Statement of the Problem

Since the Internet came to the fore of public attention around 1994, Americans have been obsessed with the scourge of easily accessed online pornography, violence, and hate speech. Newspaper and magazine articles have fed this fear with titillating stories about pornographic web sites, hate groups, and online sexual predators (Turow, 1999). This perceived abundance of harmful material, has led Congress to pass two laws, the Communications Decency Act (CDA), and the Child Online Protection Act (COPA) aimed at criminalizing Internet content deemed harmful to minors. In conjunction with these legislative solutions, the software industry has developed its own technological solution, namely content filtering software such as Cyber Patrol, SurfWatch, Net Nanny, etc. Supporters of filtering software claim that such products are effective because they block access to the majority of "objectionable" Internet content, while also being First Amendment friendly because they do not block access to benign material that children should have access to.

The goal of this project is twofold, (1. to provide concrete evidence of just how much objectionable material there truly is on the World Wide Web, and (2. to test the effectiveness of Internet content filtering software, to see if it is indeed effective in blocking access to harmful material, while also not blocking access to benign content. Stated more formally:

> ***Research Question 1:*** *What is the percentage of objectionable content on the World Wide Web?*
>
> ***Research Question 2:*** *Are Internet content blocking software programs an effective solution to limiting access to objectionable Internet content, while also not blocking access to non-objectionable content?*

These questions are extremely relevant to the ongoing debate regarding what to do about dangerous Internet content. Legislators need evidence about

the actual extent of this supposed problem, so that they can craft effective and Constitutional legislation. Unfortunately, to date Congressional leaders have based their legislative proposals on methodologically weak assessments of this problem (discussed further below). It is hoped that this study, using a sound content analysis methodology, will improve upon this situation.

These results will similarly help parents, educators, and librarians asses the risk that the Internet may pose to children, and whether or not filtering software is an effective solution for protecting children from harmful Internet material.

Objectives of the Project

The goals of this project are to estimate the amount of objectionable material on the World Wide Web, and the effectiveness of software filters in blocking such content. As was just mentioned, these results will hopefully lead to a better informed debate regarding this contentious public policy issue.

A successful completion of this project will generate two sets of interrelated results:

1. Estimates regarding the percentage of objectionable material on the World Wide Web as a whole, and estimates of objectionable material encountered through the use of common web surfing techniques.

2. Measures of the effectiveness of multiple Internet content blocking software programs in blocking access to objectionable content, while also not blocking non-objectionable material.

Background and Related Research

The desire to protect children from "new" and "dangerous" forms of content has always served as a justification for censorship. As Wartella and Reeves note, "With the development of each modern means of storytelling -- books, newspapers, movies, radio, comics, and television -- social debates regarding their effects have recurred. A prominent theme in all these debates has been a concern with media's impact on youth (1985: 119)."

Children are justifiably presumed to be different from adults, "to be more vulnerable, less able to apply critical judgmental standards, more at risk (Roberts, 1996)." As far back as Greek antiquity, Plato's *Republic* warned of the influence of storytellers:

> Children cannot distinguish between what is allegory and what isn't, and opinions formed at that age are usually difficult to eradicate or change; it is therefore of the utmost importance that the first stories they hear shall aim at producing the right moral effect.

Speaking some two thousand years later about the dangers of film, the *Christian Century* (1930) commented:

> Movies are so occupied with crime and sex stuff and are so saturating the minds of children the world over with social sewage that they have become a menace to the mental and moral life of the coming generation. (cited in Starker, 1989, p. 8)

Based on our historical fear of all forms of media content, it should be of no surprise, that our newest medium, the Internet, has been similarly attacked as an "evil influence," poised to "contaminate the health and character of the nation's children (Starker, 1989, p. 5)." The moral danger inherent in the Internet has been defined as the availability of "cyberporn."

On July 3, 1995, *Time* magazine carried the following cover article: "On a Screen Near You: Cyberporn." The article, by *Time* senior writer Phillip Elmer-

Dewitt, cited the soon to be published research of a Carnegie Mellon undergraduate student named Marty Rimm (Wallace and Mangan, 1996). Rimm's study, eventually published in the *Georgetown Law Review*, claimed to be an exhaustive look at the amount and types of pornography available on the Internet. Rimm found that 83.5 percent of Usenet images were pornographic, and that over 70 percent of the sexual images on the newsgroups surveyed "originate from adult-oriented computer bulletin-board systems (BBS) whose operators are trying to lure customers to their private collections of X-rated material (Elmer-Dewitt, 1995)." Further, he found that many of the images analyzed were exceptionally kinky and violent.

Following the publication of the *Time* article and the actual Rimm study, many Internet pundits came forward to discredit Rimm's analysis. Taking the forefront in pointing out Rimm's weak methodology were Vanderbilt University professors Donna Hoffman and Tom Novak. They argued that Rimm's selection of a few sex related newsgroups was simply not representative of the world of Usenet, or of the larger Internet. "Also, no information is provided on the degree to which these 32 newsgroups comprise the complete universe of Usenet imagery (Hoffman and Novak, 1995)." In addition to his poor sample of newsgroups (of which there are thousands mostly relating to news, recreation, and politics), he also used no clear definition of pornography to classify images as pornographic.

The onslaught of articles, emails, and Usenet posts discrediting the Rimm study led *Time* to "admit that grievous errors had slipped past their editorial staff, as their normally thorough research succumbed to a combination of deadline pressure and exclusivity agreements that barred them from showing the unpublished study to possible critics (Wilkins, 1997)." For all intents and

purposes the Rimm study had been discredited, shown to be a methodologically weak study conducted by an attention seeking undergraduate student.

Despite this discreditation, the *Time* article and the Rimm study had stirred a wild moral panic about access to pornography on the Internet. In the months that followed, several mainstream newspapers and magazines including the *New York Times, USA Today*, and *Newsweek* ran stories regarding the "threat" of Internet content. Electronic Freedom Foundation (EFF) lawyer Mike Goodwin referred to this situation as "The Great Cyberporn Panic of 1995 (1998: 206)."

Not ones to miss out on a moral crusade, several U.S. Senators and Congressmen weighed in with legislation to protect children from the scourge of easily accessed Internet pornography. As Margaret Seif notes, "the political football got blown up to gigantic proportions (1997)."

