

●黄晓斌 邱明辉

网络信息过滤中的分级体系研究*

摘 要 网络信息过滤中的分级体系,具有保护用户免受不良信息侵扰、维护作者言论自由、发挥网络推荐功能、加强网络管理等作用。国内外较有影响的分级体系有 RSACi、SafeSurf、Msproxy、Serve 2.0、“过滤王”局域网软件分级体系,等等。编制网络信息过滤分级体系,应注意充分揭示不良信息、满足用户多样化需求、易于分级、便于管理等。表 1。参考文献 8。

关键词 网络 信息过滤 分级体系 图书馆信息资源建设

分类号 G352.2

ABSTRACT In this paper, the authors summarize the functions of rating systems in network information filtering, introduce some major rating systems in foreign countries, such as RSACi, SafeSurf, MsProxy, Serve 2.0, “Filtering King”, etc., and propose some key points for the development of rating systems in network information filtering. 1 tab. 8 refs.

KEY WORDS Network. Information filtering. Rating system. Library information resource development.

CLASS NUMBER G352.2

过滤网络不良信息的方法很多,比较常用的有分级法、URL 地址列表法、动态文本分析法和图像识别技术等。其中采用分级法过滤不良信息是在现有技术条件下最为可行也最为有效的一种方法。所谓分级法,是根据一定的分级体系(rating system)把网络信息的内容属性或其他特征分门别类地揭示出来,成为分级标记,使用时与过滤模板进行比较,以决定是否过滤。在分级法中,分级体系规定了分级的类目、子类目或者类目的级别和分级的具体方法,是对网络信息进行分级的依据,其实质就是一种网络信息分类法。

1 分级体系的作用

网络信息的分级可以由作者、用户或第三方的组织机构进行。由作者实施的分级叫做自我分级,由第三方实施的叫第三方分级。根据分级体系的不同,适用对象可以分为适合自我分级的分级体系和专用分级体系两种。下面要论述的 RSACi、ICRA 和 SafeSurf 分级体系就属于自我分级体系,它们既适合作者分级,也适合第三方分级;而 N2H2 公司、Websense 公司和珠海捷朗菱公司的过滤软件所采用的分级体系只用在特定的过滤软件中,属于专用分级体系。运用分级体系对网络信息进行分级的作用如下。

1.1 保护用户免受不良信息的侵扰

网络上存在大量的色情、暴力、诽谤、诈骗、毒品交易、教唆犯罪等不良信息,对用户尤其是未成年人造成了严重的侵扰和伤害。利用一定的分级体系对这些不良信息进行分级后,网络用户包括家长和学校、公司等机构就可以根据各自的需要设置过滤模板来过滤不适宜的信息,达到安全上网的效果。保护网络用户尤其是未成年人免受不良信息的侵扰是研制分级体系的主要目的。

1.2 维护作者的言论自由

有关“不良信息”的内涵和外延,不同的社会、民族、宗教、文化、群体、个人都会有不同的界定,没有一个统一的标准。有的信息在一些用户看来是不道德、不能接受甚至是十恶不赦的,但在别的用户看来则是可以理解、可以接受、十分平常的。为了维护作者言论自由的权利,很有必要运用一定的分级体系把这些信息揭示出来,由用户决定是否访问,这在保护用户免受不良信息侵扰的同时,又可以最大限度地维护作者的权益。

1.3 发挥对网络信息的推荐功能

通过对网络信息的分级,可以在剔除用户不需要信息的同时把需要的信息推荐给用户,从而节约用户获取信息的时间,提高获取信息的效率。随着网

* 本文为教育部十五规划项目(01JA870013)的论文之一。

络信息分级范围的扩大和深化,分级体系的推荐功能将更为有效且更受关注。

1.4 限制某些用户的上网时间

在对网络信息进行分级的基础上,可以根据用户的不同情况,按类设置上网的时间或者时段。限制用户上网的时间是一种不涉及信息内容的中性管理方法,在一些组织机构尤其是在学校、图书馆等国家资助的公共机构中采用这一方法对用户进行管理能避免不少争端。

1.5 进行网络的分级管理

对网络信息进行分级后,可以按类设置访问权限,限制用户访问诸如聊天、娱乐、购物等影响效率的网站,有利于加强网络的管理,提高网络的使用效率;还可以按类分析访问日志,了解网络的使用情况,监测用户的访问行为,便于网络管理员尤其是家长和老师做好网络管理工作。

