

全文数据库增值服务模式探讨

高 冉 (北京师范大学图书馆 北京 100875)

陈淑菁 (北京教育学院图书馆 北京 100120)

摘要 文章提出全文数据库要发展增值服务的思路,并初步探讨了三种全文数据库增值服务模式,即基于文献单元整合的增值服务模式、基于元数据字段知识组织的增值服务模式、基于全文知识内容的增值服务模式。

关键词 全文数据库 增值服务 知识服务

On the Mode of Value-added Service of Full-text Database

Gao Ran (Beijing Normal University Library, Beijing, 100875)

Chen Shujing (Beijing Institute of Education Library, Beijing, 100120)

Abstract The authors put forward the suggestion that the full-text database develop value-added services. And then, the paper discusses three kinds of value-added service model, namely, the first is based on Literature units, the second is based on knowledge organization of metadata fields, the third is based on knowledge content of full-text.

Keyword full-text database, value-added service, knowledge service

1 全文数据库面临的挑战

在当今的信息时代,随着信息技术的不断发展和人们对信息需求的急剧增长,数据库业逐渐成为一个庞大的产业。而 Internet 的发展又给数据库产业注入了新的活力。当前,无论国外还是国内的全文数据库发展势头都十分迅猛,数据库已经成为信息市场的重要商品。全文数据库业的发展具有广阔的发展空间。

同时,全文数据库的发展也面临着诸多挑战。

1.1 信息环境更加复杂

除了传统的出版单位,包括商业化的出版集团和公益性的出版行为,同时还有很多其他手段,包括开放获取、预印保存以及各类学术机构仓储等。出版之后又被诸如二次数据库、网站入口、搜索引擎等进一步组织整理。在这样的信息环境中,用户往往面临着难以选择的困难。再加上网络上信息资源极为丰富,内容新颖及时,又具有免费的优势。如何让信息用户对全文数据库保持一贯的关注和信任,并乐于为之付费,是全文数据库服务商面对的最严峻的挑战。

1.2 目前全文数据库重复建设十分严重

以期刊领域全文数据库为例,20 世纪以来,除了二次世界大战之后的一段时期的年增长率为 4.7%,

期刊的数量都是以每年 3.3% 的速率匀速增长。而从 1972 年开始,科学和学术数据库数量以及这些数据库的纪录条数都呈指数增长,年增长率为 12-15%^[1],这就说明数据库中存在着大量的重复和冗余。我国全文数据库的建设重复建设也比较严重。如何在激烈的竞争环境中能够脱颖而出,是每个数据库商都需要考虑的问题。

2 全文数据库增值服务的界定

全文数据库是提供一次文献的重要工具,该领域至今也已经发展了近 40 年,目前数据库检索已经形成相对统一的模式,数据库的服务正由文献检索与文献提供向以用户最终知识需求为导向的全文数据库增值服务转变。全文数据库的重心应该由最初单纯追逐数据量的积累向重视数据质量与服务功能的改进,发展增值服务是一个可以讨论的思路。

通过对通讯、物流等其他行业对增值服务的定义,可以总结出增值服务的几个特征:

(1) 增值服务是基本服务之外的服务,是非常规服务,是能够带来收益的服务。

(2) 增值服务是基于需求的,提供的具体服务项目是基于客户需求的基础之上。针对不同的用户群和

市场需求开展的可供用户使用的非常规服务。

(3) 增值服务可以创造更高的价值。增值服务不是可有可无的,而是所有业务中重要的组成部分,并且增值服务在整个市场中的比重将会不断提高。

(4) 在为客户提供增值服务的同时,能够带动和刺激客户对基础业务的消费。

基于此,“增值服务”是指可以增加数据库服务商收益的服务,那么在全文数据库的产业中,不是指良好的检索界面或者更优质的用户体验等为产品正常运行提供的无形的东西。所谓全文数据库增值服务包括在进行的一系列分析、处理、挖掘和再生产等信息加工活动的基础上,对文献资源的潜在价值进行发掘,使其内在信息得到增值,进而形成符合用户需求的、可以带来收益的信息服务以及通过相关附加产品,带来额外收益的服务。全文数据库增值服务的基礎是传统的文献信息管理活动,方法是通过包括信息技术在内的各种手段对文献信息资源进行多维度挖掘,实现方式是通过为用户提供专业化、个性化的信息服务或产品,主要目标是提高文献的利用深度、促进知识开发、充分发掘文献的潜在价值,进一步提高文献共享程度。