Senator Charles Grassley (Republican from Iowa) proposed the Protection of Children from Pornography Act of 1995. In support of his bill, Grassley introduced the entire *Time* article into the Congressional Record, and referred to Rimm's undergraduate research as "a remarkable study conducted by researchers at Carnegie Mellon University (Wilkins, 1997)." Grassley further noted that "There is a flood of vile pornography, and we must act to stem this growing tide (June 26, 1995)." Grassley's bill did not pass, but it led to several Internet censorship bills, culminating in the Communications Decency Act (CDA) sponsored by former Senator James Exon. The bill was attached to the Telecommunications Reform Act of 1996, which was passed by Congress and signed into law by President Bill Clinton in February, 1996.

Two years after the passage of the CDA, Congress again returned to the issue of harmful Internet content. Legislators proposed the Child Online Protection Act (COPA), a bill similar to the earlier CDA, which would make it illegal to offer Internet material deemed "harmful to minors" to children under

18. In the report accompanying the bill, legislators again based the need for action on studies which claimed to show a flood of harmful material on the Internet (note: no information about the methodology behind these studies is provided in the Congressional report). As the report comments:

> As the Internet has grown, so has the availability of online pornography. In 1996, estimates that almost 50 percent of the content available on the web was unsuitable for children. Two years later, as of 1998, the estimates have increased to almost 70 percent of the traffic on the web is adult-oriented material . . . Consequently, the odds are no longer slim that a user will enter a sexually explicit site by accident. (Bliley, 1998)

To date, two studies have been put forward to repute the astonishingly high estimates of harmful Internet content cited by Congress. Computer scientists Lawrence and Giles (1999) recently published an assessment of search engine performance in indexing the World Wide Web. While their study was not focused on questions about content appropriateness, the researchers did informally code web sites in their sample for pornographic content. They found that only about 3 percent of the sites they sampled were devoted to pornography. The problem with this finding is that the researchers did not follow a systematic coding procedure, or test for reliability. Rather, they simply examined pages in their sample based on the unscientific "I know it when I see it" definition of pornography.

Attempting to improve upon this study, Zimmer and Hunter (1999) used the RSACi Internet content rating system (described in greater detail below) to rate a random sample of over 1000 web pages. They found that less than 5 percent of pages contained any objectionable material. The problem with this study is that the procedure used to achieve the random sample may not have

been powerful enough to take account of all web content. As a result, their findings can not be generalized to the entire universe of web pages.

The studies cited above, particularly the Rimm study, have serious methodological shortcomings. As a result, Congressional legislation may well be based on inaccurate estimates of just how much harmful content actually exists on the Internet. Using a strong content analysis methodology, including clear coding rules and definitions, measures of reliability, and the use of proper sampling techniques, it is hoped that this project will provide a more sound basis for public policy and Congressional debates.

Support for Filters

Over the past three years, courts have rejected both the CDA and COPA as unconstitutional restraints of First Amendment protected speech. In overturning these legislative solutions, the courts pointed to the supposedly "equally effective" but "less restrictive alternative" of Internet filtering software as the best way to keep the Internet a safe place for children (Volokh, 1997). As a result, filter technologies have been championed as *the* solution for keeping inappropriate content at the edge of cyberspace, and away from children. These self regulatory, market driven technologies are seen as First Amendment friendly, and far preferable to direct government regulation. No less than the White House has endorsed this idea, noting that "Advanced blocking and filtering technology is doing a far more effective job of shielding children from inappropriate material than could any law (Clinton, 1997)." In keeping with this statement, the White House has aggressively pushed the development and implementation of content blocking software. This push has only intensified in the wake of the Littleton, Colorado shooting tragedy.

In the days following the massacre, the news media uncovered the fact that the shooters frequently used the Internet to access Neo-Nazi and bomb making web sites. In the rush to blame something for the inexplicable killing spree, both the public and politicians cast a collective pointing finger at the Internet. A CNN/USA Today poll conducted shortly after the killings found that 64 percent of respondents said the net contributed to the tragedy (cited in McCullagh, 1999). Responding to this perceived problem, Congress and the White House drafted a flurry of new laws and proposals to curb access to "dangerous" Internet content. Several legislators are aggressively pushing the Childrens' Internet Protection Act (McCain, 1999) which will require all schools and libraries receiving federal funds for Internet access to install blocking software. Another proposed law would require any Internet Service Provider (ISP) with more than 50,000 subscribers to distribute content blocking software (Bloomberg, 1999). Similarly, the executive branch has fully endorsed filters. Speaking about Littleton at a recent conference, FCC chairman William Kennard noted "We need filtering software for families to use on their PC's. Just as you wouldn't send a child off alone in a big city, you wouldn't -- and shouldn't -- let them explore the vast landscape of the Internet without a chaperone (1999)." In a similar speech announcing a joint industry - White House "Parents Protection Site", Vice President Gore noted that filters were the best tool parents could use to protect children from the "free-fire zones and red light districts of cyberspace (1999)."

The Internet content industry has also thrown its support behind filter use. In September 1999, the Bertelsmann Foundation released a major self-regulation proposal which seeks to "protect children online as well as guarantee free speech (1999: 8)." To achieve this end, the proposal calls for the development of a voluntary international content rating and filtering system.

While the public, Congress, White House, and Internet industry may accept that content filters are the way to go, a number of scholars, civil libertarians, and journalists have asked whether these technologies are indeed the best solution to inappropriate Internet content. They point to the fact that content filtering software tends to block a great deal more speech than even government regulation would deem off limits. Further, blocking decisions can be based upon nearly any criteria, and are not open to public or institutional review. Finally, many filters do not even work as advertised, failing to block many objectionable web sites and thus giving parents a false sense of security. In short, Internet software filters championed as effective and First Amendment friendly, would seem to be anything but (Beeson and Hansen, 1997).

Are Filters Effective and First Amendment Friendly?

