

RATING THE NET

Jonathan Weinbergal

Introduction

Internet filtering software is hot. Plaintiffs in *ACLU v. Reno* relied heavily on the existence and capabilities of filtering software (also known as blocking software) in arguing that the Communications Decency Act was unconstitutional. * * * Free speech activists see this software as the answer to the dilemma of indecency regulation, making it possible “to reconcile free expression of ideas and appropriate protection for kids.” Indeed, some of the strongest supporters of such software are First Amendment activists who sharply oppose direct government censorship of the net.

Internet filtering software, further, is here. As of this writing, the Platform for Internet Content Selection (PICS) working group has developed a common language for Internet rating systems, making it much easier to create and market such ratings. Two heavily-promoted ratings systems (SafeSurf and RSACi) allow content providers to rate their own World Wide Web sites in a sophisticated manner. Microsoft’s World Wide Web browser incorporates a feature called Content Advisor that will block Web sites in accordance with the rules of any PICS-compliant ratings system, including SafeSurf and RSACi. Stand-alone blocking software—marketed under such trademarks as SurfWatch, Cyber Patrol, CYBERSitter, KinderGuard, Net Nanny, and Parental Guidance—is gaining increasing sophistication and popularity.

It is easy to understand the acclaim for filtering software. This software can do an impressive job at blocking access to sexually explicit material that a parent does not wish his or her child to see. The PICS standard for describing ratings systems is an important technical achievement, allowing the development and easy use of a variety of sophisticated ratings schemes.

In the midst of the general enthusiasm, though, it is worth trying to locate filtering technology’s limitations and drawbacks. Blocking software is a huge step forward in solving the dilemma of sexually explicit speech on the net, but it does have costs. People whose image of the net is mediated through blocking software may miss out on worthwhile speech through deliberate exclusion, through inaccuracies in labeling inherent to the filtering process, and through the restriction of unrated sites. * * *

I. Background

* * * Because nothing in the structure or syntax of the Web requires Web pages to include labels advertising their content, * * * identifying pages with sexually explicit material is not an easy task. First-generation blocking software compiled lists of off-limits Web pages through two methods. First, the rating services hired raters to work through individual Web pages by hand, following links to sexually explicit sites and compiling lists of URLs to be deemed off-limits to children. Second, they used string-recognition software to automatically proscribe any Web page that contained a forbidden word (such as “sex” or “xxx”) in its URL.

The PICS specifications contemplate that a ratings system can be more sophisticated. A rating service may rate a document along multiple dimensions. Instead of merely rating a document as “adult” or “child-safe,” it might give it separate ratings for violence, sex, nudity, and adult language. Further, along any given dimension, the rating service may choose from any number of values. Instead of simply rating a site “block” or “no-block” for violence, a rating service might assign it a rating of between one and ten for increasing amounts of violent content. * * * Finally, * * * ratings need not be assigned by the authors of filtering software. They can be assigned by the content creators themselves or by third parties. * * *

II. Accuracy

Since blocking software first came on the market, individual content providers have complained about the ratings given to their sites. Not all of those complaints relate to problems inherent to filtering software. For example, some programs tend to block entire directories of Web pages simply because they contain a single “adult” file. That means that large numbers of innocuous Web pages are blocked merely because they are located near some other page with adult content. * * * Other problems arise from the wacky antics of string-recognition software. America Online’s software, ever alert for four-letter words embedded in text, refused to let users register from the British town of “Scunthorpe.” The University of Kansas Medical Center installed Surfwatch in its Internet kiosk, and

discovered that users could not see the Web page of their own Archie R. Dykes Medical Library. For sheer wackiness, nothing can match a CYBERSitter feature that causes Web browsers to white out selected words but display the rest of the page (so that the sentence "President Clinton opposes homosexual marriage" would be rendered "President Clinton opposes marriage"). * * *

Controversies over sites actually rated by humans are less amenable to technological solution. One dispute arose when Cyber Patrol blocked animal-rights web pages because of images of animal abuse, including syphilis-infected monkeys; Cyber Patrol classed those as "gross depiction" CyberNOTs. The situation was aggravated because Cyber Patrol, following the entire-directory approach described above, blocked all of the hundred or so animal welfare, animal rights, and vegetarian pages hosted at the Animal Rights Resource Site. * * * Sites discussing gay and lesbian issues are commonly blocked, even if they contain no references to sex. * * *

One might think that a better answer lies in rating systems, such as RSACi and SafeSurf, in which content providers evaluate their own sites. An author, one might assume, could hardly disagree with a rating he chose himself. The matter, though, is not so clear. When an author evaluates his site in order to gain a rating from any PICS-compliant rating service, he must follow the algorithms and rules of that service. Jonathan Wallace, thus, in an article called *Why I Will Not Rate My Site*, asks how he is to rate "An Auschwitz Alphabet," his powerful and deeply chilling work of reportage on the Holocaust. The work contains descriptions of violence done to camp inmates' sexual organs. A self-rating system, Wallace fears, would likely force him to choose between the unsatisfactory alternatives of labeling the work as suitable for all ages, on the one hand, or "lumping it together with the Hot Nude Women page" on the other.

