

浅析网络信息检索中的问题与对策

杨 青 王瑞菊

【摘 要】 本文分析了我国网络信息检索中存在的问题,针对这些问题提出了解决的办法及对策。

【关键词】 网络信息检索 问题 对策

Abstract: In this essay, the author analyzes the problems existing in the net informational search of China's network, puts forward the solutions to the problems, and suggests countermeasures against it.

Key words: net informational search problem countermeasure

人类社会已进入信息时代,以科学技术为核心,计算机为载体,利用信息技术生产、加工和传播知识,进行财富的创造,成为这一时代的主流与特征。信息技术的发展,互联网的出现及其迅速发展,大大改变了人们的生活,人们进入了数字化的时代。电子商务、电子邮件、电子交易与电子文件等互联网产物的出现,给我们带来的不仅仅是进步与方便,问题也如影随形。

1 网络信息检索中存在的主要问题

面对网上如此浩瀚的信息海洋,人们常常会陷入两种窘迫的境地,一是信息过载,太多的信息无从消化;二是信息迷失,难以找到自己所需的信息,特别是对于网络新用户来说,从网络上获取有用信息几乎成为不可能。这种局面的形成,既有网络信息本身存在的问题,也有用户知识水平问题。

1.1 网络信息泛滥,良莠不齐

网上言论自由,出版自由在给因特网带来繁荣的同时,也导致了信息爆炸。政府、企业、学校、社会团体及个人都可以在网上发布信息。因此,网络上的信息资源不计其数,成为无所不有的庞大信息资源,基本上涵盖了人类知识的各个方面,信息资源的类型又非常庞杂。由此带来了信息无序,无用信息参杂其间,垃圾信息、虚假信息、冗余过时信息以及政治反动信息、种族歧视信息增加了信息的不确定性和用户的不安全。而且网络信息的发布具有很大的自由性和随意性,缺乏规范,使信息质量和精度低,其可靠性、权威性和利用价值受到质疑。使网络用户面对眼花缭乱的信息无所适从。

1.2 搜索引擎无统一分类标准

搜索引擎是通向互联网的必经之路,起着网络导

航的作用,随着中文网站和网上中文信息资源数量的急剧增加,中文搜索网站也应运而生。最著名的中文搜索引擎如搜狐、悠游和中文 Yohao 等为用户上网检索信息提供了便利,但存在的问题也不少。

1.2.1 搜索引擎大多采用自建的分类体系,具有很大随意性

网络信息分类与学科及知识体系之间缺乏必要的内在联系,体现在网络信息的分类体系对学科知识或学科的覆盖不到位。由于分类工作的不规范、不专业,造成类目设置缺乏规律性,类目归属不合理,同位类排列不能揭示类间关系,横向关系揭示上存在不一致及类不确切等一系列影响检准率的问题。如:在著名的 Infoseek 中,出现了按“儿童和家庭——教育——图书馆——大学图书馆”展开的序列,如果用户要查找关于“大学图书馆”的信息,怎么会想到从“儿童和家庭”入手呢?

1.2.2 关键词检索

用户只要把想要查找的关键词或短语输入查询栏中点击 Search(查询)即可,速度快,返回结果数量也很大,但由于许多网站所用术语不规范,利用关键词检索返回的结果有许多,虽会有关键词却与主题无关,如要查找“AB”返回的却有许多与“AB”无关的其他“*B”,有时甚至无法确认主题。

1.2.3 目录检索

从大类直至最终相关信息网址都是依靠树形链接组成的,查找很方便,但因搜索引擎很多,它们之间无统一严格的分类标准,如查《兰州大学学报》,可从“兰州、大学、学报”三个类别去查询,给查找特定主题带来了一定困难。

1.3 检索查全率和查准率低

特定主题的查找一般依赖于关键词检索,可是目前搜索引擎采用的排序方法有所不同,大多考虑关键词出现的位置和频次,所以会出现许多结果。更重要的是用户对特定主题的检索往往难以定义其关键词,即使定义了关键词,命中记录数往往很多,不易查准。若再进一步限定,命中记录数会减少却又很容易漏检而不能查全。

1.4 部分网站质量差

近几年来我国的网络发展速度快,随之也出现了一些问题,各种网站蜂涌而现,这其中不乏一些优质的网站,如新浪、搜狐、中国经济网、中国专利信息网等等。但也有许多网站质量差,其栏目设置混乱,根本无导航措施,或者内容毫无特色,甚至内容陈旧,长期不更新,给网络用户查找信息特别是普通主题,如金融、时政等带来了麻烦。