相对传统的以文献实体为中心的管理方式,在全文数据库的增值服务中,应引入知识管理思想和客户服务意识,实现以用户为中心的管理,并在市场分析的基础上进一步扩大用户群,提供专业化、针对性的服务。大力进行文献信息资源的开发,发挥全文数据库服务商的资源优势,进行知识再生产,从而最终形成全新的知识产品,增加数据库服务商的收益、创造更高的价值,为更多的用户提供多方位、多角度、多途径的知识服务和技术服务,并带动一次文献的销售。

3 全文数据库增值服务模式

这里对文献信息本身从知识组织的不同角度,将全文数据库增值服务分成基于文献单元整合的增值服务模式、基于元数据字段知识组织的增值服务模式和基于全文知识组织的增值服务模式三类。需要说明的是,关于增值方式的界定不是封闭空间,随着数据库产业的发展,增值服务的内涵也会随之不断变化。

3.1 基于文献单元层面的增值服务模式

由于一般的收录较为完全的数据库数据量都非常大,而不同用户群的需求却是有差异的,有些只需要全部数据库的一部分,比如:医疗、教育、农业等不同行业的机构用户,却要支付全额的收费,这样的性价比会遏制一些潜在的用户。那么针对这样的市场需求,就可以开展基于文献单元层面的整合类型增值服务模式。针对不同的用户群和市场需求开展可供用户使用的其他服务,是对市场细分的结果。基于文献粒度知识组织增值服务就是围绕增值服务的这个特点

展开的。此类增值服务需要进行充分的市场调研、发挥本企业的文献资源优势并结合市场,推出强有力的、有针对性地增值服务。

基于文献单元重新整合的增值服务模式是指针对文献进行重新整合,深入不同行业的不同需求,整合不同领域的知识库或者知识系统。一般的全文数据库是以完整收录文献原有形态,经数字化加工,多重整序而成的专类文献数据库,如期刊全文数据库、学位论文全文数据库、会议论文全文数据库、专利全文数据库、标准全文数据库、报纸全文数据库等。而新的面向不同行业的专业知识库或者知识平台则是面向特定类型的机构和群体。针对其具体用途,系统地采集机构中各种人群需要的各层次范围的知识信息,将原有的面向数据进行信息组织变为面向用户的信息组织。针对某一行业特殊需求,打破原有的根据文献类型进行分库的格式,改为新的整序方式,从源数据库中提取该行业相关文献资源,再补充本行业专有资源,与原有文献共同组成、根据行业特点重新整序的知识库。

此类知识库流程如图3所示。首先是了解用户需求信息,在此基础上对全文数据库资源进行过滤和筛选并对其进行分类。由于资源广泛,还需要对其建立统一的元数据描述,最终形成针对某一客户群的行业资源知识库。

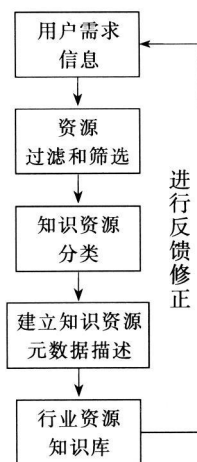


图1 个性化知识资源库构建过程模型

目前,国内的中国知网和万方数据都开展了类似的增值服务。如万方的万方医学网、冶金企业技术竞争情报系统、电力行业知识服务等;中国知网的中国医院数字图书馆、中国城建数字图书馆、中国农业数字图书馆等。这两者都是针对各自资源的特色,发布针对不同用户群的新产品。

