Hidden Universes of Information on the Internet

Google bing similarweb

Day #1 & Day #2

SearchSystems.net

LinkedIn

MarineTraffic.com

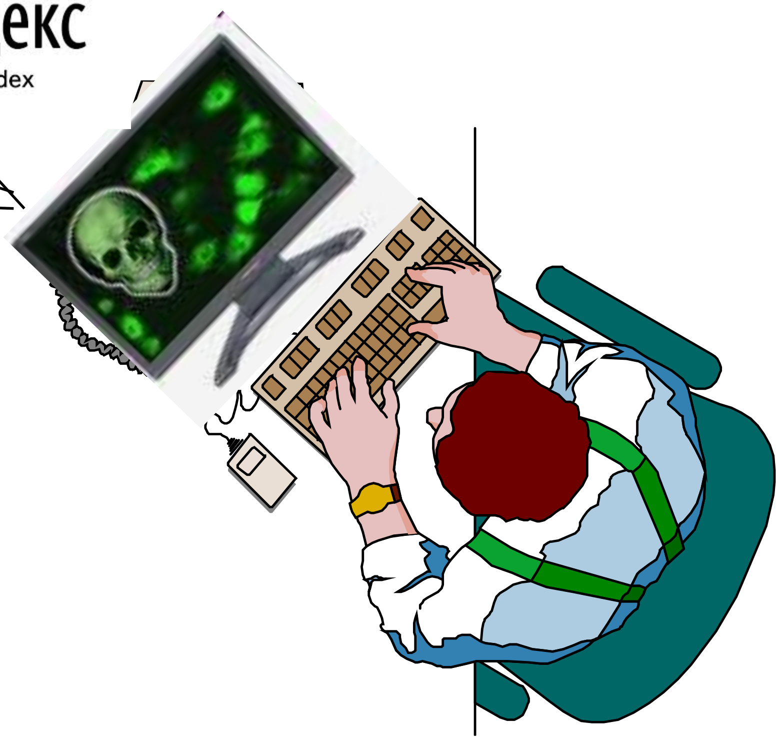
Wikipedia The Free Encyclopedia

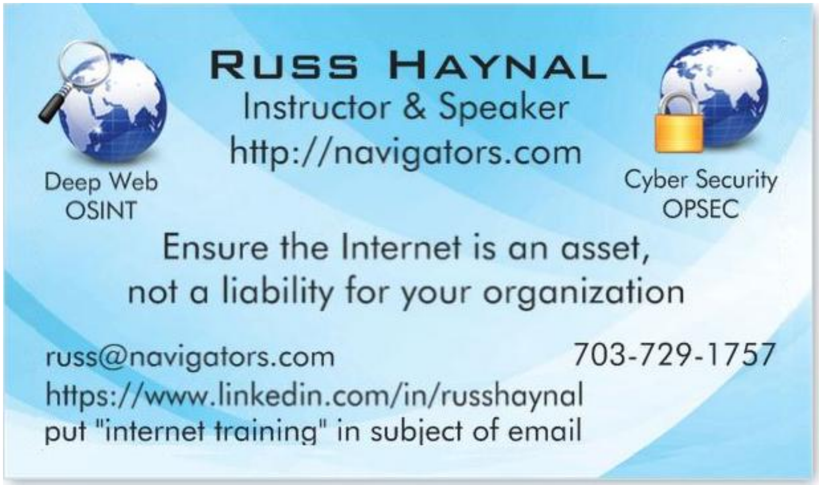
Яндекс

Yandex

wikimapia

abyznewslinks.com





Russ Haynal

Instructor & Speaker

http://navigators.com

Deep Web
OSINT

Cyber Security
OPSEC

Ensure the Internet is an asset,
not a liability for your organization

russ@navigators.com 703-729-1757

https://www.linkedin.com/in/russhaynal

put "internet training" in subject of email

Revision 07/2025

Note: If you send me an email, put “internet training” in the e-mail's subject

Course Outline

- Introduction to Internet Architecture
- “Persona” issues
- Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

Online Web page = http://navigators.com/opensource.html

Disclaimer

- This session illustrates a wide variety of search tools, techniques and research methods
- Consult your organization’s policies to verify if these methods are approved for your types of Internet connections (including visits to navigators.com)

Internet Definition

“A large collection of inter-connected networks and computers”

“A new fundamental form of communication that will absorb other communication channels”

Internet represents a
once per thousand year event
Last such event = Gutenberg printing press

Are You Literate in Today’s Online World?

Number of Hosts in each Domain

Top Level Domains

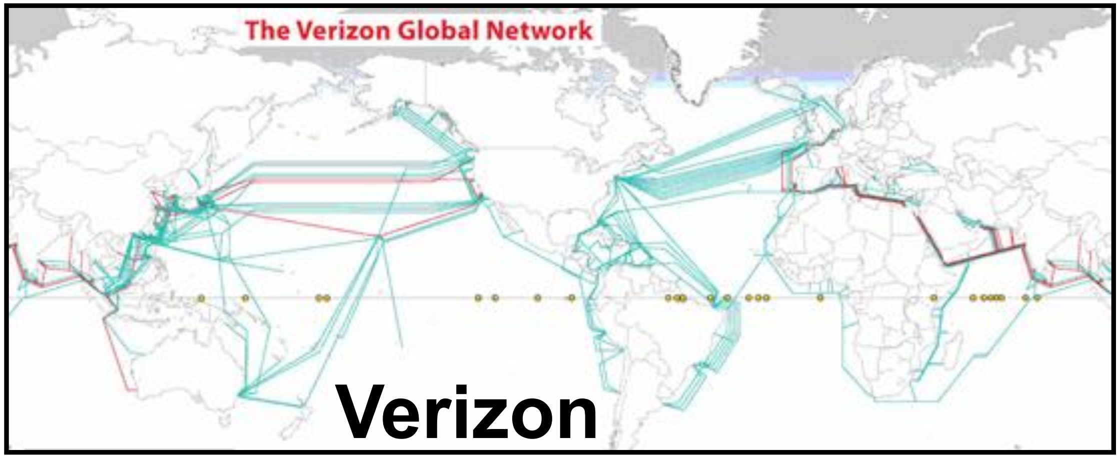
net	386,970,568
com	169,975,462
edu	11,424,990
gov	2,276,632
org	2,161,611
mil	1,443,379

jp	79,002,746
de	48,087,619
br	46,023,691
it	28,538,734
fr	23,529,249
cn	20,196,732
mx	19,298,175
au	16,792,160
ar	14,737,149
nl	13,188.872
ru	13,183,783
pl	12,897,921
ca	10,242,678
in	8,337,038
tr	6,998,966
co	6,851,655

tw	6,811,801
za	6,005,425
uk	5,740,402
be	5,520,698
se	5,473,537
ch	5,230,015
eg	5,044,567
es	4,798,915
fi	4,548,069
th	3,879,942
no	3,798,249
	pt,at,cl,cz
	hu,dk,gr,nz
	il,ro,ua,sg
us	2,025,370