1.5 网络信息变化大,重复率高

网络信息瞬息万变,除每天都有大量新信息在网上发布外,还有大量原有网站、网页的内容在不断更新、变换、增加和删除,且无规律可循。搜索引擎往往返回符合检索条件的所有信息,检索结果的重复率相当高。同一个网页,常常被重复显示十几次甚至几十次,所给结果并非用户所需信息本身,而是一系列网址,获得所需确切信息需以牺牲大量时间为代价。

1.6 用户使用网络的不适应性和局限性

网络信息资源在数量、结构、分布、类型和控制机制、传递手段等方面,都与传统的文献信息资源有着显著差异,人们对传统文献的检索行为习惯,对用户还有着广泛和深刻的影响。用户对网络信息资源的了解及其利用还存在着一个认识、接受和熟悉的过程,加之用户的受教育程度、知识结构等原因,也造成用户利用网络信息的局限性。

2 解决网络信息检索中问题之对策。

因特网是一个虚拟世界。涉及许多环节和主体,针对网络信息检索中存在的问题,应当从多个方面入手,多途径解决以上问题。

加强宏观调控

必须加强政策调控,建立一个完善的网络安全管理制度,通过立法程序把一些最基本的规范上升为法律、法规,强制性地使人遵守,才能保证起码的网络社会秩序,促进网络信息健康有序的发展。

首先,按照国家有关规定和本单位具体情况,制定网络信息资源利用的相关条例、制度和纪律,约束用户信息行为,教育用户在使用网络时注意信息的选择和

遵守规则。

其次,积极开展和编制有关软件及应用网络新技术,对网络上的信息进行技术上的筛选过滤,建立安全屏障,最大限度地防止有害信息的侵入。

再次,加强对网络信息资源的系统评价工作。组织相关专业人员,编制评价网址的相对标准,然后在对各种网址进行广泛浏览的基础上对各个网站或主页进行深入评价和详细介绍,进而向用户推荐,提高用户利用网络信息资源的质量和效益。

2.2 统一搜索引擎分类标准

网上信息纷繁杂乱,需要对数据进行挖掘以实现精确分类。由于我国目前存在的搜索引擎数量较大,必须制定一个统一严格的分类方法来管理,新的网络信息分类标准应是传统文献分类标准与当前网络信息分类方法相互融合的产物,兼备传统文献分类标准应是传统文献分类与科学性及目前网络信息分类方法的灵活性。可以参照国际标准,也可以制定国内标准,还要规范网络术语,使其与常规术语接轨,提高资源共享程度。

2.3 提高网站质量和规范程度

国家应制定相关法律加强网站管理监督,严格把关,对质量较差的网络加以限制,甚至可以取缔。同时制定相关政策,提高网络规范程度,以保证用户的检索效率。各网站设计者与开发者理应加强自身素质,掌握信息检索的发展动向,虚心听取用户建议和要求,客观地评价自己的网络,有目的有重点地不断改进自己的网站,努力为用户奉献一个优质的网络,提供高效实用的信息资源。

2.4 提高检全率和检准率

检全率指搜索引擎提供的检索结果中相关信息文档数与网络信息中存在的相关信息文档之比。检准率是搜索引擎提供的检索结果与用户信息需求的匹配程度,也是检索结果中有效信息的文档数量与搜索引擎提供的全部文档之比。在检索信息特别是专业信息时利用关键词检索很困难,其命中记录数很多不易查准,如果进一步限定则命中记录又很容易漏检。解决这个矛盾可以利用相关的数据库,做到两者兼顾,相互制约,彼此约束。为了查全,先在一个限制较少的数据库中尽可能全地检索该主题;为了查准,在另一个数据库中用在上一个数据库中内容出现较多的内容进行检索,最后用相关检索词进行限定,从而得出准确的文献记录。从另一个角度看,网上检索必须以数据格式的标准化为前提,各检索系统应该具有规范控制,包括主题规

(下转第94页)