3.2 基于元数据字段知识组织的增值服务模式

存储于数据库中的文献用元数据来揭示其特征,基本要素包括题名、作者署名、摘要、关键词、分类号、正文和参考文献等部分。为了检索的准确、方便,元数

据是数据加工底层非常重要的环节。而除此之外,在已经标准化的元数据基础上还可以开展其他类型的增值服务。

基于元数据相关字段知识组织的增值服务的核心则是把元数据作为节点,将各数据库的文献从微观上组织成为内容关联的知识网络,便于读者通过关联线索发现知识内容。使得全文数据库群通过各种各样的链接,形成知识层次上关联,编织起一个巨大的、具有可扩展性的知识网络。知识网络根据知识的体系结构和文献的内容关联组织文献,形成根据各种特征组织的文献网络和知识网络,如现在已经很完善的主题网络、分类网络等。在网络环境下,一次检索结果,就可营造无数个主题数字化学习环境,有助于人们深入学习和研究,同时发现和获得相同、相似、相关的文献和知识。这样,人们可以不必局限在某数据库中查找所需的某方面的知识,而是可以通过这个知识网络得到围绕某方面知识的更为广泛的相关内容,为文献提供更多的线索。对文献的标引深度越深,可提供的文献节点就越多,此类增值服务可开发的余地就愈大,是一次文献和二次文献的有效结合。下面就以引文库、机构库和作者库为例,说明在节点基础上开发出的数据库的作用。

(1) 引文数据库

例如大家所熟知的引文索引正是通过参考文献,即引文这一元数据作为二次加工的对象,以文献为节点,引文为联系把海量文献连成网络,才产生新的组织方式,并开发出更多的服务。这也是全文数据库可开发的重要增值服务之一。被引用情况可以作为极好的学术水平的测度,而引文分析法也作为一种重要的文献计量法被广泛应用于许多领域。因此,在此基础上而开发出的引文库则可以作为新的产品出现,可以实现多种全文数据库不能实现的价值,例如目前CNKI设计研发的新的中国期刊引文库以及Elsevier的Scopus数据库都具有比全文数据库更加多样化的功能。

但是,需要补充的是,全文库的收录范围如果不是非常全面的话,这样的评价则有失偏颇,可以在相对全面的基础上与索引库进行必要的合作。

(2) 机构库

以机构为单位,展示发表文献内容并统计发表文献总量,可以较为连续、集中、系统地反映某一机构(如高校)的科研成果,结合引文库统计被引文献数量。作为机构评价的主要依据,这当然是建立在某个领域文献收录较为完全的基础之上的,并且也会吸引被收录文献较少机构更乐于将其学术成果(题录或全文)收录至该全文数据库。

除此之外,机构库还可以配合对于机构的介绍,形成独特、权威、类别层次清晰的黄页。

(3) 作者库

通过作者库的构建,可以使文献通过作者节点串

联起来,不同于以往的单纯的作者姓名的匹配,而是结合机构库或者主题范围,准确定位作者所属机构、主要研究方向或课题,从而真正找到某一作者的全部作品,减少同名同姓的噪声问题。由于作者是文献的主要责任者,对文献内容负有主要知识责任,而通常一位作者的研究领域及其研究内容是比较固定或是比较相关的,通过专家学者库的构建,使文献形成一个以“作者”为节点的知识网络。结合其他数据源,从机构、学科、科研方向、基金项目等各个角度查找权威的专家学者,并有发表文献数量以及其他统计信息作为佐证,详细了解学者信息。还可以从多个角度,如同一机构、同一学科领域、合作者等来揭示专家学者之间的关系。

可以较为连续、集中、系统地反映某一作者的科研成果,结合引文库统计被引文献数量,以方便了解自己的科研成果被同行或相关领域研究者认可的程度,促进其进一步的科研创作。这样也可以吸引作者更乐于将其学术成果(题录或全文)收录至该全文数据库。

