

# Computers and Media: Privacy

---

CSCI 1200  
JAREK SZLICHTA

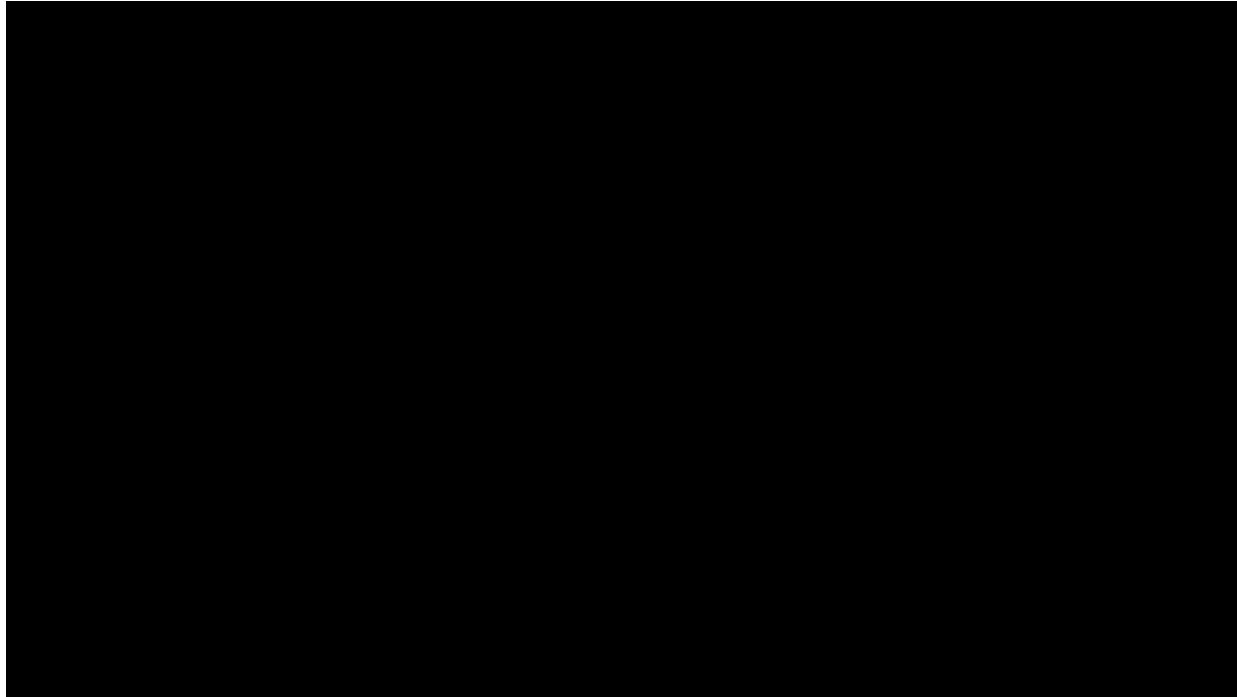
---

# Privacy

# Privacy and Cybercrime

---

Mikko Hypponen TED Talk



# Terminology

---

# What is Privacy?

---

## Informational privacy

- Rules governing the collection and handling of personal data (credit, medical, government)

## Bodily / Sexual

- Protect physical self from search, genetic tests, drug testing, sexual activities and preferences

## Communication privacy

- Security of mail, telephones, e-mail

## Territory

- Limits of intrusion into domestic and workplace environments, video surveillance, ID checks, searches

# Privacy Laws in Canada

---

1974 - Privacy Act (PA, Federal)

1990 - Municipal Freedom of Information and Protection of Privacy Act (MFIPPA, Municipal)

- Regulation 372 (Institutions)
- Regulation 823 (General)

1990 - Freedom of Information and Protection of Privacy Act (FIPPA, Provincial)

- Regulation 460 (General)
- Regulation 459 (Disposal)

2001 - Personal Information Protection and Electronic Documents Act (PIPEDA, Federal)

2004 - Personal Health Information Protection Act (PHIPA, Provincial)

- Regulation 329 (General)

# Information Privacy

---

Recorded or oral information about an identifiable individual

Ties in with law, ethics, security, information and records management

# Collecting Information

---

Reasonableness test – a measurement to determine the validity of an action or process

Notice & Consent



# Key Concept #1: Reasonable

---

1. Was the collection **necessary** to meet a specific need?
2. Was the collection **effective** in meeting the need?
3. Was the loss of privacy **proportional** to the benefit gained?
4. Is there a **better way** to meet the need?

# Key Concept #2: Notice & Consent

---

The private sector asks; the public sector tells (generally)

Notice of collection, privacy policy, acceptable use policies, end user license agreements ...



## THE REAL MEANING BEHIND ALL THOSE CORPORATE PRIVACY EXPRESSIONS

OCTOBER 30, 2011 BY *SIMON*

We've all read them a hundred times before. Broad statements of privacy good will by large organizations. But what do they mean? You can bet that the real intent is coded.