The majority of reports of Internet content filters being both underinclusive (failing to block the worst pornography, hate speech, violence, etc.), and overinclusive (blocking non-sexual, non-violent content), have come from journalists and anti-censorship groups who have used largely unscientific methods to arrive at the conclusion that filters are deeply flawed. A common method used by such groups has been to select a purposive sample of interesting sites and simply see if they are blocked or not by a particular filtering product. For example, the Censorware Project has used this method to expose unjustified blocking of benign web sites by Cyber Patrol (1998) and X-Stop (1998). Similarly, the Center for Media Education tested several filtering programs against a sample of 45 alcohol and tobacco related web sites. Their study found underinclusive performance, and concluded that "stand-alone filters do not effectively screen promotional alcohol and tobacco content (1999: 3)." While such studies are informative, they are limited to narrow areas of the web, and

generally suffer from a lack of methodological rigor. The goal of this project is to improve upon the above studies by applying social science methods of randomization and content analysis to examine the effectiveness of Internet software filters.

Proposed Methodology

The Data and the Target

The data for this project are individual, English language web pages, **not** web sites, drawn from several samples of World Wide Web content. The unit of analysis is also individual web pages. One could examine secondary pages that are linked to from a particular page, but this would create an enormous and unwieldy sample. As a result, coders will only look at individual pages and not follow links to subsidiary pages. The only exception to this rule will be on pages that have no other content than an "Enter this site" link, in which case the link will be followed, and the first fully developed page will be rated.

The target is the presence, absence, and degree of "objectionable material" on individual web pages. This will be measured using the RSACi rating system described below, which identifies four categories of potentially harmful content: language, nudity, sex, and violence.

Sampling Procedures

Internet users, including children, come across content through numerous surfing techniques. Perhaps the two most prevalent of which are serendipitous surfing from page to page, and the use of search engines to identify specific types of desired content. Users also access the Internet in different locations, for example the home, in school, and at public libraries. This study will seek to

reflect the range of people's surfing techniques, and surfing contexts by using four different samples of web content.

The first sample, roughly analogous to serendipitous surfing will be a set of 1000 randomly generated web pages. Obtaining a truly random sample from the universe of all web pages presents a somewhat difficult problem. For example, Zimmer and Hunter (1999) attempted to accomplish this by using a random link feature provided by the WebCrawler search engine. Although these sites were randomly provided by WebCrawler, this does not mean they were a random sample of all web content. Because of the web's vast size, currently estimated at some 800 million individual pages, even the most powerful search engine only indexes about 16 percent of the web's content (Lawrence and Giles, 1999). As such, the random sample produced by WebCrawler was only representative of the percentage of the web indexed by the search engine (about 50 million pages). Fortunately, Lawrence and Giles, who made the above estimation of the web's size, have come up with a better sampling technique which this study will adopt. This technique is the use of random Internet Protocol (IP) address sampling. Since all web sites have an IP address which corresponds with its URL (for example, the IP address 204.91.138.50 corresponds with the URL http://www.epic.org/), if you write a routine which randomly produces IP addresses, you can produce a sample that is truly representative of the entire web.

The second sample of web pages, roughly analogous to typical search engine use, is a set of 1000 popular search term results. To obtain a list of words to search with, we will turn to Searchterms.com, a site which tracks the most frequently searched for terms on popular search engines. We will take Searchterms.com's 100 most popular search terms and enter them into the AltaVista search engine. For each search result, we will only take the first ten

links generated by AltaVista, thus producing an overall sample of 1000 web pages. In terms of "objectionable" material, it is interesting to note that search engines have been singled out as one way that children are intentionally and unintentionally exposed to adult material. If children are actively seeking out pornography they need only enter the terms "sex" or "porn" into any search engine to receive thousands of links to such sites. However, search engines also expose children to adult material through the most innocuous of searches. For example, searching for the term "toys" on several major search engines produces links to sex toy stores and pornography web pages. Similar objectionable and unintended results are even produced by searches for current events topics such as "Monica Lewinski" or "Columbine." These results occur due to the way that search engines index content. Most search engines read a web page's "meta tag," a piece of HTML which describes the content of a particular page or site. Unfortunately, several less reputable site owners place keywords they know to be popular in their meta tags, even if such words bear no relevance to the content of their site. Therefore, pornography sites seeking increased traffic will include keywords such as toys, play, Columbine, etc. in their meta tags. Thus, when a child searches for toys, expecting to be transported to Toys-R-Us, he/she might be linked off to a pornography web site that included toys in its meta tag. As a result of these characteristics, rating search engine results for objectionable content will provide an interesting window onto the "danger" of using a common web surfing technique.

Our next sample is meant to be representative of what web sites students attempt to access when using the Internet at school. In the spring of 1999, the Annenberg Public Policy Center tracked the Internet use of students at a West Philadelphia high school. This resulted in a huge log file of what sites students visited over the course of a semester. We will use this data set to obtain a sample

of pages that students viewed over the course of one week during the spring semester. Only one week will be analyzed do to the tremendous number of sites accessed over the course of the entire semester. The data set will first have to be cleaned of identical site requests, and of sites which are no longer active. However, once cleaned, this data set should provide a unique way to gauge if students are accessing pornography and other types of inappropriate content as many parents and legislators fear. Useful guidance in generating a sample from this data set comes from the Censorware Project, who conducted a similar analysis on one month's worth of a Utah public school's student Internet use (Censorware, 1999).

Our final sample, meant to reflect what sites people visit at public libraries, will be obtained by approaching a large public library system and asking if they will volunteer their web access log files. While many libraries have expressed reservations about releasing log files, the American Library Association has volunteered to help connect me with libraries willing to share such information. As with the school log file, this data set will have to be cleaned before it can be used. Once cleaned, we will again rate one week's worth of web sites accessed.

Coding System, Training Coders, and Reliability Measures

To answer the question of how much objectionable material there is on the World Wide Web, and in our search term, school, and library samples we will use the Recreational Software Advisory Council's Internet rating system or RSACi. RSAC was originally developed by Stanford Communication professor, Donald F. Roberts, to rate the content of video games, and provide parents with a way to protect their children from excessive violence. However with the

advent of the Internet, the system was adapted to allow web site owners to self rate their content. Currently, RSACi is the most popular system for rating content on the Internet, with more than 100,000 web sites using it to self rate (RSAC, 1999).

RSACi contains four content categories (language, nudity, sex, and violence) each with five levels of severity (0, 1, 2, 3, 4). So for example, within the language category, a site may be rated 0 if it contains no objectionable language, 1 if it contains mild expletives, 2 if it has profanity, 3 with strong language, and 4 if it contains crude, vulgar language. Table 1 gives a summary of RSACi's rating categories. The full definitions for each category and level are provided in the Appendix.