同样的基金、课题、项目等也都可以起到这样的作用。微观上,这类知识网络揭示知识间的短程关联,主要反映知识间客观存在的直接关系;宏观上,这类知识网络可以揭示知识间的长程关联,反映学科知识间的交叉、渗透、融合等关系。是对知识有效的识别、获取、开发、分解、使用、存储和共享,是显性知识和隐性知识构建转化和共享的途径。

3.3 基于全文知识内容的增值服务模式

除了描述文献形式和内容的元数据之外,在全文数据库中,还存储着文献中最核心的资源——全文数据。全文中蕴含着更丰富的知识,如何让用户更便捷的获得文献中的某个知识点,或者从海量文献中分析出更深入的隐性内容则是目前增值服务要考虑的问题。

此类增值服务模式是在针对全文数据库最核心的资源——全文来展开的,开发这类知识产品,对技术的要求更高,但却是几种增值服务模式中最具有全文数据库特色的。

(1) 基于文献知识单元组织的增值服务

元数据对文献的内容特征的描述仍然停留在主题和关键词的层次上,着重描述文献的一般属性,对文献知识主体部分的描述能力相对比较弱。然而,这种服务已远不能满足用户对特定知识的需要,现实情况中,各类用户需求总体特征的日渐深入,有实用化、微观化和浓缩化的趋势。

那么如何更深入的描述正文,并为用户提供知识服务成了增值服务需要关注的主要任务之一。一篇文献中所蕴含的知识是丰富的。不同的读者在阅读科技文献的过程中,由于个体的不同,所关注论文的知识对象侧重是不同的,有些注重于解决问题的过程,即某些具体案例;有些注重于某些结论或者某些概念;

有些则需要其中的某些数据。

可以通过将原有的单纯以文献为单元的组织方式,改为对文献的知识单元进行组织实现增值服务。从全文数据库中提炼出能够明确表达一个知识内容的知识元,在增值服务中有选择地对这些知识单元进行合理挖掘和组织,找到文献中的精华部分,即精粹信息^[3],从噪声信息的包裹下剥离出这部分知识,进行合理组织,在严格的标准化规范下,无数知识元经过系统化、序列化加工后整合而建成的知识元库,是一个海量知识信息资源的集合,在知识挖掘和整合管理技术的支持下,知识资源加工趋向细化和智能化。并进一步形成脱离文献载体的新的网络化的知识关联,在此基础上,与各种类型的全文进行链接,构成内容广泛的知识网络,为用户提供更方便的知识获取途径。由于改变了知识因子间的联系,所以其结果是可以提供新知识,还可以将文献的提供与知识单元的提供联结在一起,使用户在获得知识单元的同时了解其知识的来源,体现这些知识因子的准确性,也可以引导用户进一步去察看一次文献中更完整的论述。

向用户提供经过这样知识重组的知识产品,如各种事实库、概念库、数值库、图表库等,并在此基础上建立便于用户检索的新平台。这种知识服务是向用户直接提供他们所需的知识或信息,大大提高知识的利用率和针对性,节省科研人员的大量时间。其根本目的在于使读者以最快的速度接触到最需要的信息,直接支持用户的知识获取和知识创新,从根本上提升对用户的贡献力度,由此创造重要的社会价值、利润来源和市场地位。

(2) 基于海量文献内容分析的增值服务

内容分析作为一种基于定量分析的研究方法,对文献资料的依赖性很大,要求来源广泛并且准确可靠,全文数据库则提供了便利的条件,可用数据挖掘和知识发现技术使全文数据库中的大量文献在这方面发挥优势。国外的计算机辅助内容分析法的软件很多,国内则还极少。全文数据库完全可以着力于其翔实的文献基础,结合自动抽取有用信息的计算机程序以及基于本体的内容分析法的应用,借着开展增值的契机,根据内容分析理论,开展以实现高效决策、计划和管理的面向决策支持的内容分析系统的开发。此类系统是运用内容分析方法及计算机处理技术,全面拓展用户类型,满足全新需求,促进行业发展。

