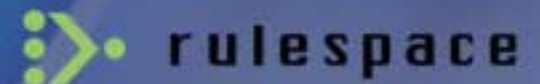


# Please Unblock Playboy.com

Technical Issues  
Behind the  
Responsible  
Deployment of  
Content Filtering

Search Engine Meeting  
Dan Lulich, CTO  
April 2002



# *Agenda*


- **Introduction to RuleSpace**
- **Framing the debate**
- **Does content filtering work?**
- **RuleSpace data from the field**
- **Second generation solutions**
- **Issues for responsible deployment**

# *Introduction to RuleSpace*

- **Technology vendor powering Parental Controls and Enterprise Filtering Applications**
  - Patented automatic filtering solution based on Machine Learning
  - Our product is deployed in several ways
    - Online Lookup Service and/or Embedded Software Toolkit
- **Emerging as a leader in this business**
  - We have been selected by many of you here
    - Most successful ISPs, Portals, and Search Engines
  - Winner of 2001 Technology Excellence Award, PC Magazine



# Framing the Debate



**WIRED NEWS** BUSINESS POLITICS WIRE SERVICE  
CULTURE TECHNOLOGY TOP STORIES

LOOK FOR  Wired News  [Get Wired News Your Way](#)  
Newsletters, handheld versions, alerts ...

**Inc** Subscribe now & get your Inc bag & pen set FREE!

**POLITICS** **Judges Blast Library Filtering**  
By Declan McCullagh

Today's Headlines  
10:55 a.m. April 12, 2002 PDT [Print this](#) + [E-mail it](#)

[The Human Clone: What Gives?](#)  
[Report: MS Fees Bribe Attorneys](#)  
[Judges Blast Library](#)

4:31 p.m. April 4, 2002 PST  
WASHINGTON -- The U.S. Congress' third assault on Internet pornography appears likely to meet the same ignoble end as the previous two.

A two-week trial over library filtering ended Thursday with a trio of judges criticizing the Children's Internet Protection Act ([CIPA](#)) as an unreasonable intrusion into the rights of Americans to view legal material online.



**techlive**

Search:  [Site Map](#)

You are here: [Home](#) > [Tech News](#) > [Politics And Law](#) > **Experts Say Porn Filters Don't Work**

- HOME
- TV SHOWS
- TECH NEWS
- PRODUCTS & REVIEWS
- HELP & HOW TO
- ENTERTAINMENT
- Interact
- Video
- Shopping

**Tech Live: Tech News and Information**  
Weekdays at 9 a.m., 4 p.m., and 8:30 p.m. Eastern

**Experts Say Porn Filters Don't Work**  
**CIPA trial witnesses say filters block info sites along with smut.**

By David Morgan, Reuters  
March 26, 2002

## *Framing the Debate*

- **The Children's Internet Protection Act (CIPA) and the Neighborhood Internet Protection Act (NCIPA) went into effect on April 20, 2001**
- **The laws place restrictions on the use of funding that is available through the Library Services and Technology Act, Title III of the Elementary and Secondary Education Act, and on the Universal Service discount program known as the E-rate (Public Law 106-554)**
- **These restrictions take the form of requirements for Internet safety policies and technology which blocks or filters certain material from being accessed through the Internet**



# *Does Content Filtering Work?*

- Nunberg, Geoffery, [The Internet Filter Farce](#), *The American Prospect*, January 1-15, 2001
- **Filtering architectures are fundamentally flawed**
  - Categories are too broad
  - Filters don't come close to human language capabilities
  - Filters do not possess "Adult Judgment"
  - Filters can not handle non-textual data? Images, Java Script etc.
- **Web is not a place that lends itself to police sweeps**
  - Sites change frequently
  - Enforcement is hard
    - Owners are hard to find
    - Pornographers roam outside of US legal jurisdiction
  - Pornographers game the systems
  - Policing inhibits young people from seeking useful information