都注明“未经授权,禁止复制”或“利用本网站稿件,须经书面授权”的字样,然而图书馆对电子出版物和网站上的作品或按学科门类,如文字类、财经类、社会科学类;或按主题;或按作者进行整序(电子剪报),然后放到图书馆的主页上,供读者随意浏览或下载,甚至被制成光盘或打印成资料出售,而均未得到授权许可,这种行为很明显地超出了合理使用的范围,尤其将电子剪报用于有偿服务,产生营利行为,构成了侵权。

我国著作权法规定了12种情况下使用他人作品为合理使用,这也同样适用于网络环境下的电子剪报业务。事实表明,凡在报刊上刊登或者网络上传播的作品,除著作权人声明或者下载该作品的网络服务商受著作权人委托声明不得转载、摘编的以外,图书馆网站予以转载、摘编,开展电子剪报业务,并按规定支付报酬,注明出处的,不构成侵权。但若网站转载、摘编超过原文三分之二的,则认定为侵权。由此可见,电子剪报不能随意所为,必须得到授权许可,并且要注明原文名称及作者姓名,要支付报酬,否则,可能会导致造成直接侵权。

5 购买电子出版物提供服务易产生的著作权问题

5.1 提供合法电子文献服务造成的侵权

有人提出,图书馆从正规的书市、新华书店、音像总汇等市场上购买的合法电子出版物作为馆藏资源,然后向读者提供服务,就不会构成对他人著作权的侵犯,事实并非如此简单,市场上出售的合法电子出版物,一般是经著作权人授权许可后复制的数字化作品

或数字化制品,不存在侵犯他人复制权的问题。但应当指出的是,这些合法电子出版物是仅供购买者“阅读”和在“合理使用”范围内正常利用,并非允许购买者可以随意“复制”或“汇编”。图书馆若将这些电子出版物以电子剪报的形式纳入自己的数据库,则是一种侵权行为,侵犯的是他人对于作品的汇编权。

5.2 提供合法数据库服务造成的侵权

也有人提出,图书馆从正规市场上购买的合法数据库作品作为馆藏,为读者提供优质服务,不存在侵权问题,情况也非如此。我国新著作权法把数据库作为汇编作品纳入了保护范围,汇编作品的著作权由汇编人享有。图书馆未经数据库著作权人许可,擅自将他人的数据库作品并入自己库中,同样是侵犯了他人的著作汇编权。

参考文献

- 1 卢松苗.入世后图书馆的知识产权策略.图书与情报,2002(4)
- 2 李顺德.数字图书馆与知识产权保护.图书情报工作,2002(12)
- 3 邵葵.2002年我国图书馆数字化建设中的著作权管理现状调查.图书馆杂志,2002(12)
- 4 陈志宏.知识产权课程教学中的几个问题思考.图书馆论坛,2002(5)
- 5 陈志宏.论高校图书馆开展信息服务中的知识产权问题.图书馆学研究,2002(4)

甘清瑛 广州体育学院图书馆。

(上接第83页)

范、名称规范,从而保证检索的查全率和查准率。

2.5 正确应对网络信息的不稳定

定期或不定期地搜集一些新增或更新的网络信息资源网站域名及搜索引擎并及时公布,使用户及时了解最新检索信息。其次在各自网站的网页上设置按学科、主页、专题组织的信息引导系统,为用户提供网络导航性质类的宣传指导。

2.6 提供网络信息导航,完善用户指导

为了给用户提供清晰的思路,检索界面应有层次性,措词应符合明确性要求。同时,为了解决检索方法的简便性和检索效果的有效性这对矛盾,可采取有较高检索率的检索方法与加强用户指导相结合的方法。

在网络环境下,面对日益激烈的信息大战,我们必须奋起直追,提高普通用户的信息检索能力,生产出更多的信息精品,在网络信息检索方面脱颖而出,使我国

的网络信息检索朝着大众化、资源共享化、简便化的方向发展。

参考文献

- 1 贾芳华.网络搜索引擎与联机检索系统的比较分析.图书情报工作,2002(5)
- 2 李曼.网络信息检索的检准率研究.图书情报工作,2003(3)
- 3 凌美秀.关于搜索引擎当前存在的主要问题及其发展趋势的探讨.高校图书馆工作,2001(5)
- 4 郭家文.网络信息检索效率研究.图书情报,2002(3)
- 5 陆刚.图书馆网络信息资源利用的障碍及对策研究.现代情报,2000(6)
- 6 焦玉英,索传军.网络环境中信息检索理论与实际的发展.图书情报知识,2001(3)

杨青 王瑞菊 兰州大学图书馆。