2 国内外目前较有影响的几个分级体系

从上个世纪90年代开始,国外已有不少分级机构为过滤网络不良信息而研制了各种各样的分级体系;我国在这方面的研究虽然要晚一些,但现在也有珠海捷朗菱公司、清华同方思科等编制了自己的分级体系,并开发出相应的过滤软件投放市场。这里列举国内外影响较大的几个分级体系。

2.1 RSACi 分级体系

1996年美国RSAC(娱乐软件咨询委员会)为了应对美国国会的CDA法案,组织有关方面的专家编制了这个适合自我分级的分级体系,目的是要保护儿童,避免其受到不良信息的侵扰,维护网页作者的言论自由^[1]。其立类以“客观中性”为原则,设暴力、性、裸体和语言四个类目,每个类目下按危害程度分为五个级别,如语言类分为非冒犯性的语言、轻微的脏话等,类目体系简单,十分适合自我分级。为了保证自我分级的客观性,尽量减少“主观”的成分,RSACi对一些虽有科学、文学、艺术价值的不良信息也不作特殊处理,如暴力类不区分合法或非法的暴力,裸体类也不区分是艺术还是色情的裸体。作者要对网页进行分级时可以登录RSACi网站,根据要求填写一份表单后就可以得到该网页的分级标记。

RSACi推出后,其适合自我分级的特点得到了许多网页作者和网站的支持,从而极大地扩大了网络信息分级的覆盖面;而且它的分级标记支持PICS规范,家庭或学校可以在不增加开支的情况下通过设

置浏览器过滤不良信息。所以RSACi一经推出就受到网络行业、家长和老师的热烈欢迎。

但是,网络毕竟是一个多元化的领域,RSACi的立类原则及其所体现出来的思想观点和价值观念也受到不少批评^[2]。首先,它所秉承的“客观中性”的立类原则实际上是倡导以欧美为中心的价值体系;其次,它的类目体系过于简单,难以满足用户多样化的需求;再次,它的类目设置、类目的级别和分级的方法也颇多争议:因为类目体系十分简单使得一个类目要包括许多内容,如语言类就涉及种族仇视、侮辱秽语、恶意诽谤、下流手势等内容,而烟酒、毒品、武器等危害严重的信息又无类可入;类目的级别按危害程度划分,但量化标准却有失客观,比如语言类中的种族仇视与粗话、下流手势的级别相同,性类中两个男人间的亲吻与男女间的亲吻同属一个级别就不够妥当。

1999年ICRA(互联网内容分级协会)兼并了RSAC后,针对RSACi的缺陷于2000年12月推出了新的自我分级体系——ICRA分级体系^[3]。ICRA分级体系设裸体与性、暴力、语言、其他主题、聊天五个一级类目,下设若干个子类目。它的最大特点是以子类目取代了RSACi类目的级别,扩充了类目和子类目的数量,增加了不少对儿童有严重危害的类目,同时把适合儿童的具有艺术、教育、医学价值的不良信息独立出来。ICRA分级体系出台后,在欧美和亚洲的许多国家得到了推广使用,特别是在2002年6月网络巨头AOL、微软、雅虎表示它们已采用ICRA分级体系对自己网站上93%的网页进行分级后^[4],这进一步扩大了它在网络界的影响力。

2.2 SafeSurf 分级体系

这是由美国的SafeSurf公司开发的、互联网上最早的一种自我分级体系之一,目的是为了保护儿童、维护成年人的言论自由。该分级体系和同期的RSACi一样强调立类的客观中性,但类目详细得多^[5]。它把涉及成人主题的内容分为年龄段、亵渎、异性主题、同性主题、裸体、暴力等11个类目,在每个类目下又分出九个级别,其中年龄段类的级别按儿童到成人的次序排列,其他类目的级别除3—5级揭示参考或艺术价值的内容外基本上按内容描述的具体程度排列。据SafeSurf公司称网络上已有5万多个网站用它进行自我分级。

2.3 N2H2公司的MS Proxy Server 2.0过滤软件的分级体系

N2H2 分级体系只有一个层次,设有 35 个限制类目,6 个例外类目,与 RSACi、SafeSurf 相比类目要详细得多^[6]。N2H2 过滤软件的主要客户是学校、公司、家庭和图书馆,所以其分级体系除了有不良信息的类目如毒品、暴力外,还有许多限制学生、员工访问的类目如内衣、泳衣、学生作弊信息等。该分级体系最有特色的是区分出六个例外类目,包括教育、儿童适用、历史、医学、适度、口头表达,这六类内容即使归入 35 个类目中用户也可以访问。例外类目的设置使整个分级体系变得较为简洁,同时又具有一定的延展性。

