

网络信息过滤

3.1 网络信息过滤的基本概念

目前有关信息过滤的定义比较多,概括起来,比较有代表性的主要有如下几种。

- (1) 信息过滤指的是把信息传递给需求者的一系列过程的总称^①。
- (2) 信息过滤主要是运用一定的工具和根据一定的标准设置过滤条件,在网络运行过程中一旦触发条件则将不需要的信息拒之门外,而其他信息仍可以通过网络传播进来^②。
- (3) 信息过滤是一种系统化的方法,用来从大量信息流中抽取某个人希望得到的信息^③。
- (4) 信息过滤就是信息的选择性传播^④。

从上述不同角度的定义,可以看出信息过滤的内涵主要包括如下几个方面。

- (1) 信息过滤的目的是要满足特定用户的信息需求。
- (2) 信息过滤的依据是信息与特定用户信息需求的相关性。
- (3) 信息过滤的对象是动态的网络信息流。
- (4) 信息过滤不仅从信息流中选取需要的信息,也从中剔除不需要的信息。

综上所述,网络信息过滤就是根据一定的标准和运用一定的工具从动态的网络信息流中选取用户需要的信息或剔除用户不需要的信息的方法或过程。

3.2 网络信息过滤的原理

图 3-1 是信息过滤系统的一般模型。图中,一个或一组用户由于工作、学习、生活的需要产生了信息需求,这种需求在较长的一段时间里保持相对的稳定。用户对某种信息的选择需求是信息过滤的前提。用户的信息需求以计算机能够识别的形式揭示出来,就是用户

① Belkin, Nicholas J., et al. Information Filtering and Information Retrieval: Two Sides of the Same Coin? Communications of the ACM, 1992, 12.

② 黄晓斌. 因特网信息过滤研究. 图书情报工作, 2000(7): 42~45.

③ 何军, 周明天. 信息网络中的信息过滤技术. 系统工程与电子技术, 2001(11): 76~79.

④ 梅海燕. 信息过滤问题的研究. 现代图书情报技术, 2002(2): 44~47.

需求模板(profile,也叫过滤模板),它是进行信息过滤的预定条件。对于用户需求模板,可以是正向的,也可以是反向的,也就是说既可以揭示用户需要的感兴趣的信息,也可以描述用户不需要的不感兴趣的信息。

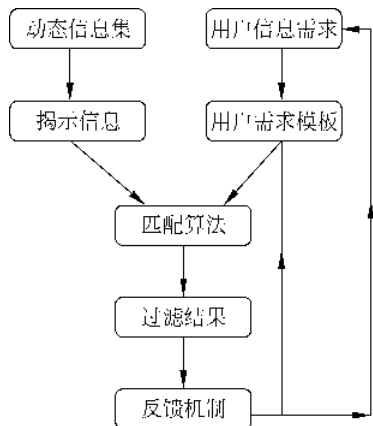


图 3-1 信息过滤系统一般模型

在系统中,对动态的网络信息集不作预处理,只是当信息流经过系统时才运用一定的算法把信息揭示出来。匹配算法与用户需求模板的描述方法、信息的揭示方法是相互联系的,常用的匹配模型有布尔模型、向量空间模型、概率模型、基于知识的表示模型以及混合模型等,主要任务是剔除不需要的信息、选取需要的信息并按相关性的大小程度提供给用户。

为了提高信息过滤的效率,系统还根据用户对过滤结果的评价通过反馈机制作用于用户和用户需求模板,使用户逐渐清晰表示自己的信息需求,使用户需求模板的描述变得越来越具体和明确。

在整个信息过滤系统中,用户需求模板的构建、信息的揭示、匹配算法和反馈机制是最为关键的部分。在现有技术条件下,全自动的信息过滤系统还处于试验阶段,为了提高实用性,往往会在这些关键部分进行必要的人工干预,如对动态的信息集先作预处理、人工修改用户需求模板等。

3.3 网络信息过滤的种类

(1) 根据信息过滤系统的结构,可以分为基于内容的过滤和协作过滤两种。

① 基于内容的过滤(content-based filtering)。又叫认知过滤,是利用用户需求模板与信息的相似程度进行的过滤,能够为用户提供已感兴趣的相似的信息,但不能为用户发现新的感兴趣的信息。在反馈机制的作用下,用户的信息需求处于循序渐进的变化过程中。

