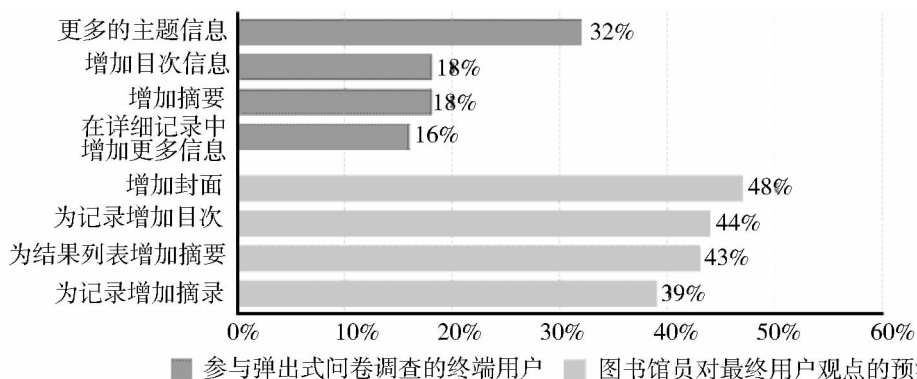
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查)

为检索结果增加封面,为记录增加目录,为检索结果增加摘要,以及为书目记录增加摘要,这些都是图书馆员所预测的对用户有帮助的措施。为佐证这点,终端用户提出他们希望有更多的“主题信息”,其次是增加评价性信息,它们与图书馆员所预测的“增加目录和摘要”相似。

最期望的与信息发现相关的数据质量提高措施

在您确认所需内容时,哪些措施对您帮助最大?(终端用户弹出式问卷调查)

以下提高措施中,您认为哪个对您的用户最有帮助?(图书馆员调查)



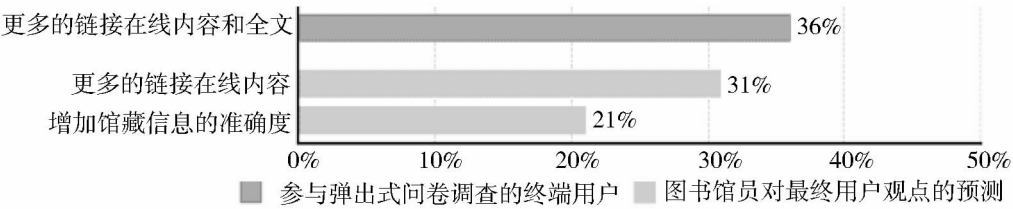
来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查与图书馆调查)

在信息传递方面,“更多的在线内容的链接”是图书馆员和终端用户调查中选择最高的措施。图书馆员也预测到“增加图书馆馆藏信息的精确度”对用户来说有帮助,因为这无疑会帮助用户提高其信息传递体验,但终端用户对此并不知情。

最期望的与信息传递相关的数据质量提高措施

哪些措施对您确定所需内容帮助最大?(终端用户弹出式问卷调查)

以下措施中,您认为哪个对您的用户最有帮助?(图书馆员调查)



来源:联机目录:用户和图书馆员需要什么,OCLC,2009(终端用户弹出式问卷调查与图书馆调查)

结 论

图书馆员和终端用户

研究表明,终端用户和图书馆员对目录数据质量的重点存在较大的差异。终端用户和图书馆员使用目录的目的各不相同。终端用户使用目录来满足自身信息需求;图书馆员是为完成工作任务而使用目录数据。这些目的差异造成对编目数据质量的不同需求。

研究同时发现,工作角色、图书馆类型和地区的不同也会显著影响对编目数据质量的需求。在各种类型的图书馆员中,工作角色和地域差距似乎是造成差别的首要原因。

编目人员对数据结构的认知和终端用户对此方面知识的欠缺也是造成图书馆员和终端用户对数据质量需求不同的一个原因。极少终端用户了解编目数据的结构,尽管在本次研究中,部分终端用户对依赖结构化数据的某些功能(例如分面检索、高级检索)表示欢迎。