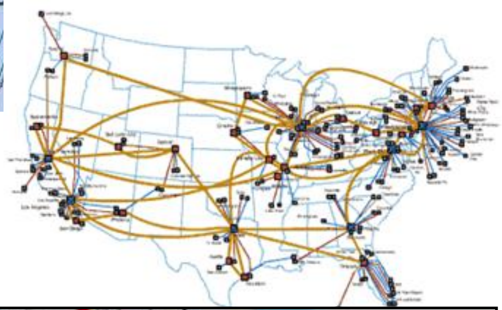
Source: www.isc.org

Example Network Maps

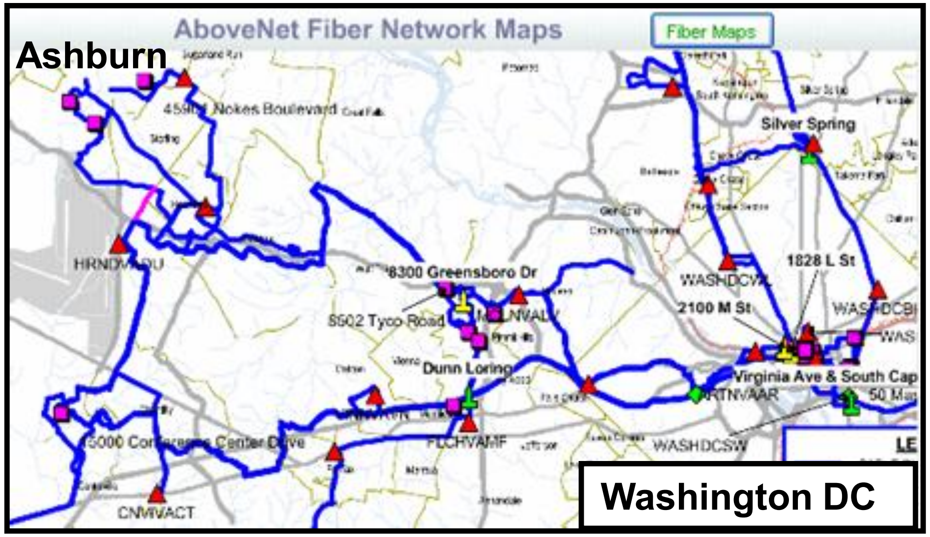


Verizon

Sprint

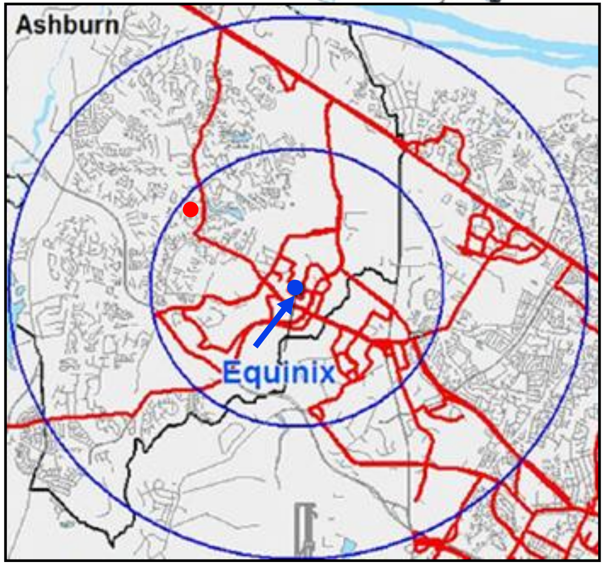


AT&T



Ashburn

Washington DC

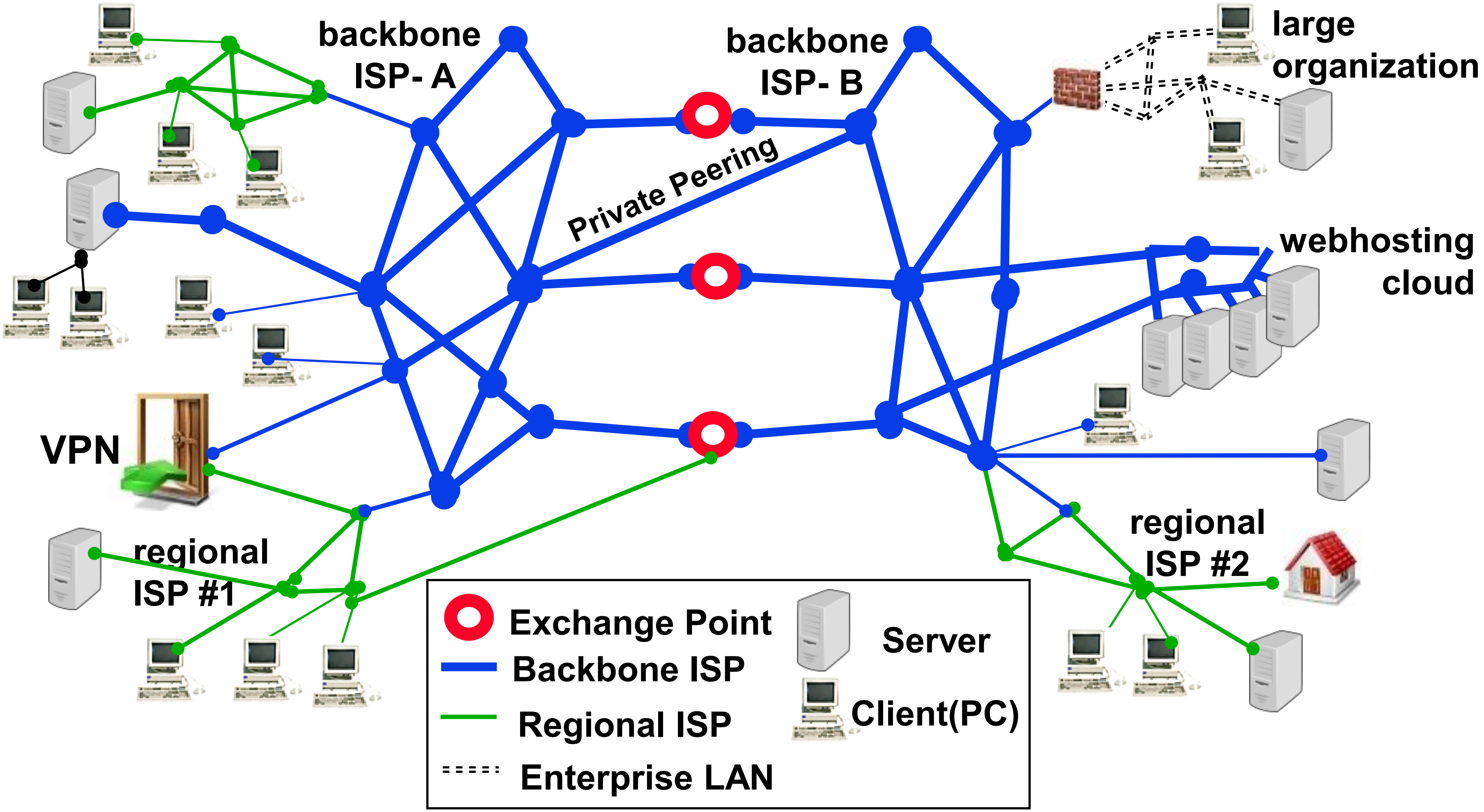


Ashburn

"The Bullseye of America's Internet"

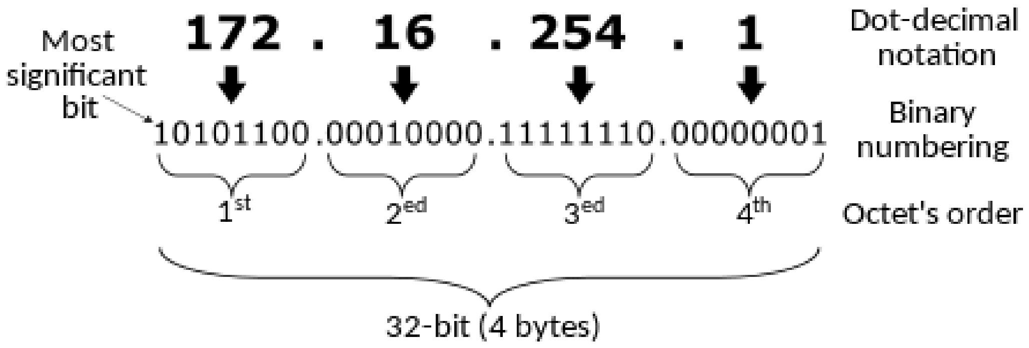
Many People Can Observe Your Internet Usage

- Network traffic flows through multiple Internet providers
- Routers direct packets of traffic along the “preferred” path



Internet Protocol Address (IPv4)

- Has 32 bits of information (a binary sequence of 32 zeroes and ones)
- Expressed as a set of 4 numbers, each number has a range of 0-255



172 . 16 . 254 . 1

Dot-decimal notation

Most significant bit

10101100.00010000.11111110.00000001

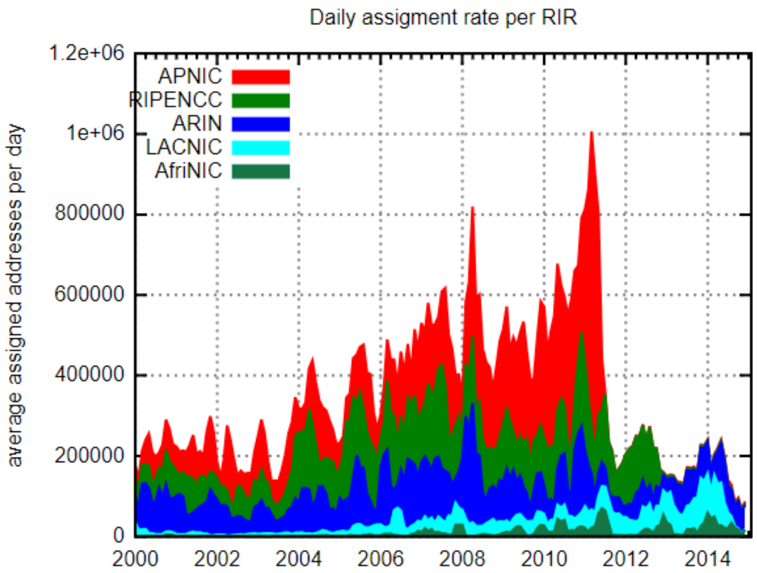
Binary numbering

1st 2ed 3ed 4th

Octet's order

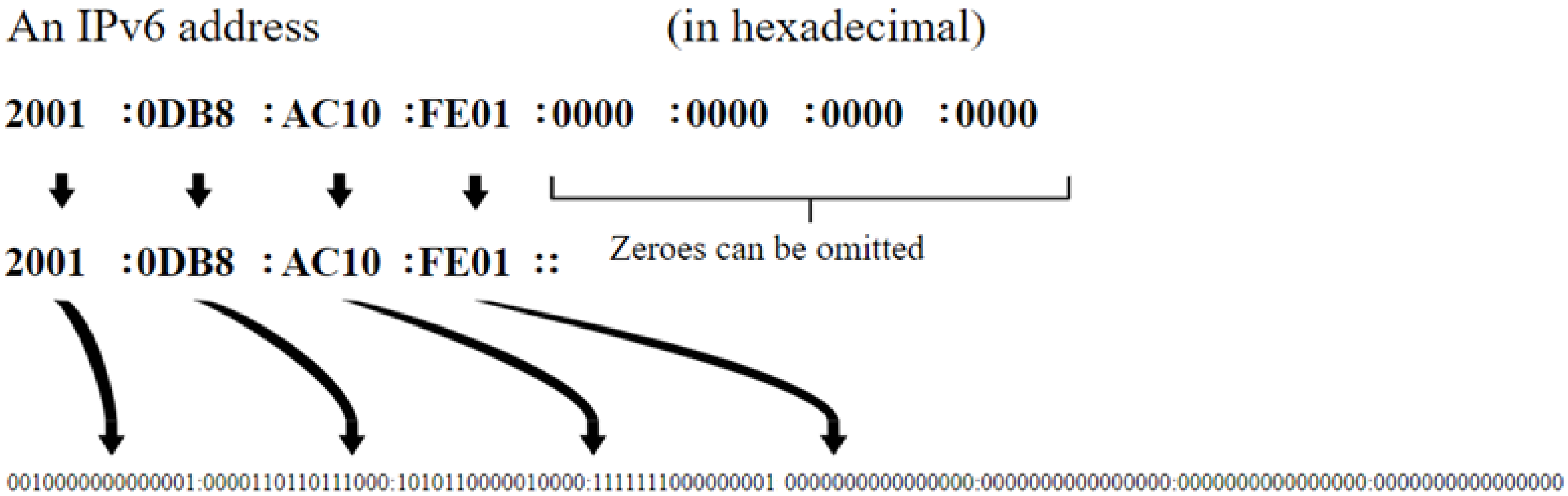
32-bit (4 bytes)