Have a quick read of this one and see if you can fathom what the organization is on about:

"Here at company XXX your privacy is our paramount concern. We believe we've got the customer experience just right and we've consulted widely on our new technology. That's why we've reached out to you to help us build strong user controls in an environment of Privacy by Design. However we know how important it is to achieve the right privacy balance and so we actively support Data Protection reform and modernization. We believe in informed consent, that all organizations should be fully compliant and that that processing should be Fair and Lawful."

Yes indeed. A masterful exercise in saying nothing.

Privacy International believes we all need some clarity, so here's our handy alternative glossary of all those privacy and data protection expressions you've stomached for so long:

 [Search](#)[Add new comment](#)

# Examples

---



Register  
Log in  
Web Mail  
We Are..

## Categories

ELECTRONICS  
Computers  
HACKING.  
FASHION  
GENERAL  
HEALTH  
INVESTMENT  
LIFESTYLE  
LINKS  
MALAYALAM  
AMMETER  
NEW PRODUCTS  
TECHNOLOGIES [Ancient]  
TECHNOLOGIES [latest]  
Tourism  
WOMAN

« Want to Find Love? Be Giving.

SELF ACTUALIZATION. »

# How to use Google for Hacking.

Google serves almost 80 percent of all search queries on the Internet, proving itself as the most popular search engine. However Google makes it possible to reach not only the publicly available information resources, but also gives access to some of the most confidential information that should never have been revealed. In this post I will show how to use Google for exploiting security vulnerabilities within websites. The following are some of the hacks that can be accomplished using Google.

## 1. Hacking Security Cameras

There exists many security cameras used for monitoring places like parking lots, college campus, road traffic etc. which can be hacked using Google so that you can view the images captured by those cameras in real time. All you have to do is use the following search query in Google. Type in Google search box exactly as follows and hit enter

**inurl:"viewerframe?mode=motion"**

Click on any of the search results (Top 5 recommended) and you will gain access to the live camera which has full controls.

you now have access to the Live cameras which work in real-time. You can also move the cameras in all the four directions, perform actions such as zoom in and zoom out. This camera has really a less refresh rate. But there are other search queries through which you can gain access to other cameras which have faster refresh rates. So to access them just use the following search query.

**intitle:"Live View / - AXIS"**

Click on any of the search results to access a different set of live cameras. Thus you have hacked Security Cameras using Google.

## 2. Hacking Personal and Confidential Documents

Using Google it is possible to gain access to an email repository containing CV of hundreds of people which were created when applying for their jobs. The documents containing their Address, Phone, DOB, Education, Work experience etc. can be found just in seconds.

# Questions

---

How does the Internet help privacy, and how does it hurt?

How will changes and evolutions in technology impact privacy?



# ACLU

AMERICAN CIVIL LIBERTIES UNION

BECAUSE FREEDOM CAN'T PROTECT ITSELF.

[KEY ISSUES](#)

[TAKE ACTION](#)

[MULTIMEDIA](#)

[ABOUT US](#)

[DONATE](#)

[HOME](#) › [KEEP AMERICA SAFE & FREE](#) › [SURVEILLANCE & PRIVACY](#)

## Surveillance Under the Patriot Act

[EMAIL](#)

[SHARE](#)

[PRINT](#)

October 24, 2011

Hastily passed 45 days after 9/11 in the name of national security, the Patriot Act was the first of many changes to surveillance laws that made it easier for the government to spy on ordinary Americans by expanding the authority to monitor phone and email communications, collect bank and credit reporting records, and track the activity of innocent Americans on the Internet. While most Americans think it was created to catch terrorists, the Patriot Act actually turns regular citizens into suspects.

[Share on Facebook »](#)

[Share on Twitter »](#)

[Download the graphic »](#)  
[Download the PDF »](#)



### ACT NOW

[Oppose cybersecurity legislation that violates Americans' rights »](#)

### Learn More:

- [› Keep America Safe and Free](#)
- [› Reform the Patriot Act](#)
- [› The Patriot Act, 10 Years Later](#)



# Suicide of Tyler Clementi

From Wikipedia, the free encyclopedia

**Tyler Clementi** was an eighteen-year-old student at [Rutgers University](#) in [Piscataway, New Jersey](#), who jumped to his death from the [George Washington Bridge](#) on September 22, 2010. His roommate Dharun Ravi had [video streamed](#) Clementi kissing another man over the Internet without Clementi's knowledge, and this encounter was viewed briefly by Ravi and a fellow hallmate, Molly Wei. Ravi later made a second attempt to view Clementi's sexual encounters.<sup>[2][3]</sup> Ravi has been charged with invasion of privacy and bias intimidation, while Wei is not being charged, in exchange for testifying against Ravi.<sup>[4]</sup>

## Contents [hide]

- 1 Background and incident
- 2 Prosecutions
- 3 Reaction
  - 3.1 Government
  - 3.2 Other
- 4 See also
- 5 References
- 6 External links