2.4 Websense 公司的 Websense Enterprise v4.4.1 过滤软件的分级体系

由于 Websense 软件的市场定位是员工网络管理方案,主要用于防止员工滥用网络,所以在该分级体系的 31 个一级类目中,除了有不良信息的类目如成人资料、药品、赌博等外,更多的类目如娱乐、职位搜索等是从防止网络滥用的角度而设置的。这些类目有的还分出子类目,如“成人资料”类细分为“成人内容、裸体画、性、性教育、女式内衣与游泳衣”五个子类目。整个分级体系的类目和子类目有 80 多个,体系十分周详^[7]。

2.5 珠海捷朗菱科技有限公司的“过滤王”局域网管理软件的分级体系

该分级体系把可能影响上网者身心健康的站点分为不良站点(包括色情、暴力等)、写真站点(包括性健康、性教育、写真集等)、游戏站点、交友站点四大类,不设子类目也不分级别,作为一个过滤软件的分级体系,其类目明显过于粗糙^[8]。

比较上述五种分级体系,可以发现:从类目体系来看,基本上是以主题对象聚类为主,辅之以其他标准。在设置类目时,对危害严重的不良信息如性、暴力、亵渎、裸体、毒品等大都设为一级类目,突出揭示。同时重视类目设置的实用性,如适合自我分级的分级体系较为简单,一级类目控制在十个左右,以便于网页作者分级;过滤软件的类目设置强调用户的需要,类目体系相对要复杂些,即既要考虑到最终用户(如小孩、学生、员工)的可能,又要便于管理方(如家长、学校、公司)的管理。为了适应用户多样化的需求,各种分级体系的一级类目都有不断增加的趋势。从类目展开来看,除 N2H2 的分级体系只有一个层次外,其他分级体系都在一级类目下细分出子类目或类目的级别,但各种分级体系都控制在两层以内。子类目仍以主题对象聚类为主,而类目级别的划分则有主观和客

观两种标准,如 RSACi 类目的级别划分标准比较客观,可以保证分级结果的一致性;而 SafeSurf 类目的级别划分标准比较主观,分级的结果就会因人而异。从 RSACi 被 ICRA 取代来看,对类目划分级别已不能满足分级的需要,也无法满足用户过滤的要求。

3 编制适合我国图书馆的网络信息分级体系

图书馆在馆藏资源数字化建设过程中,可能会涉及一些并不适合所有读者的信息,如居于色情边缘的裸照、恐怖片、艳情小说等;在组织网络信息资源时,必然也会遇到各种各样的网络不良信息,图书馆很有必要把这些信息揭示出来,实行分级管理,以满足读者的不同需要。而要把这些信息揭示出来,其中关键的一环就是如何制定或选择分级体系的问题。从上面介绍的分级体系来看,它们并不适合我国图书馆用于分级网络信息资源,这是因为:①任何一种分级体系都是一定的思想观点和价值观念的反映,即使是声称“客观中性”的 RSACi、ICRA、SafeSurf 也不例外。思想观点和价值观念的差异正是这些分级体系不能很好地满足我国图书馆需要的主要原因。②上述分级体系中,有的类目设置过于粗糙,无法充分地揭示网络信息资源;有的则过于复杂,不可避免地造成同主题信息的分散和分级成本的上升。③图书馆作为人们获取信息的主要场所,它必须在保护用户避免受到不良信息的侵扰与维护作者和读者的言论自由之间达到一定的平衡,所以在编制或选择分级体系时必须更多地考虑我国的法律法规对不良信息的界定。

因此,笔者认为很有必要编制一种适合图书馆分级网络信息资源的分级体系。另外,由于我国目前还没有一种适合自我分级的分级体系,从提高分级体系的影响、扩大网络信息分级的覆盖面出发,如果所编制的分级体系既能满足图书馆的需要又能兼顾自我分级和要求,这将会是一件很有意义的事。要编制这样的分级体系,应突出要求如下几点:

(1)充分地揭示网络不良信息。我国制定和颁布的一系列与规范网络信息内容、维护网络信息安全有关的法律法规中,较有代表性且适用于图书馆的是 2000 年国务院颁布的《互联网信息服务管理办法》,在其第十五条中十分全面地规定了九种禁止性的不良信息,涉及面非常广泛。除此之外,还有许多界于违法边缘的临界信息。这就决定了要编制的分级体系类目会比较多,才能充分地揭示各类网络不良信息。