终端用户对高级检索功能的有用性的认可这一事实,说明增加字段索引和受控词汇能再下一代联机目录有效提高用户的检索体验。

在一定程度上,图书馆员通过相同的方法是使用本馆目录系统和 WorldCat。他们优先考虑的数据质量因素因其角色差异而不同。相对于终端用户和其他类别的图书馆员,编目与采访馆员更加看重数据库的纠错能力。他们也认为目录的规范化的结构数据非常重要,例如,索引字段和权威规范标目能支持高级检索、检索限制条件、分类浏览和其他依赖于编目数据结构的功能。他们还比较重视重复条目的合并以及不准确数据的修订。

平衡终端用户和图书馆员的需求

在图书馆内外,目录系统都有不同的用户群。研究表明,终端用户更重视扩展数据(如目录、简介等)及在线内容链接,包括文本和多媒体。图书馆员也是目录的主要使用者,他们对数据质量的要求与终端用户有很大不同。如果数据满足他们的需求,他们就可以更加快速、高效地完成工作。而且,一些受到图书馆员重视但还没有引起终端用户注意的数据质量需求(例如,可纠错和控制词汇),也有助于满足用户需求。

在人力和经济资源不受限制的情况下,联机目录数据质量计划可以满足终端用户和图书馆员的所有需求。但在资源受限的情况下,图书馆决策者必须做出选择,创造性地部署资源,在数据质量的需求之间做出平衡。一个致力于在终端用户和图书馆员需求间取得平衡、但给予终端用户需求更多重视的数据质量计划,将有利于图书馆在所服务的用户社区更好发展。

可用性、用户为中心的设计及信息组织原则

在阅读 Fran Miksa 博士论文中关于卡特《字典型目录的规则》的发展与思考的那章后^[1],再阅读《资源描述和获取》草稿^[2]的第一章(它继承了“目录描述功能需求”^[3]的理念),我们能够得到一些有趣的结论。尽管卡特的规则和 RDA 出现在不同的历史时期,但他们的信息组织原则是一致的,从卡特到 RDA 有清晰的发展脉络。

尽管在 FRBR 概念形成的过程中,也做过少数严密的用户研究^[4]。但是没有证据显示卡特

或者 FRBR 或 RDA 的制定者曾经针对信息系统终端用户系统测试过他们的假设。Fran Miksa 在其职业生涯中曾多次指出,图书馆的信息组织原则研究和用户与使用研究似乎总是相对独立的。Fran Miksa 说:“我认为信息用户和信息使用对我们来说仍然是未知的,就像我们在大雾中开车所能看见的景象一样”^[5]。

当 Web 信息服务成为主流时,它们吸引了各类信息检索公众的注意,而那些打造了这些 Web 信息服务的人们看起来并不在意(如果他们知道的话)图书馆所信奉的信息组织概念框架,大量的试验和反复的试误似乎成为常规。

当 Web 的开发者们越来越清楚地知道什么有效和什么没有效时,很多人开始学习利用 Web 虚拟世界所提供的新机会,例如 Flickr 和 Facebook 这样的以不同方式组织大量信息的系统。在网络中,可用性和用户为中心的设计原则似乎取代了传统的信息组织原则,至少是图书馆员曾实践的信息组织原则。David Weinberger 在研究组织大量数字信息的新方法时,提出“所有的事物都具有关联,因此,所有的事物都是元数据”^[6]。基于他们在流行网站上的体验,熟悉互联网的终端用户期望能够检索到来自多种来源的丰富的元数据,方便地找到并获取所需的内容。

考虑这两种为用户组织信息的方式——一种来自图书馆,另一种来自 Web——本研究得到的结果就不令人惊讶了。图书馆员对数据质量的观点仍在很大程度上受到信息组织经典原则的影响,而终端用户对数据质量的期望来自与他们在流行网站上获取的信息组织经验。现在所需要做的是整合两种方式中的最好方面,从而得到一种新的扩展的定义,说明图书馆联机目录的“质量”涵义,以及谁有义务来提供这样质量的目录服务。

元数据和内容

目录的传递、链接和更多的在线内容

本研究中,终端用户(而非图书馆员)认为增加更多的在线资源链接来提高目录质量是最为重要的措施。终端用户认为,从发现到获取是一条不间断的工作流,而图书馆员倾向于认为用户任务是分离且独立的。在关于 FRBR 的一篇博文中,Karen Coyle 甚至提出,“FRBR 的用户任务(查找、确认、选择、获取)被限制在一个范围内,这同样会限制我们对用户和目录的理解”^[7]。

