

国内外文献数据库个性化服务研究^{*}

廖 凤^{1,2} 周静怡¹

¹中国科学院国家科学图书馆 北京 100190 ²中国科学院研究生院 北京 100190

〔摘要〕选择国内外典型学术文献数据库进行调研,总结当前文献数据库个性化服务的主要方式;概述各文献数据库提供的主要个性化服务功能,并对其特色及不足之处进行分析;探讨当前信息环境下用户信息需求的变化趋势,分析淘宝和 Google 个性化服务模式中可供文献数据库借鉴之处,对文献数据库深层次个性化服务的开展进行思考并提出建议。

〔关键词〕文献数据库 个性化服务 推送技术 个性化推荐

〔分类号〕G252

Research on the Personalized Service of Literature Databases at Home and Abroad

Liao Feng^{1,2} Zhou Jingyi¹

¹National Science Library, Chinese Academy of Sciences, Beijing 100190

²Graduate University of Chinese Academy of Sciences, Beijing 100190

〔Abstract〕Through researching on the personalized service of the typical literature databases, several personalized service modes are summarized, including the main function, features and deficiencies of the personalized service. And then, changes of user information demand during current information environment are explored, and personalized service model of Taobao and Google are discussed for references. At last, the authors put forward some suggestions on how to develop deep level personalization services.

〔Keywords〕literature database personalized service push technology personalized recommendation

1 引言

在数字化信息环境下,文献数据库商通过网络平台将其集成的学术刊物资源提供给用户。然而,如果没有有效的服务将资源传递给有需要的用户,资源的价值就不能实现。如何才能按需服务?这正是个性化服务要解决的问题。数据库平台的个性化服务是指以用户为中心,以网络和信息技术为平台,通过获取并分析用户的偏好、背景以及信息行为模式,提供充分满足用户信息需求的一种集成性信息服务^[1]。目前各大数据库平台在传统的检索、浏览服务之外,纷纷推出了各自的个性化服务功能。

为了了解个性化服务的应用现状,本文根据文献数据库的学科类别、数据规模、文献类型、文献语种等特征,选择了 ISI Web of Science^[2]、ScienceDirect^[3]、Emerald^[4]、SpringerLink^[5]、Wiley InterScience^[6]、

PubMed^[7]、DP^[8]、维普中文期刊数据库(以下简称“维普”)^[9]等典型平台进行研究,力求既有一定的代表性,又具备一定的覆盖度。其中既有全文期刊数据库,也包含文摘目次数据库;既有综合学科数据库,也包含特定专业领域数据库。

2 个性化服务主要方式

目前国内外文献数据库提供的个性化服务可概括为:用户管理服务、存储服务、定制推送服务和个性化界面服务。

2.1 用户管理服务

用户管理服务是针对用户个人信息收集和管理的服 务,具体表现为提供用户注册、偏好设置、信息行为自动捕捉等。用户信息是开展个性化服务的前提,各数据库目前主要通过用户注册来获得用户基本信息。

系统获得的 用户信息越全面,就意味着越了解用

* 本文系 2008 年国家科学图书馆青年人才领域前沿项目“CSCD 的深层次个性化定制服务研究”(项目编号: YP08Z61001)研究成果之一。

户需求,提供的个性化服务也将越有针对性。调研结果也印证了这一点:Web of Knowledge要求用户注册时填写了姓名、密码、邮箱、职称、学科等项目,这些信息很好地支撑了其多样化个性化服务功能的实现;维普在用户注册时只要求用户名和密码,这与其不具备邮件定制功能也是吻合的。

图1显示各数据库用户注册的字段信息,可以看出:一般情况下Email、用户名、密码是必须的;学科、兴趣领域等更进一步的信息并没有引起很大关注,说明数据库平台在挖掘用户的偏好和兴趣方面还有很大空间。

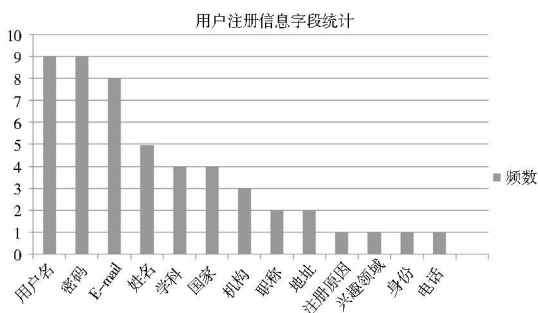


图1 用户注册信息字段统计

2.2 存储服务

存储服务是指数据库系统通过一定的技术手段帮助用户保存下操作过程中有用的信息,为用户再次查找信息提供便利。根据存储内容可以将存储服务分为以下几种类型:

- 检索结果保存:一般在检索结果页面上提供多种检索结果保存方式:PDF格式、HTML链接方式以及导出至Endnote等文献管理工具。

- 检索策略保存:用户可以保存经常使用的检索策略,任何时候无需花费额外的时间重复输入检索式,即可执行检索获取检索结果。在Web of Knowledge和Science Direct中,还可以对保存的检索式进行逻辑组合构造新的检索式。

- 操作历史保存:操作历史是用户操作行为的记录,为用户回溯自己的操作提供了线索,也为用户行为特点分析提供了依据。Science Direct提供了记录和跟踪用户最近100次操作的服务。

- 快速链接保存:快速链接相当于用户在特定数据库下的一个收藏夹,可以把常用网站保存起来,如Science Direct的Quick Links服务。

- 常用出版物保存^[10]:把用户最关注的期刊或图书存储起来,可以随时点击题名浏览,如Wiley提供基于题名的刊物保存。

2.3 定制推送服务

定制推送服务是根据用户预先定制的需求,把符合要求的内容发送到用户邮箱或RSS阅读器,及时为用户传递最新数据,省去用户登陆数据库的繁琐。

从推送方式上来看,目前主要有两种方式:①传统的Email推送:用户自己设置订阅主题、更新周期以及发送格式等,系统按照这些参数以Email的方式给用户推送相关内容;②利用RSS方式进行推送:用户在相关页面上点击RSS Feed图标把Feed添加到阅读器中,就能随时随地通过阅读器查看定制的信息。目前大部分国外数据库平台都支持Email和RSS两种定制方式。

从推送内容来看,主要有以下类型:

- 专题推送:根据用户设定的学科、主题、关键词等,把最新内容推送给用户,如Science Direct的Topic Alerts, JOP的Enhanced Alerts。

- 检索式推送:用户在对检索策略进行保存的同时,可以进行邮件或者RSS定制,每当数据库有新增文献时,系统会筛选出与保存检索式匹配的文章发送给用户;Science Direct Email, Wiley, SpringerLink均提供这种服务。

- 目次推送:用户选择自己需要定制的期刊,当所选期刊的最新期次可获得时,系统将最新内容目录发送给用户,Science Direct Emerald Wiley, SpringerLink, DP等提供此类推送。

- 引文推送:当数据库中有新文献引用了用户定制的文章时,系统通过邮件通知用户,如Science Direct和Web of Knowledge的Citation Alerts功能。

- 产品信息推送:用户可选择是否定制数据库产品与服务相关信息,如数据库更新信息、新增功能以及用户培训资料等。

2.4 个性化界面服务

个性化界面服务主要涉及界面结构和界面内容。界面结构指系统界面的模块类别、布局形式、语言定制以及界面色彩等;界面内容指各模块包含的具体信息内容。

由于Web of Knowledge包括了多个数据库产品,用户可以在个人起始页定制所关心的产品;Science Direct为注册用户提供了个性化主页,默认包括4大模块:资源浏览、最近操作、最新资讯以及快速链接;SpringerLinker则可以对界面语言进行定制。

3 文献数据库个性化服务概述

3.1 ISI Web of Knowledge

Web of Knowledge是由 Thomson Scientific建立的新一代学术信息资源整合平台,其个性化服务包括:

- ①用户管理服务。
- ②存储服务:检索策略保存、检索结果保存、期刊保存。
- ③定制推送服务:目次推送、引文推送、产品信息推送。引文推送将选定文献最新被引用情况通报给用户,推送方式包括 Email和 RSS两种。
- ④个性化界面服务:选择起始页和语种定制。

Web of Knowledge的个性化服务功能能够满足用户的基本需求,引文推送功能较有特色。但更多深层次的服务尚未提供,比如可以结合用户的学科类别和检索行为记录,从引文指标库中为用户推荐相关的 TOP N期刊、学者或者热门文章等。

3.2 Elsevier ScienceDirect

ScienceDirect是 Elsevier出版集团的核心产品,其个性化服务包括:①用户管理服务:用户注册、个人信息管理及偏好设置。②存储服务:检索策略保存、检索结果保存、操作历史保存、快速链接保存。快速链接主要有:期刊 图书收藏链接,通告页、检索式调用页、热门文章推荐页等常用内部链接,投稿页、Scopus等外部网页链接。③定制推送服务:专题推送、检索式推送、目次推送、引文推送。④个性化界面服务:个性化主页、结果显示设置。⑤其他:电子馆藏报告。