② 协作过滤(collaborative filtering)。又叫社会过滤,是利用用户信息需求之间的相似性或用户对信息的评价进行的过滤。由于价值观念、思想观点、知识水平或需求偏好相同或相似的用户,他们的信息需求往往也具有相似性,基于这一思路,通过比较用户需求模板的相似程度或者根据用户对信息的评价而进行的过滤,既可以为用户提供正感兴趣的信息,又可以提供新的感兴趣的信息。在这种系统中,用户的信息需求有可能呈现跃进式的变化。

(2) 根据过滤操作的不同位置,又可以分为服务器过滤和客户端过滤。或依据用户需求模板所在的位置可以分为上游过滤和下游过滤两种。

① 上游过滤(upstream filtering)。将用户需求模板存放在服务器端或代理端上。服务器过滤可以在将信息下载到客户端之前进行过滤,外来的信息要经过它才能进入本地局域网,而内部信息也要经过它才能传送出去。因而可以设置一些相应的限制,对一些网址和信息进行控制。这种方式的优点是被过滤掉的信息决不会被下载到客户端。优点是支持基于内容的过滤,也支持协作过滤;缺点是模板不能用于不同的网络应用中,用户和过滤系统之间的交流变得越来越复杂,适用于大型单位。

② 下游过滤(downstream filtering)。过滤软件也可以是客户端的一部分,用户需求模

板存放在客户端上。优点是模板可用于不同的网络应用；缺点是只能实现基于内容的过滤。软件下载到客户端再进行过滤，需要大量时间进行维护，适用于个人或小的单位。

(3) 根据信息过滤的作用，可以分为推荐系统、阻挡系统和一般过滤系统三种。

① 推荐系统(recommended system)。根据用户对信息的需求特点把信息推荐给合适的接收者，属于协作过滤系统的一部分。

② 阻挡系统(blocking system)。通过设置一定的条件剔除用户不需要的信息，而对其他信息用户仍然可以获取，它主要用于过滤网络不良信息或者是用户不需要的信息。我们主要研究这种系统。其中，在下面论述的过滤方法中，分级法、URL 地址列表法主要用于协作过滤系统，而自动文本分析法既可用于基于内容的过滤系统，也可用于协作过滤系统。

③ 一般过滤系统。是除推荐系统、阻挡系统之外的其他过滤系统或兼有多种功能的系统。

(4) 根据网络信息的应用工具，可以分为 Web 的信息过滤、E-mail 的信息过滤、FTP 的信息过滤、BBS 的信息过滤和 Blog 的信息过滤等。

(5) 根据网络信息过滤的实现方式，可以分为基于专门过滤软件的信息过滤、基于浏览器的分级审查系统和专门过滤引擎的信息过滤、基于防火墙的信息过滤以及基于硬件系统的信息过滤等。基于防火墙的数据包过滤通过静态和动态的制定安全策略，控制进出内部网的数据包，保护内部网络及各结点主机的正常工作，同时防止内部保密数据被非法窃取，以及外部有害数据的流入。包过滤防火墙在网络层上实现对数据包的处理和过滤，没有获得较完整的基于应用协议的数据，因此，无法实现对应用层数据内容的过滤。基于网间的应用层数据过滤技术通过比较完整地获取数据包中应用层的数据内容，实现基于应用层协议的内容过滤，当前很多网络信息过滤系统的研究主要采用这种方法。

(6) 根据网络信息过滤的目的，可以分为两个方面：一是过滤不良信息，主要目的在于维护网络信息的健康，净化网络环境，可称之为不良信息过滤；二是过滤掉无用、不相关的信息，主要目的在于获取与用户需求密切相关的信息，在此可称之为获取相关信息过滤。