Table 1: The RSACi Rating System

	Violence	Nudity	Sex	Language
Level 4	rape or wanton, gratuitous violence	provocative frontal nudity	explicit sexual acts or sex crimes	crude, vulgar language or extreme hate speech
Level 3	aggressive violence or death to humans	frontal nudity	non-explicit sexual acts	strong language or hate speech

Level 2	destruction of realistic objects	partial nudity	clothed sexual touching	moderate expletives or profanity
Level 1	injury to human beings	revealing attire	passionate kissing	mild expletives
Level 0	none of the above or sports related	none of the above	none of the above or innocent kissing; romance	none of the above

I would like to train 10 coders in the use of the RSACi rating system. I have used this system in two pilot studies, and found that coders quickly (usually in less than 20 minutes) understood how to apply RSACi. As a result, I am confident that the 10 coders (likely Penn undergrads) I recruit will quickly understand the system.

Reliability will be measured using a reproducibility, test-test design, as outlined by Krippendorff (2000). We will measure the reliability of each RSACi category individually, and the reliability of the system as a whole. In the two pilot studies mentioned above, very high alpha's (above .80) were achieved for each RSACi category and for the system as a whole. As such, in this study results will be considered reliable if alpha's of above .80 are achieved in all categories and for the system as a whole.

One reliability diagnostic that should be particularly useful in this study, will be the test of systematic disagreements. Because interpretations about what constitutes pornography, hate speech, extreme violence, etc. vary widely among conservatives and liberals, systematic biases may appear in the use of RSACi. Identifying such systematic disagreements will allow the data to be recoded to account for deviant coders (Krippendorff, 2000).

Filter Effectiveness Test

Once the web pages from our four samples have been reliably coded using the RSACi system, we can then combine these results with the performance of several popular filtering programs in blocking or not blocking these same web pages. We will test two sets of filtering software: client-based programs which are primarily used on individual computers by home users, and server-based programs, primarily used by schools and libraries, which monitor Internet access for large numbers of networked workstations. Below is a list of the client and server-based filtering programs to be tested:

Client-based Programs	**Server-based Programs**
CYBERsitter	Bess
Cyber Patrol	Cyber Patrol
Net Nanny	I-Gear
SurfWatch	Smart Filter
	X-Stop

Based on the assertions made by anti-censorship groups and journalists who claim that filters fail to block many "dangerous" sites, while conversely frequently blocking benign content, we will test the following hypotheses:

> ***Hypothesis 1:*** *Internet content blocking software will be underinclusive. They will fail to block access to sites with "objectionable material."*
>
> ***Hypothesis 2:*** *Internet content blocking software will be overinclusive. They will block access to sites with no "objectionable material."*

To test these hypotheses, RSACi rating decisions will be compared to the actual filter performance -- i.e. site blocked, site not blocked -- of the 9 programs mentioned above. A site will be considered blocked if the filter programs

completely deny access to it. Partial blocks, such as word masking will not be considered, as they still allow access to the majority of a web page.

Each of the client-based filter products will be purchased or downloaded, and all left with their default settings on. Default settings will be used due to the theory that few parents customize filter software. The only change made to these programs will be to download the latest blocked sites lists from each company.

The server-based programs will be tested in a manner that reflects the fact that many school and library systems attempt to "tweak" the programs blocking categories. For example, Cyber Patrol has 12 broad content categories which it filters for. Each of these categories can be turned on or off to reflect the concerns of system administrators. We will attempt to replicate this type of customization by testing each of the server-based filters at all levels of customization. For Cyber Patrol, this would mean conducting tests with all categories turned on, with the Violence category turned off, with the Sexual Education category turned off, and so on until all possible combinations are tested. This will result in reports that would say "with x categories enabled, Cyber Patrol blocks x-percent of sites, but with y categories enabled, the program only blocks y-percent of sites."

A site will be deemed to contain "objectionable" material if any of its content receives an RSACi rating of 2, 3, or 4. Such sites should theoretically be blocked by filter software. Conversely, a site will be deemed "not objectionable" if the highest score in all content categories is either 0 or 1. Such sites should theoretically not be blocked. For example, a site with an RSACi score of 0 - language, 4 - nudity, 3 - sex, and 1 - violence, would be deemed "objectionable" because its highest rating was a 4, and it should therefore be blocked.

Using these RSACi-based definitions of "objectionable - not objectionable" our inclusiveness hypotheses can be clarified. A filter will be deemed

underinclusive if it fails to block a site with a 2, 3, or 4 RSACi rating. A filter will be deemed overinclusive if it blocks a site with only a 0 or 1 as its highest RSACi rating. Tables 2 and 3 summarize these rules:

Table 2: Objectionable Definitions

	Highest RSACi score in any category
objectionable	2, 3, or 4
not objectionable	0 or 1

Table 3: Inclusiveness

	Filter Performance
underinclusive	fails to block "objectionable" content
overinclusive	blocks "not objectionable" content

Computations

This study will not use advanced statistical or algebraic techniques. Rather, simple frequencies and crosstabs will be sufficient to answer research questions regarding the percentage of objectionable Internet content, and filter performance.

Sample Results

Tables 4 and 5 present sample results from a pilot study which used the method described above (Hunter, 1999). Table 4 shows the frequency of objectionable material as measured by RSACi in a sample of 200 web pages.

Table 5 then compares those ratings to the combined performance of the client versions of CYBERsitter, Cyber Patrol, Net Nanny, and SurfWatch.