# *Does Content Filtering Work?*

- **Most vendors deploy Control Lists**
  - Overblocking and underblocking is rampant
  - The Web is too big. Impossible to rate it all.
  - The Web changes too fast. 2M new pages everyday
- **A few vendors dynamically filter with Keywords**
  - Overblocking is inevitable
  - Really bad at Violence detection
- **Green Spaces**
  - Restrict to much valuable information
- **Vendors are insensitive to privacy issues**
  - Keep logs and facilitate automatic notification
  - Attack those with an different agendas
    - Block anonymizers and critics
    - Don't make control lists available to the public
- **Vendors misrepresent performance**
  - Measuring performance based on user traffic is misleading
  - Randomly chosen addresses is only way to measure

# *RuleSpace Data from the Field*

- On average we have ~1B hits per day on our deployed servers. This translates to approximately 100M pageviews per day.
- We receive on average 120 customer override requests per day
  - Between 20-30 are URL ratings that need correction
  - The rest violate our *customers policies*
    - Every single day someone asks us to unblock playboy.com
- **Since our service has been live, we have seen the parental controls opt-in rate double**
  - This is a strong testament to the performance of our service. If users don't like it they can opt-out at anytime. They appear to be keeping it turned on.

# *Observations from the field*

- **Parental Controls has become a “Must Feature” for ISPs**
  - 2000: We don't filter and proud of it
  - 2001: Download a third party client side app from us
  - 2002: We want to build it in and offer it as a service
- **Customers accuracy expectations start at 100% and begrudgingly moves down from there**
  - Parental Controls customers
    - Happy with broad definitions
    - Demand perfect sensitivity
    - Will always trade overblocking for underblocking
  - Enterprise Customers
    - A narrow category definitions
    - Demand perfect Specificity
    - Will never trade a overblock for underblock
- **Most customers seem to only care about Pornography**
  - Yet they don't want to be embarrassed by not having coverage in other categories
  - They don't criticize our performance on Alcohol or Tobacco for example