## Background and incident

[edit]

Clementi was born on April 29, 1992 in [Buffalo, New York](#), and raised in [Ridgewood, New Jersey](#). A graduate of [Ridgewood High School](#), he was a talented [violinist](#); he played with the Ridgewood Symphony Orchestra and participated in the Bergen Youth Orchestra as concertmaster.<sup>[5]</sup> Ravi and Wei both graduated from [West Windsor-Plainsboro High School North](#); Ravi lived in [Plainsboro, New Jersey](#) and Wei in [Princeton, New Jersey](#).<sup>[6]</sup>

According to Clementi's instant messaging chats released in court documents filed by Ravi's attorney, Clementi told his parents that he was gay before leaving home for college.<sup>[4]</sup> While his father supported him, Clementi said that his mother had basically "rejected" him.<sup>[4]</sup>

Clementi had complained to his [resident assistant](#)<sup>[7]</sup> and two other officials that Ravi, his roommate, had videostreamed him kissing another man using a webcam. He wrote in detail on the [Just Us Boys](#) message board<sup>[8]</sup> and [Yahoo!](#) message board about complaints he filed through university channels about his roommate. His posts indicated that he did not want to share a room with Ravi.

Clementi wrote that he had asked the resident assistant for a new room after having learned about the first incident and then discovering that Ravi invited his [Twitter](#) followers to watch a second sexual encounter. "He [the resident assistant] seemed to take it seriously," Clementi wrote in a post about 15 hours before his jump from the [George Washington Bridge](#). He wrote that he also reported the incident to two unnamed "higher-ups."<sup>[2]</sup> According to court documents filed by Ravi's attorney, on September 20, 2010, Clementi wrote to a friend online that he was not really bothered by what Ravi had seen.<sup>[4]</sup>

Clementi's wallet was found on September 22 on the walkway adjacent to the George Washington Bridge's New York-bound lanes after witnesses reported seeing someone jump. His car, cell phone, and computer were also found near the bridge. Police recovered a body on September 29 in the [Hudson River](#) just north of the bridge. It was confirmed the next day that the body recovered was that of Clementi.<sup>[9]</sup>

**Tyler Clementi**



Facebook profile self portrait of Tyler Clementi<sup>[1]</sup>

<b>Born</b>	April 29, 1992 <sup>[citation needed]</sup> <div>Buffalo, New York, United States<sup>[citation needed]</sup></div>
<b>Died</b>	September 22, 2010 (aged 18) <div>New York City, New York, United States</div>
<b>Cause of death</b>	Suicide by jumping
<b>Occupation</b>	<span>Student</span>



# Questions

---

What was the role of social media in this case?

Was this a privacy violation? Why?

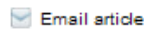
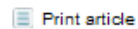
Could privacy laws have prevented Dharun Ravi and Molly Wei doing what they did?

## Accusations of 'Inappropriate Behavior' Resurface for Herman Cain

PRESENTED BY



AP



DASHIELL BENNETT | 970 Views | 5:08 AM ET

Politico reported late on Sunday that Herman Cain was twice accused of "inappropriate behavior" by women who worked with him when he ran the National Restaurant Association in late 1990s. The story, which gives few details of the incidents, says that in both cases the women were given financial settlements and left the company. The settlements included non-disclosure agreements. Politico

### DEPARTMENTS

#### ■ TODAY'S BEST

Five Best Friday Columns

More Today's Best

[Get Five Best by email](#)



#### ■ THE SMART SET

Frances Bean the Bride; Sheen's Show Finds a Buyer

More The Smart Set



#### ■ TRIMMING THE TIMES

The End of Octsnowber

More Trimming the Times



#### ■ TODAY IN RESEARCH

Yes, You Can Eat Too Much Black Licorice

More Today in Research



[All Departments](#)



[tell us anything](#)

Have a story we missed? A link we have to click? A sharp opinion about the news? Instead of waiting for us to post it, tell us on the Open Wire.

[Submit your news and ideas](#) | [See all reader posts](#)

# Questions

---

Is it possible to 'start over' in the online world the same way it is in the real world? Should it be?

If I delete a tweet / posting / photo, is it still accessible? How?

What can you still find online that's 'old'?