3.4 网络信息过滤与网络信息检索

网络信息检索的目的是使用户从经过专门的组织整理的网络信息集合中获取特定需要的信息，它包括了信息的揭示、存储、查找和传递以满足特定用户信息需求的方法和过程。相对来说，信息检索比较重视静态信息的揭示和用户需求的匹配问题，对于动态信息的处理和用户需求的变化则重视不够。随着网络的日益普及和网络信息总量的激增，信息过载、信息污染的问题正严重困扰着用户，单靠信息检索的方法要从动态的庞大信息流中获取相关的信息已是力不从心，因此信息过滤作为一种能选取用户需要的信息、剔除用户不需要的信息的有效方法就应运而生了。比较上面所述的信息过滤的概念和信息过滤系统的基本原理，从中可以看出信息过滤与信息检索有许多共同之处。

(1) 目的相同。信息过滤与信息检索都是根据用户的信息需求从信息集中获取信息，目的是要满足用户的信息需求。在信息检索中，描述用户信息需求的是检索式，而在信息过滤中则为用户需求模板。

(2) 原理基本相同。信息过滤系统和信息检索系统的原理基本相同，只是在信息检索

系统中必须预先对信息集进行收集、揭示,揭示的结果保存在数据库中,而信息过滤系统不对信息集进行预处理。信息检索的技术方法广泛应用于信息过滤系统中,信息检索长期发展积累起来的技术方法,如用户需求的描述方法、信息的描述方法和匹配算法等都在信息过滤系统中得到了应用,但信息过滤更重视反馈机制和用户需求模板的构建和维护。

(3) 效率的评价基本相同。对于信息过滤的效率评价,不同领域有不同的评价指标。过滤软件开发商往往用过滤正确率(应该过滤的信息被正确过滤的比例)、过滤错误率(不应该过滤的信息被错误过滤的比例)来评价过滤软件的效率,计算公式为:

① 过滤正确率 = 被正确过滤的信息/应该过滤的信息×100%

② 过滤错误率 = 被错误过滤的信息/不应该过滤的信息×100%

而计算机信息管理学者 Donna Harman 则认为应该把过滤效率和检索效率的评价指标统一起来,像信息检索分为检准率和检全率一样,信息过滤也分为滤准率和滤全率。计算公式如下:

① 滤准率 = 过滤得到的用户需要的信息/过滤得到的信息×100%

② 滤全率 = 过滤得到的用户需要的信息/用户需要的所有信息×100%

这样,信息过滤和信息检索的效率就有了可比性,而且像检准率和检全率一样,滤准率和滤全率在一定情况下也存在互逆的关系。

信息过滤和信息检索也有许多不同之处(见表 3-1)。信息过滤更关注一个或一组用户长期的信息需求,同时跟踪用户需求的变化;过滤的是动态的半结构化或非结构化的数据,而且不对信息集进行预处理;过滤的结果着重于剔除不需要的信息,带有即时性。

表 3-1 网络信息检索与网络信息过滤的区别

网络信息检索		网络信息过滤
用户	一般是对一个用户	针对一个或一组用户
信息需求	用户一次性的或短期的信息需求;信息需求的描述为检索表达式	用户长期的信息需求;信息需求的描述为用户需求模板
信息集合	必须对信息集进行标引,结果存储在数据库中	对信息集不作预处理,过滤时自动分析
对象	检索的是相对静态的结构化的数据	过滤的是动态的半结构化或非结构化的数据
结果	输出的是相关的信息,带有回溯性	不仅输出需要的信息,更着重于剔除不需要的信息,带有即时性

3.5 网络信息过滤的作用与局限性

3.5.1 网络信息过滤的作用

1. 减轻认知压力的减压阀

因特网的信息不仅数量庞大而且增长十分迅速,在方便人们利用的同时,也对人们的认知能力提出了严峻的挑战。当人们用搜索引擎去检索信息时,这种缺乏智能的以字符匹配

为主的检索工具往往会因返回的结果太多、质量良莠不齐而令人无从下手；而且在这些检索结果中有许多并不是用户真正感兴趣的，用户要找到自己需要的信息无异于大海捞针。信息过滤是减轻用户认知压力的较好方法之一。它根据用户信息需求的变化提供长期的信息服务，能够节约用户获取信息的时间；它在为用户提供所需要信息的同时，更着重于剔除用户不需要的信息，从而提高了用户获取信息的效率，减轻了用户的认知负担，起到了减压阀的作用。

