Testimony of Dr. Lorrie Faith Cranor before the COPA Commission, 20 July 2000

http://www.research.att.com/~lorrie/

I have been asked to provide a technical overview of the technologies that are the subject of this hearing. I am going to describe to you a taxonomy[1] that I developed with my colleague Paul Resnick, which should provide a useful way for thinking about the different technologies that are available.

Technologies that promote safe and appropriate online experiences for children generally provide mechanisms for 1) identifying or describing content of a particular type, and 2) taking an action based on the type of content. A wide variety of techniques can be used to classify content, and a wide variety of actions can be taken based on the classification. But the common features that I think you will see in most, if not all, the technologies you hear about during this hearing, is that they all employ some classification technique, and some mechanism for taking an action based on that classification.

Classification

Regardless of what actions are taken, mechanisms are needed to identify content of a particular type. If we want to promote or restrict access to a particular kind of content, we must first figure out what that content is. How do we find the educational content to recommend or the inappropriate content to restrict? What criteria do we use to determine what should be recommended or restricted? And who or what actually does the work to identify each kind of content?

Who/how

Classification may be done by a variety of different parties:

- **Third-party experts** may be employed to label content. For example, many filtering companies use teams of information specialists, parents, and teachers to assist in classifying content.
- **Automated tools** may be employed to classify online content. Some of these tools are used to classify content dynamically, as the user requests it. Other tools are used to assist human classifiers in finding suspect sites.
- **Local administrators** such as parents or teachers may personally decide what content should be accessible to children under their supervision. Some tools allow the person who configures the software to provide their own lists of acceptable or unacceptable content by URL or by providing a list of key words or phrases to be searched for automatically.
- **Content providers** may rate or label themselves. For example, many adult content providers post notices on their sites stating that their content is not suitable for

children. In addition, the Platform for Internet Content Selection (PICS) was designed to support self-labeling (in addition to third-party labeling).

- **Surveys or votes** are often used to rate restaurants or movies. This technique has seen some limited use for rating online content.

Classification scheme

Classification schemes may be designed to identify content that is "good for kids" or content that is "bad for kids" or both. The content may be classified on the basis of

- its age suitability;
- specific characteristics or elements of the content, such as what language it is written in or whether it contains nudity or violence;
- or who created the content, such as a distinction between government and non-government sources.

Classification schemes can be designed to be fairly descriptive or very simple. A sophisticated rating system might have 20 variables that must be set, while a simple rating system might have a single "thumbs up" variable. Each system has its benefits. The sophisticated system provides more information, but requires more work to label content and to interpret the labels for each application. The simple rating system is quite easy to use, but conveys less information.

Besides having fewer variables, simple rating systems are also often more subjective. Rating systems may include both descriptive information and subjective opinions about, for example, the appropriateness of content. However, subjective systems can be problematic when users do not know if the bias inherent in the system matches their own. Also, from descriptive information one can always derive a new set of "subjective" opinions. If you are told about the content of a site in terms of violence, language, nudity, sex, and who paid to produce it, one can make a thumbs-up or thumbs-down decision. Given only someone else's thumbs down, however, one cannot recapture the descriptive information. Once opinions replace descriptions, information is lost.[2]

Scope

Internet content is provided through a variety of protocols including HTTP (Web sites), FTP, gopher, chat, telnet, instant messaging, and email. Some products and services focus on one or a small number of these protocols, while others provide more comprehensive solutions, monitoring everything a child does online. In addition, some products and services monitor only incoming communications, while others monitor both incoming and outgoing communications. Tools that monitor outgoing communications can often be configured to prevent children from giving out personal information that could be used to harm them such as their home address or phone number.

Actions

We have found tools that take six types of actions based on content labels or characteristics of online content: *suggest, search, inform, monitor, warn,* and *block.*

- **Suggest:** recommend appropriate content for children. A wide variety of Web sites, pamphlets, and books provide lists of child-appropriate content. In addition, some filtering software includes lists of suggested sites for children to explore.
- **Search:** select content that is appropriate for children and matches a query. Internet search engines allow people to enter a query and find all the indexed content that matches that query. Some search engines can be configured to filter the query matches and show the user only those matches that are appropriate for children. Other search mechanisms perform searches over databases that contain only sites deemed appropriate for children
- **Inform:** provide information about the content. Labels, reviews, and other descriptions of content can help parents and other supervisors guide children towards appropriate Internet content. However, in order for this information to be useful, it must be easily accessible. Some tools are designed to provide information about content when a user begins to access that content. This information may be displayed in the form of a graphic or banner on a Web page, or as part of the browser or other software.
- **Warn:** provide information about content and recommend against accessing that content before it is displayed. While mechanisms that inform provide information about content that is being viewed, warning mechanism indicate that content is not recommended, before the content is displayed. These tools can be useful for protecting against children accidentally downloading content that could be upsetting to them. Many adult Web sites include a prominent warning on an introductory page that content on other pages at the site is inappropriate for those under 18 (some include mechanisms for making sure those under 18 cannot access their content, but many rely on the warning as the only deterrent). A tool designed to block content that includes a password override could be used as a warning mechanism as well. Parents could provide a password that their children could use to access content that would otherwise be blocked. Thus children would be warned that the content may not be appropriate, but can proceed to access it anyway if they so desired.
- **Block:** prevent children from accessing content. A wide variety of tools prevent children from accessing inappropriate content. Some filter out specific Web sites that have been classified as inappropriate. Others filter out content that contains words or phrases that have been deemed inappropriate. And others filter out all content unless it appears on a "good for kids list."
- **Monitor:** record for later inspection a list of the content accessed or attempted to be accessed by a user; this may be a complete list or include only the content deemed inappropriate for children. Many filtering tools also include monitoring mechanisms that allow an adult "administrator" to review the log to determine what Web sites the child visited, what email the child sent, or what kinds of chats the child was participating in. Some filtering tools also log all attempts to access content in violation of the administrator's policy.