# Explore the readings online

---

[8 Facebook Privacy Flaps](#)

[USB Sticks Lost with Voter Information](#)

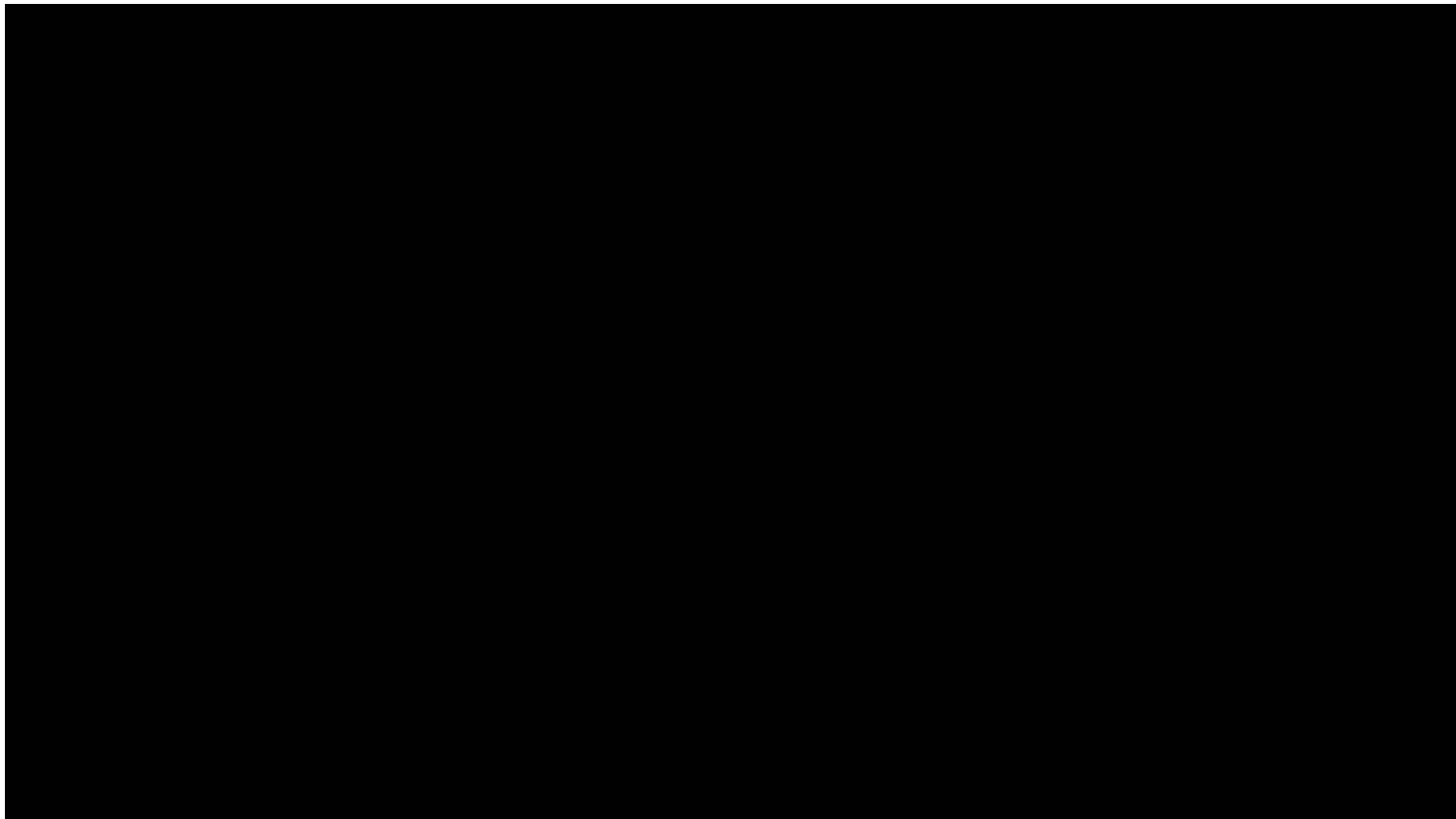
[Is your printer spying on you?](#)

[iOS location tracking lawsuit](#)

# The Power of Anonymity Online

---

Christopher “Moot” Poole TED Talk



---

# The Internet

# Mechanics of The Internet

---

MEDIA TYPES

# World Wide Web $\neq$ The Internet

---

The Web *started* the Internet revolution

- The *killer app* for the Internet, the reason why most people got connected

Really only a part of the Internet, but many people view it as the same



# Who am I?

---



# Tim Berners-Lee

---



# History

---

Originally developed Web, so high energy physicists could interchange data and information on experiments by Tim Berners-Lee at CERN in 1990

- No patents, idea made freely available!
- Later founded the W3C, a consortium to oversee the Web, which adopted the same approach



World's first website:

<http://info.cern.ch/>

---

What if the Web was  
patented?

# History

---

Physics experiments involved a large number of researchers, approx. 250, hard to keep everyone informed on progress and results

Wanted a computer-based system that could manage and communicate information

# Hypertext

---

Early experiments with a [hypertext](#) system on a [single computer](#)

A hypertext system supports the linking together of documents using [hyperlinks](#), selecting the hyperlink takes you to the information

Many experiments with hypertext systems in the 1980s, a popular idea

But no way of sharing it with other people

# Sharing Hypertext: The Web

---

The idea that Berners-Lee had was to use the Internet to share the data

The [hyperlink](#) refers to information (text) on the Internet, not on the same computer

Berners-Lee invented a protocol, [http](#), for transferring this information over the Internet and a language, [html](#), for encoding this information so it can be displayed