# *The Reporter Test - Sensitivity*

RuleSpace Proprietary

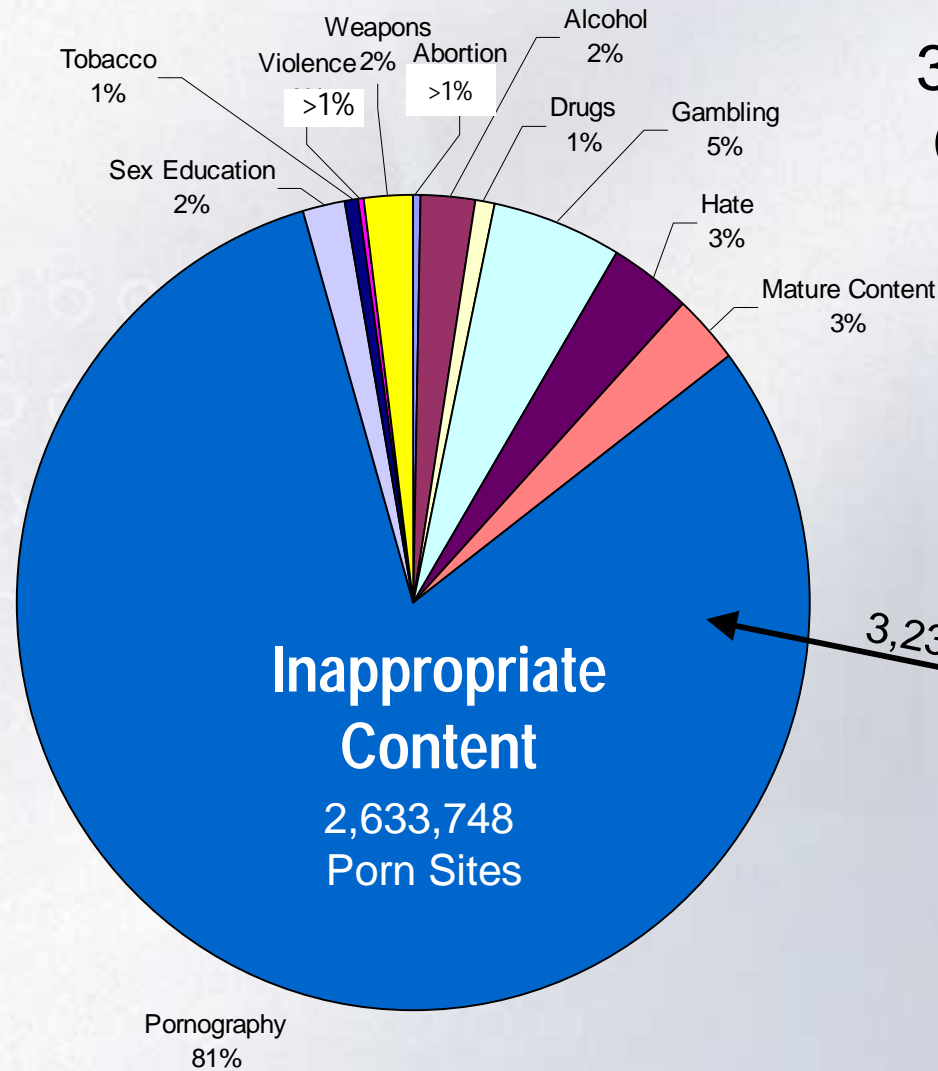
## *Control Lists – “First Generation”*

- **Huberman – The Laws of the Web – Power Laws everywhere you look**
- **Top 50 properties get 90% of the traffic – MediaMetrix**
- **If you have a control list that rates these properties correctly you should do well on the average. Easy to do with human editors**
- **It’s not so simple. It turns out these sites are the gateways to the web and generate dynamic content. The portals, search engines, and news services routinely serve up content inappropriate for kids.**
- **You have to rate it all dynamically.**

# *RuleSpace “Second Generation”*

RuleSpace Proprietary

# We Sweep the Entire Web Every 60 Days



# *The Training Problem*

RuleSpace Proprietary

# *Capturing the Context*

- **Web Site Analysis**
  - Infers the category based on website organization
- **Link Analysis**
  - Induces category from links to previously categorized pages
- **URL Analysis**
  - Scans the URL itself using learning machine technology
- **Traffic Analysis**
  - Category chosen by “suggestions from” and “behavior of” users
- **Web Authorities**
  - Asks other web experts how to assign the category label
- **Human Analysis**
  - Hand grading of web content into categories

# *Context Sensitive and Adaptive*

RuleSpace Proprietary

# *Categorization Affects Policy*

- **Categorization implies policy:** at its deepest level, a category is a policy in itself
- **Categories create policies:** defining a category establishes the depth and breadth of a policy.
- **Categories inform policies:** a clearly defined category has an innate authority which undifferentiated content lacks; that authority colors its perception and use