由于终端用户来自一个拥有大量在线信息的世界,很自然他们希望能够获取内容,而不仅仅是发现、选择、被指导怎样去获取内容(图书馆目录的惯用做法)。当 Google 图书检索抓住了那些使用图书馆馆藏用户的注意力时,终端用户对于立刻链接到图书在线内容(至少是预览片段)的想法会越来越普遍。而且,终端用户“链接到内容”的希望将超越文字,还将包括影像和音乐的片段。

发现、传递和丰富数据

结果显示,终端用户希望通过扩充基本的编目记录(例如增加目次、简介/摘要、封面、摘录和内容片段),减少馆藏描述(编目数据)和馆藏本身的差别。对一些用户来说(本研究中为大学生),类似书评、评级、标签等的社会性内容也很有帮助。图书馆员至少同意目次信息应被加入到编目记录中。

本研究发现,获取至少与发现同等重要,甚至更加重要。用户不愿意去发现那些他们不能

获取的信息,或者获得那些不能满足需要的信息。为节省时间和精力,终端用户需要这些扩充内容帮助他们确定是否值得花费时间去获取发现的资料。

“实物传递在图书馆界一直都是不受重视的庶子”^[8]。这句话出自 Valerie Horton 在 2009 年第二届关于图书馆资源实物传递的 Moving Mountains 会议上的报告。她通过一系列的演示描述了图书馆传递的现状(馆际互借、联合借阅、家庭传递),为思考如何提高实物传递在图书馆工作中的地位提供了一个良好的开端。

主题词和主题信息

当终端用户选择“更多的主题信息”作为一个重要的目录改进措施时,他们的意思是什么?考虑到在编目中往往只用到来自受控主题词表的少量不重复主题词^[9],终端用户不太可能是想要更多的受控主题词。参考终端用户在提高目录质量时的优先选择以及终端用户聚焦小组的讨论,“更多主题信息”应该被解释为那些没有包含在标准目录描述中的具有主题涵义的数据元素。

同时,受控主题词也能够从多个途径为终端用户提供帮助:例如作为主题索引条目,(当记录的主题词包含多种语言时)支持多语言检索,用于优化和扩展检索,用于浏览,作为词或短语链接到分类表或其他术语表,支持相关度排序计算等。为了支持这些特性,今天的编目依靠劳动密集型的方法来生产受控主题词。人们越来越担心这种方法不能持续下去,图书馆必须找到更为经济的方法来为用户提供受控词所能提供的便利。

标准编号

研究发现,类似于 ISBN 的标准编号对于图书馆员的工作非常重要。但除了被用户用来确认已知资料,标准编号对于支持终端用户的其他工作也是至关重要的。标准编号能够为不同仓储中的相同条目提供可靠链接,同时也支持机器到机器的交互,这有助于改善发现到传递的用户体验。

何处获取数据

扩充数据

本研究之前,已经有人发现,扩充数据(例如摘要和内容注释)对终端用户检索有重要影响^[10],用户需要扩充记录^[11],扩展的数据提高了图书馆资源的利用率^[12]。然而,手工为每个记录增加扩展数据并不现实。图书馆需要通过合作来共同承担丰富发现系统、关键词和索引所耗费的成本。

现有的编目数据是第一个可以挖掘的宝藏。利用 FRBR 组织目录数据是获取如摘要、主题标签云、封面和目次等更多有价值的信息的有前景的方法。OCLC 研制的“小说搜索器”(FictionFinder)^[13]在挖掘 FRBR 数据方面取得了较大的成功,如下图所示。通过挖掘 FRBR 的信息簇,“小说搜索器”能够在检索结果页面上提供作品摘要,如下图中“小说搜索器”检索“弃婴”一词,检索结果中列出了关于此主题的一些名著,如《呼啸山庄》。数据挖掘是获取扩展数据加入基本编目描述的最有潜力的方法。