At least some of the rating services' problems in assigning ratings to individual documents are inherent. It is the nature of the process that no ratings system can classify documents in a perfectly satisfactory manner. Consider first how a ratings system designer might construct a ratings algorithm. She might provide an algorithm made up entirely of simple, focused questions, in which each question has a relatively easily ascertainable "yes" or "no" answer. (Example: "Does the file contain a photographic image depicting exposed male or female genitalia?") Alternatively, she might seek to afford evaluators more freedom to apply broad, informal, situationally sensitive guidelines so as to capture the overall feel of each site. (Example: "Is the site suitable for a child below the age of 13?")

In jurisprudential terms, the first approach relies on "rules" and the second on "standards." The RSACi system attempts to be rule-based. In coding its violence levels, for example, to include "harmless conflict; some damage to objects"; "creatures injured or killed; damage to objects, fighting"; "humans injured or killed with small amount of blood"; "humans injured or killed; blood and gore"; and "wanton and gratuitous violence; torture; rape," its designers have striven to devise simple, hard-edged rules, with results turning mechanically on a limited number of facts. * * * Other rating systems rely more heavily on standards. The SafeSurf questionnaire, for example, requires the self-rater to determine whether nudity is "artistic" (levels 4 through 6), "erotic" (level 7), "pornographic" (level 8), or "explicit and crude" pornographic (level 9). * * *

A problem with standards is that they are less constraining; relatively speaking, a standards-based system will lack consistency and predictability. Rules become increasingly necessary as the universe of law-appliers becomes larger, less able to rely on shared culture and values as a guide to applying standards in a relatively consistent and coherent way. * * *

Let's return to the choices facing a ratings system designer as she constructs blocking software. * * * [W]hat sort of material should trigger ratings consequences? Should children have access to material about weapons making? How about hate speech? Or artistic depictions of nudity? Again, she can take two different approaches. First, she can decide all such questions herself, so that the home user need only turn the system on and all choices as to what is blocked are already made. CYBERSitter adopts this approach. This has the benefit of simplicity, but seems appropriate only if members of the target audience are in basic agreement with the rating service (and each other) respecting what sort of speech should and should not be blocked.

Alternatively, she can leave those questions for the user to answer. The ratings system designer need not decide whether to block Web sites featuring bomb-making recipes or hate speech. She can instead design the system so that the user has the power to block those sites if he chooses. Microsoft's implementation of the RSACi labels allows parents to select the levels of adult language, nudity, sex and violence that the browser will let through. Cyber Patrol allows parents to select which of the twelve CyberNOT categories to block.

Either approach, though, imposes restrictions on the categories chosen by the ratings system designer. If the system designer wishes to leave substantive choices to parents, she must create categories that correspond to the different sides of the relevant substantive questions. That is, if the designer wishes to leave users the choice whether to block sites featuring hate speech, she must break out sites featuring hate speech into a separate category or categories. If she wishes to leave the user the choice whether to block sites that depict explicit sexual behavior but

nonetheless have artistic value, she must categorize those sites differently from those that do not have artistic value. On the other hand, if the system designer makes those substantive decisions herself, making her own value choices as to what material should and should not be blocked, she must create categories that correspond to those value choices. The problem is that many of these questions cleave on lines defined by standards. Many users, for example, might like to block "pornography," but allow other, more worthy, speech, even if it is sexually explicit. * * * The problem increases with the heterogeneity of the service's audience: the more heterogeneous the audience, the more categories a rating system must include to accommodate different user preferences.

With this perspective, one can better appreciate the limitations of RSACi's attempt to be rule-bound. RSACi ignores much content that some other ratings systems classify as potentially unsuitable, including speech relating to drug use, alcohol, tobacco, gambling, scatology, computer hacking and software piracy, devil worship, religious cults, militant or extremist groups, weapon making, tattooing and body piercing, and speech "grossly deficient in civility or behavior." For many observers (myself included), RSACi's narrow scope is good news because it limits the ability to block access to controversial political speech. My point, though, is that RSACi had to confine its reach if it was to maintain its rule-bounded nature. * * *

In sum, rating system designers face a dilemma. If a rating service seeks to map the Web in a relatively comprehensive manner, it must rely on a relatively large group of evaluators. Such a group of evaluators can achieve fairness and consistency only if the ratings system uses simple, hard-edged categories relying on a few, easily ascertainable characteristics of each site. Such categories, though, will not categorize the Net along the lines that home users will find most useful, and will not empower those users to heed their own values in deciding what speech should and should not be blocked. To the extent that ratings system designers allow evaluators to consider more factors in a more situationally specific manner to capture the essence of each site, they will ensure inconsistency and hidden value choices as the system is applied.