2. 开展个性化信息服务的推动者

把用户需要的感兴趣的信息主动地推荐给用户，这就是个性化信息服务。随着网络信息数量的日益庞大和飞速增长，信息服务业要满足不同背景、不同目的、不同时期用户的信息需求，必须针对不同的用户提供不同的信息服务，信息过滤作为一种有效的技术手段对个性化信息服务起到了巨大的推动作用。在个性化信息服务中，最重要的是收集和分析用户的信息需求，由于信息过滤的反馈机制具有自我学习和自我适应的能力，可以动态地了解用户兴趣的变化，可以越来越明确、越来越具体地掌握用户的信息需求，从而为用户提供更有针对性的信息。在协作过滤系统中，还可以根据用户需求之间的相似性来推荐信息，有可能为用户提供新的感兴趣的信息，拓宽用户的视野。

3. 保持网络畅顺的把关人

美国心理学家库尔特·劳因 1947 年提出，在群体传播中存在一些把关人 (gatekeeper)，只有符合群体规范或把关人价值标准的信息内容才能进入传播的渠道^①。在这里，把关人刻画的是一个博览群书、交流广泛、善解危难的智者形象，他根据自己的价值观念和立场观点把信息传递给需求者。用把关人来形容信息过滤在解决网络挤塞、保持网络顺畅方面所起的作用是再合适不过了。通过信息过滤，可以减少不必要的网络信息传递，节约宝贵的信道资源；利用信息过滤，可以对网络信息的流向、流量和流速进行合理的配置，使网络更加畅顺。而对于用户来说，信息过滤由于剔除了大量不需要信息的流入，可以避免塞车的现象。

4. 清除不良信息的净化器

信息过滤与网络不良信息的关系可谓渊源深厚，早在 1982 年 Denning 在“电子垃圾”一文中就把解决电子垃圾的问题与信息过滤联系起来^②。20 世纪 90 年代以后，随着网络不良信息的日益严重，信息过滤作为解决不良信息问题的技术手段之一更是受到社会各方面的广泛关注，许多国家也支持过滤网络不良信息的研究。过滤网络不良信息是信息过滤最重要的应用之一，通过分级类目、关键词和规则等描述用户的信息需求，以分级、URL 地址列表、自动文本分析等方法过滤不良信息，同时运用一些人工干预的方法提高信息过滤的效率，在保护网络用户尤其是未成年用户免受不良信息侵扰方面发挥了很好的作用。

① 彭兰. 网络传播概论. 北京：中国人民大学出版社，2001.

② Peter J. Denning. ACM president's letter; Electronic Junk. Communications of the ACM, 1982, 3.

3.5.2 网络信息过滤的局限性

(1) 影响正常的信息利用。目前的过滤技术还存在比较多的问题,过滤效果不甚理想,通常会有一些不该过滤的信息过滤了,而一些应该过滤的信息却仍然在继续传播,起不到应有的效果,因此,反而影响正常的信息利用。

(2) 涉及到个人隐私。信息过滤关注用户长期的信息需求、动态跟踪用户需求变化的做法通常会涉及到用户的个人隐私问题。许多用户反对这种做法,一些国家也立法禁止网站使用软件跟踪用户的行为举止,认为这样做影响到网络信息的自由利用。

(3) 影响到网络的速度。在实时过滤网络信息时,网络的响应速度会受到一定的影响,特别是在过滤算法比较复杂或者过滤所使用的数据库比较庞大时,影响较为明显。

(4) 限制了用户的认知。信息过滤系统既然担负着把关人的角色,必然是根据过滤标准的制定者的立场观点和价值观念向用户提供信息。用户如果只能接触到把关人所选择和推荐的信息,长此以往,接触的范围和认知能力都会受到制约。