来源:联机目录:用户和图书馆员需要什么,OCLC,2009

另一个现在应用较为普遍的方法,就是从图书馆自动化系统供应商那里购买扩展数据。可以参考 Cherie Madarash – Hill 和 J. B. Hill 在 2005 年出版的关于从图书馆自动化管理系统商获取数据的研究^[14]。第三种方法就是使用应用程序接口,把扩展数据加入到图书馆目录中,例如通过亚马逊的 Web 服务^[15],也可以通过其他 API 获取服务。

相关排序、分类和用户事务数据

目前,图书馆目录提供的相关排序并不能满足用户需求。考虑到 Melvin Maron 和 John Kuhns 在 1960 年发表的关于概率索引 (probabilistic indexing) 的论文^[16],以及后来很多信息科学家们的研究和成果,在这方面还没有显著的进步很令人惊讶。Maron 和 Kuhns 作为早期研究自动信息检索者,通过实验来检测概率索引这种自动文献检索技术。他们的实验抽取用户的检索词,将其与分配给文献的带有权重的关键索引词相匹配,通过统计推理,然后按照与用户检索相关的概率将结果排序。这种技术产生了大有希望的结果,而且在信息领域引发了一系列相关的研究。九十年代后期,Ray R. Larson (拉尔森) 等人在相关排序技术方面也做了很有成效的研究^[17],但没有得到图书馆管理系统软件商的支持。拉尔森的相关排序技术的特别之处在于他以分类聚类技术来帮助聚焦主题检索。拉尔森和他的团队使用与分类号相关的字段来辅助检索结果的排序,新近澳大利亚国家图书馆的工作人员通过挖掘目录中的结构化数据为图书馆目录的使用者提供更好的排序结果^[18]。自 2008 年 11 月运行新的目录系统以来,他们使用 MARC 数据中的字段加上馆藏信息对澳大利亚图书馆检索结果进行相关排序,已经获得了较大的成功。

研究结果显示,终端用户对图书馆检索结果按相关性排序的期望受到其在谷歌和亚马逊这样的网站上检索经验的影响。谷歌巨大的成功很大程度上取决于其对检索结果排序的能力,这种能力是建立在对链接关系的复杂算法之上。亚马逊对检索结果的出色的相关排序则是基于检索和购买趋势上,以及利用用户提供的评价数据。

建议

正如内容提要中所述,这个报告是写给那些寻求改进目录数据的人们和那些致力于下一代

图书馆目录对多种资源信息的提供、摄取、聚合、同步或链接的人们。下面是提供给这些读者的一些建议：

- 检查比较图书馆在编目工作、目录管理、链接功能和扩充内容(例如目次、摘要等)的投入,重新平衡他们之间的关系,更好地满足终端用户对目录数据质量的需求。
- 与图书馆界和相关机构探索如何通过数据挖掘、应用程序接口、与出版商和其他供货商的合作以及图书馆合作项目来获取或生产扩展数据(例如目次、摘要等)。
- 鼓励相应的组织完善联机目录相关排序的研究和发展。通过重新部署分类数据(与分类号相关的类目词)和其他已有数据来改进相关性排序。
- 更加重视图书馆传递服务和能够改善终端用户体验的数据元素。
- 探索将馆藏传递到家庭或办公室的可行性,或者其他更迅速便利的实物馆藏传递方法。愿意改变规则和流程,部署人力、财力来完成这一工作。
- 发掘图书馆联盟化的馆际互借,在适当情况下,提供从目录链接到资源实时购买和连接数字资源的功能。考虑设置一个标准(例如 24 小时之内)来帮助用户获取所需的资源。
- 努力完善图书馆链接元数据的管理和与订购数据库、开放资源间的互操作。
- 考虑增加更多的数字内容,这样终端用户可以方便地从目录链接到内容,把对印本的描述与它们的数字化版本相连接。
- 如果可能,将目录链接到摘要或内容预览,包括文本和多媒体。
- 检查本地编目记录修改情况,分析哪些直接影响终端用户的发现和获取,哪些并没有什么帮助。重新设计工作流程,将人力资源集中于那些终端用户最为迫切的需求和那些必须手工完成的任务。
- 图书馆不可能独自完成全部任务。必须加强与更多机构(甚至是终端用户)的合作和协调。正如在国会图书馆工作组最终报告第一节中所提到,传统图书馆的工作流程重复性大且成本高,例如,同样的修订会在多个图书馆重复多次^[19]。恰当的合作机制可以分担数据的改进工作,使得图书馆在开销最小的情况下保证用户最好的目录数据使用体验。
- 在图书馆界内,最大程度地寻找自动生产和维护结构化数据的方法,用于支持字段组配检索、分面检索和高级检索等功能。
- 探索图书馆界与相关组织合作的方法,通过标准号码的引用和互连来支持图书馆和终端用户的任务。

