

●马费成 陈 锐

面向高速信息网络的信息资源管理(一)^{*}

——从技术角度的分析

ABSTRACT Internet is a base for fast information networks. Information organisation management of fast information networks includes primary organisation management - such as free-text, hypertext, Home-Page, etc - and secondary organisation management - such as search engine, menu, etc. Information services of fast information networks include virtual libraries for enterprises, electronic libraries, information development for users, etc. The authors think that we should put emphasis on international cooperation, user training and standardisation. 6 refs

KEY WORDS Information resources management. Information superhighway. Internet. Information organisation.

CLASS NUMBER G350

信息资源管理(IRM)被认为是现代信息管理的制高点^[1]。可以毫不夸张地说,面向高速信息网络的信息资源管理是整个IRM的核心,是新的世纪人类对自身创造的信息资源进行有效管理和开发利用的基础及前提。这一领域的研究已经远远超出了情报学的范畴,而需要综合应用信息科学、管理学、社会学、经济学、法学、心理学等多学科的理论方法,从技术、经济、人文三个不同的角度进行全方位研究,从而形成现代IRM研究的三维结构,才能最终解决信息资源的管理和开发利用问题。

从技术角度,人们除了利用信息科学的原理研究解决大系统的稳定性、网络结构的有序性和高速率传输中的各种问题外,主要是用情报学的理论方法研究高速信息网络上的信息组织方法和信息服务模式,探索各种适合网络

特点的信息系统、信息媒介和利用方式。从经济角度,人们主要研究以网络化为基础的信息市场、信息产业、信息经济的形成、发展、特征和运行模式,研究网络技术的评价选择以及网络信息经济效益评价方法等方面的问题。从人文角度,人们试图通过政策、法规、伦理道德实现高速信息网络的规范化和有序化管理,使高速信息网络能健康成长。

本文从技术角度,或者说更侧重在情报学范式下讨论面向高速网络的信息资源管理、组织和利用问题。此后,我们还将分别从经济和人文角度讨论面向高速网络的信息资源管理。

1 高速信息网络及其信息资源构成

高速信息网络即“全球信息基础结构”(GII)

* 本文属国家教委“九五”人文、社会科学规划项目“面向高速信息网络的信息资源管理”的研究成果。

Global Information Infrastructure), 它由全球通信基础设施和全球“信息上层建筑”构成。前者指专用网络和公用网络的物理通信平台, 包括线路交换功能和网络路由功能, 后者指信息资源和信息服务的应用系统。目前人们所熟悉的因特网(Internet)作为世界上最大的全球信息网, 已经成为建设高速信息网不可替代的基础, 被视为未来高速信息网的雏形和信息时代的重要支柱。因特网对未来高速信息网的贡献体现在两个方面: 一是因特网目前所依赖的高速骨干网将成为高速信息网的重要通信基础结构, 二是因特网上极为丰富的信息资源和种类繁多的网络化信息服务方式将成为未来高速信息网络信息上层建筑的重要组成部分^[2]。因此, 目前所称的高速信息网或信息高速公路实质上是指因特网而言。面向因特网的信息资源管理与开发利用是未来面向高速信息网络的信息资源管理和开发利用的基础和重要内容, 也是今天可供研究的实际对象。本项研究也主要围绕因特网展开。

高速信息网上的信息资源种类繁多, 分布在不同的网络节点上, 都是数字化信息及其信息媒介为基础的电子信息资源, 它可以利用现代信息技术进行制作、加工、传播、转换和二次开发。信息媒介大体上可以划分为通信媒介、广播媒介和制品媒介三大类。通信媒介包括话音通信、信息通讯网以及与通信有关的信息服务, 广播媒介主要指广播和电视, 制品媒介指封装在一定形式的载体内的信息媒介, 包括印刷、电子和声像形式等^[3]。可以认为, 投入高速信息网的一切电子信息均为高速信息网的信息资源。