(5) 需要一定的投入。网络信息过滤需要一定的工具和维护的经济成本,因此,增加了网络信息服务和利用者的负担。

3.6 网络信息过滤的主要方法

网络信息过滤的方法很多。从过滤的对象来看,信息过滤可以分为内容过滤、网址过滤和混合过滤三类。内容过滤是通过文本分析、图像识别等方法扫描信息的内容,阻挡不适宜的信息;网址过滤是对有问题的网址进行控制,不允许用户访问其信息;混合过滤是将内容与网址结合起来控制不适宜信息的传播。从是否对网络信息进行预处理来看,信息过滤可以分为主动过滤和被动过滤两类:主动过滤是先对网络信息进行预处理,如对网页或网站预先分级、建立允许或禁止访问的地址列表等,在过滤时根据分级或地址列表决定能否访问;被动过滤是对网络信息不作预先的处理,过滤时才对地址、文本内容或图像等信息进行分析以决定是否过滤。常见的网络不良信息过滤方法主要有如下几种。

3.6.1 分级法

分级其实就是按照等级进行分类,它是根据网页的内容属性或其他特征,按照一定的标准分门别类地揭示和组织网页的一种方法。分级的结果以一定的形式描述出来成为分级标记,这是对网页进行分级过滤的依据。分级标记可以附在网页上也可以保存在文件或数据库中,使用时与过滤模板进行比较以决定是否过滤。网页的分级可以由网页作者、用户或第三方进行,由网页作者实施的分级叫做自我分级,由第三方的组织或机构实施的分级叫第三方分级。下面将分别介绍分级体系、PICS 规范、分级标记的产生、存储、传输和使用。

1. 分级体系

就像对电影、电视进行分级必须依据一定的标准一样,对网页的分级也必须有一定的标

准,这个标准就是分级体系。分级体系是对网络信息进行分级的依据,它规定了分级的类目、子类目或者类目的级别以及分级的方法,实际上就是一种网络信息分类法。分级体系可以分为适合自我分级的分级体系和专用分级体系(主要是过滤软件用的分级体系)两类。这里列举几种国内外比较有名的分级体系。

(1) RSACi 分级体系是为了应对美国国会的“儿童保护法案”,由美国娱乐软件咨询委员会(RSAC)于 1996 年组织有关方面的专家编制的一个自我分级体系^①。目的是既要保护儿童避免受到不良信息的侵扰,又要维护网页作者的言论自由。它的立类以“客观中性”为原则,设暴力、性、裸体和语言 4 个类目,每个类目下按危害程度分为 5 个级别,如语言类分为非冒犯性的语言、轻微的脏话、一般性的脏话、下流的手势、暴露或粗鲁的语言,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,RSACi 的类目尽量排除了“主观”的因素,对一些具有科学、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。网页作者要分级时可以登录 RSACi 网站,根据网页的内容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,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,家长和老师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、家长和老师们的欢迎。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。首先,它所谓的“客观中性”的立类原则实际上是以欧美的价值体系为中心的。其次,它的类目设置过于简单,无法满足用户多样化的需求。再次,它的类目体系、类目级别和分级方法也颇有争议。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、侮言秽语、恶意诽谤、下流手势等内容,而一些危害严重的内容如毒品、烟酒、武器的信息却又无类可入。类目级别按危害程度划分,但量化标准却有失客观,例如语言类中的种族仇视与粗话、下流手势同列为第 4 级,性类中两个男人之间的亲吻与男女之间的亲吻同属一个级别等。ICRA(互联网内容分级协会)于 1999 年兼并了 RSAC 后,在 RSACi 的基础上于 2000 年 12 月推出了新的自我分级体系——ICRA 分级体系。ICRA 分级体系设裸体与性资料、暴力、语言、其他主题和聊天 5 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,同时把适合幼童的具有艺术、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,2002 年 6 月网络巨头 AOL、微软、雅虎表示,它们已采用 ICRA 分级体系对网站上 93% 的网页进行了分级,这进一步扩大了 ICRA 分级体系在网络界的影响力。

(2) SafeSurf 分级体系^②是美国 SafeSurf 公司开发的、因特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。它和同期的 RSACi 一样,强调立类的客观中性,但类目详细得多。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、裸体和暴力等 11 个类目。每个类目下又分为 9 个级别,其中年龄段类按幼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,其余各级别基本上

^① <http://www.icra.org>(2008-08-08).