III. Unrated Sites

Blocking software can work perfectly only if all sites are rated. Otherwise, the software must either exclude all unrated sites, barring innocuous speech, or allow unrated sites, letting in speech that the user would prefer to exclude. What are the prospects that a rating service will be able to label even a large percentage of the millions of pages on the Web? What are the consequences if it cannot?

First, consider rating services associated with individual manufacturers of blocking software, such as CYBERSitter and Cyber Patrol. These services hire raters to label the entire Web, site by site. The limits on their ability to do so are obvious. As the services get bigger, hiring more and more employees to rate sites, their consistency will degrade. * * * As a practical matter, providing access to all unrated sites is not an option for these rating services; it would let through too much for them to be able to market themselves as reliable screeners. Instead, they must offer users other options, dealing with unrated sites in one of two ways. First, they can seek to catch questionable content through string-recognition software. CYBERSitter, for example, offers this option. The problem with this approach, though, is that at least under current technology, string-recognition software simply doesn't work very well. This article has already mentioned America Online's travails with the town of Scunthorpe and the word "breast";⁷⁸ other examples are easy to find. Surfwatch, for example, blocked a page on the White House web site because its URL contains the forbidden word "couples" (<http://www.whitehouse.gov/WH/kids/html/couples.html>). The second option is for the rating services simply to block all unrated sites. Industry members seem to contemplate this as the necessary solution. Microsoft, for example, cautions Internet content providers that "if a rating system to be useful, the browser application must deny access to sites that are unrated."⁸⁰ Other observers reach the same result.

What about self-rating approaches, like those of SafeSurf and RSACi? These services have the potential for near-universal reach, since they can draw on the services of an effectively unlimited number of evaluators. * * * On the other hand, self-rating services will not achieve their potential unless content providers have a sufficient incentive to participate in the ratings process in the first place. That incentive is highly uneven. Mass-market commercial providers seeking to maximize their audience reach will participate in any significant self-rating system, so as not to be shut out of homes in which parents have configured their browsers to reject all unrated sites.⁸³ Many noncommercial site owners, though, may not participate. They may be indifferent to their under-18 visitors and may not wish to incur the costs of self-rating. It is still early to predict what those costs may be. For the owner of a large site containing many documents, supplying a rating for each page may be a time-consuming pain in the neck. * * *

The Internet is justly celebrated as "the most participatory form of mass speech yet developed." * * * But this prospect is threatened if widespread adoption of blocking software ends up removing much of the speech of ordinary citizens, leaving the viewer little to surf but mass-market commercial programming. One hardly needs the Internet for that; we get it already from the conventional media.

In sum, blocking software could end up blocking access to a significant amount of the individual, idiosyncratic speech that makes the Internet a unique medium of mass communication. Filtering software, touted as a speech-protective technology, may instead contribute to the flattening of speech on the Internet.

IV. Children, Adults, and Blocking Software

You may protest that I am making much of little here. After all, blocking software is intended to restrict children's access to questionable sites. It won't affect what adults can see on the Internet—or will it? It seems to me that, in important respects, it will. The desire to restrict children's access has spurred the recent development of filtering technology. Widespread adoption of that software, though, will not likely leave adults unaffected.

In a variety of contexts, we can expect to see adults reaching the Internet through approaches monitored by blocking software. In the home, parents may set up filters at levels appropriate for their children, and not disable them for their own use. * * * Other people get their Internet connections through libraries; indeed, some policymakers tout libraries and other community institutions as the most promising vehicle for ensuring universal access to the Internet. The American Library Association takes the position that libraries should provide unrestricted access to information resources; it characterizes the use of blocking programs as censorship. This policy, however, is not binding on member libraries. It is likely that a substantial number of public libraries will install blocking software on their public-access terminals, including terminals intended for use by adults; indeed, some have already done so. * * * Still other people get Internet access through their employers. Corporations too, wary of risk and wasted work time, may put stringent filters in place. Some large companies worry about the possibility of being cited for sexual harassment by virtue of material that came into the office via the Internet. Even more are concerned about sports and leisure information they feel may detract from business productivity. * * * In sum, we may see home computers blocked for reasons of convenience, library computers blocked for reasons of politics, and workplace computers blocked for reasons of profit. * * *

This should affect the way we think about filtering software. Any filtering system necessarily incorporates value judgments about the speech being blocked. These value judgments are not so controversial if we think of the typical user of blocking software as a parent restricting his children's access. It is part of a parent's job, after all, to make value judgments regarding his own child's upbringing. The value judgments are much more controversial, though, if we think of the typical "blockee" as an adult using a library computer, or using a corporate computer after hours. If we are concerned about these users' access to speech, then we need to think hard about the way blocking software works, the extent to which it can be accurate, and the extent to which it is likely to exclude the sort of speech that makes the Internet worthwhile.