2 情报学意义上的高速网络信息资源管理

在高速网络环境下, 情报学的基本任务和目标没有变, 仍然是组织知识信息, 以便提

供利用^[4], 但是它被赋予了新的内涵。

高速信息网把方便的信息服务带进了家庭和办公室, 使人们可以像使用自来水和煤气一样方便地使用全球信息资源。但是高速信息网络建设并没有带来一个真正有序的信息空间, 相反, 网络的迅速扩张带来的信息污染、信息混乱、信息犯罪、信息侵权远远超出了传统意义上的“情报危机”。首先是信息量急剧增长, 使网络无序扩张, 网上信息陷入严重混乱, 使人们难以通过网络获取所需的信息; 其次是信息污染, 由于各种联入因特网的网络和信息资源没有质量控制标准, 一些质量混乱、粗制滥造的网络联入, 使得互联网通信混乱, 加之淫秽信息泛滥和病毒感染, 使得网上信息污染日趋严重; 第三是信息侵权和安全问题, 由于高速信息网的巨大开放性, 进出方便、存取自由, 任何个人或团体均可以把自己的计算机和局域网联入, 同时可以通过网络终端获取网上信息, 使得网络信息安全大大下降, 信息产权保护十分困难。

面对这种形势, 情报学需要与其他学科协同, 研究解决由于网络无序扩张带来的网络信息量急剧增长、信息混乱、信息污染、信息侵权、犯罪和网络信息利用之间的尖锐矛盾, 用科学有效的方法组织网上的信息, 使之有序化, 成为人们便于利用的形式, 并利用网络和其他信息技术以最快的速度向用户提供所需信息。这实质上包括两个方面: 一是网络信息组织; 二是网络信息服务。二者共同构成情报学意义上的网络信息资源管理与开发利用的基本任务和内容。在网络环境下, 计算机终端进入了家庭和办公室, 菜单式的查询工具应用得越来越多, 曾使人们产生一种错误的认识, 以为由专职信息工作者提供的信息服务变得不那么重要了, 用户可以通过终端直接从网络上获取所需要的信息, 自我服务。但人们很快发现, 他们通过自己的终端一般只能获得网上的表层信息, 要进行深层次的开发, 还需要掌握许多专门的网络信息组织、

开发和获取方面的技术,这便是专职信息工作者的职责。

高速网络信息资源管理与开发利用虽然使情报学面临多方面的挑战,但同时也给情报学带来了新的发展机遇。由于人类迄今没有找到语义信息的表示方法和测度单位,过去也不具备对信息进行处理和组织的技术条件,我们只能借助文献单元和信息的字面含义来组织信息。这是一种典型的线性结构,是一种低效率的信息系统,用户往往不能从这类系统中获得针对性很强的信息。建立在高度发达的信息技术基础上的各类全文检索系统、多媒体异构信息系统、超文本系统和超媒体系统,从各种不同角度、不同侧面展示信息结构,为知识信息的组织提供了技术支持,使人们有可能超越知识信息组织理论的约束,按知识和信息的逻辑结构来组织知识信息^[5],并通过网络信息导航、信息表现使信息可视化,对信息进行深层次的开采,逼近用户真正的需求。这完全改变了传统的知识信息组织模式,形成了知识信息组织、获取、传输和利用的新格局,大大拓展了情报学的用武之地。

3 面向高速信息网络的信息组织方式

根据因特网上信息资源的特征和构成,同时也根据人们对网络信息开发利用的需要,我们可以把网络信息划分为不同层次。人们正是按这些层次在组织和管理网上信息。

3.1 网上一次信息的组织与管理

面向高速信息网的信息组织,首先是要将网外丰富的信息资源电子化,投入高速信息网络,如同一次文献的生产一样,我们也可以把这种经过加工、组织入网的电子化信息称为“网上一次信息”。网上一次信息来源广泛、种类繁多、内容复杂、包罗万象。

网上一次信息主要有以下组织方法:

(1) 自由文本方式(Freetext)。

这种信息组织方式主要用于全文数据库

建造,是对非结构化的文本信息进行组织和处理的一种方式,它不是对文献特征的格式化描述,而是用自然语言深入揭示文献中的知识单元。根据文献中的自然语言揭示文献所含的知识单元,按文献全文的自然状况直接设置检索点。它所组织的是人们创造或采集到的网外全文信息,是输入网络的新资源。