Table 4: All Samples Combined (N=200) Objectionable Content

	objectionable	**not objectionable**
language	27 (13.5%)	173 (86.5%)
nudity	19 (9.5%)	181 (90.5%)
sex	17 (8.5%)	183 (91.5%)
violence	7 (3.5%)	193 (96.5%)
any objectionable (all categories)	36 (18%)	164 (82%)

Table 5: All Filters Combined Over - Underinclusive

	not objectionable	**objectionable**	**total**
not blocked	129 (78.7%)	9 (25%)	138 (69%)
blocked	***35*** ***(21.3%)***	***27*** ***(75%)***	62 (31%)
total	164 (100%)	36 (100%)	200 (100%)

Analytical Constructs and Inferences to be Made

The underlying analytical construct in this study is society's overwhelming fear of new media content. As described in the literature review, this fear is a recurrent historical phenomena which often leads to calls for censorship. But are these fears grounded in reality? The results of our content analysis will help answer this question. While it would be difficult to say that society's fears are justified if "x percent" of content is objectionable, it is reasonable to compare the estimates that legislators are using with the results from this study. Thus, if Congress believes that as much as 70 percent of the Internet's content is unsuitable for children, but our study finds that less than 5 percent of content may meet this criteria, then we clearly have a problem. This is particularly the case if Congress passes a law which would limit access to some percentage of the Internet that is above the percent found using a methodologically strong content analysis of Internet material. In this scenario, Congressional legislation would be burning down the barn to roast the pig. Therefore, the prime inference that will be drawn from our proposed content analysis of web content, is whether or not the current policy climate reflects a realistic assessment of the amount of objectionable material on the Internet.

A similar construct is at work in the filter effectiveness study, where fear of new media content combined with faith in technology has lead to blind endorsements of filter software. Again, comparing this construct with our filter analysis will better allow us to asses whether these programs are a real solution for keeping children away from harmful material, while not limiting their access to valuable content. Baseline under and overinclusive percentages which would "prove" that a filter is effective, are difficult to determine. Nevertheless, our proposed filter analysis will give parents, educators, and librarians, at least some effectiveness results to base their filter purchasing decisions upon.

Time and Resource Needs

With the help of 10 undergraduate coders, I expect that it will take roughly six months to complete this project. All work will be conducted at the Annenberg School using Annenberg rooms and computing resources. As such, if accepted, this would be a largely "virtual" fellowship. However, once the primary study is completed, I would be free to finish the remainder of my fellowship in Cambridge, presenting my results to Berkman Center faculty and students. Below are project time and cost estimates:

Rough Time Line

January - February 2001

- Purchase, install, and test all filter programs
- Recruit and train undergraduate coders
- Develop an SPSS data file which will allow student coders to enter coding decisions directly into the computer
- Write the random IP address sampling program
- Obtain 1000 link sample using the random IP address program
- Obtain 1000 link search results sample

March - April 2001

- Content analyze the random IP address and search result samples
- Test all filters against random IP address and search result samples
- Obtain and clean school and library access log files
- Content analyze school and library access log files
- Test all filters against school and library access log files

May - June 2001

- Write up preliminary results for all samples and all filters
- Begin planning for a conference to present findings to the public

July - December 2001

- Present results, give lectures, etc. at the Berkman Center
- Complete and distribute a detailed project white paper
- Convene a one day "Filtering Summit" at the Berkman Center involving panelists from schools, libraries, and filter manufacturers

Cost Estimates

Client Filtering Software (for home use test)

Cyber Patrol	30
CYBERsitter	60
Net Nanny	30
SurfWatch	40
Total	**$160**

Server-based Filtering Software (for library and school use tests)

Note: All server-based filtering solutions are priced on a per user per terminal basis. As a result, higher or lower prices than listed below may be negotiated

depending upon the server configuration adopted for this project. The prices listed are general qoutes for a 50 or fewer user liscense.

Bess	1,200
Cyber Patrol	1,200
I-Gear	1,200
Smart Filter	2,000
X-Stop	500
Total	**$6,100**

Dedicated Project Server (the above server-based filtering programs will be installed on this machine)

Dell Workgroup Server with
Windows NT and Microsoft Proxy 2.0 **$3,700**

Archiving Software (to archive web pages from the 4 samples)

Teleport Pro **$40**

Student Labor (undergraduate coders)

1600 man hours at $7.00/hour
(10 coders working 20 hours
per week for two month) **$11,200**

Total Project Costs

Client Filtering Software	160
Server-based Filtering Software	6,100
Dedicated Project Server	3,700
Archiving Software	40
Student Labor	11,200
Total	**$21,200**

Critical Appraisal

While this proposed project is quite ambitious, I am highly confident that it can be completed and yield valuable new policy information regarding the contentious issue of harmful Internet content and the use of software filters. I base my confidence on the fact that I have conducted two smaller studies using the same methodology outlined above. Both studies took roughly one month to complete, with an additional month devoted to writing up the results. These smaller studies have been received favorably, with one to be published in an upcoming journal article (Hunter, 1999), and the other to be published as a book chapter (Zimmer and Hunter, 1999).

Combining this background experience with the resources of the Berkman Center and the Annenberg Public Policy Center, I believe that this research will produce credible and informative results.

The only intellectual reservation I have about this project, is that the RSACi system may not be capturing the full range of Internet content that people find objectionable. For example, the RSACi system has no categories for gambling web sites, or for alcohol, tobacco, and drug related sites. Many parents probably would not want their children seeing such material, but this would not be reflected in our use of the RSACi system. This would point to the development of a more inclusive rating system, that better reflects parents fears about Internet content.

References

Beeson, A. & Hansen, C. (1997). Fahrenheit 451.2: Is cyberspace burning? ACLU. Available via the World Wide Web at http://www.aclu.org/issues/cyber/burning.html .

Bertelsmann Foundation. (1999, September). Self-regulation of internet content. Available via the World Wide Web at http://www.stiftung.bertelsmann.de/internetcontent/english/download/Memorandum.pdf .

Bliley, T. (1998). Child Online Protection Act Congressional report. Available via the World Wide Web at http://www.eff.org/pub/Censorship/Internet_censorship_bills/1998_bills/s1482_hr3783_19981005.report .

Bloomberg News. (1999, 13 May). Senate unanimously passes filtering bill. Bloomberg News. Available via the World Wide Web at http://www.news.com/News/Item/0,4,36540,00.html .

Censorware. (1999). Censored internet access in Utah public schools and libraries. The Censorware Project. Available via the World Wide Web at http://censorware.org/reports/utah/main.shtml .

Censorware. (1998). Cyber Patrol and Deja News. The Censorware Project. Available via the World Wide Web at http://www.censorware.org/reports/dejanews.html .

Censorware. (1998). The X-Stop files: Deja voodo. The Censorware Project. Available via the World Wide Web at http://www.censorware.org/reports/x-stop.html .

Center for Media Education. (1999). Youth access to alcohol and tobacco web marketing: The Filtering and rating debate. Available via the World Wide Web at http://www.cme.org/ .