^② <http://www.safesurf.com/>(2008-08-08).

按内容描述的具体程度排列。现有 50 000 个网站用它进行自我分级。

(3) N2H2 公司^①的 Microsoft Proxy Server 2.0 过滤软件的分级体系只有一个层次, 设 35 个类目, 另有 6 个例外类目, 类目按字母序排列, 和 RSACi、SafeSurf 相比, 类目更为详细。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成人、毒品、非法、裸体、色情、亵渎、性、暴力外, 还有许多需要限制学生、员工访问的类目, 如新闻、招聘、娱乐、邮件、股市、内衣、泳衣和学生作弊信息等。N2H2 的分级体系最有特色的是例外类目, 它专门细分出 6 类虽归入 35 个类目中但也可以访问的内容, 这 6 类内容是教育、儿童适用、历史、医学、适度 and 口头表达, 这种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。

(4) Websense 公司的 Websense Enterprise v4.4.1 过滤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。它的分级体系设 31 个一级类目, 除了不良信息的类目如成人资料、药品、赌博、违法行为、极端主义、暴力之外, 更多的类目如娱乐、职位搜索、购物、运动、旅游、汽车是从防止网络滥用的角度而设置的。有些类目下还分出子类目, 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 5 个子类目。整个分级体系的类目和子类目有 80 多个, 类目体系十分周详^②。

(5) 珠海捷朗菱科技有限公司的“过滤王”局域网络管理软件把有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、毒品、邪教和赌博等)、写真站点(包括性健康、性教育和写真集等)、游戏站点和交友站点 4 大类, 不设子类目也不分级别^③。作为一个过滤软件的分级体系, 其类目明显过于粗糙。

比较这些分级体系, 可以发现:

(1) 大类结构。从类目体系来看, 基本上是以主题对象聚类为主, 辅之于其他标准。从类目设置来看, 对不良信息中的非法信息如性、暴力、亵渎、裸体和毒品等大都设为一级类目, 突出揭示。同时类目的设置重视实用性, 如适合自我分级的分级体系较为简单, 一级类目控制在 10 个左右, 以便于网页作者分级。过滤软件的类目体系相对要复杂些, 类目设置更重视用户的需要, 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多的趋势。

(2) 类目展开。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, 但各种分级体系都保持在两层以内。子类目仍然是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。

总而言之, 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; 同时分级体系也在一定程度上决定了过滤软件

① http://www.n2h2.com/products/bess.php?os=filtering_info&content=categories(2008-08-08).

② <http://www.prc.websense.com/products/about/database/categories.cfm>(2008-08-08).

③ <http://www.zetronic.com.cn/product.htm>(2008-08-08).

的功能和效率。因此,用户在分级网络信息或选用过滤软件时,应该根据自己的需要选择分级体系、配置分级类目。

2. PICS 规范

在讨论 PICS 的原理之前,先来明确如下几个概念。

(1) 分级服务(rating service)。是指根据分级体系为网络信息提供和发布分级标记的服务。也有人把 rating service 译为分级机构。

(2) 分级体系(rating system)。规定了分级的类目、类目的级别以及分级的标准。每一种分级体系可以包括一个或多个类目,如微软浏览器 IE 从 1996 年起就采用的分级体系 RSACi(the Recreational Software Advisory Council on the internet)把游戏软件和网站分为暴力、性、裸体和语言 4 个类目。每个类目还可以继续分出细目或赋予一定的级别,如在 RSACi 的暴力类目下面,还分为无暴力、打斗、杀戮、带血腥的杀戮场面、恣意的而且非常无理的暴力行为 5 个级别。

(3) 分级标记(content rating)。分级机构根据分级体系对网络信息进行分级后产生的数据集。它可以嵌入到网页文件中,也可以独立存在,是对网络信息进行过滤的主要依据。分级标记主要由三部分组成:分级机构标识、标记选项和分级。分级机构标识一般以分级机构的 URL 作为识别符;标记选项揭示所分级文件或者是标记本身的附加属性,如分级的时间和有效期等;分级是描述文件内容的一个或多个类目或级别的属性。