(2) 超文本方式(Hypertext)。

这种信息组织方式将网络上相关文本的信息有机地编织在一起,以节点为基本单位,节点间以链路相连,将文本信息组织为某种网状结构,使用户可以从任一节点开始,根据网络中信息间的联系,从不同角度浏览和查询信息。目前因特网上绝大部分一次信息均采用这种组织方式。

(3) 主页方式(Home page)。

这种信息组织方式类似于档案全宗的组织方式,它将有关某机构或个人的信息集中组织在一起,是对某机构或个人的全面介绍。目前因特网上关于机构或个人的信息几乎毫无例外地均采用这种组织方式。

3.2 网上二次信息的组织与管理

一次信息入网后,如何快捷、高效地从浩如烟海、急剧增长的网上一次信息中找到用户所需要的信息,是一个十分棘手的问题。高速信息网虽是现代信息技术集大成者,但它却不具备将网上信息资源自动转换为用户所需特定信息的情报功能。需要将情报学关于知识信息组织的理论、原理和方法应用于高速网络上的信息层次控制,构建网上一次信息检索工具,如同传统的文献体系中的“文献链”原理^[6]一样。

在传统的文献体系结构中,各种类型的文献组成了一个“文献链”。文献链的概念揭示了各种文献彼此的依存关系及其产生演变的时间顺序,从而帮助人们认识文献情报流的特性,并注意链中各环节、各层次的平衡问题。图书、期刊、会议文献、科技报告等各种类型的一次文献是人们创造性的成果或对原始

知识进行组织后的产物, 经过替代、重组、综合、浓缩后进入二次文献领域, 成为形形色色的检索工具、手册、名录、词典、文摘、题录等等; 二次文献经过二次替代, 进入三次文献领域, 成为书目之书目、文献指南、综述、述评之类的文献, 形成文献的不同层次, 构成了文献链的基本环节。同样, 我们可以将文献链的原理用于网络信息组织控制。以自由文本、超文本和主页方式进入网络的网上一次信息的工具。

目前网上二次信息的组织主要有列形式:

(1) 查询器(Search engine)形式。

这是目前因特网上对网上二次信息进行组织的主要形式, 如 Hotbot, yahoo, Altavista, Archie 等, 其实质是一种报导、存贮网上一次信息的检索工具, 其性质、作用与文献链中检索工具毫无二致。用户输入自己的检索式, 由 Search engine 自动将其与存贮在网上的一次信息特征进行比较匹配, 将符合用户要求的网上一次信息的描述记录以超文本方式显示出来, 供用户检索式要求的网上一次信息很多, 动辄成千上万, 因此越来越多的 Search engine 提供二次检索、法定数检索、法定数检索功能, 以及对检索结果进行再处理功能, 如按用户要求进行排序, 或者按符合用户检索式要求的程度由大到小排列输出检索结果, 或者允许用户限定检索的时间、地域范围, 或者提供并行检索功能等等。

(2) 指示数据库(referral database)。

作为网上二次信息形式之一的指示数据库, 其存贮的是有关网上一次信息的地址以及相关的关于信息的描述信息, 如 <http://www. Interl iom>; <http://www. Harvard edu> 之类。其建库方式, 数据库结构与普通的存贮结构化数据的书目数据相似, 除存贮网上一次信息的地址和描述信息外, 还要依据一定的索引语言抽取或赋予关于该一次信息的标识作为检索点。用户输入检索式, 计算机

自动进行扫描匹配, 将符合用户要求的记录检索出来供用户选择。用户若对某一网上一次信息感兴趣, 可再转到 Netscape, 输入该一次信息的地址, 即可获得所需的网上一次信息。这种方式与 Search engine 方式相比较, 缺点是须经过两个步骤, 第一步在数据库中获得地址, 第二步在 Netscape 中输入地址进行查找, 而不象 Search engine 那样一次检索的结果便是超文本方式, 直接 click 便可获得所需一次信息。优点是入库记录都经严格选择, 具有较强的针对性和较高的可靠性, 检索结果适用性强, 而不象 search engine 那样检索结果过于庞杂, 常常使用户无所适从。因此, 指示数据库方式常用来组织专题性的或专用网上二次信息。