- Total number of possible IPv4 Numbers: 4,294,967,296 (2^{32})
- 1 million IPv4 numbers per day were being allocated, until depletions began in 2011



Internet Protocol Address (IPv6)

- **Has 128 bits of information (a binary sequence of 128 zeroes and ones)**
- **Expressed as 8 groups of 4 hexadecimal digits (0-9, ABCDEF)**



340,282,366,920,938,463,463,374,607,431,768,211,456 (2^{128}) unique IPv6 addresses

3,911,873,538,269,506,102 addresses per square meter of the Earth's surface

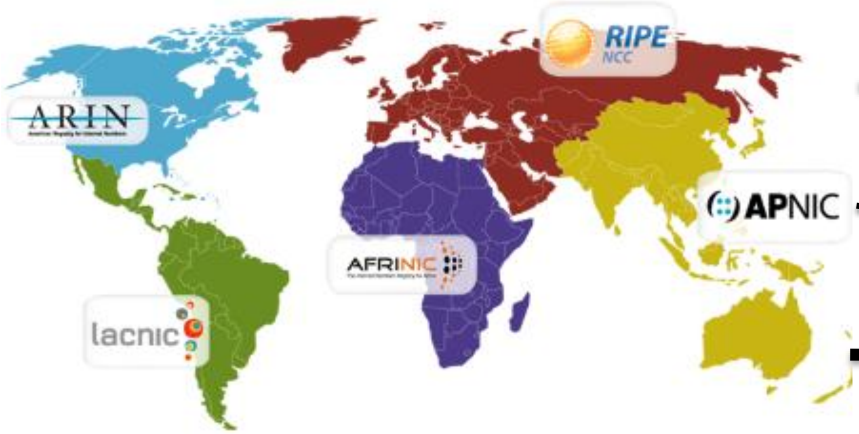
4,500,000,000,000,000 addresses for every star in the entire universe

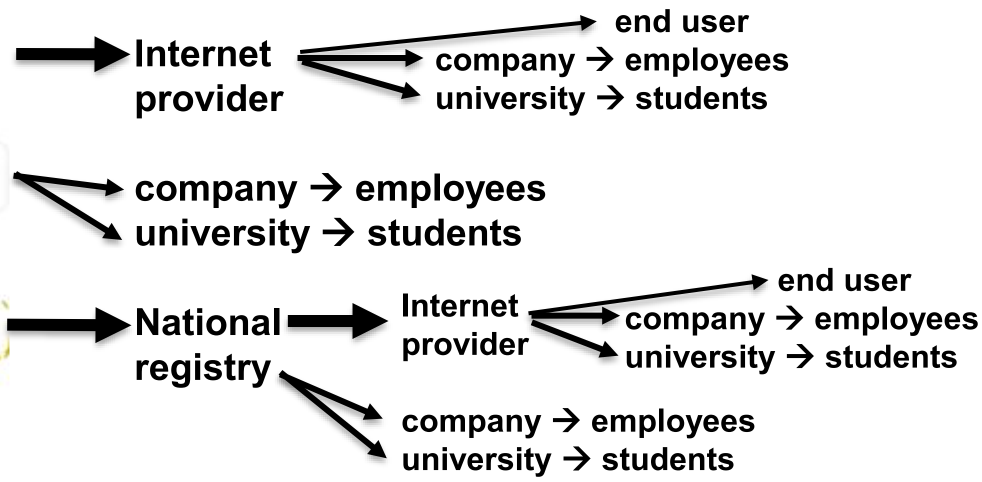
- **IPv6 supports prioritization of traffic and simplifies route addressing**

IP address Allocation

- **Every Internet connection has a unique IP address**
- **IP addresses are initially allocated through a hierarchy, and can "migrate" via multi-national companies, mergers, acquisitions**

IANA → Regional Internet Registry → Local Internet Registry

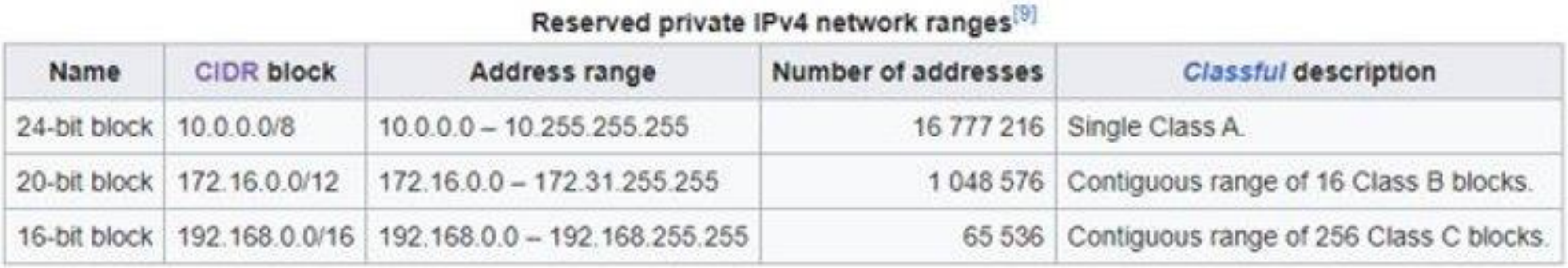




Internet provider → end user
Internet provider → company → employees
Internet provider → university → students

company → employees
university → students

National registry → Internet provider → end user
National registry → Internet provider → company → employees
National registry → Internet provider → university → students
National registry → company → employees
National registry → university → students



Reserved private IPv4 network ranges[9]

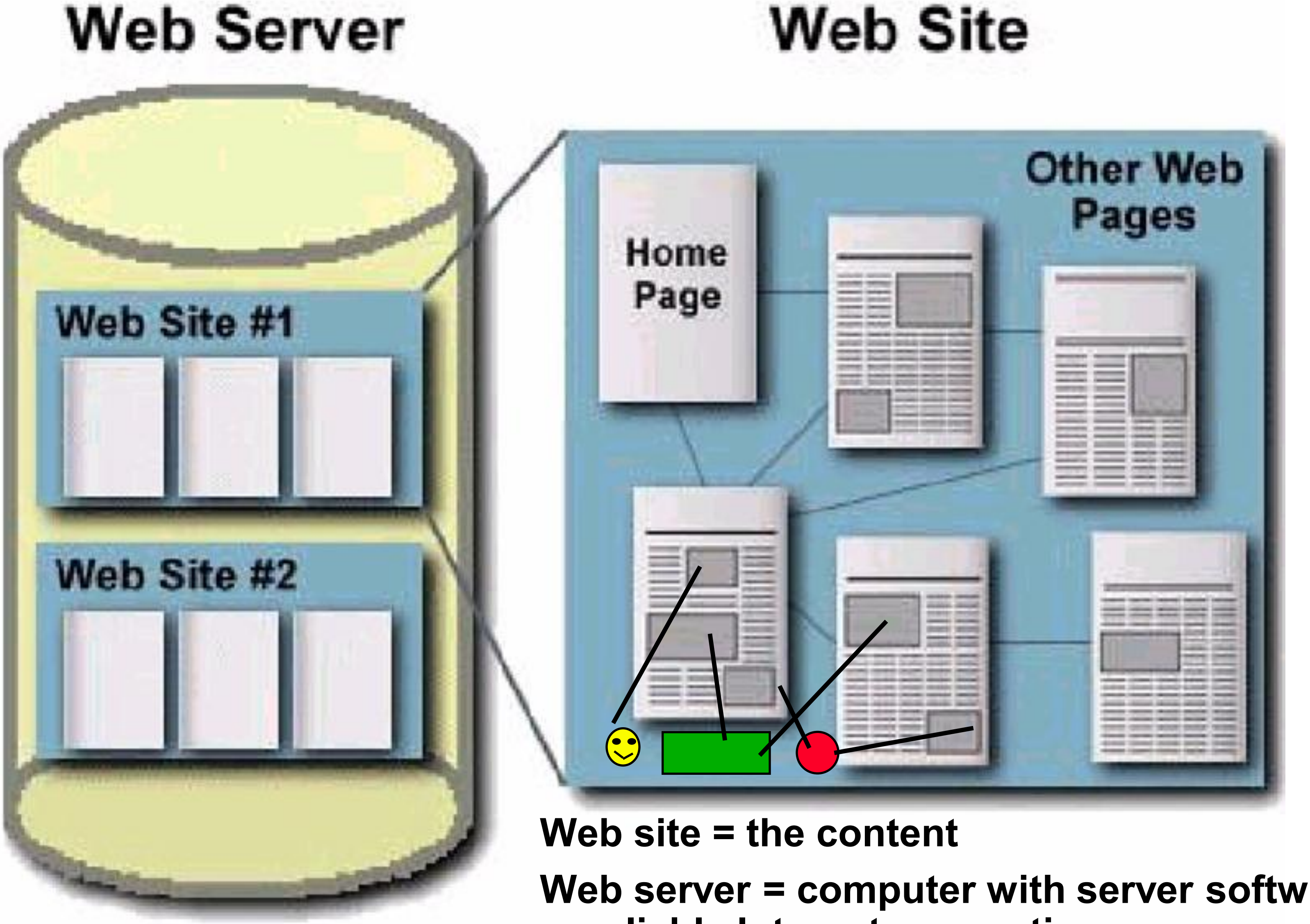
Name	CIDR block	Address range	Number of addresses	*Classful* description
24-bit block	10.0.0.0/8	10.0.0.0 – 10.255.255.255	16 777 216	Single Class A.
20-bit block	172.16.0.0/12	172.16.0.0 – 172.31.255.255	1 048 576	Contiguous range of 16 Class B blocks.
16-bit block	192.168.0.0/16	192.168.0.0 – 192.168.255.255	65 536	Contiguous range of 256 Class C blocks.