Clinton, W. (1997, 16 July). Remarks by the President at event on the e-chip for the internet. The White House Office of the Press Secretary. Available via the World Wide Web at http://www.whitehouse.gov/WH/News/Ratings/remarks.html .

Elmer-Dewitt, P. (1995, 3 July). On a screen near you: cyberporn. Time Magazine.

Goodwin, M. (1998). Cyber Rights. New York: Random House.

Gore, A. (1999, 5 May). Remarks on the Internet. White House Office of the Press Secretary. Available via the World Wide Web at http://www.whitehouse.gov/WH/New/html/19990505-4219.html .

Grassley, C. (1995, 26 June). Senate testimony. Congressional Record. S9017-S9023.

Hoffman, D. and Novak, T. (1995). Rimm commentary. Project 2000. Available via the World Wide Web at http://www2000.ogsm.vanderbilt.edu .

Hunter, C.D. (1999). Internet filter effectiveness: Testing over and underinclusive blocking decisions of four popular filters." Social Science Computer Review, 18 (2).

Kennard, W. (1999, May 4). Remarks of William Kennard at the Annenberg Public Policy Center conference on Internet and the family. Available via the World Wide Web at http://www.fcc.gov/Speeches/Kennard/spwek916.html .

Krippendorff, K. (2000). Content analysis : An introduction to its methodology. Beverly Hills : Sage.

Lawrence, S. and Giles, C.L. (1999, July). Accessibility of information on the web. Nature.

McCain, J. (1999). Childrens' Internet protection act. Senate Bill 97. Available via the World Wide Web at http://www.thomas.loc.gov/ .

McCullagh, D. (1999, 23 April). Looking for something to blame. Wired News. Available via the World Wide Web at http://www.wired.com/.

Recreational Software Advisory Council. (1999). About RSACi. Available via the World Wide Web at http://www.rsac.org/ .

Roberts, D. (1996, 28 August). Media content rating systems. The Wally Langenschmidt Memorial Lecture. Available via the World Wide Web at http://www.rsac.org/ .

Seif, M. (1997). Symposium comments cited in B.U. Journal of Science and Technology Law, Vol 3.

Starker, S. (1989). Evil Influences. New Brunswick, NJ: Transaction Publishers.

Turow, J. (1999, May). The Internet and the family: The View from parents the view from the press. Annenberg Public Policy Center of the University of Pennsylvania, Report No. 27.

Wallace, J. and Mangan, M. (1996). Sex, Laws, and Cyberspace. New York, NY: Henry Holt & Company.

Wartella, E. & Reeves, B. (1985). Historical trends in research on children and the media: 1900-1960. Journal of Communication, 35 (2), 118-133.

Wilkins, J. (1997, 17 September). Protecting our children from Internet smut: moral duty or moral panic? The Humanist, 57, (5), 4.

Volokh, E. (1997). Freedom of speech, shielding children, and transcending balancing. Supreme Court Review, 31.

Zimmer, E. and Hunter, C.D. (1999). Risk and the internet: Perception and reality. Presented at the Citizens at the Crossroads Conference, New London Ontario, October 1999.

Appendix: RSACi Content Categories and Definitions

RSACi Rating Language

In order to determine the appropriate advisory level for language, you will be asked to go through a checklist of very specific terms to determine whether or not your content contains language, expressions, images, portrayals, etc., which some viewers might potentially consider objectionable. The RSACi rating addresses two kinds of speech; 'hate speech' and 'objectionable speech'; that is, language ranging from mild expletives or profanity to crude, vulgar, and

obscene statements and gestures. You are urged to review the Definitions before submitting your answer.

Moving through the list below in order from top to bottom, please click the first button of the content descriptor that applies to your content. Does your content portray:

(4) crude, vulgar language
(4) explicit sexual references
(4) extreme hate speech
(4) epithets that advocate violence or harm against a person or group
(3) strong language
(3) obscene gestures
(3) hate speech or strong epithets against any person or group
(2) profanity
(2) moderate expletives
(1) non-sexual anatomical reference
(1) mild expletives
(1) mild terms for body functions
(1) slang
(0) none of the above

Definitions for RSACi Language Questions

GUIDING PRINCIPLE
The construction of a list of every word, action, innuendo, and gesture that a reasonable person would consider as crude, slang, profane or explicit is a never-ending task. Times change. Words change. Gestures change. New street slang is constantly evolving. Language considered inoffensive in one culture may be considered vulgar in another culture. It is therefore your responsibility to properly interpret and classify any slang, profanity or vulgarity according to the usage in the title and the general category definitions below. Words or expressions in the title that fit a definition or categorization, but do not appear on a word list, should be treated as if they do appear on the list.

CONTAIN
The inclusion of specific content in any form or manner, including but not limited to printed words, written descriptions, oral recitations, and other audio sounds.

CRUDE LANGUAGE; EXPLICIT SEXUAL REFERENCES

Crude references, direct or indirect to intercourse: Fuck, bugger, mother-fucker, cock-sucker, penis-breath. Crude references to genitalia: prick, cock, pussy, twat, cunt. Explicit street slang for intercourse or genitalia.

EXTREME HATE SPEECH
The combination of vulgar language with hate speech or epithets; advocating violence or harm against a person or group.

HATE SPEECH
Any portrayal (words, speech, pictures, etc.) which strongly denigrates, defames, or otherwise devalues a person or group on the basis of race, ethnicity, religion, nationality, gender, sexual orientation, or disability is considered to be hate speech. Any use of an epithet is considered hate
speech. Any description of one of these groups or group members that uses strong language, crude language, explicit sexual references, or obscene gestures is considered hate speech.

EPITHET
A disparaging or abusive word or phrase used in the place of the name of any person or group. There are many examples of slang terms which, in any given historical period, function almost exclusively as epithets: e.g., honky, nigger, coon, spic, greaser, chink, slant, faggot, etc. In addition, sometimes a word which is not in and of itself an epithet functions as one because of context. For example, in some contexts the word "pig" may be used in place of "police officer," thus becoming an epithet. In other contexts, and at different times, the word "monkey" has been used as an epithet to refer to individuals of Asian descent and to individuals of African descent.