Mechanisms and Interface

Tools for selecting content can be implemented in a variety of ways, through a number of different technical mechanisms and with a wide variety of user interfaces. Some of the important differentiators between tools include where the tool is located, how it can be updated, and how customizable it is.

Location

The mechanisms that implement the actions described above may be located in a variety of places in a computer system including on the user's personal computer, on a local area network (LAN) or local proxy server, at an Internet Service Provider (ISP), on a remote proxy server, or as part of a search engine or Web site.

- **Personal computer:** Placing mechanisms on personal computers can facilitate their configuration and reconfiguration by parents, teachers, or other administrators. On the other hand, it may also facilitate the reconfiguration of these mechanisms by children, against their parents' wishes and possibly without their parents' knowledge. Some PC-based products have been designed with mechanisms to prevent tampering. Many PC-based products require frequent updates; some can update themselves automatically when the PC is connected to the Internet.
- **LAN or local proxy:** Placing mechanisms on LANs or implementing local proxy servers can be a useful solution in situations involving networked PCs, such as schools and libraries. Centralized configuration is easier for system administrators and harder for individuals to tamper with.
- **Internet Service Provider:** Some ISPs offer services designed especially for children. ISPs may provide filtered Internet access or restrict access to chat rooms, newsgroups, or other types of services.
- **Remote proxy:** Subscribers to remote proxy servers configure their browser software to pass all requests through the proxy server. Some of these services include mechanisms that prevent children from getting around the proxy server.
- **Search engine:** Some search engines return only pointers to content that is appropriate for children.
- **Web site:** A variety of Web sites list content appropriate for children.

Updates

As new content appears, it must be classified if tools that make use of classification information are to stay up to date (this is not an issue for tools that classify content on-the-fly.) Some products and services are continuously updated and include mechanisms for users to easily and quickly take advantages of updates. Others require users to manually download updates.

Customizability

Internet filtering products provide a large range of customization options including: mechanisms for customizing allow and block lists; specifying key words or phrases to

trigger actions; specifying categories of content to allow or block; and specifying whether inappropriate content should trigger a block, a warning message, a log entry, or other action. While highly customized products can often address a wide variety of customer needs, unless they are carefully designed they may be quite complicated to configure.

Other Features

A variety of other features are available, including the ability to limit the time of day or the amount of time children are online, provide separate settings and passwords for different children, and prevent children from accessing directories on the computer where their parents store important data. Some blocking tools can block individual words and images on a page, while others block whole pages or even entire sites. Some provide explanations about why they have filtered content, while others block silently. In addition, some tools have child-friendly user interfaces (and parent-friendly configuration procedures).

Discussion

I have outlined the range of tools that support parents' ability to choose online content appropriate for their children. When I first inventoried these tools in 1997, I found about three dozen tools that were available at that time. At last check, GetNetWise.org had found over 120 tools that are currently available. The proliferation of tools in this area has lead to increased innovation and the availability of tools to meet a wide variety of needs. Every community, and indeed every family, has their own standards for what types of content are appropriate for their children. Even within a family, different content may be deemed appropriate for children of different ages. These differences lead to a need for a variety of different tools. The type of computer a family owns, where it is located in the house, and how comfortable the parents are in using the computer may also impact the kinds of tools the family may choose to use. I think it is important that we continue to see a diversity of tools offered in the marketplace. In addition, I would like to see increased transparency from vendors about the criteria they use to classify content so that parents can more easily select the tool that best matches their family's values. This information should be easily obtainable from each vendor's web site as well as on the software packaging.

I hope this brief overview has proven valuable to the members of the Commission, and I would be pleased to answer any questions you might have.

---

[1] Lorrie Faith Cranor, Paul Resnick, and Danielle Gallo. Technology Inventory: A Catalog of Tools that Support Parents' Ability to Choose Online Content Appropriate for their Children. September 1998. http://www.research.att.com/projects/tech4kids/

[2] Lorrie Faith Cranor and Joseph Reagle Jr. Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences. In Jeffrey K. MacKie-Mason and David Waterman, eds., *Telephony, the Internet, and the Meda*. Mahwah: Lawrence Erlbaum Associates, 1998. http://www.research.att.com/~lorrie/pubs/dsp/