- **IPV6 Unique Local Addresses = FC00::/8 and FC00::/7**

Domain Name System

- The Domain Name System (DNS) associates alpha-numeric names with IP addresses
- Names are registered with country-specific registrars or commercial registrars such as Go Daddy
- DNS servers are distributed throughout the Internet
  - They function as a set of inter-linked phone books
- You enter “www.navigators.com”
  DNS servers match it to “209.59.210.79”
- Historical meaning for domain names
  - .com=commercial .net= Internet Provider .org = non-profit
  - .uk = United kingdom .pk= Pakistan (= $16/year) .ru = Russia
- Reality…. Many country domain names are for sale to ANYONE from ANYWHERE

Web Server / Web Site



Web pages
= html

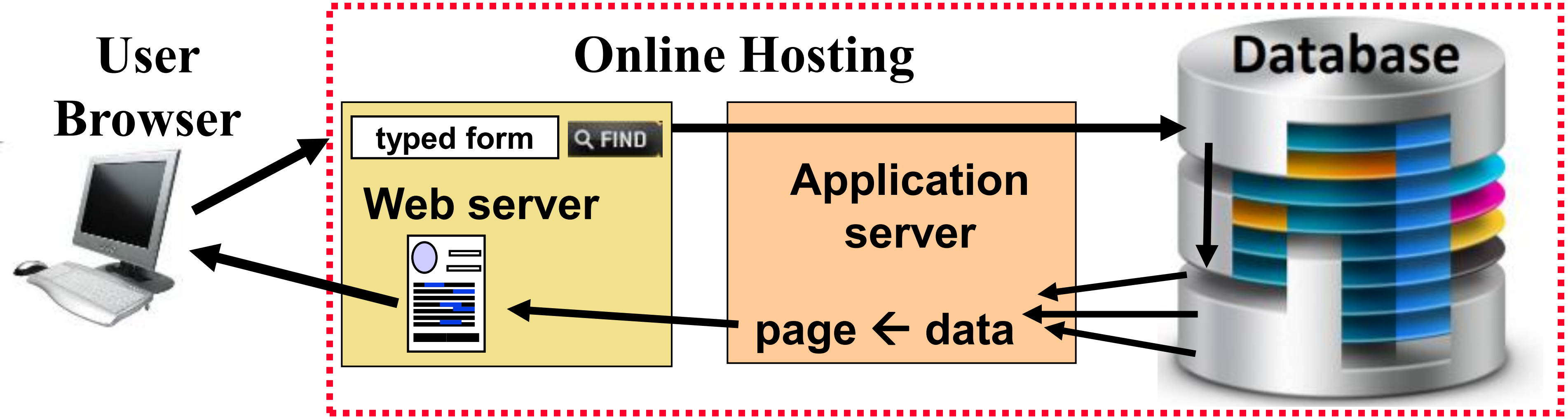
Graphics
= gif, jpg

Other files
=pdf, ppt, doc, txt, exe, zip

Web site = the content

Web server = computer with server software and reliable Internet connection

A More Complex Environment

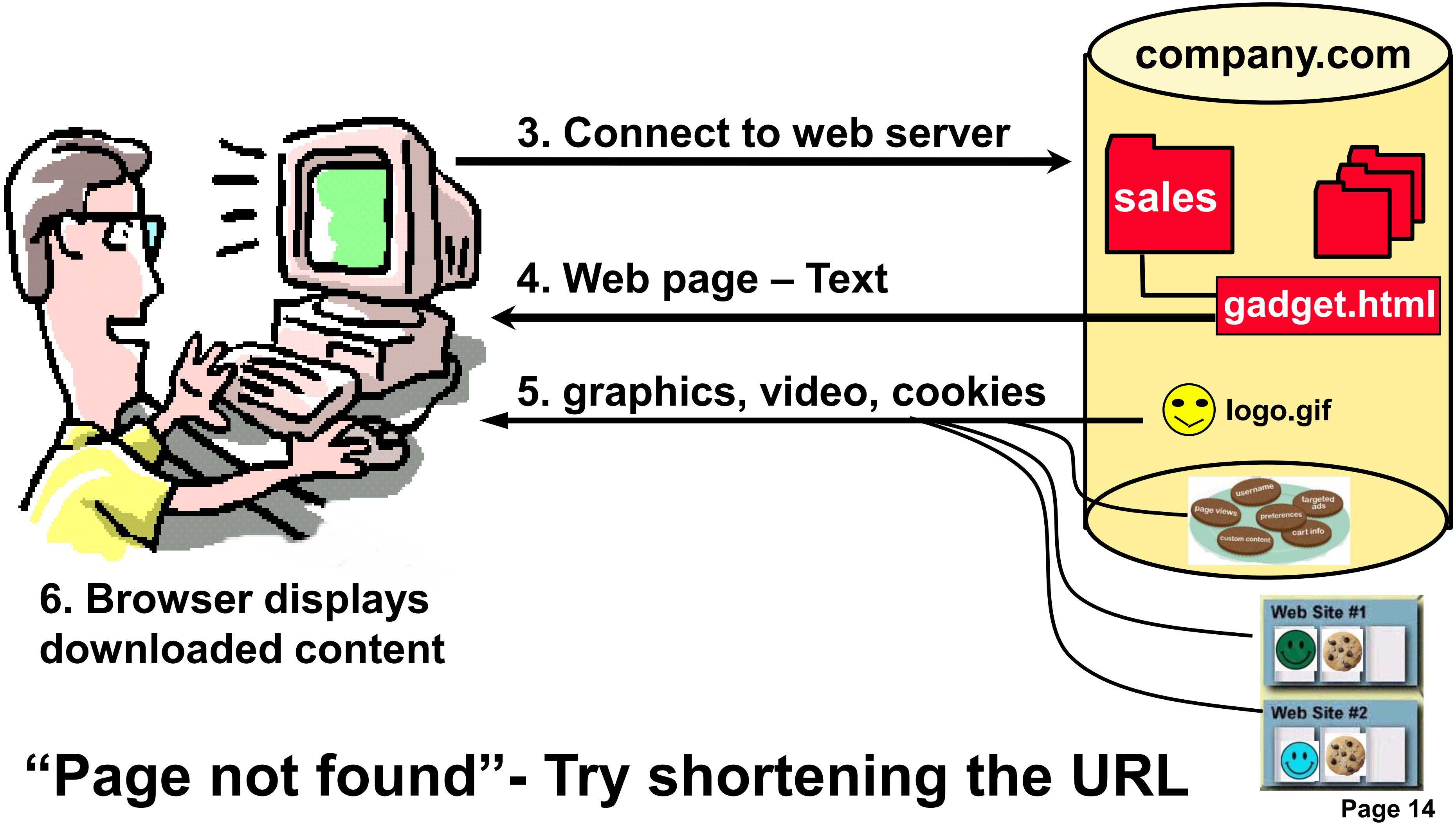


- Internet users interact with web server
- Web server query is passed to a database
- Database content is displayed in a TEMPORARY web page, created in response to USER-actions
- Most database content is unreachable by search engines

Accessing a Web Page

1. Browser requests URL: http://www.company.com/sales/gadget.html

2. Domain name look-up: root → .com → company.com → IP #



3. Connect to web server

4. Web page – Text

5. graphics, video, cookies

6. Browser displays downloaded content

“Page not found”- Try shortening the URL

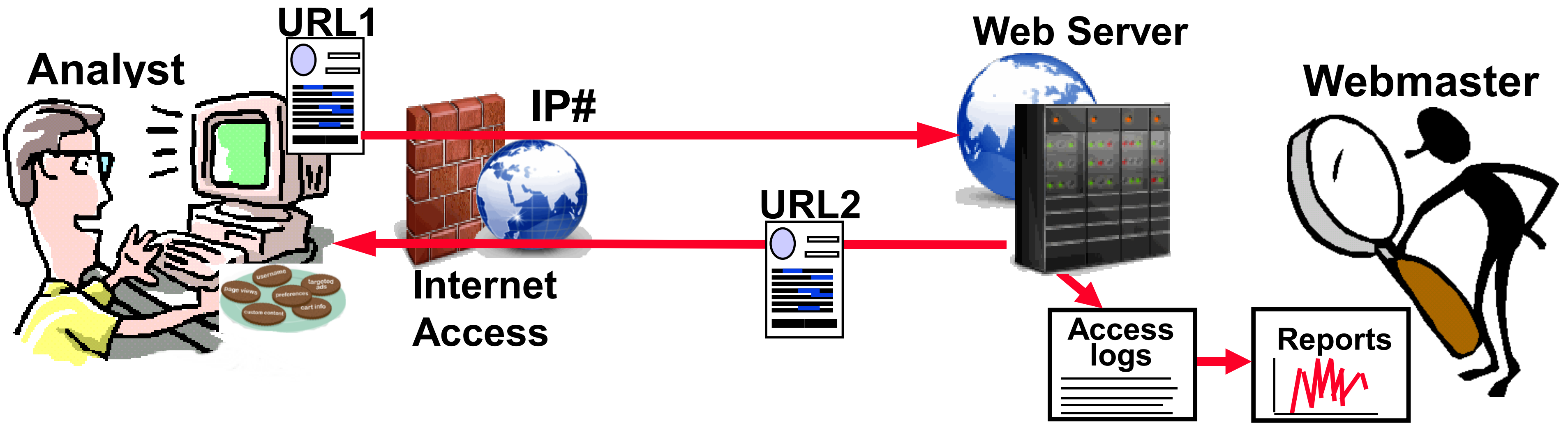
Course Outline

- Introduction to Internet Architecture
- → “Persona” issues
- Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

Online Web page = http://navigators.com/opensource.html

Introduction to “Persona”

As users surf the Internet, persona details are transmitted



- User is viewing a page (URL1), then clicks to another page (URL2)
- The web browser sends “environment variables” to the web server
- Webmasters use this information to learn about users, and their organization (physical location, interests, software)