(3) 菜单方式。

这种方式主要用来组织用于浏览的网上二次信息。在网络环境下, 浏览检索的重要性日渐突出, 用户越来越多地通过浏览检索来确定其尚不清晰的情报需求, 并初步了解网上一次信息的内容, 根据需及时调整检索策略。以菜单方式组织的网上二次信息本来是一个超文本文件, 一般是围绕某一专题, 采用分类法、地序法、时序法、主题法等方式, 将与该专题有关的网上一次信息的线索(一般是其地址)和有关描述信息依次罗列, 供用户浏览选择, 用户若对其中一项感兴趣, 直接用鼠标在其上 click 便可。这种方式类似于手工检索中在某一专题下对款目的浏览。由于菜单方式组织的网上二次信息专题性较强, 且能较好地满足族性检索的要求, 因而受到用户的欢迎。如目前因特网上的 Jobhunter, resources for Librarian on the Internet, resources for small business 等, 均是以菜单方式组织网上二次信息范例。目前这种以菜单方式组织的网上二次信息大多出自因特网爱好者或自愿者之手, 也有部分由机构或政府部门制作提供。

以上对网上一次信息和网上二次信息组

织方式的划分,有的是信息存贮形式,有的是查询方式,不一定完全符合逻辑,但这些方式目前都是网上信息存在或表现的主要类型,以此作为划分标准比较简明。

对网上一次信息进行加工,运用 Search engine,指示数据库和菜单方式组织成网上二次信息,从而使网上一次信息进入二次信息领域,实现了对网上一次信息的控制,在逻辑上序化和优化了网络信息资源,为充分开发利用这种资源提供了前提条件和可能。这种加工实质是抽取一次信息的地址,抽取或赋予一次信息以描述和标识,这三者构成二次信息的内容,其实质是文献链中的“替代”方法。替代就是根据一定的规则描述网上一次信息的特征,对其内容进行各种不同程度的压缩,并通过二次信息组织方式使描述结果有序化,是对网上一次信息的再次传递,体现了信息工作者有价值的劳动,并成为网上信息流的控制闸门,使信息流的流量得以测度,流向得到合理调节。正是这种“替代”成为一次信息过渡到二次信息的桥梁。

若有必要,完全可以对网上二次信息进行二次替代形成“网上三次信息”。帮助用户快捷、高效地找到合适的 Search engine,指示数据库和专题菜单,以进一步提高检索效率和网络信息资源开发利用水平。实质上人们已经这样做了,如 Netscape 附带的超文本式说明,某些大型机构的 Home page 也有很多这样的网上三次信息,以帮助其局域网内用户更好地掌握和使用因特网。

从“网上一次信息”到“网上二次信息”,再到“网上三次信息”,进入网络的信息资源的可控性、有序性、易用性一步步增强,人们对高速信息网上信息资源的技术管理能力也一步步提高,对其进行一切利用的基础也一步步扎实。三个不同层次的信息分别代表着对信息的不同加工程度,构成一个类似于文献链的“网络信息链”;网络信息组织程度也大大提高。“网络信息链”的概念,是从整体

有序和动态的高度来探讨网络信息的产生、演变和结构规律的,有助于人们正确了解,即首先借助于三次信息,进而掌握二次信息,最后获得一次信息。因此,网络信息链有助于人们高效、充分地利用网络信息资源。

4 基于高速信息网络的信息服务

情报学不仅关注信息的组织和开发,同时更关注信息的需求与利用,以便最大限度实现知识信息的效益。高速信息网络虽有丰富的信息资源及快速、便捷的信息通道,但用户需要的信息却不能自动到达其手中。如同丰富的资源和便利的高速公路网需要专职服务部门进行组织调配才能使用户获得所需要的物品一样,专职的信息服务在网络时代也是不可少的,而且变得更为重要。这也是网络信息资源管理的重要内容。