OBSCENE GESTURES
Any visual or described gestures, body movements, such as flipping the bird, mooning, non-verbal indications of sexual insult, etc., indicating any of the above. Any visual or described innuendo, euphemisms, street slang, double-entendre for any of the above.

STRONG LANGUAGE
Strong, but not crude, language for genitalia: asshole, butthole, dork, dong, pecker, schlong, dick. Strong language for bodily functions or elimination: Shit, piss, cum, asswipe, buttwipe. Strong language for sexual functions or intercourse: jerk-off, balling, shtupping, screwing, boffing, cumming. References to genitalia used in a sexual setting including the use of penis, vagina, rectum, semen.

PROFANITY

To treat something regarded as sacred with abuse, irreverence, or contempt. To use the name of a deity with contempt or as a curse.

MODERATE EXPLETIVES
The words bastard and bitch (when used as epithets rather than biological terms), son-of-a-bitch, turd, crap.

MILD EXPLETIVES
The words hell and damn, ass and horse's ass, BUT NOT asshole, assface, asswipe; butthead and buttface BUT NOT butthole and buttwipe.

NON-SEXUAL ANATOMICAL REFERENCES
Words such as penis, vagina, rectum, semen used in a non-sexual context.

MILD TERMS FOR BODY FUNCTIONS
Words such as piss and poop not used in a sexual context.

SLANG
No profanity, expletives, vulgar gestures, innuendo, double-entendre, vulgar street slang other than listed below.

A. Inoffensive slang: darn, drat, golly, gosh, dang, rats, sheesh, geeze, gee wiz.
B. Screw to indicate cheated or harmed, BUT NOT screw in any sexual context.
C. Butt to indicate one's rear end as in "get your butt out of here, or "I'm going to paddle your butt," or "he fell on his butt.," BUT NOT butthead, butthole, buttface, buttwipe, etc.
D. Ass when referring to the animal, but not "Horse's ass."
E. Dork used in a non-sexual context as in, "He's a dork."
F. Sucks used in a non-sexual contest as in, "That sucks," or "He sucks."

RSACi Rating Nudity

In order to determine the level of nudity, if any, in your content, please go through the checklist of very specific terms about how nudity is portrayed. Definitions are provided for all terms that must be understood to make the

determinations necessary to answer the questions. The definitions are highly specific and the objectivity of the labeling system depends on using them correctly.

Moving through the list below in order from top to bottom, please click the first button of the content descriptor that applies to your content. Does your content portray:

(4) frontal nudity that qualifies as a provocative display of nudity
(3) frontal nudity
(2) partial nudity
(1) revealing attire
(0) none of the above

Definitions for RSACi Nudity Questions

PORTRAYAL
Any presentation including, but not limited to, pictures, no matter how crudely drawn or depicted, written descriptions, oral recitations and or audio sounds.

HUMAN or HUMAN-LIKE BEINGS
Any sentient being, no matter how portrayed (photographed or drawn) or how crudely drawn (including stick figures) understood by any reasonable person as human or humanoids in form including alien sentient beings that have human-like form (head AND arms AND torso AND legs AND walking upright).

NUDITY
Any portrayal of a human's buttocks (other than the exception below), genitalia, or female breasts, or of humanoid genitalia or female breast(s), including such portrayals as see-through blouses, the pasties of a topless dancer, or other types of clothing which do not prevent exposure of those parts of the body. This definition also includes nudity in widely recognized works of art and nudity in documentary context. NOTE: An exception is made for portrayals of the buttocks of characters which a reasonable person would consider as BOTH (a) something other than a true human being or representation thereof, AND (b) a character that normally is expected to be unclothed and whose natural state is undressed. If the portrayal is such that it would not cause a reasonable person to comment upon or take notice of the exposed buttocks, then, for this one exception, the characters require no rating for nudity.

PROVOCATIVE DISPLAY OF FRONTAL NUDITY

Any portrayal of genitalia that might reasonably imply sexual arousal, or the display of frontal nudity in what might be reasonably considered a sexual context.

FRONTAL NUDITY
Any portrayal of a nude sentient being which shows pubic hair or genitalia, excluding known animals in their natural state of undress.

PARTIAL NUDITY
Partial nudity is a subset of nudity. Any portrayal of a human buttocks or female breasts, or of humanoid female breast(s), including such portrayals as see-through blouses and other types of clothing which do not prevent exposure of the body and portrayals with minimal covering, such as pasties on the breasts of a topless dancer. In the case of non-humans, portraying buttocks does not constitute partial nudity IF AND ONLY IF one can surmise that the creature is natural state is undressed.

REVEALING ATTIRE
Any portrayal of a human/humanoid that does not portray nudity, yet portrays outlines through tight clothing, or clothing that otherwise emphasizes male or female genitalia, female nipples or breasts (including the display of cleavage that is more than one half of the possible length of such cleavage), or clothing on a male or female which a reasonable person would consider to be sexually suggestive and alluring.

RSACi Rating Sex

In order to determine the level of sexual activity, if any, in your content, you will be asked to go through a checklist of very specific terms about how sex is portrayed. Definitions are provided for all terms that must be understood to make the determinations necessary to answer the questions. The definitions are highly specific and the objectivity of the labeling system depends on using them correctly. You are urged to review the definitions before submitting your answer.

Moving through the list below in order from top to bottom, please click the first button of the content descriptor that applies to your content. Does your content portray:

(4) sex crimes
(4) explicit sexual acts
(3) non-explicit sexual acts
(2) non-explicit sexual touching
(2) clothed sexual touching
(1) passionate kissing
(0) innocent kissing or romance
(0) none of the above

Definitions for RSACi Sex Questions

PORTRAYAL
Any presentation including, but not limited to, pictures, no matter how crudely drawn or depicted, written descriptions, oral recitations, and or audio sounds.

HUMAN or HUMAN-LIKE BEINGS
Any sentient being, no matter how portrayed (photographed or drawn) or how crudely drawn (including stick figures) understood by any reasonable person as human or humanoids in form including alien sentient beings that have human-like form (head AND arms AND torso AND legs AND walking upright).

SEX CRIMES
Any portrayal of unwanted, unauthorized, or otherwise non-consensual sexual acts forced upon one sentient being by another sentient being (rape).
Any portrayal of explicit or non-explicit sexual acts, consensual or not, between a human or human-like being that a reasonable person would consider as being under the age of 18, and another human or human-like being that a reasonable person would consider over the age of 18. Any portrayal of sex, consensual or not, between an animal and a human or human-like being (bestiality).