未来研究设想

希望基于此报告的发现进行研究的学者可以从以下几个观点考虑。这些建议包括相当简单、实用的项目,也包括比较深入的研究。

- 为图书馆增加编目扩展数据(如,目录、摘要等)提供指导。从不同的方面对比其开销和收益(手工扩充、数据挖掘、购买数据、通过 API 获得等)。
- 研究怎样应用分类表、主题词表和结构化数据来改进联机目录的相关性排序。
- 实验和评价支持分面检索和其他高级检索的更为经济的方法(自动化、数据挖掘、用户提供数据等)。Greta de Groat 在 2009 年数字图书馆联盟委托的报告提供了丰富的内容,介绍了关于使用数据挖掘、聚类、映射和其他工具来自动增强元数据,支持寻找和使用数字馆藏^[20]。
- 本研究主要关注终端用户和图书馆员口头表达的需求。如果能够设计适当的日志分析

来衡量被调查对象的实际行为,可以对本研究报告提供有效的补充。

参考文献:

- [1] Cutter, Charles A. , W. P. Cutter, Worthington Chauncey Ford, Philip Lee Phillips, and Oscar George Theodore Sonneck, Rules for a dictionary catalog (Washington: Government Printing Office, 1904).
- [2] Joint Steering Committee for the Development of RDA, Full Draft of RDA, Chapter 1, 2008, <http://www.collectionscanada.gc.ca/jsc/rdafulldraft.html>.
- [3] IFLA Study Group on the Functional Requirements for Bibliographic Records, Functional requirements for bibliographic records: final report, UBCIM publications, new ser. , v. 19 (München: K. G. Saur, 1998).
- [4] Indiana University, " Variations/FRBR: Variations as a Testbed for the FRBR Conceptual Model," <http://www.dlib.indiana.edu/projects/vfrbr/>. FRBR project Web site, Kent State University School of Library and Information Science, <http://frbr.slis.kent.edu/projectinfo.htm>.
- [5] Miksa, Fran, " Information organization and the mysterious information user" (lecture delivered at the School of Information, University of Texas – Austin), forthcoming in Libraries and the Cultural Record.
- [6] Weinberger, David, Everything Is Miscellaneous (New York: Times Books, 2007), 105.
- [7] Coyle, Karen, blog post February 21, 2007, <http://kcoyle.blogspot.com/2007/02/i-have-longhad-hard-time-with-frbr.html>.
- [8] Horton, Valerie, " Moving Mountains and Crossing Rivers" (report from the second conference on library physical delivery, Cincinnati, Ohio, September 2008), Collaborative Librarianship 1 (1) (2009): 18 – 26, <http://www.collaborativelibrarianship.org/index.php/jocl/article/viewArticle/3>.
- [9] Author correspondence with Diane Vizin – Goetz, Consulting Research Scientist, OCLC, January 9, 2009.
- [10] Markey, Karen, and Karen Calhoun, " Unique words contributed by MARC records with summary and/or contents notes," in ASIS 87: Proceedings of the 50th ASIS Annual Meeting, Boston, Mass. October 4 – 8, 1987 (Medford, NJ: Published for the American Society for Information Science by Learned Information, 1987).
- [11] Markey, Karen, and Karen Calhoun, " Unique words contributed by MARC records with summary and/or contents notes," in ASIS 87: Proceedings of the 50th ASIS Annual Meeting, Boston, Mass. October 4 – 8, 1987 (Medford, NJ: Published for the American Society for Information Science by Learned Information, 1987).
- [12] See the literature review in Markey, Karen, " The online library catalog: paradise lost and paradise regained?" D – Lib Magazine 13 (12) (2007), www.dlib.org/dlib/january07/markey/01markey.html.
- [13] FictionFinder, <http://www.oclc.org/research/projects/frbr/fictionfinder.htm>.
- [14] Madarash – Hill, Cherie, and J. B. Hill, " Electronically enriched enhancements in catalog records: A use study of books described on records with URL enhancements versus those without," Technical Services Quarterly 23 (2) (2005): 19 – 32.
- [15] As described by Ferraro, Joshua, blog entry " Amazon.com Web Services and library catalogs," April 16, 2007, <http://blogs.liblime.com/developers/2007/04/16/amazoncom-web-services-andlibrary-catalogs/>.
- [16] Maron, M. E. and J. L. Kuhns, " On relevance, probabilistic indexing and information retrieval," Journal of the ACM 7 (3) (1960): 216 – 244.
- [17] Larson, Ray R. , Jerome McDonough, Paul O'Leary, Lucy Kuntz, Ralph Moon, " Cheshire II: Designing a next – generation online catalog," Journal of the American Society for Information Science 47 (7) (1996): 555 – .
- [18] Dellitt, Alison and Tony Boston, " Relevance ranking of search results from MARC – based catalogues: from guidelines to implementation, exploiting structured data," National Library of Australia staff paper, 2007, http://www.nla.gov.au/nla/staffpaper/2007/documents/Boston_Dellitrelevance-ranking.pdf.
- [19] Library of Congress, On the Record: Report of the Library of Congress Working Group on the Future of Bibliographic Control. Washington, D. C. : Library of Congress, recommendation 1.1, " Eliminate redundancies," 2008, <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecordjan08-final.pdf>.
- [20] de Groat, Greta, " Future directions in metadata remediation for metadata aggregators," Digital Library Federation, 2009, <http://www.diglib.org/aquifer/dlfi10.pdf>.

