



扫码关注深读周刊
微信“红蓝铅笔”

看到时常浏览的网页一个接一个无法打开,不少宅男就能清楚,网络扫黄的下一把火将烧向哪里。

自4月13日“扫黄打非·净网2014”专项行动启动以来,从淫秽色情网站、网络文学网站到最近个别视频播放器乃至门户网站,相继被有关部门处理。这并非让网络褪“色”的第一次尝试,但密集通报处理、剑指知名网站都让本次行动注定不再是一场网民围观的“热闹”。

那就让我们严肃点儿,谈谈网络扫黄中的门道。

网络扫黄“技术活”

本报深度记者 刘德峰

黄不黄,谁说了算?

继“91熊猫看书网”等知名网络文学网站之后,新浪网遭遇了全国“扫黄打非”办公室的精确打击。

4月24日,“扫黄打非”办向社会通报相关情况时称:经核实,新浪公司在其开办的新浪网读书频道中,登载了《全村女人的梦中情人:极品小村医》、《霸占全村美妇:山村美娇娘》、《美少妇的不伦之恋:小镇情人》等20部淫秽色情互联网作品;在新浪网视频节目中,登载了《女子交响乐团》、《比基尼美女表演》等4部色情互联网视听节目。

这些公布的书名和节目名,黄色风格可见一斑。但究竟什么信息能够被划归到黄色一类?目前我国仍没有精细化的界定。

4月14日,一名网络小说作者在知乎网公开了晋江文学城编辑发来的消息,这名编辑在消息中称,“所有乱伦题材,完结的作品如果能改也改一下设定,未完结正在连载的作品要求必须修改”,“所有未成年人发生性关系的内容,恋童、虐童情节必须不可以有,如果有这样的情节请一定修改。”

接连出现的“所有”、“必须”、“一定”这种语气强调、不容商量的词汇,可以看出如今网络文学网站对尺度的拿捏程度和紧张程度。当然,该编辑也不忘在消息中补充一下:“KISS什么的还是可以有的,请大家不要冤枉过正。”

当然,依据这些信息作为判断标准仍显不足,因此,更权威、更全面的标准,还要看我国的相关法律规定。

我国刑法第三百六十七条第一款规定,本法所称“淫秽物品”,是指具体

描绘性行为或者露骨宣扬色情的诲淫性书刊、影片、录像带、录音带、图片及其他淫秽物品。有关人体生理、医学知识的科学著作不是淫秽物品,含有色情内容的有艺术价值的文学、艺术作品不视为淫秽物品。

2004年9月,“两高”又出台《关于办理利用互联网、移动通讯终端、声讯台制作、复制、出版、贩卖、传播淫秽电子信息刑事案件具体应用法律若干问题的解释》(下简称“解释”)。

“解释”对刑法第三百六十七条第一款规定的“其他淫秽物品”做出了更进一步的解释。2009年,国新办、工信部、公安部等部门曾联合开展整治互联网低俗之风专项行动,列出了网络低俗内容的十三项标准,包括“直接暴露和描写人体性部位的内容”、“全身或者隐私部位未着衣物,仅用肢体掩盖隐私部位的内容”以及“带有侵犯个人隐私性质的走光、偷拍、漏点等内容”等。

尽管界定“淫秽”、“低俗”的法律及规定日臻完善,但如果想依此判断出黄色信息,显然会让人失望。

2013年,由百度、腾讯、金山等十多家互联网公司组成的“安全联盟”对外宣布,公开招聘“首席淫秽色情鉴定官”。在此后被公开的考卷中,一道文字题这样提问:“科技处女干事每月经过下属科室都要亲口交代24口交换机等技术性器件的安装工作”这句话中一共有多少个淫秽色情词汇?众多围观者、讨论者始终没能形成一致性答案,该题目也一度被称为“神级面试题”。

曾有媒体报道指出,全国“扫黄打非”办的工作人员也曾经类似判断上的难处。对于裸体艺术图片,按照刑法及“解释”的规定,不属于淫秽物品范畴。而按照“十三项标准”,它又属于低俗内容。

即使相关法律规定渐渐为黄色信息划定了红线,可具体到一个网页、一篇小说或一张图片,但衡量其是否为黄色信息的标准,每个人心中有不同刻度的标尺。

正因如此,判断一个信息是否为“黄色”,“肯定还是以‘人的判断’作为最终的标准。”近日,在某大型互联网公司技术部工作的张强(化名)说。

人机配合各有所长

“在互联网海量信息面前,如果仅靠工作人员来判断是否黄色信息,那得需要多少人力成本啊。”在一家安全软件公司工作的李明(化名)告诉齐鲁晚报记者,人的判断以及人为制定的标准,需要工具的配合。

张强向齐鲁晚报记者介绍,比如黄色图片识别系统,经过多年的研究,一些学术机构和商业机构已经开发出较为成熟的系统模式,供有关部门使用。“它可以通过识别图片中色彩的信息,或是和人体器官特征相匹配的信息,识别出黄色图片。”

这种识别技术与交警部门的车牌识别技术类似。车牌识别系统在获取拍摄的机动车照片后,经过去除背景、定位车牌、分割字符、抽取字符特征等一系列过程后,最终得到所拍摄机动车的车牌号。

与黄色图片识别系统略有不同,相关技术部门对黄色词汇的识别显得相对直接。“基本包括两种方式,一种是建立一个词库,另一种是建立某种规则。”张强说,像“色情”一词,按照第一种方式,只要工作人员在词库中设定了这个词,那这个词就会被系统自动筛选出来。而按照第二种方式,“即使一个句子前面有个‘色’,后面有个‘情’,中间隔

了一堆乱七八糟的词,只要工作人员制定规则说这种的也要筛选,那它肯定也会被呈现出来。”

“但工具或者说系统的识别,还是没法达到人的智能水平。”李明说,这就需要扫黄工作人员与识别系统更好地配合。

实际上,不管是计算机对图像的识别还是对文字的识别,都是基于一种运算方式。而这种运算方式的基础,又是人为设定的参数。“在某个参数下,筛选出的图片或文字,都需要工作人员来判断,这个系统的标准是否合适。”张强告诉齐鲁晚报记者,在国家没有明确规定的情况下,工作人员只能通过这种方式调试出一种最好的参数标准。“机器的判断与工作人员的主观判断越接近,就说明这个识别系统的效果越好。”

在张强看来,黄色信息的鉴别难点在于视频文件的识别与筛选。“主要是因为视频数据的传输是经过压缩编码的。”他说,用户下载的视频文件,只有用播放器打开时才呈现为一段视频,如果不借助播放器,实际上只是一堆乱七八糟的编码。

“视频这个东西的解码是很花时间的。”张强介绍,不像图片识别系统,可能一秒钟就能区分出几百张图片是不是黄色图片,视频文件的识别,则可能需要花费半个小时才能看出一个视频是不是黄色视频。“每天在互联网上上传和下载的视频那么多,相关部门没有足够的运算资源去做这个。”他说,虽然目前围绕视频识别技术也有一些前沿的研究,但这项技术还没有达到商用水平。于是,近年来频频见诸报端的“鉴黄师”,也成为鉴别黄色视频的主要力量。

(下转B02版)