EXPLICIT SEXUAL ACTS
Any portrayal of sexual activity that a reasonable person would consider as more than just non-explicit sexual activity because it shows genitalia. This includes any portrayal of sexual activity by one human or human-like being, or among multiple humans, including, but not limited to masturbation and sexual intercourse of any kind (oral, anal vaginal), that shows genitalia.

NON-EXPLICIT SEXUAL ACTS
Any portrayal of sexual activity that a reasonable person would consider as more than just clothed sexual touching or non-explicit sexual touching, either by one human or human-like being or among multiple humans, including, but not limited to masturbation and sexual intercourse of any kind (oral, anal, vaginal), that may show nudity, but does not show genitalia. Non-explicit sexual activity includes sound on an audio track, such as the kinds of groans, moans, and other sounds that to a reasonable person would imply sexual activity was taking place.

NON-EXPLICIT SEXUAL TOUCHING
Any portrayal of any touching between or among humans or human-like beings, that a reasonable person would consider more than just passionate kissing, including but not limited to such things as groping, petting, licking, and rubbing, that falls short of intercourse (sexual, oral, or otherwise), and that does show bare buttocks or female breasts, but does NOT show genitalia. Non-explicit sexual touching does NOT include non-explicit or explicit sexual acts as defined above and does NOT include masturbation.

CLOTHED SEXUAL TOUCHING
Any portrayal of any activity or touching between or among humans or human-like beings, other than innocent kissing and passionate kissing, that falls short of intercourse (sexual, oral, or otherwise) or masturbation, and that does NOT show bare buttocks, female breasts, or genitalia, but that any reasonable adult would perceive as sexual in nature. This includes but is not limited to such things as groping, petting, licking, rubbing. Non-explicit sexual touching does NOT include non-explicit or explicit sexual acts as defined below and does NOT include masturbation.

PASSIONATE KISSING
Any portrayal of humans or human-like creatures kissing that a reasonable person would consider more than just innocent kissing. This includes any kissing during which tongues touch (or mouths are obviously open), and any kissing on, but not limited to, the neck, torso, breasts, buttocks, legs.

INNOCENT KISSING
Any portrayal of humans or human-like creatures which a reasonable person would consider as just kissing on lips (without touching of tongues), head, shoulder, hands or arms, but not any other areas including but not limited to neck, breasts, torso, or legs. Innocent kissing shows affection and/or love, but creates no reasonable perception of stronger sexual activity as defined in this methodology.

ROMANCE
Portrayals of activity showing love and affection with NO stronger sexual contact as defined in this methodology. This might include embraces, hugging, innocent kissing, holding hands, etc.

RSACi Rating Violence

In order to determine the level and type of violence, if any, in your content, you will be asked to go through a checklist of very specific terms about whether and how violence or its consequences are depicted. Definitions are provided for all terms which you need to understand in order to make the determinations necessary to answer the questions. The definitions are highly specific, and the objectivity of the system depends on using them carefully and correctly. You are urged to review the Definitions before submitting your answer.

Moving through the list below in order from top to bottom, please click the first button of the content descriptor that applies to your content. Does your content portray:

(4) wanton, gratuitous violence
(4) extreme blood and gore
(4) rape
(3) blood and gore
(3) intentional aggressive violence
(3) death to human beings
(2) the destruction of realistic objects with an implied social presence
(1) injury to human beings
(1) the death of non-human beings resulting from natural acts or accidents
(1) damage to or disappearance of realistic objects?
(0) sports violence
(0) none of the above

Definitions for RSACi Violence Questions

PORTRAYAL
Any presentation including, but not limited to pictures, no matter how crudely drawn or depicted, written descriptions, and/or oral recitations, and/or audio sounds.

THREATENING
The portrayal of the intention to inflict harm, injury, evil on another being. Something that a reasonable person would consider to be menacing to another's safety or well-being.

WANTON, GRATUITOUS VIOLENCE

The visual portrayal of the continuation of intentional aggressive violence that causes damage/harm/death to any sentient being once that being has been rendered helpless and/or non-threatening, such as physical torture, continued attacks on or damage to corpses, dismembering or eating a corpse.
EXTREME BLOOD/GORE:
The visual portrayal of living beings being torn apart or slaughtered, dismembered body parts.

RAPE
The portrayal (video, audio, or written) of any unwanted/unauthorized, non-consensual sexual intercourse, whether vaginal, anal, oral, or fondling, forced upon a sentient being by another sentient being(s). In any sexual or sexually suggestive interaction, "No" is assumed to mean "No."

BLOOD / GORE
The visual portrayal of blood splashing, pools of blood on the ground, objects or persons smeared or stained with blood.

INTENTIONAL AGGRESSIVE VIOLENCE
The existence of a threat or the actual carrying out of threatening actions that directly or indirectly cause, or if successful would cause, physical harm, damage, destruction, or injury to a sentient being or realistic object. This includes the visual portrayal of the results of aggressive violence including, but not limited to dead bodies, damage, audio distress, etc., even if the violent act itself is not shown. It does not include psychological attacks. but is limited to physical harm, damage, destruction, and injury. possible to have a credible threat which does not cause a change in behavior.

IMPLIED SOCIAL PRESENCE
The presumption, unless a reasonable person would clearly think otherwise, that a realistic object is inhabited, or carrying, or concealing humans, even though the humans have not been seen or heard.

SPORTS VIOLENCE
Competitive sports games such as football, basketball, car racing, sumo wrestling, etc. have may elements of violence but are not intentional aggressive violence. It is still sports violence if players or participants are shown carried off the field, conscious or unconscious, even though on a stretcher, unless there is death, dismemberment, or blood and gore involved.
Note: Sports violence does NOT include wrestling, boxing, street fighting, karate, etc. games if the intended goal is to hurt or render the opponent unable to function. These actions are considered as intentional aggressive violence. A fight within a sports game, such as during a hockey game, would also be considered intentional aggressive violence. Definitions for Violence

Rating: PORTRAYAL: Any presentation including, but not limited to pictures, no matter how crudely drawn or depicted, written descriptions, and/or oral recitations, and/or audio sounds.