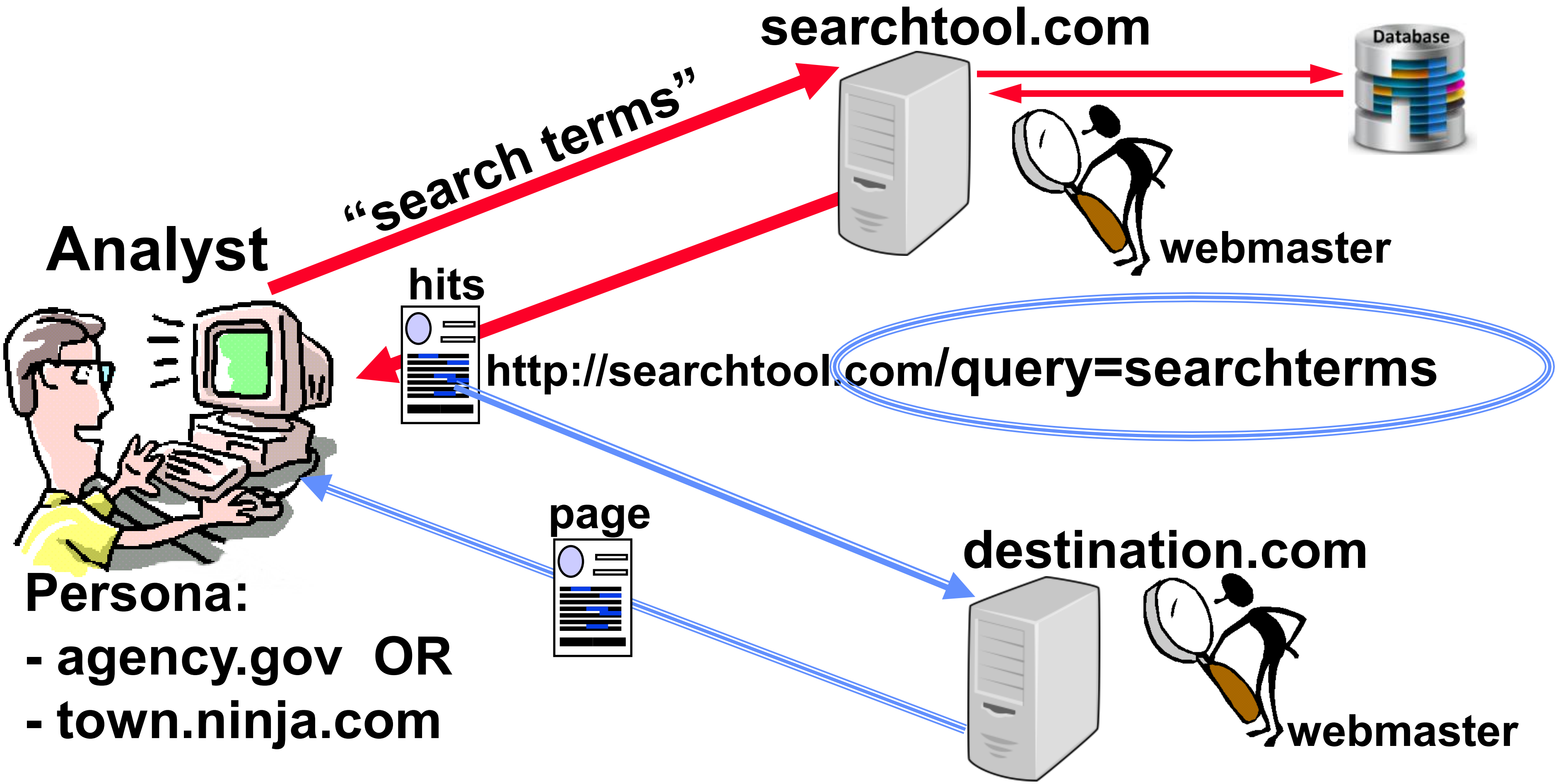
You should understand what websites know about you

Persona Details

- Know your persona <u>before</u> you visit any website
- Should you visit:
  - badguy.com from agency.gov
- Your persona is communicated via “environment variables” such as:
- REMOTE_ADDR = IP number of your computer or proxy
- REMOTE_HOST = Domain name associated with your IP Number
- HTTP_REFERER = URL of the <u>previous</u> page you clicked within

- Be careful how you create web pages
  Do you want to reveal the following :
  - http://badguy.com is listed on
    http://intranet.agency.gov/joe_smith/investigation_targets.html
- Persona details are also tracked via cookies, beacons, Javascripts, browser plugins, remotely stored objects

A Typical Scenario...

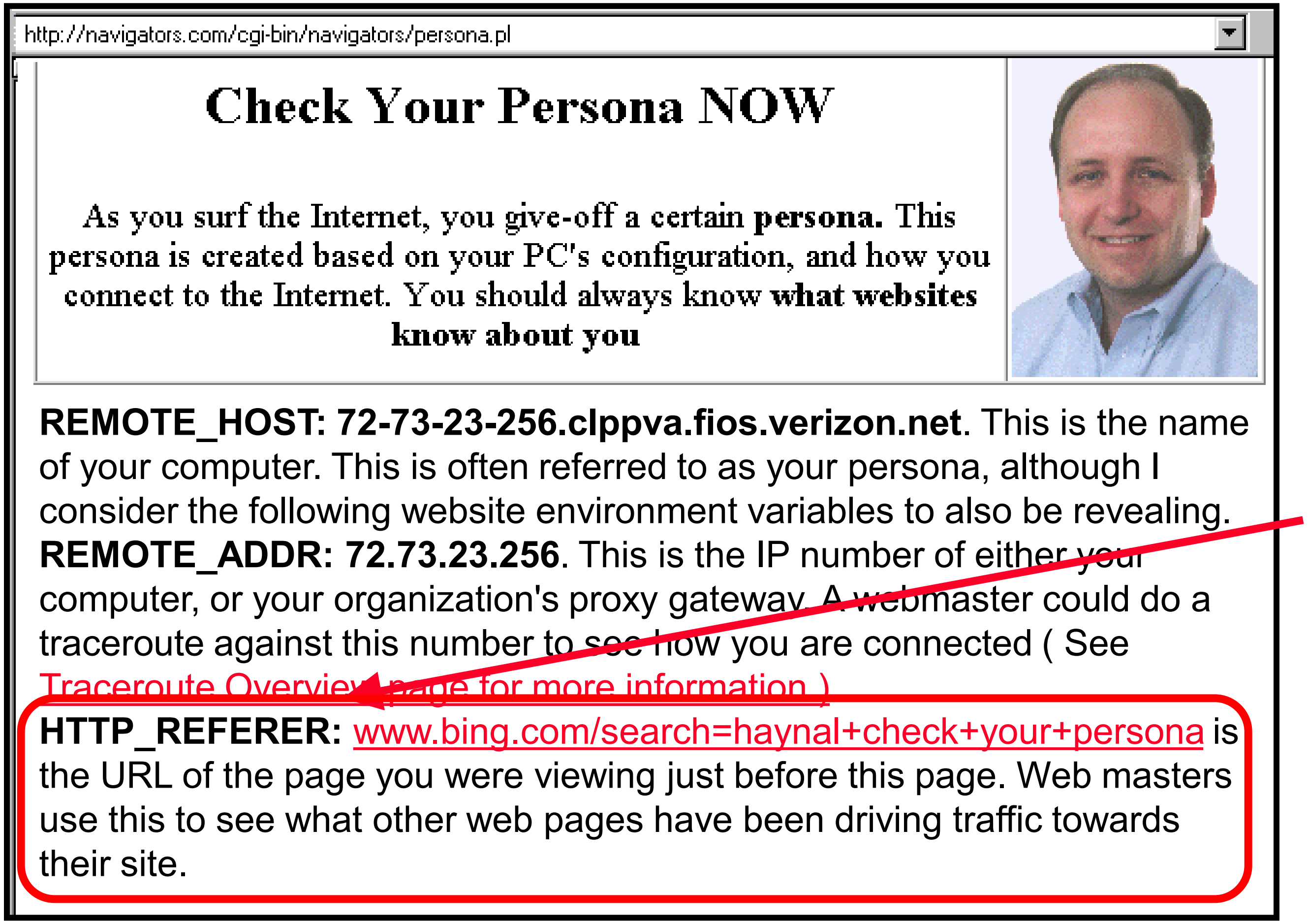


Persona:
- agency.gov OR
- town.ninja.com

━ searchtool.com webmaster knows your "search terms"

═ destination.com webmaster knows the "search terms" and search techniques used to find them

Always Check Your Persona



http://navigators.com/cgi-bin/navigators/persona.pl

Check Your Persona NOW

As you surf the Internet, you give-off a certain **persona.** This persona is created based on your PC's configuration, and how you connect to the Internet. You should always know **what websites know about you**

REMOTE_HOST: 72-73-23-256.clppva.fios.verizon.net. This is the name of your computer. This is often referred to as your persona, although I consider the following website environment variables to also be revealing.
REMOTE_ADDR: 72.73.23.256. This is the IP number of either your computer, or your organization's proxy gateway. A webmaster could do a traceroute against this number to see how you are connected (See Traceroute Overview page for more information)
HTTP_REFERER: www.bing.com/search=haynal+check+your+persona is the URL of the page you were viewing just before this page. Web masters use this to see what other web pages have been driving traffic towards their site.