The whole system was produced [in several months](#) completed by Christmas 1990

# History

---

Started distributing the system in 1991, mainly in the high energy physics community

By 1992 some universities were beginning to use the Web, mainly for research

There were some competing systems, e.g. [gopher](#), but not as good



# Mechanics

---

Two programs involved:

- **server** - maintains all of the information
- **browser** - allow users to view the information

This is a fairly standard Internet architecture, many services are built this way

Any browser can attach to any server and retrieve information

Servers are relatively simple, browsers are now quite complicated

# Mechanics

---

Information organized into [pages](#), each web page was stored in a separate file

So server was basically [sending files to the browser](#)

Simple file format, html, was used to encode information (a form of XML)

Designed to be created with a simple text editor, this was the main way of creating web pages for the first 5 years

# Early Websites

---

Initial information on html was a short web page, when printed less than 10 pages

Basically a list of the html formatting commands, with a short description of each command

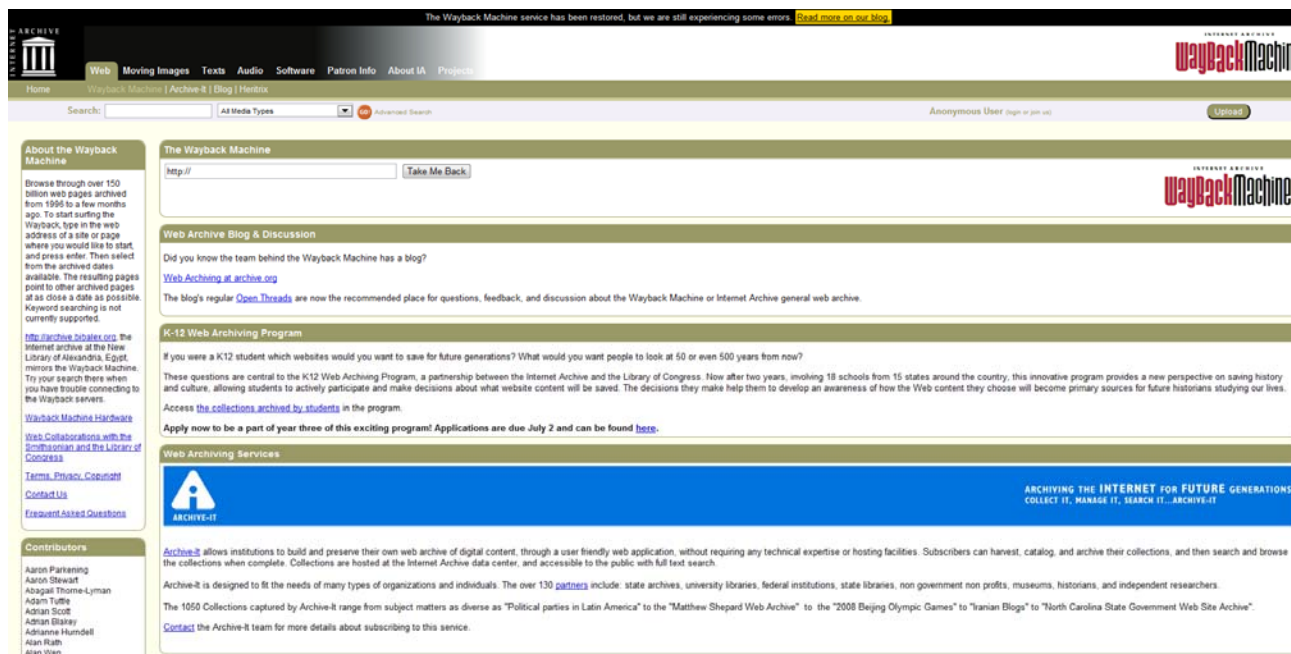
Very simple set of commands, aimed at easy processing in the browser

Later images were added and some fancier formatting, such as tables

Not much interest outside of the university community until around 1994

# www.archive.org

Use the “Way Back Machine” to look at an old version of a popular website... what has changed?



# MIME

---

Http uses MIME types (Multipurpose Internet Mail Extensions) to determine the information in a file, whether its text, image, movie, etc.

MIME developed to send media in mail messages, ability to add images, video, etc

Based on that we can associate the file type with the program used to display the information

The media is displayed by a separate program running in its own window

# History

---

Early browsers followed the MIME approach

When they encountered data they couldn't display, used MIME file to determine program used to display data

Data not shown as part of the web page, displayed in a separate window

Hard to work with images, since they weren't embedded in the web page

# History

---

For some data this is okay, for example PDF files

For other data would like it **integrated** with the web page

Browser **plugins** were developed to handle this, first appeared in Netscape around 1995

A plugin is given a window and region for display, usually this is part of the browser window: rich media became integrated