Science Direct的个性化服务是多角度、多层次的,其特色之处:①尽量收集了用户的机构、身份、兴趣偏好等信息;②电子馆藏报告使得普通用户可以了解到有权限访问的期刊详情,检索时更有针对性;③提供了引文推送功能。

3.3 Emerald

Emerald数据库主要涵盖管理学、图书馆学、工程学等领域期刊,其提供的个性化服务如下:①用户管理服务:用户注册、成员关系(根据用户角色提供成员关系选择,进而推送相关资讯);②存储服务:检索策略保存、检索结果保存、期刊保存;③定制推送服务:检索式推送、目次推送、每周汇编推送(将最近一周发布的期刊论文内容汇编,每周向用户通报一次);④其他:时事通讯(如网站的升级、更新等信息)。

Emerald擅长根据用户的角色提供有针对性的服务(成员关系),而且能够主动替用户汇总信息(每周汇编),用户不必逐一浏览便可获得内容梗概。其个性化服务着重于用户的资讯保证,但个性化界面相对简单。

3.4 SpringerLink

SpringerLink网络平台提供 1 600余种学术期刊的全文服务,其具备基本的个性化服务功能:①用户管理

服务;②存储服务:检索策略保存、检索结果保存、期刊保存;③定制推送服务:检索式推送、目次推送;④个性化界面服务:可以对界面语言进行定制。

SpringerLink提供标签服务,用户可以根据需要为保存的内容创建标签进行管理;但是在功能设计上不够简洁、直观,部分功能易用性不足。

3.5 Wiley InterScience

Wiley InterScience是一个综合性的网络出版及文献服务平台,Wiley提供的个性化服务功能包括:①用户管理服务;②存储服务:检索策略保存、检索结果保存、期刊保存;③定制推送服务:检索式推送、目次推送、产品信息推送。

Wiley的个性化服务传统且简单;如果想在个性化服务上走得更远,Wiley必须考虑深入挖掘用户需求,在此基础上探索深层次的个性化服务功能。

3.6 PubMed

PubMed作为世界上最大的生物医学数据库,其个性化服务通过 MyNCBI实现,功能主要包括:①用户管理服务:用户注册、用户参数设置、用户设置分享、链接账户设置。②存储服务:检索式保存、检索结果保存、书目信息保存、操作历史保存。③定制推送服务:检索式推送。④个性化界面服务:链接显示设置、突出显示设置。⑤其他服务:过滤器、外部链接服务。过滤器实质是检索式定制,只是这种推送发生在用户每次执行检索时,系统会自动加入过滤条件。

用户设置分享将共享理念应用到了个性化服务中;通过链接账户设置,用户在跨系统访问 NCBI平台下其他数据库时就不必再重新申请账号;个性化定制功能齐全,用户可以自己设置结果显示格式;此外还提供过滤器和外部链接服务。PubMed个性化服务全方位地考虑了用户需求,兼具广度和深度,值得其他系统借鉴。

3.7 IOP

IOP覆盖的学科范围主要包括物理、能源以及一些交叉学科,其个性化服务包括:①用户管理服务;②存储服务:检索策略保存、检索结果保存、期刊保存;③定制推送服务:目次推送、专题推送。

IOP的个性化服务功能虽然简单,但是可用性较好。

3.8 维普

维普中文期刊数据库包括 7 953种中文科技全文期刊,内容涉及自然科学、工程技术、医学卫生等学科的 27个专题。维普提供的个性化服务如下:①用户管

理; ②存储服务: 检索式保存、检索结果保存、期刊收藏; ③个性化界面服务: 个性化主页。

维普目前还未提供定制和推送服务, 个性化主页也基本是静态设置; 但其分类定制和关键词定制可以帮助实现快速检索; 系统简单易操作, 可用性较好。

4 对文献数据库深层次个性化服务的思考

目前, 文献数据库的个性化服务多数停留在传统的保存、定制功能, 并没有随着用户信息环境的变化, 探索一些深层次的服务项目。

4.1 追本溯源——用户需要什么样的个性化服务

4.1.1 知识化服务 当前用户信息活动的重点和瓶颈正从文献获取转变为知识发现, 用户所要求的信息服务将不再是简单地检索和获取文献, 而是帮助他们将知识内容从众多信息对象中挖掘出来, 根据其内在特征进行鉴别、关联、重组, 帮助他们识别和创造新的知识^[11]。