首先由用户提出研究问题,并从文献信息的总体中抽取研究问题样本的主题领域和时间范围。然后系统自动挖掘研究所需考察的各项因素,这些因素都应与分析目的有必然的联系,且便于抽取操作。分析单元可以是单词、符号、主题、人物,以及意义独立的词组、句子或段落乃至整篇文献。对这些分析单元进行分类或者聚类,并实现百分比、平均值、相关分析、回

归分析等统计分析工作。

系统只提供量化的数据,最终由用户对结果进行解释与检验,由用户对量化数据做出合理的解释和分析,并与文献的定性描述判断结合,提出自己的观点和结论。

参考文献

- [1] Regazzi J. The Battle for Mindshare[EB/OL]. [2007-03-14]. <http://www.nfaiss.org/RegazziFinalMilesCorrad2004.pdf>.
- [2] Hedidi Tolliver- Nigro. Diversifying with value- Added Services [M]. NAPL, 2007.
- [3] 倪晓建. 精粹信息鉴定研究[M]. 21世纪图书馆发展与变革. 北京: 图书馆出版社, 2000: 67- 71.
- [4] 张晓林. 走向知识服务: 寻找新世纪图书馆情报工作的生长点[J]. 中国图书馆学报, 2000(5): 32- 37.
- [5] 姜永常. 论知识服务与信息服务[J]. 情报学报, 2001(5): 572- 577.
- [6] 范并思. 发展面向决策的内容分析[J]. 图书情报工作, 2005(6): 9- 13.
- [7] 韩丽凤, 杨毅, 林佳. 文摘索引数据库增值功能研究[J]. 中国图书馆学报, 2005(5): 45- 49.
- [8] 严英. 高校档案的管理现状及增值服务研究[J]. 安徽农业大学学报: 社会科学版, 2007(6): 140- 142.
- [9] 曹锦丹. 基于文献知识单元的知识组织——文献知识库建设研究[J]. 情报科学, 2002(11): 1187- 1189.
- [10] 胡发连. 基于概念短语的科技文献知识管理系统的设计与实现[D]. 长沙: 中南大学, 2007.
- [11] 张蕊, 邱均平, 周黎明. 计算机辅助内容分析软件进展研究[J]. 图书情报工作, 2005(6): 29- 32.
- [12] Mary Ellen Bates. Free, Fee- Based and Value- Added Information Services[EB/OL]. [2008-03-20]. http://factiva.com/collateral/files/whitepaper_feevsfree_032002.pdf.
- [13] 曹琴仙, 于森. 基于内容分析法的专利文献应用研究[J]. 现代情报, 2007(12): 147- 150.
- [14] 中国知网专家学者库 demo 版[EB/OL]. [2008-03-21]. <http://192.168.20.66/scholar/index.aspx>.
- [15] 张鸣. 学科专题知识库的知识组织及实现研究[D]. 武汉: 武汉大学, 2005.
- [16] 邱均平, 邹菲. 关于内容分析法的研究[J]. 中国图书馆学报, 2004(2): 12- 17.
- [17] 李本乾. 描述传播内容特征, 检验传播研究假设: 内容分析法简介(下)[J]. 当代传播, 2000(1): 47- 51.
- [18] 姜永常. CNKI 数字图书馆知识服务研究[J]. 情报学报, 2004(3): 265- 274.
- [19] 朱素兰, 符雄, 蒲瑞芬. 从 CNKI 看全文检索系统的发展[J]. 现代情报, 2005(11): 35- 37.

[作者简介] 高冉, 女, 1983年生, 北京师范大学图书馆馆员。

陈淑菁, 女, 1984年生, 北京教育学院图书馆馆员。

收稿日期: 2009-10-20