# Browsers

---

The first popular browser was Mosaic

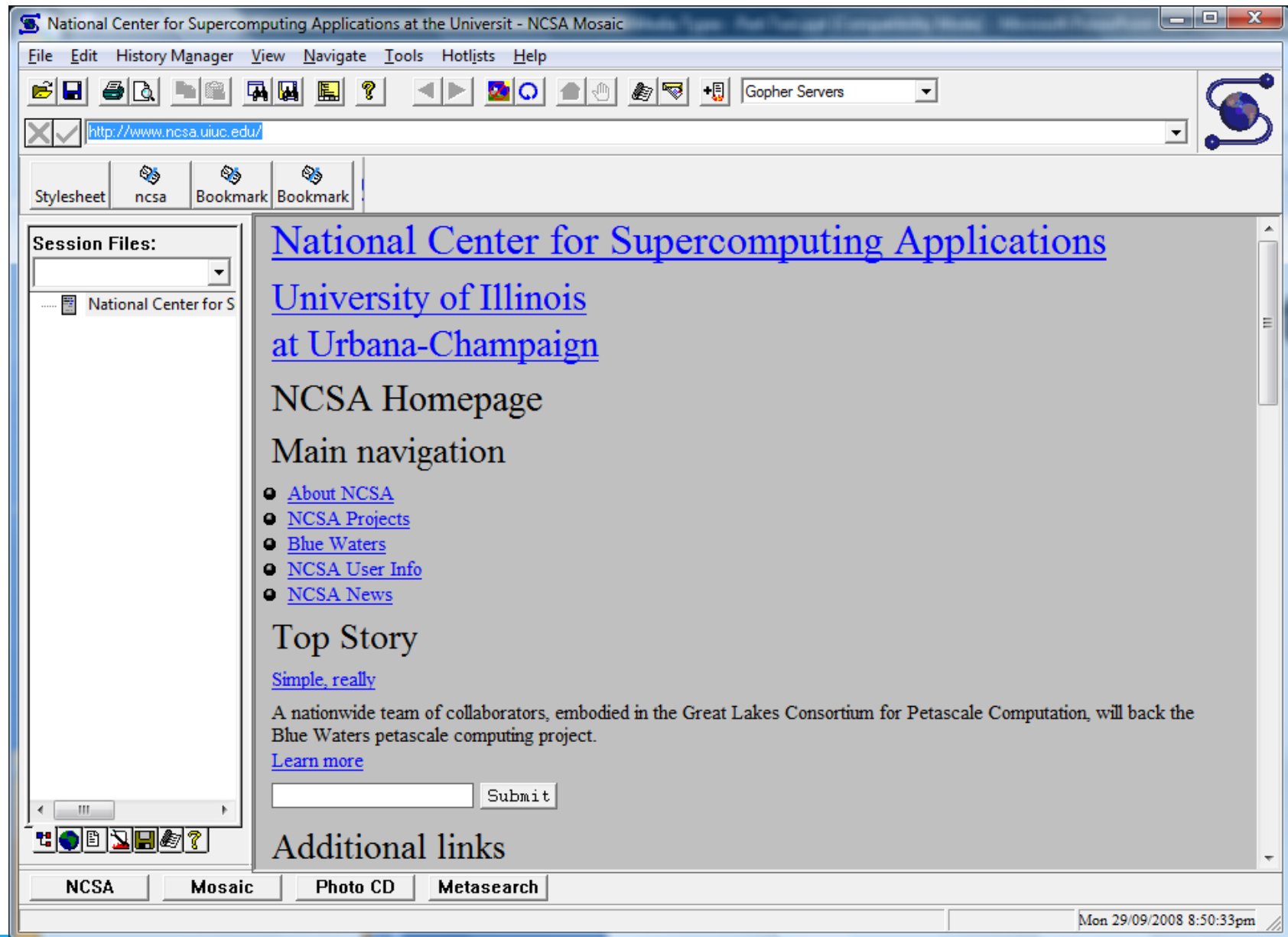
Developed at Supercomputer Center at University of Illinois

Distributed for free, popular 1993-1997, can still be downloaded

Netscape was founded by some of the developers of Mosaic to develop commercial browsers

Thought people would be willing to pay a lot for browsers, bad business model





# Browsers

---

Initially two versions of Netscape:

- free download, compete with Mosaic
- commercial package with extra features

Initial cost of commercial version was \$60 per computer

Not significantly better than Mosaic, so not many copies of commercial version were sold

# Browsers

---

Generally the quality of Netscape software was quite poor, very slow and buggy

Tried to make a conceptually simple program complicated so they could charge for it

Resulted in overly complex program, didn't perform well and was very hard to maintain

This was clear when the source code was released in 1998

Some of this source code lives on in [Mozilla based](#) browsers (for example Firefox)

# Microsoft Internet Explorer

---

Microsoft entered the game late, had to catch up quickly

Result: Internet Explorer was a simple program developed in 6 months, partly based on a commercial product being developed by a company they bought

Much more efficient than Netscape, and had fewer bugs

Distributed for free by Microsoft

# Browser Wars

---

The start of browser wars, led to the anti-trust suits against Microsoft

Netscape never made money and eventually got into financial trouble

Bought by AOL in 1999, which then sold part of Netscape to Sun

Official support for Netscape ended on March 1, 2008, realized that Microsoft won on the commercial front, Firefox on the open source front

# Browser Trends

---

Both Firefox and IE are large complex programs

Some browsers, such as Chrome, are much smaller

- don't do as much, but do a better job of it

Why couldn't Netscape do this?