(2)满足用户多样化的信息需求。由于用户信息需求的多样性和对不良信息认定的差异性,设置分级体系类目时,一定要考虑到用户思想观点和价值观念的多样性,尽可能多设置一些类目或子类目。

(3)易于分级。如果分级体系的类目比较多,则有利于充分地揭示网络不良信息和满足用户多样化的需求。但是,如果分级体系既要适合图书馆分级又要兼顾自我分级,则每一级的类目不宣太多,控制在十个左右比较合适。类目多了,不仅会增加分级的成本,也会分散应该集中的主题,更有可能使网页

作者失去自我分级的积极性。

(4)便于管理。这是从加强图书馆管理的角度提出来的。分级体系中除了要设有不良信息的类目外,对一些并非不良信息但又需要适当控制的信息如聊天、游戏、娱乐站点等也应该设类揭示出来,以满足管理的需要。

根据我国法律法规对不良信息的界定,借鉴国内外各种分级体系的编制方法,笔者提出了一个初步的适合我国图书馆分级管理网络信息资源的分级体系大纲(见表1)。

表1 图书馆适用分级体系大纲

1 政治 包括攻击社会主义制度、危害国家安全、泄露国家秘密、颠覆国家政权、破坏国家统一、损害国家利益、破坏民族团结、扰乱社会秩序等信息	4 与犯罪有关的信息 包括赌博、洗钱、恐吓、敲诈、枪支贩运、毒品交易、教唆犯罪等信息
2 文化 包括邪教、封建迷信、种族歧视等信息	5 信息传播 包括失实信息、垃圾信息、过时信息、黑客信息、计算机病毒等
3 道德 裸体、性、色情、暴力、诽谤、谣言、侵犯版权和商业秘密等信息	6 其他 包括烟、酒、粗俗、聊天、游戏、娱乐、广告等信息
	7 例外 (属于上述类目但又具有下列性质的内容) 具有教育、历史、医学、科学、文学、艺术价值的信息

在分级体系大纲中,按主题把不良信息和其他不适宜的信息划分为政治、文化、道德、与犯罪有关的信息、信息传播和其他信息六类。前四类基本上是属于《互联网信息服务管理办法》所规定的不良信息的范围,第五、六类则是出于保护未成年读者、便于管理的需要而设立的。这六个类目下,再按主题划分出子类目,各类的子类目基本上在十个左右,类目设置较为均衡,整个分级体系控制在两层以内。

分级体系大纲中的第七类是例外类,主要是把归入1—6类但又具有教育、历史、医学、文学、科学、艺术价值的信息再次揭示出来,给用户更多的选择自主权。例如关于性教育方面的资料,在分入第三类的“性”子类目中后,由于具有教育价值,同时又分入第七类的“教育”子类目中。过滤时,例外类的权限最高,当用户选择过滤“性”子类目的内容但不过滤“教育”子类目的内容时,有关性教育方面的资料将不会被过滤。所以,例外类目的设置可以使整个分级体系变得较为简洁,同时又具有一定的延展性。

图书馆利用分级体系对网络信息资源进行分级后,分级的结果最好能生成 PICS 标记,这样用户可以方便地利用 IE 浏览器的内容审查功能过滤不良信

息。同时,为了便于网页作者进行自我分级,图书馆还可以在自己的网站上开设分级服务,为网页作者自动生成 PICS 标记,由此可以扩大分级体系的影响力,并使分级标记在更大范围内得到共享。

参考文献

- 1,3 <http://www.icra.org>
- 2 JM Balkin, et al. Filtering the Internet: A Best Practices Model. www.copacommission.org/papers/yale-isp.pdf
- 4 http://computer.online.sh.cn/computer/gh/content/2002-06/28/content_381556.htm
- 5 <http://www.safesurf.com>
- 6 http://www.n2h2.com/products/bess.php?os=filtering_info&content=categories
- 7 <http://www.prc.websense.com/products/about/database/categories.cfm>
- 8 <http://www.zetronic.com.cn/product.htm>

黄晓斌 中山大学信息管理系副教授,博士。通讯地址:广州市。邮编 510275。

邱明辉 海军广州舰艇学院图书馆馆员。通讯地址:广州市。邮编 510275。

(来稿时间:2004-04-19)