Important note:
This test page is most accurate when clicking on a link to arrive at this page

Look for this variable,
If this is missing, then no referring URL is being passed via http_referer

- **Several persona testers are listed at: navigators.com/persona.html**

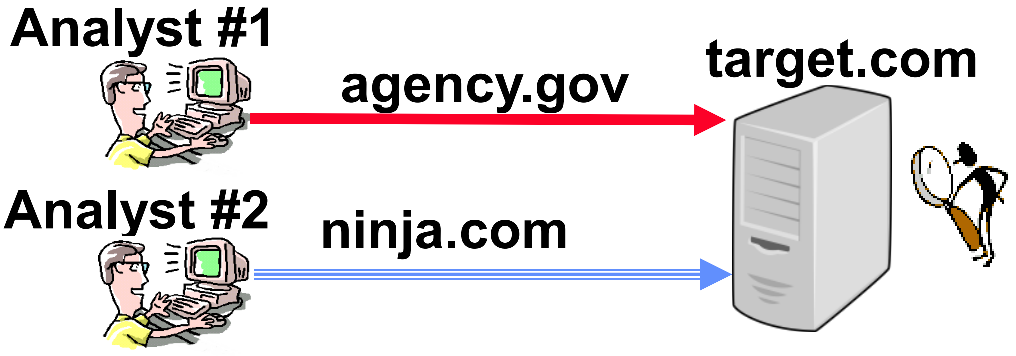
Exposing a “less recognizable” persona

━ Analyst #1: uses “agency.gov” persona to visit “targets”

═ Analyst #2: uses “ninja.com” persona to visit “targets”

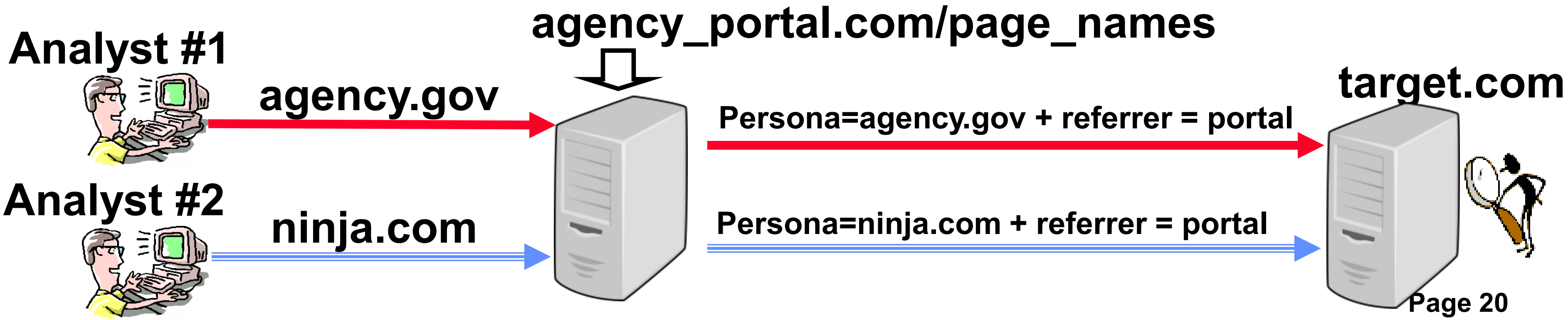
Result: “ninja” persona may be recognized as “agency.gov” visitor

The “parallel visit” Problem...



Even with no http_referer, a webmaster can still make the association due to high volume hits, usage patterns, software footprint, etc.

The “portal” Problem...


agency_portal.com/page_names
Analyst #1
agency.gov
Analyst #2
ninja.com
Persona=agency.gov + referrer = portal
Persona=ninja.com + referrer = portal
target.com

Course Outline

- Introduction to Internet Architecture
- “Persona” issues
- → Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

Online Web page = http://navigators.com/opensource.html

Course Exercises – Choose your Topic

Search for the same topic throughout the course

This enables comparison of results among the various search tools / techniques

Pick a topic you can focus on for 2 days

A combination of lecture, demo, and hands-on exercises will occur for each major search tool:

Lecture – I introduce the search tool/technique

(Please refrain from using your computer)

Demo - I demonstrate the tool/technique

(Please refrain from using your computer)

Individual search – You search your chosen topic

- Be an "explorer", not a "camper"

- Make bookmarks/favorites, and keep going

Plan Your Internet Research

- **Spell it Out - Define the topic, key words, acronyms, "what" and "who"**
- **Strategize - Choose the approach, online resources, specific search tools**
- **Search - Get online, stay focused, use advanced search features**
- **Sift - Filter the results, follow the leads**
- **Save – Make bookmarks, take notes, organize results, share with co-workers**

Spell out the topic...

1. Name of topic, what do you want to learn / desired end-goal

____

____

2. Spell out the topic (search terms, acronyms, abbreviations)

common, simple terms	obscure, specific terms
______________________	______________________
______________________	______________________
______________________	______________________
______________________	______________________

3. Make a list of "who" might publish such information (industry association, government agency, NGO's, user group, etc.)

____

____

Overview of Search Tools

Russ Haynal
Internet Instructor & Speaker
http://navigators.com/search_tool_intro.html

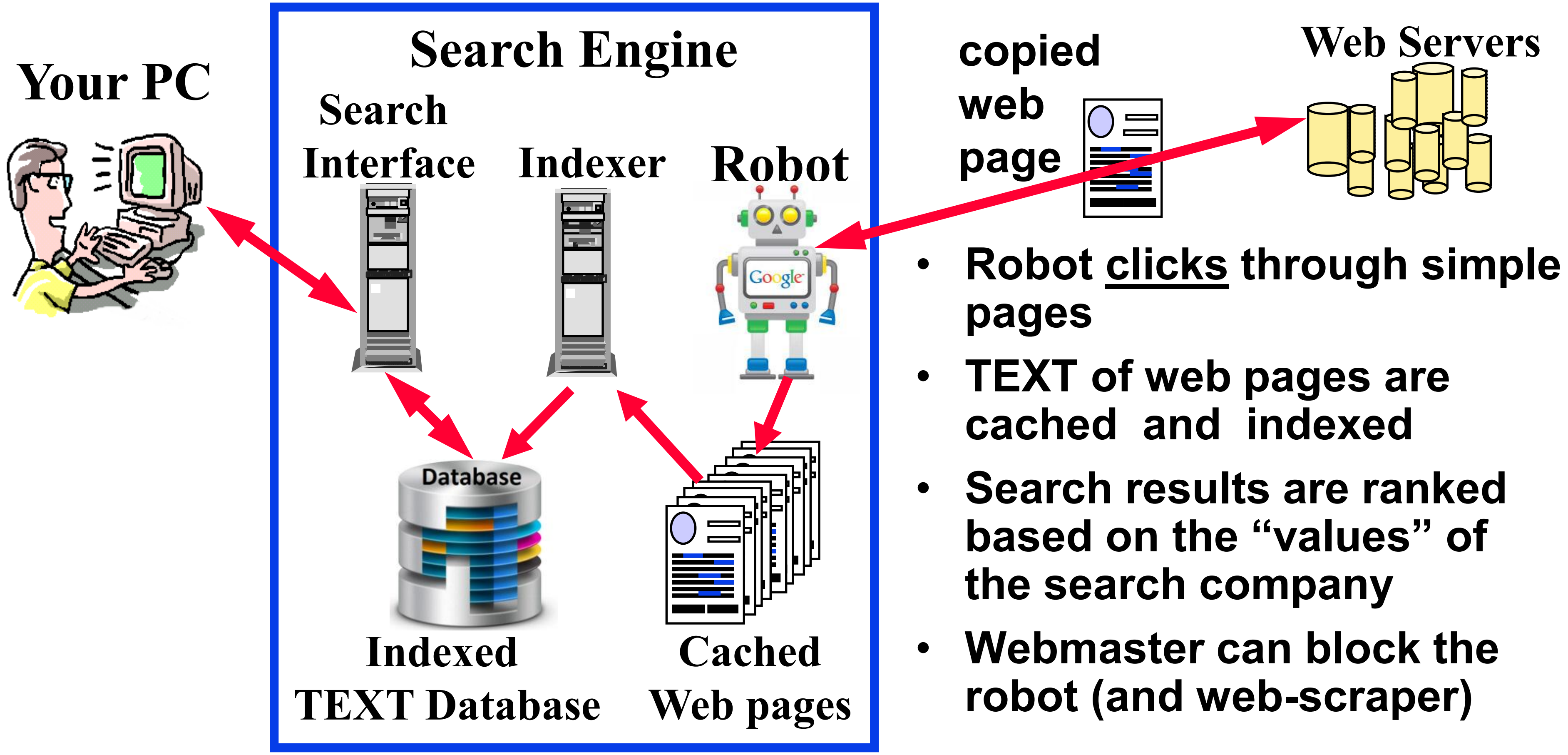
- **Search Engine** (Google, Bing)
  - large database – text from billions of clickable pages
- **“User Pages”** people who “care” about the topic
  - hundreds of topic-related: links, posts, documents
- **Specialized Tools**
  - database focused on a specific topic

Pick the right tool...

Every tool has strengths and weaknesses

Search Engines

(google.com , bing.com, ChatGPT)