# History

---

Netscape really wanted to replace the desktop, so it had lots of functionality

Replace desktop with browser, this is what Google is currently trying to do – will history repeat itself?

It included email, news, web page design and several other programs

Why it ended up being so complex, and one of the main reasons why it failed

Couldn't be general enough

# Servers

---

Early web servers were very simple, only a few pages of code

All they needed to do was find the file the browser requested and send it to the browser

Also sent information about the file

Know whether you really wanted to download it, important in the early days when network connections were slow



# Static Websites

---

Problem: the early web was *static*, no way to automatically modify or generate a page

Once a page was produced it stayed that way

Need for dynamic content, the ability to generate a page based on current data

# Search Engines: Dynamic Content

---

Example: search engines need dynamic content

They maintain a database of web pages, when user asks about a topic they search the database to find the relevant pages

Need to generate the web page to be sent to the user showing the search results

Also needed for services like eBay and Amazon

# CGI: Common Gateway Interface

---

Earliest dynamic content used CGI, a way of executing a program on the server

Browser collected information using forms

Information packaged and sent to server, along with name of CGI program

# Servers

---

CGI program then determined what the user requested, retrieved the information, and then generated the web page

Server then sent it to the client

CGI is still widely used today, but there are now many other ways of doing the same thing

# Servers

---

Haven't evolved as much as browsers

Standard server is Apache, an open source project, Microsoft and others have tried to compete, but have made little progress

Main concern on servers is security, make sure people can't break into the web site

CGI can be a security problem, can potentially execute any program on server

# Cookies

---

Another tool for dynamic web content

Each http request is a separate transaction, browser knows which computer it comes from, but little else

Server may not know that the same user is producing a set of requests, that they are in fact related

This is required for ecommerce and similar services, like Amazon

# Cookies

---

A server can **send a cookie to a browser**, which then **stores it** on its local disk

The cookie identifies the site that sent it, plus any information that site wants saved

When you go back to that site, it asks the browser to send back its cookie, and the information stored in it

# Cookies

---

If the cookie exists, server knows you have been at that web site before, and can retrieve information about you

This is how personalized web sites are produced

Can be used to link together a set of transactions, like in Amazon

It can also be used to track the places that you visit and how you arrived at the current web site



# Search Engines

---

In the first few years of the web there were no search engines

Problem: how do you find interesting web sites?

Not many sites, so could use links or email messages

Tried to use links to keep track of all the sites on the web

NCSA maintained a web page that had links to all the top level websites

# Search Engines

---

This quickly broke down, too many web sites, practical up to about 400 or 500 websites

Search engine: visit all of the web sites, collect keywords of each page, store this information in a large database

Could actually be done in early years of the web, popular search engines were [Altavista](#) and Yahoo, even Netscape tried

# Search Engines

---

Early search engines could search all of the web in a few weeks, relatively up to date

As web grew problems started to occur

First the web got large, and many pages weren't of much interest

Hard to know what terms to index, many web pages now have keywords that help with indexing

# Explore vs. Search: Yahoo!

When it first started, Yahoo was a *manually-created* index of the Web!

To find a site, you would navigate through the hierarchy



# Bots and Web Crawlers

---

A [search engine](#) is a type of program that is called a bot or [web crawler](#)

These programs automatically search the web for interesting things

Could be trying to index the web, or could be searching ecommerce sites to see who has the best deal on certain products

# Bots

---

Generate millions of http requests, not necessarily very smart, but very fast

If all of these requests go to the same server, they can cause lots of problems

For example, bot may request a page for every item in amazon.com catalogue, slowing Amazon servers

# Bots

---

Several things have been done to make bots easier to control

Rules for bot programming:

- can only send a small number of requests to the same site each second
- must obey the bot control information
- retrieve only the information that you need

# Bots

---

Bot control information guides the bots to the parts of the site that are of interest to them

Control which information becomes public through search engines



# The Search Problem

---

Search could produce thousands or even millions of web pages

How do we find the page that we want?

# Search Engines

---

Search engines attempt to rank the pages, sort the pages based on “interest”, send the most interesting pages first

---

What is “interesting”?