(4) 分级服务系统(label bureau)。通过网络提供分级服务的计算机系统。有的分级服务系统除了可以提供分级标记外,还可以提供原文件。许多分级机构都有自己的分级服务系统,有些分级服务系统也可提供多个分级机构的分级标记。

PICS 分级过滤的模式如图 3-2 所示。

因特网具有开放性、动态性和国际性等特点,要从源头上控制网络不良信息的做法是很不现实的。但在网络信息与用户之间增加一层控制不良信息的机制,则相对比较容易实现对信息内容的控制,这便是许多分级过滤软件开发的基本思路。但 PICS 并没有制定一个具体的分级体系或者是开发出一套过滤软件,它只是规定了分级标记的一般格式和发布方法,从而为各种分级体系和过滤软件之间建立起一个平台,使各种与 PICS 规范兼容的分级标记易于被运用

PICS 规范开发的过滤软件处理。各个分级机构可以运用自己的分级体系对网络信息进行分级,分级所得的结果形成 PICS 标记。过滤软件的用户或监管者可以相对独立地选择分级机构和分级体系,十分灵活地定制过滤模板,从而实现对网络信息的过滤。在这种模式下,还可以把过滤软件的开发从分级机构中剥离出来,使分级机构只需以一定的观点和方法对网络信息进行分级,而不必涉及软件开发的问题;软件公司可以专注于技术开发,无须考虑信息过滤的价值取向问题。这是过滤网络不良信息的一种十分有效而且理想的模式。

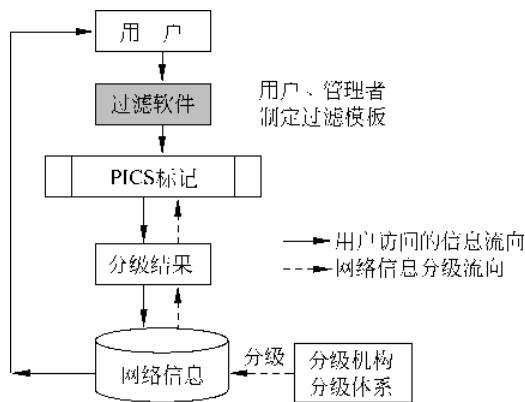


图 3-2 PICS 分级过滤的模式

根据一定的分级体系对网页进行分级后,分级结果可以用 PICS 标记、RDF 格式的元数据或数据库的记录这三种形式描述出来。由于 PICS 规范在网络信息分级方面具有较大的影响力,下面简要介绍一下 PICS 规范的分级结果描述。

PICS 系列规范中的 PICS Label Distribution Label Syntax and Communication Protocols 规定了 PICS 标记的一般格式和 3 种传输标记的方法^①。一个网页的 PICS 标记包括以 URL 作为标识的分级机构;与分级信息有关的属性一值对;提供实际分级信息的属性一值对。其一般格式为:

```
(PICS-1.1
  <service url> [option...]
  labels [option...] ratings (<category> <value>...)
  [option...] ratings (<category> <value>...)...)
```

如采用 RSACi 分级体系对 PICS 站点进行分级后,产生的 PICS 标记可以描述为:

```
(PICS-1.1 "http://www.rsac.org/ratingsv01.html"
labels by "martinm@raleigh.ibm.com"
on "1996.06.24T10:11-0500" exp "1998.06.06T08:00-0500"
for "http://www.w3.org/PICS/" generic true
ratings (n 0 s 0 v 0 l 0))
```

上例中前三行揭示了 PICS 标记的版本、分级体系、分级人员和分级的具体时间及时效时间;其作用范围是 URL 为 <http://www.w3.org/PICS/> 站点下的所有网页;最后一行说明根据 RSACi 分级体系,对该站点的分级是裸体(n)、性(s)、暴力(v)、语言(l)的级别均为 0,也即无裸体、无性、无暴力、非冒犯性语言的内容。