以高速网络为基础可以开展丰富多采的信息服务,信息服务与信息组织开发总是密不可分、互为条件。目前围绕因特网开展的信息服务主要有以下几种:

4.1 为企业建立虚拟资料室或信息中心

一个企业或组织很难有充足的人力、物力和财力投入信息工作,尤其是对于信息资源,企业不可能也没有必要大量购进。可以充分利用网上丰富的信息资源,围绕企业的需求建立虚拟资料室或信息中心为企业所用。目前许多信息机构或信息工作者正在用这种方式帮助企业开发万维网(WWW)服务。

万维网上具有大量商用信息和企业所需的技术信息,约占该网信息总量的 80% 左右,可以采用“镜象跟踪”和“浏览书签”的方式从该网收集企业所需要的信息。镜象跟踪能够动态跟踪用户选中的节点,并把该节点上新增加的信息自动拷贝到用户的计算机终端中;浏览器书签则可以使用户对自己感兴趣的节点嵌入“书签”;需要时将其取回。同时

还可以利用网络的其他功能, 如 FTP、E-mail、Telnet 等广泛收集企业所需要的信息, 为企业建立虚拟资料室或信息中心。图书馆、情报机构和专职部门也可采用这种方式来增加自己的馆藏或开展专门服务。

4.2 建设电子图书馆

电子图书馆是指一个特定组织范围内用电子方式存贮图书、期刊的全文内容, 为用户提供访问、检索、查阅和全文提供服务的信息系统。电子图书馆是网络时代图书馆的技术模式和发展方向。信息服务必须重视电子图书馆的建设, 将现有的非电子信息资源最大限度地转换成电子信息资源, 以增加高速信息网上信息资源的存量, 并向用户提供电子图书馆的新型服务。另外要重视以高速信息网为基础的用户常备信息库的建设, 即供用户长期使用的专业或专题信息库。

4.3 利用网络功能开展多样化服务

如利用电子函件开展馆际互借, 利用文件传送功能为用户提供图书馆的联机目录, 利用网络通信为用户提供大型联机系统的数据库检索, 还可开展电子论坛、图文信息电视广播、数据广播、电视远程会议、语言信箱、电子出版等服务。

4.4 为用户进行信息挖掘服务

“信息挖掘”(Information Mining)已逐渐成为网络化时代信息服务的一个专用概念, 它与传统的信息检索不同, 能够在异构数据(如多媒体等)组成的信息库中, 从概念及相关因素的延伸比较上找出用户需要的深层次的信息, 既可进行数据挖掘, 也可进行文本挖掘。对任何一个特定的用户的特定需求来说, 网络上的任何一个信息库对他来说, 冗余信息总是大量存在, 而信息库几乎都是异构数据组成的, 如何从中将最有针对性的信息找出来, 则只能用信息挖掘这种方式。信息挖掘还可改革传统的定题服务、专项服务、委托服务、信息保证的内容和方式, 形成一个全新的适合网络时代要求的信息服务组合。

5 几个值得重视的问题

面向高速网络的信息资源管理与开发利用这一综合性的课题, 分配给情报学的任务主要是从技术角度提供知识信息组织的理论和方法, 以便有效地获取和利用网上信息。要完成这一任务, 还应当注意下列一些问题。

5.1 加强国际合作

网上二次信息和三次信息的加工制作是一项涉及面广、需要付出巨大代价的持久性工作, 没有各方面的合作, 没有众多的信息工作者参与就无法实现, 正如 CA、EISA、SCI 等国际知名的二次文献信息编制需要世界范围内的有关方面合作一样。由于网上一次信息数量庞杂, 新陈代谢快, 网上信息的组织就更需要世界范围内的合作, 保证准确、及时地报道网上一次信息及其变动情况, 提高二次信息和三次信息的质量并实现规模效益。