# Search Engines

---

Types of interest:

- closest match to user's request
- sites willing to pay the most money
- knowledge of user's previous requests, how successful we have been in the past

There are books on positioning your web site within the lists returned by search engines, this is very important for commercial sites

# Page Rank

---

Google algorithm for ranking results

Based on how many sites link to a page

- Also, the content of the words linking to your site
- And the popularity of the sites linking to your site

# User-specific Ranking

---

Another approach is to let explore and rank the results

- develop a profile of user preferences and use it to re-rank in future

Refine the search based on feedback from the user, user provides some guidance to the search process

# Portals

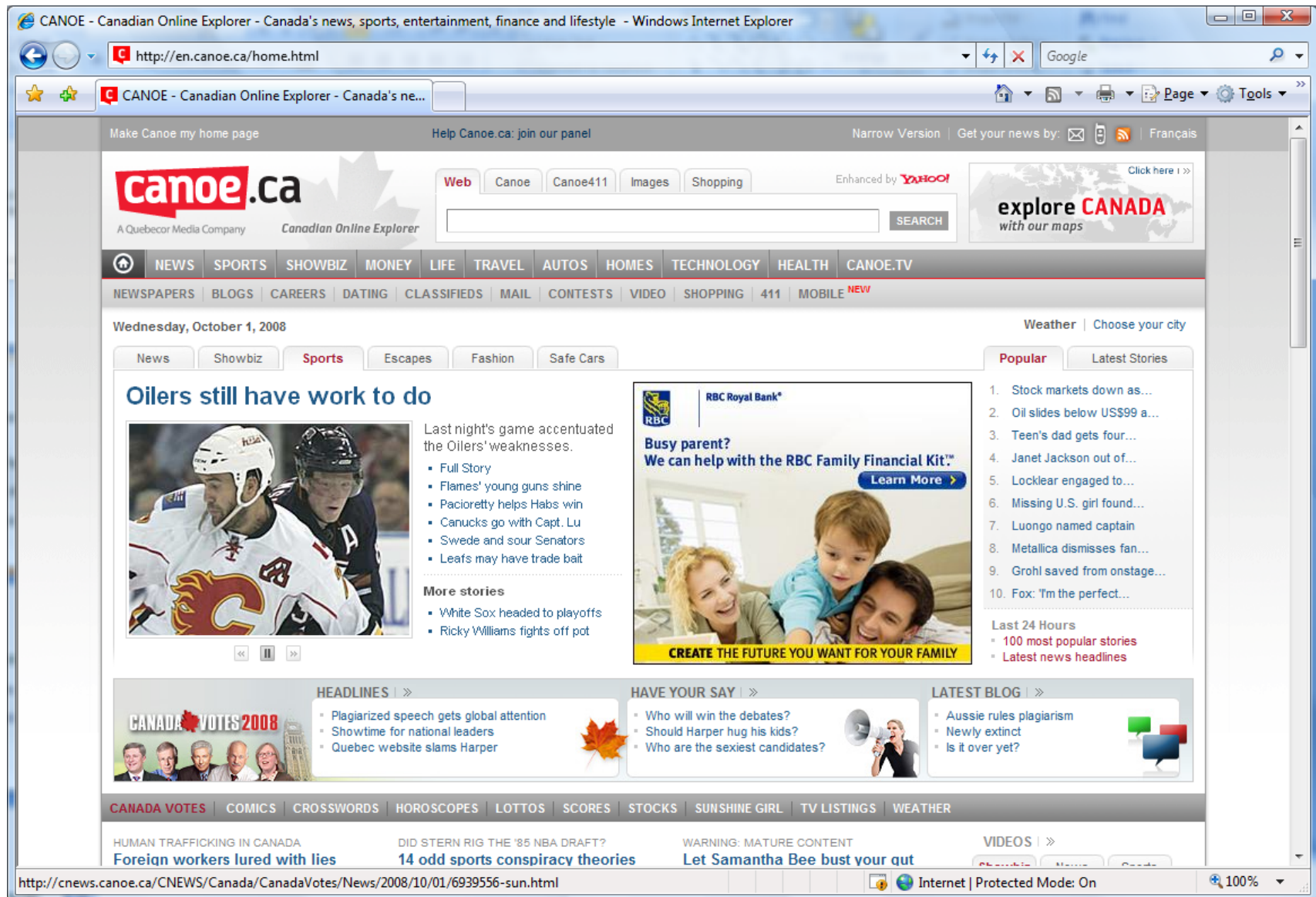
---

Portal: a site that maintains links to other sites

- Yahoo!
- MSN.ca
- Canoe.ca

Like search engine, goal is to help people find information they need

Links are grouped into categories, so its easier to find a particular site





# Portals

---

Portals have several advantages

They are at least partially constructed by people, serve an editing function:

- classification of web sites makes sense, not based on simple keywords
- the linked sites are usually relevant and useful
- the links actually work

Portals aim at quality over quantity

# Portals: Yahoo

---

Yahoo started as one of the most sophisticated search engines

Moved to a portal, better business model, can charge to have site listed on portal

Portals can be more profitable than search engines

Have since expanded to other web services, such as streaming media

Yahoo is now a web services company, not just a search engine

# Summary

---

Today we reviewed:

- Mechanics of the internet