译者后记

互联网时代,新的工具、服务的出现如雨后春笋,发展迅猛,已成为用户信息环境的重要组成部分,并深刻影响用户的信息行为。这就要求图书馆和信息服务机构需要不断研究和了解新技术、用户行为、态度等等方面的变化和趋势,重新审视图书馆的定位、角色和职能。

OCLC 一直致力于关注发展变化的用户信息环境、研究分析用户的需求和信息行为,提供了一系列优秀的研究成果。从《审视发展环境,认清前景趋势》(OCLC Environmental Scan: Pattern Recognition, 2003 年)到《信息格式趋势:内容、而非传统的载体形式》(Information Format Trends: Content, Not Containers, 2004 年);从《对图书馆和信息资源的认知》(Perceptions of Libraries and Information Resources, 2005 年)、《大学生对图书馆和信息资源的认知》(College Students' Perceptions of Libraries and Information Resources, 2005 年)到《网络世界中的共享、隐私与信任》(Sharing, Privacy and Trust in our Networked World, 2007 年)、《从意识到资助:美国图书馆经费政策研究》(Awareness to Funding: A study of library support in America, 2008 年)……。这些报告为图书馆和信息服务机构了解发展环境、了解用户需求提供了严谨、详实的资料,具有重要的参考价值,也成为《图书情报工作动态》长期固定跟踪和报道的重要核心信息源之一。

2009 年 4 月, OCLC 发布了最新研究报告《联机目录:用户和图书馆员需要什么》(Online Catalogs: What Users and Librarians Want, 2009 年),在 OCLC 和 OCLC 中国办事处的支持下,由我馆组织翻译,其中第 1-3 章由张红丽编译,第 4 章由苏娜编译,第 5 章由周萍编译,第 6 章由孙敏杰编译,第 7 章由王欣编译。鲁宁、丁艳君负责进行一校,李麟进行了二校,张晓林进行了最终的三校定稿。

鉴于时间紧迫、译校者水平有限,编译不当和错漏之处在所难免,敬请读者不吝指正。译文商榷请与《图书情报工作动态》编辑部联系:lilin@mail.las.ac.cn。

国家科学图书馆

《图书情报工作动态》OCLC - 2009 年报告编译小组



6565 Kilgour Place
Dublin, Ohio 43017-3395
1-800-848-5878 +1-614-764-6000
Fax: +1-614-764-6096
www.oclc.org

ISO 9001 Certified