5.2 重视用户培训

由于网络日益普及, 越来越多的用户将直接通过自己的网络终端获取信息, 急需对用户进行培训, 使他们掌握网络操作技术, 了解网络信息的构成和分布状况, 熟悉网上信息查寻工具和获取渠道, 以便及时、准确获取自己所需要的信息。

5.3 加强信息工作标准化

在网络环境下, 信息服务业中各部门相互依赖更强、关系更复杂, 信息系统日趋社会化和国际化, 为确保信息资源共享并最大限度地开发利用, 必须高度重视信息工作标准化、规范化, 实现信息的最大效益。

参考文献

- 1 卢太宏 国家信息政策 北京: 科学技术文献出版社, 1993
- 2 张钟 当前 Internet 的某些热点问题 情报学报, 1996, 15(1)
- 3 曾民族 面向电子信息资源的 (下转第 45 页)

一个以上书目记录, 不然就造成书目记录重复, 影响检索。因此对于书目数据库必须进行查重与剔重的维护作业。

在数据库建设过程中, 总会有更新的记录进入库中, 或存在库中的是出版前预编记录(在版编目记录), 当出版物正式出版后, 根据出版物编制的正式完整记录要登录到库中来。上述这些记录都必须取代库中原有记录, 以保证库中的数据质量和记录的惟一性。

进行这项检验是根据记录头标中第 5 位(从 0 记位)记录状态的代码来判定的。当该状态为 c(经修改的记录)、d(被删除的记录)、p(出版前记录, 现已为完整的正式记录)时, 要将这些状态的书目记录取代库中原有状态为“n”的记录。当然, 在检验记录状态代码的同时, 还必须核对记录中的题名或 ISBN 或责任者等项, 看两个记录是否为同一出版物的书目记录, 相同时才可替代, 以便保证准确地替代记录。当一个书目数据库经常接受外来数据时, 这项检验与替代作业是不可避免的。

利用软件对书目数据校验, 可以采取批处理方式, 也可以采取实时处理方式进行。接受外来数据或接收本系统成批的在编记录经过审校正式进库时, 均应采取批处理方式对

数据进行校验。凡校验中发现数据有误或有疑问, 均应加以提示, 供人工核对与修改。对于正确的记录也应统计出数量提供数据库管理者备案。所谓实时处理, 即在编目过程中编目员每送一个字段(或子字段)或一个记录, 软件就对其检查, 检查结果当时就提示给编目员, 以便及时修改。采取哪种方式, 由系统设计者根据系统用户的实际需求决定。但无论哪种方式, 软件检验都应对数据中的问题与差错有简洁、明了而又具体的提示, 这样才便于编目人员判断、改正。一个成熟的图书馆编目软件系统都应具备这种数据自动校验的功能, 由于经验不足可能短时间校验功能还不周全, 但应根据实践的不断总结、积累逐步完善起来。这样才可减轻编目人员繁琐的校对劳动, 提高数据的质量, 受到用户的欢迎。

参考文献

- 1 中国机读目录格式使用手册编委会 中国机读目录格式使用手册 北京: 华艺出版社, 1995
- 2 ISDS 记录的校验和数据库管理 1986 12

朱 岩 北京图书馆教育中心主任。通讯地址: 北京白石桥路 39 号。邮编: 100081。

(来稿时间: 1997. 8. 21 编发者: 赵薇)

(上接第 17 页)

- 信息服务业及其技术发展动向 情报学报, 1996, 15(1)
- 4 马费成 情报学的进展与深化 情报学报, 1996, 15(1)
 - 5 马费成 论网络时代的图书情报教育 图书情报知识, 1996(4)
 - 6 陈光祚等 科技文献检索 武汉: 武汉大学出版社, 1987

马费成 武汉大学图书情报学院院长, 教授, 博士生导师。主要从事情报学理论方法和信息经济学领域的教学科研工作。出版著作 7 部(含合作), 在国内外发表论文 60 余篇。通讯地址: 武汉市, 邮编 430072

陈 锐 武汉大学图书情报学院情报学专业博士生。在国内外发表论文 40 余篇。通讯地址同上

(来稿时间: 1997. 8. 12。编发者: 李万健)