4.1.2 集成化服务 在这个效率制胜的时代, 用户总是希望以最小的付出获得最大的收获。如果可以通过单一入口让用户获得所需的大部分信息, 或者能够将用户的各类资源进行有效的整合和组织管理, 那么将大大提高用户的工作效率。

4.1.3 注重分享与交流的服务 文献数据库平台有能力为用户发掘更多的交流渠道, 支持科研工作者对于文献资源和数据资源的链接和共享, 为其提供更宽松的学术交流环境。

4.1.4 关注用户潜在需求的服务 潜在需求是用户自己没有意识到却实际存在的需求。基于对用户基本信息的掌握和行为信息的分析, 借助一定的技术手段, 可以挖掘到用户尚未意识到的信息需求^[12]。

总之, 文献数据库平台只有持续关注用户的信息需求变化, 才能提供有效的个性化服务。

4.2 他山之石——其他领域个性化服务模式

上面提及的需求变化, 其他领域平台已经开始关注并付诸实践了。

4.2.1 淘宝网 淘宝网^[13]是亚洲最大网络零售商圈, 其个性化收藏功能可以设置是否与别人共享; 还提供收藏品的对比分析, 为购物者提供决策支持, 这是一种深层次的个性化收藏。

通过跟踪用户的浏览、点击行为以及页面停留时间, 系统为用户提供“淘宝猜你可能喜欢”的个性化推荐功能。其中相同收藏者推荐和相同浏览者推荐颇有

借鉴意义; 因为有着相同的收藏爱好和浏览焦点的用户, 通常有着相似的信息需求。

4.2.2 iGoogle Google^[14]是由 Google 公司提供的个性化首页服务, 通过添加各种 widget 用户可以定制各种资讯、博客或论坛内容, 也可以在页面上集成各种功能的小工具。同时用户还可以通过鼠标拖动小工具放置在不同的位置, 随意布局自己的页面。

信息的广度及定制的灵活度, 是 Google 给予文献数据库的重要启示。信息的广度指除了传统的文献资源, 可以考虑其他类型资源 (如科学数据、专利等) 的嵌入和推荐; 定制的灵活度指可以把界面的布局和显示权利交给用户。

4.3 何以攻玉——对于文献数据库个性化服务的建议

4.3.1 个性化建模推荐 个性化建模推荐是一种“用户建模——分析需求——信息存储——信息推送”的模式, 根据信息内容与用户需求的匹配程度由系统自动推荐给用户相关信息, 这是挖掘和满足用户潜在信息需求的有效途径。

4.3.2 更开放的收藏和定制 Web 2.0 提倡交流、分享和自组织, 文献数据库在个性化服务中可以融入这些理念。如提供基于标签的收藏管理; 设置收藏共享, 让有共同研究兴趣的用户分享资源。此外, 用户需要访问多个系统的信息资源, 因此更开放的个性化定制机制也是用户所希望的。

4.3.3 更自主和集成的个性化主页 就各数据库目前提供的个性化主页来看, 普遍存在赋予用户主动性不够的问题。就集成功能来看, 收藏、定制等功能均以资源内容为中心, 可以考虑对资源的分析、关联、重组等深层次功能, 帮助用户进行知识发现。

4.3.4 个性化检索 基于关键字的检索方式存在检索结果大量重复、不相关信息较多等缺陷; 而个性化检索通过搜集用户信息需求建立用户模型, 在检索过程中根据用户模型自动修正检索策略, 并根据用户需求对检索结果进行排序, 是改善检索效果的有效方式。

5 结 语

个性化服务满足了用户的个性化信息需求, 使资源发挥出更大功效, 对于文献数据库的市场前景非常重要。数据库厂商应该主动去发现和了解用户的个性化需求, 同时借鉴一些前进的技术与模式完善个性化

(下转第 146 页)

3 马括其人

马括的生平事迹各种历史文献皆不详载,笔者根据黄序、马括自序及其他材料推测出以下方面:

马括字季机。由上文所引黄昇序中称马括为季机可知。

马括为建安(今福建省建瓯)人。马序中言伯父为古洲,考马古洲,《宋史》亦无载。魏庆之《诗人玉屑》卷一九“马古洲”条下记“古洲马庄父”^[13],黄昇《中兴词选》卷六记载:“马庄父,字子严,自号古洲居士,建安人”^[14]。从马古洲为“建安人”可推知马括亦为建安人。黄昇为福建建阳人,常与建安人魏庆之等交游,那么与这位爱好朱熹之书的建安表兄马括交游当为可信之事。