在这个规范中还规定了分级标记的 3 种传输方法。

(1) 分级标记嵌入到 HTML 文件的<META>标签中,随同 HTML 文件一同传输。例如,嵌入 PICS 主页的分级标记是:

```
...<META http-equiv=PICS-Labels
content='(PICS-1.1 "http://www.rsac.org/ratingsv01.html"
l by "martinm@raleigh.ibm.com"
on "1996.06.24T10:11-0500" exp "1998.06.06T08:00-0500"
r (n 0 s 0 v 0 l 0))'>...
```

(2) 在各种网络协议中只要遵循 RFC-822 邮件格式,分级标记就可以和网页一同传输。RFC-822 定义的邮件由邮件头和邮件内容两部分组成,当用户向服务器提交访问网页及其分级标记的请求后,服务器把网页当作邮件内容,把分级标记放在邮件头中一并传递给用户。

(3) 分级标记与所分级的网页分离通过分级服务系统(通过网络提供分级服务的计算机系统)进行传输。在这种方法中,分级服务系统接到用户需要某网页分级标记的请求后,返回给用户一个包含 PICS 标记的文件。

3. PICS 技术规范介绍

从 1996 年至今,PICS 工作小组开发出了一系列的技术规范,其中有些已成为 W3C 的

^① [http://www.w3.org/PICS/#Specs\(2008-08-08\)](http://www.w3.org/PICS/#Specs(2008-08-08))。

推荐标准,它们中比较重要的有如下几个。

1) Rating Services and Rating Systems and Their Machine Readable Descriptions

一种分级体系可能被多个分级机构选用,而一个文件也可能被多个分级机构标记,为了识别分级机构及其所采用的分级体系,本规范制定了一套 MIME(多媒体因特网邮件扩展)的数据类型 application/pics-service 语句,用于描述分级机构及其分级体系。这种语句能够被分级过滤软件识别,从中解析出分级标记的内容,也可用于生成用户接口以定制过滤软件。下面以 IE6.0 中文版中采用的 RSACi 为例,简要介绍一下 PICS 对分级机构及其分级体系的描述方法。

```
((PICS - version 1.0)
  (rating - system "http://www.rsac.org/Ratings/Description/")
  (rating - service "http://www.rsac.org/ratingsv01.html")
  (name "RSACi")
  (description "Recreational Software Advisory Council 的 Internet 分级服务是以 Stanford 大学有近 20 年研究媒体效果经验的 Donald F. Roberts 博士的研究成果为基础的。")
  (category
    (transmit - as "v")
    (name "暴力")
    (label
      (name "级别 0: 无暴力")
      (description "没有侵犯性的暴力行为,没有自然的或意外的暴力事件。")
      (value 0) )
    (label (name "级别 1: 打斗") (description "...") (value 1) )
    (label (name "级别 2: 杀戮") (description "...") (value 2) )
    (label (name "级别 3: 带血腥的杀戮场面") (description "...") (value 3) )
    (label (name "级别 4: 恣意的而且非常无理的暴力行为") (description "...")
      (value 4) ))
  (category (transmit - as "s") (name "性")...)
  (category (transmit - as "n") (name "裸体")...)
  (category (transmit - as "l") (name "语言")... ))
```

例子中,分级体系采用其 URL(<http://www.rsac.org/Ratings/Description/>)作为识别符,通过具有唯一性的 URL 可以了解该分级体系的类目设置、类目的取值范围以及分级的标准。

同样,分级机构以其 URL(<http://www.rsac.org/ratingsv01.html>)作为识别符标明分级标记的来源,通过该 URL 可以得到该分级机构的详细信息。

其后的 name 以简洁的语词来命名分级机构,而 description 则是对该分级机构的详细描述。

RSACi 的 4 个类目暴力、性、裸体、语言的传输用名分别为 v、s、n、l。传输用名是类目的标识,在一种分级体系中具有唯一性。当分级体系用于标记多语种的内容时,传输用名必须相同,以保证软件能够处理用不同语种表示的类目名称。

在每个类目下列出了级别名称和级值,如暴力类的级值 0~4,分别代表无暴力、打斗、杀戮、带血腥的杀戮场面、恣意的而且非常无理的暴力行为这几个级别。

2) PICS Label Distribution Label Syntax and Communication Protocols

本规范规定了分级标记的一般格式和 3 种传输标记的方法。一个文件的分级标记包括