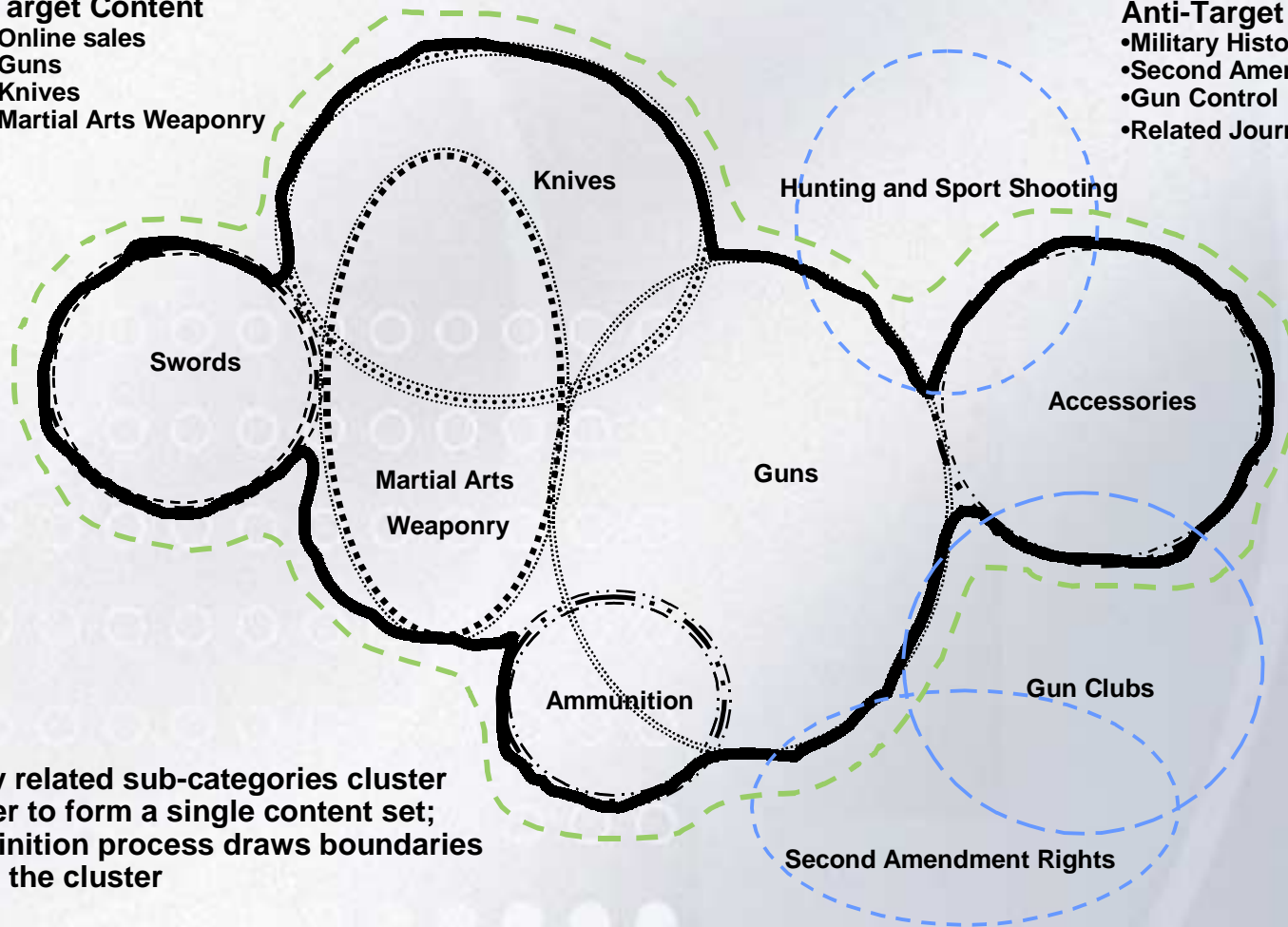
# Defining Categories is Messy

## Target Content

- Online sales
- Guns
- Knives
- Martial Arts Weaponry

## Anti-Target

- Military History
- Second Amendment Debates
- Gun Control
- Related Journalism



# *Policy and Responsibility*

- What does responsible access really mean, and how should it be achieved?
  - Does a filtering company stand *in loco parentis* for purposes of parental controls applications?
  - What is Possible vs. Profitable
  - It is clear that the needs of users are infinite and individual
- **Bring categorization policies to the surface**; too often the policies created by the behavior of a categorization or filtering product are concealed behind the rhetoric of proprietary information
- Categories change over time; **managing drift** is constant, not occasional: if left to themselves categories will begin to conflict with themselves and the policies they create will become unmanageable
- More users = more complex needs; **system must be scalable and flexible**

## *Unblock Playboy.com? What's the Context?*

- I need the 1976 Jimmy Carter interview for my school project
- Playboy.com is our business partner
- I need to check our ad placement
- In our country, female nudity of the type found on Playboy.com is not considered pornographic
- I see more raunchy sex on HBO than I do on Playboy
- I am an investor in Playboy Enterprises (PLA - traded on NYSE). I need to check on my investment

# *Looking Forward*

- **We do not take full advantage of all of the available context**
- **We pay too little attention to semantics; our shallow methods are at their limits for some categories like “hate mongering”**
- **Embrace Change, Close the Loop, and Learn**
  - Monitor categories
  - Monitor users but respect their privacy
  - Monitor the system; surprising things will come out
  - If it's not broken, fix it anyway: it will be broken sooner or later.



**rulespace**