马括约生于淳熙末绍熙初,1251年仍在世,卒年不详。马括为黄昇之表兄,今人张健考证黄昇生年在1196年左右^[15],方彦寿认为黄昇生活年代在1190至1252年间^[16]。由此推定马括也大约生于淳熙末年绍熙初年。他1251年为《文衡》写序,说明此年仍在世。

最后为使马括生平更为清楚,笔者为马括做一小传:马括,约生于淳熙末年绍熙初年,淳祐辛亥(1251)仍在世,卒年不详。字季机,建州建安(今福建建瓯)人。伯父马古洲为南宋孝宗、宁宗时著名诗人、词人。儿时得伯父、父亲以朱熹之书教授,20岁左右着力研究朱熹著作,与表弟黄昇等人交游。有《经济文衡》。淳祐辛亥(1251)曾经在崇川居住,是时初刻《经济文衡》黄昇序,马括自序。

〔作者简介〕张秋娥,女,1963年生,教授,博士,发表论文近30篇,出版著作5部。

(上接第70页)

服务功能,真正做到以用户为中心,这样才能提升用户满意度。

参考文献:

[1] 曹树金,司徒俊峰.数据库平台的个性化服务研究.图书馆论坛,2007,27(6):126-130.

[2] Web of Knowledge. [2009-08-10]. <http://www.isiknowledge.com/>.

[3] ScienceDirect [2009-08-10]. <http://www.sciencedirect.com/>.

[4] Emerald [2009-08-10]. <http://www.emeraldinsight.com/>.

[5] SpringerLink [2009-08-10]. <http://www.springerlink.com/>.

[6] InterScience [2009-08-10]. <http://www3.interscience.wiley.com/>.

〔作者简介〕廖凤,女,1986年生,硕士研究生;周静怡,女,1981年生,馆员,发表论文10余篇。

通过以上考证可得到以下启示:利用善本来进行文献考辨的方法,应当持久运用。中华再造善本是进行文献考辨应当采用的新文献库。

参考文献:

[1] 滕珙.类编标注文公先生经济文衡.中华再造善本.北京:北京图书馆出版社,2006.

[2] 大庭修.江戸时代中国典籍流播日本之研究.戚印平,王勇,王宝平,译.杭州:杭州大学出版社,1998:327.

[3] 中国古籍善本书目编辑委员会.中国古籍善本书目(子部·上).上海:上海古籍出版社,1989:61-62.

[4] 胡玉缙.四库全书总目提要补正.王欣夫,辑.北京:中华书局,1964:749.

[5] 王重民.中国古籍善本提要.上海:上海古籍出版社,1983:224.

[6] 赵连稳,朱耀廷.中国古代的学校、书院及其刻书研究.北京:光明日报出版社,2007:154.

[7] 真德秀.西山先生真文忠公文集.北京:商务印书馆,1937:825-827.

[8] 黄虞稷.千顷堂书目.上海:上海古籍出版社,1990:300.

[9] 于敏中.天禄琳琅书目.北京:中华书局,1995:114-115.

[10] 永瑢.四库全书总目.北京:中华书局,1965:784-785,805.

[11] 滕珙.经济文衡/景印文渊阁四库全书.台北:台湾商务印书馆,1983.

[12] 曹之.中国古籍版本学.武汉:武汉大学出版社,1992:513-523.

[13] 魏庆之.诗人玉屑.上海:上海古籍出版社,1978:425.

[14] 黄昇.中兴词选.中华再造善本.北京:北京图书馆出版社,2004.

[15] 张健.魏庆之及《诗人玉屑》考/香港浸会大学中文系.人文中国学报第十辑.上海:上海古籍出版社,2004.

[16] 方彦寿.黄昇《花庵词选》新论.泉州师范学院学报(社会科学版),2006(1):85-90.

com.

[7] PubMed [2009-08-10]. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=pubmed>

[8] IoP. [2009-08-10]. <http://www.iop.org/>

[9] 维普. [2009-08-10]. <http://159.226.100.28/>

[10] 范爱红.国外数据库产品的个性化服务.现代图书情报技术,2004(8):22-24.

[11] 张晓林.走向知识服务:寻找新世纪图书情报工作的生长点.中国图书馆学报,2000(5):32-37.

[12] 陈成鑫.E-science环境下用户信息需求与信息服务研究.情报科学,2009,27(1):108-112.

[13] 淘宝. [2009-09-12]. www.taobao.com/

[14] Google [2009-09-12]. <http://www.google.cn/ig/china>