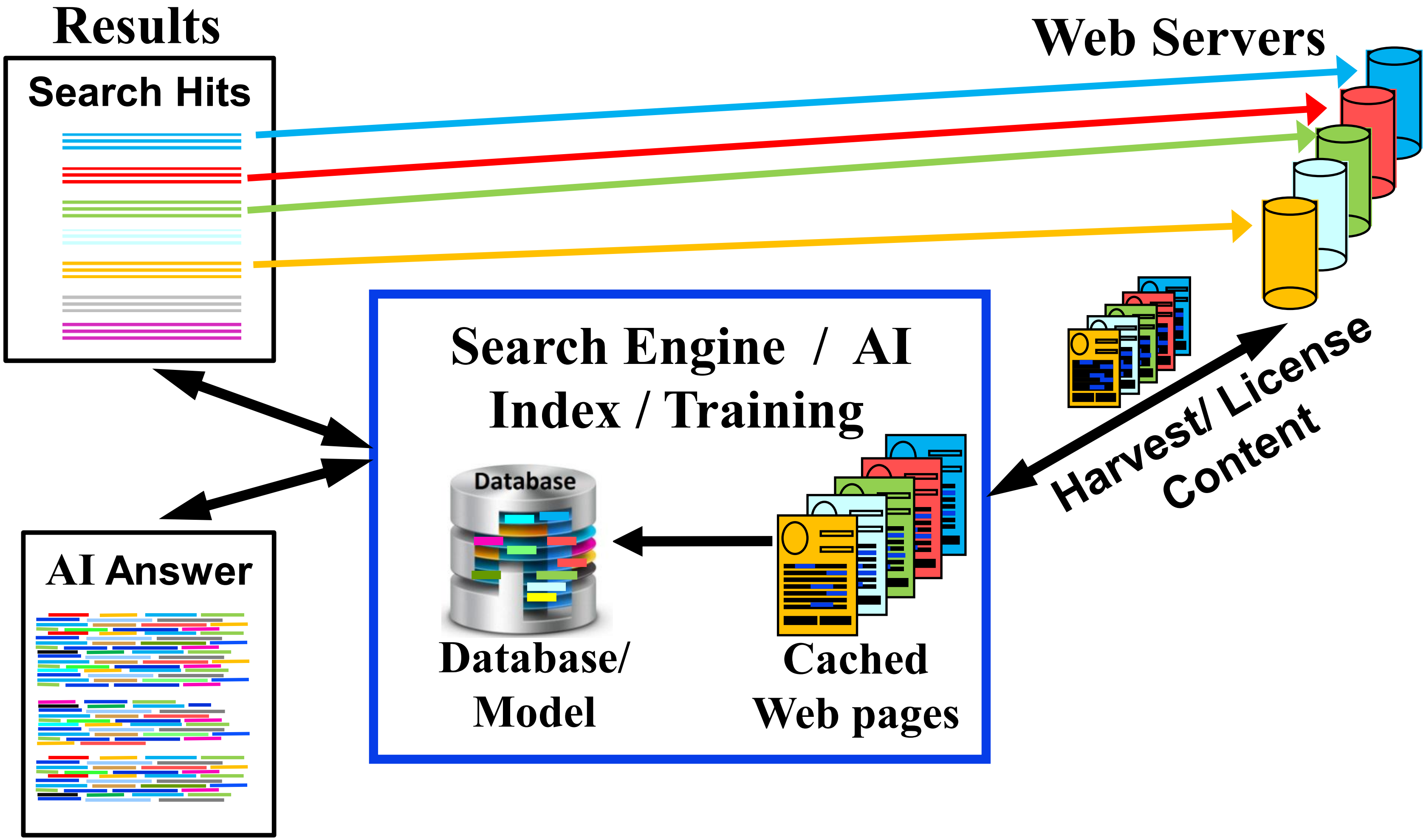


- Robot <u>clicks</u> through simple pages
- TEXT of web pages are cached and indexed
- Search results are ranked based on the “values” of the search company
- Webmaster can block the robot (and web-scraper)

Envision the target page “Use your imagination”

- Settings → “search help”, “search settings”, “advanced search”, etc.

Consider Economics of “search hits” vs “AI answers”



How will content creators make money?
(example: nytimes.com/robots.txt)

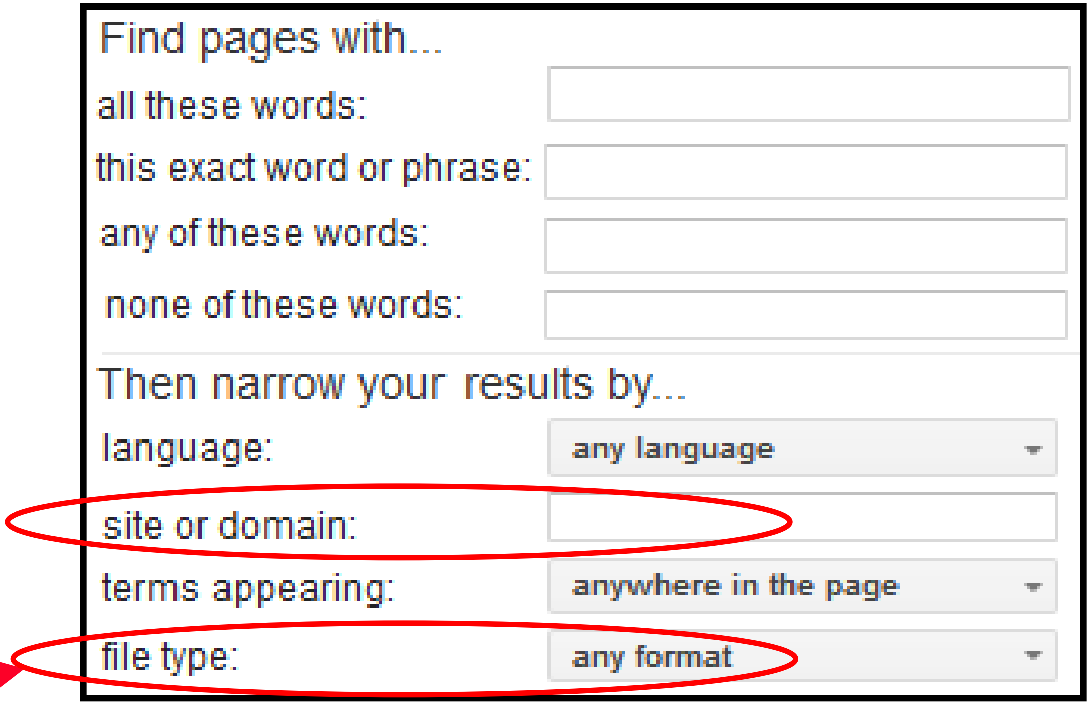
Class Exercise: Using a Search Engine

- Go to google.com and bing.com
- Enter identical terms into both search engines (make sure search terms remain unchanged)
- Look through the search results
  - Which gave more hits?
  - Are top-ten hits the same?
- Add additional specific search terms as needed to focus the search results
- Make bookmarks of useful sites

Advanced Search = Efficient Search !

basic search

advanced search


Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

Then narrow your results by...

language: any language

site or domain:

terms appearing: anywhere in the page

file type: any format

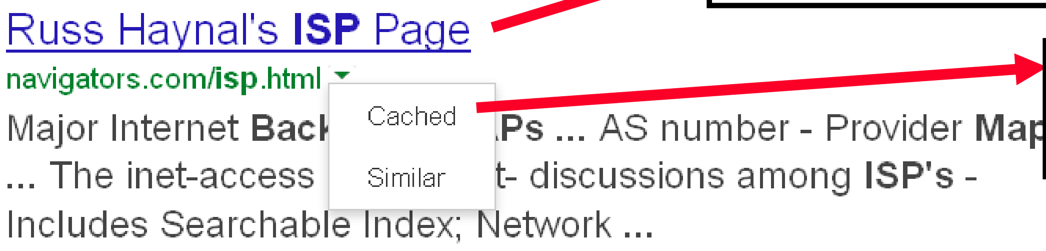
Limit search to specific sites or domains

filetype:pdf = detailed content from great web sites

- **Bottom right of Google home page: Settings → Advanced Search**
- **Top right of Google search results: ⚙ → Advanced Search**

OPSEC Concerns (OPerations SECurity)

Viewing cached page at Google = poor OPSEC



Russ Haynal's **ISP** Page

navigators.com/**isp**.html ▾

Major Internet **Back** … **Ps** ... AS number - Provider **Map** ... The inet-access … t- discussions among **ISP's** - Includes Searchable Index; Network ...

Target's site for text and media

Google for <u>text</u> of cached page + Target's site for embedded media

Viewing "text only" cached web page = improved OPSEC

1) Cut and paste this text into browser address bar:

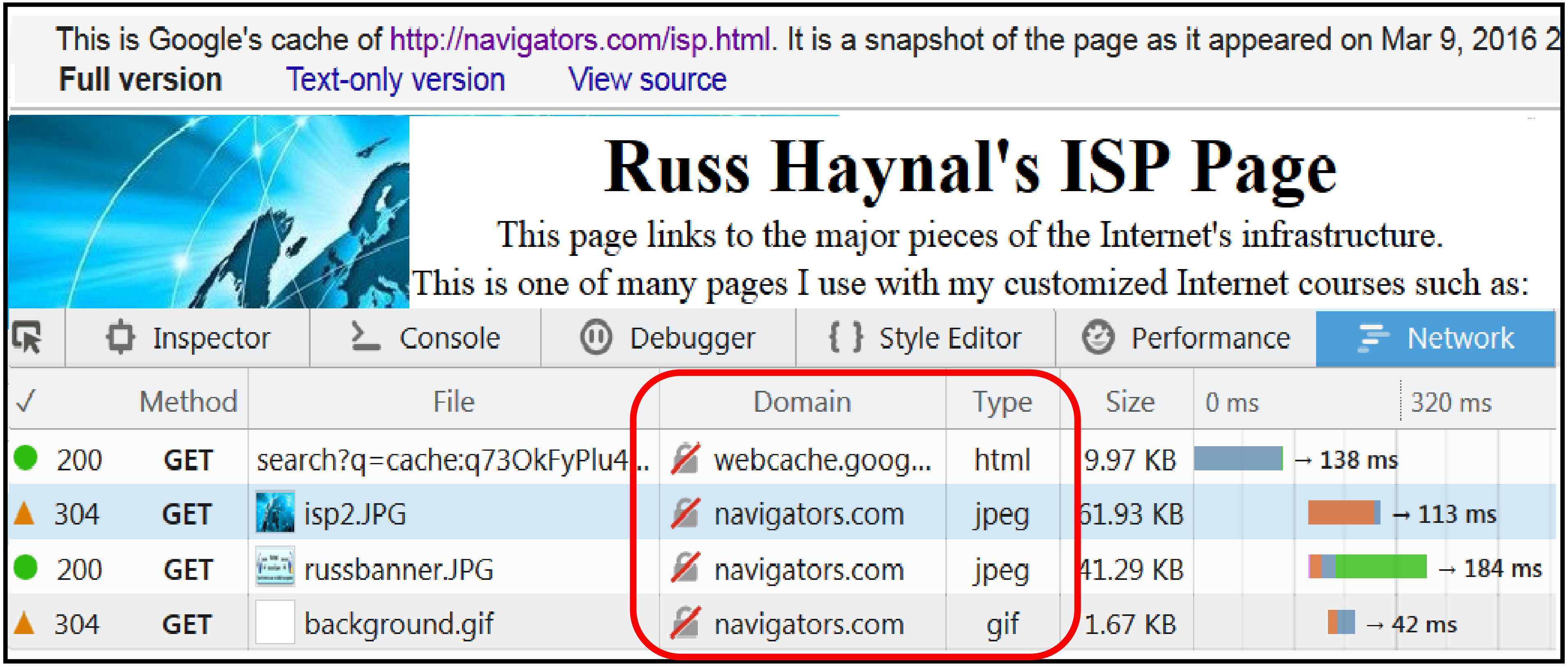
webcache.googleusercontent.com/search?strip=1&q=cache:

2) Add the target address onto the end of the above string:

webcache.googleusercontent.com/search?strip=1&q=cache:navigators.com/isp.html

Feb 2024: Google <u>removed</u> Cached link !!!
Several months later: webcache address stopped working

Monitoring the browser's activity



This is Google's cache of http://navigators.com/isp.html. It is a snapshot of the page as it appeared on Mar 9, 2016 2

Full version Text-only version View source

Russ Haynal's ISP Page

This page links to the major pieces of the Internet's infrastructure.
This is one of many pages I use with my customized Internet courses such as:

✓	Method	File	Domain	Type	Size	0 ms 320 ms
● 200	GET	search?q=cache:q73OkFyPlu4...	webcache.goog...	html	9.97 KB	→ 138 ms
▲ 304	GET	isp2.JPG	navigators.com	jpeg	61.93 KB	→ 113 ms
● 200	GET	russbanner.JPG	navigators.com	jpeg	41.29 KB	→ 184 ms
▲ 304	GET	background.gif	navigators.com	gif	1.67 KB	→ 42 ms

- **First line gets text (html) from webcache.google.com**
- **Next 3 lines get graphics (jpeg & gif) from navigators.com**

Firefox → Tools → Browser Tools→ Web Developer Tools → Network

CTRL – SHIFT – I

Course Outline

- Introduction to Internet Architecture
- “Persona” issues
- Search: Search Engines
- → Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

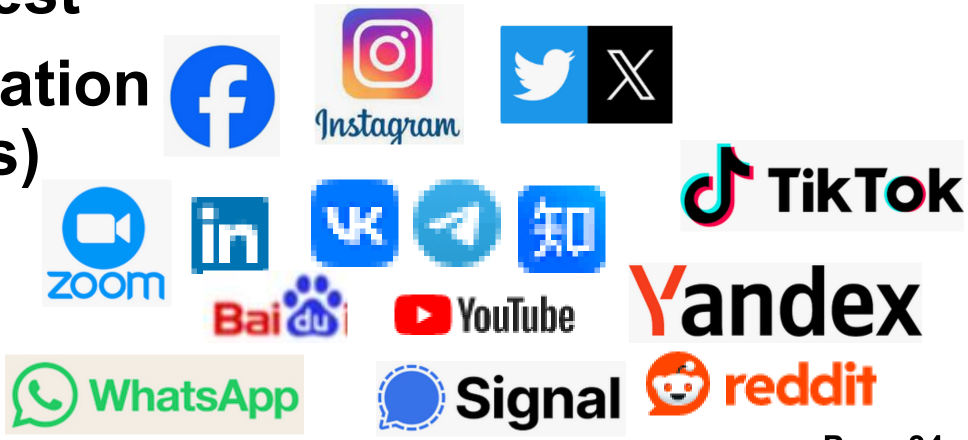
Online Web page = http://navigators.com/opensource.html

“User Pages”

- Focused on a specific subject
- Developed by “experts” in that field
(or a person with passion for the subject)
- Often contains “the best” online resources

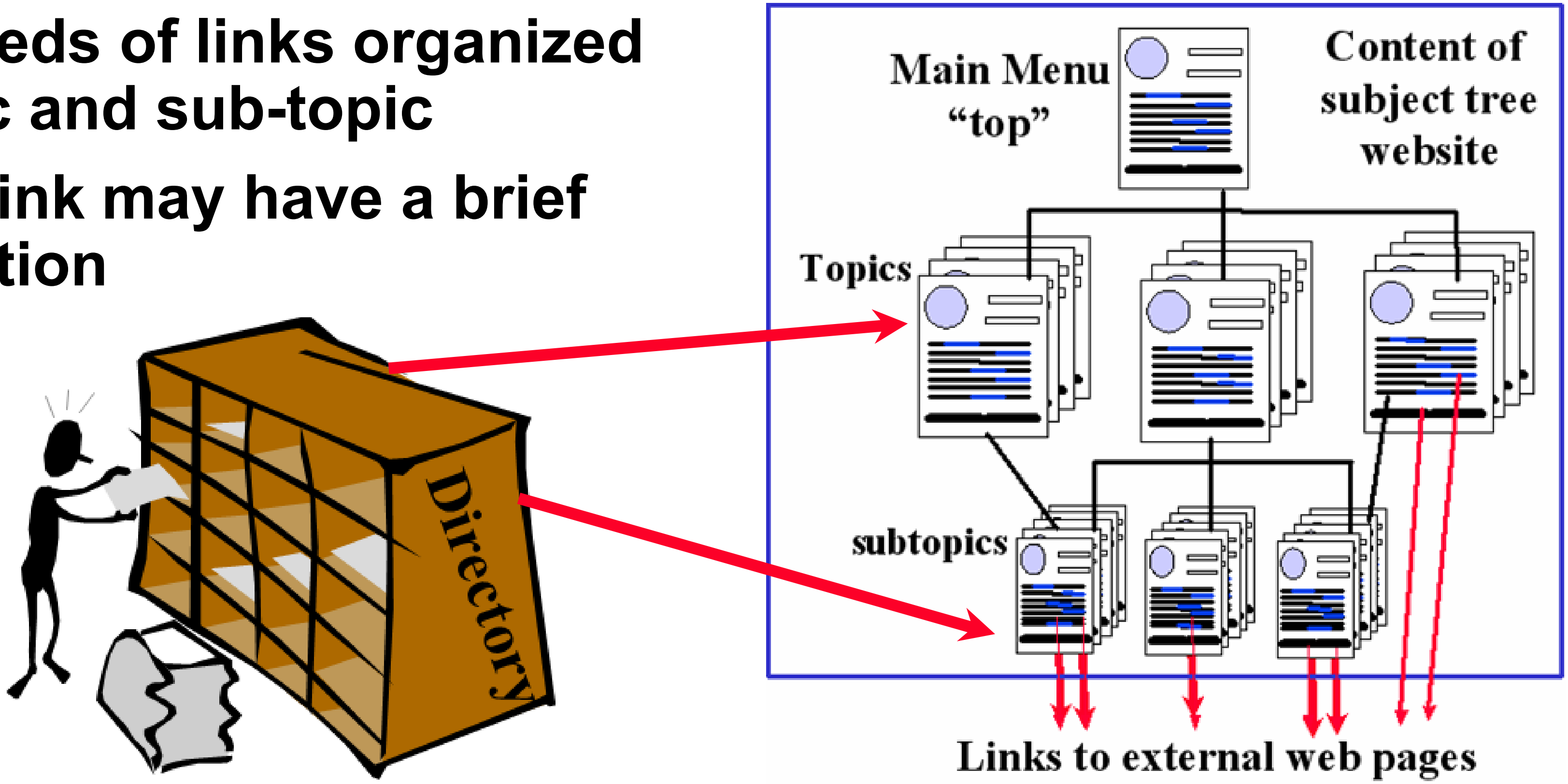
Finding “User pages”

- Subject directory
- Groups of users in a forum, conference, journal, club
- Contribute to wikipedia, wikimapia
- “User pages” point to other “user pages”
- Watch for sites labeled:
  “Joe’s ultimate guide to widgets”
- “Surfing Upstream” from several related sites
- Ask other researchers – there are several sites that everyone knows as “the best”
- Interactive, live communication
  (Chat, VOIP, virtual worlds)



Subject Directory

- Hundreds of links organized by topic and sub-topic
- Each link may have a brief description



Search: your_topic directory

e.g. golf course directory → www.thegolfcourses.net

public records directory → searchsystems.net

General directory: curlie.org

Finding Groups of Users

- Forum – discussion focused on a particular topic
- Many users can participate by posting messages
- Search : your_detailed_topic forum post replies
  = threads and posts that discuss your topic

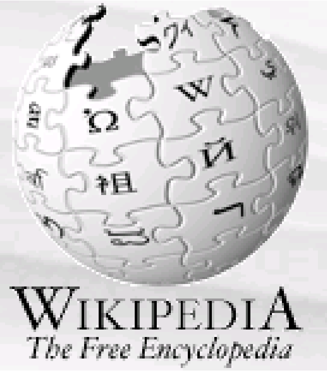
- Other ways for users to communicate/collaborate…
- Gatherings: conference , convention, symposium , summit, seminar, expo, “trade show”, festival , workshop
- Publications: journal , magazine , “white paper”, thesis
- Membership: consortium, association, federation, society, club, league, “user group”, alumni
- Search: your_topic conference , your_topic festival, journal, etc.

- Individual: resume, Curriculum Vitae, CV, biography

Reminder: membership requirements are a barrier to search engines

Wiki ____

- A Wiki allows immediate creation and editing of pages by “anyone”



- Wikipedia.org – encyclopedia that can be instantly edited by ANY Internet user
- Good starting point for many subjects to gain an overview of the topic
- Page can be biased from the most recent editor
- Some entries get “locked-down” due to vandalism

wikimapia ← Owners based in Russia

- old.wikimapia.org – same concept applied to maps
- “map type”→ google map: zoom to the right location
- “map type” → “wikimapia classic” : to see comments
- To learn about the author: click on a comment box:
  menu → history → the user’s name → stats →
  click on the statistics numbers = places that user has added/edited

Will The Target Notice You?

Assess a website's popularity and demographics before visiting the site

- How many hits can be made on the target's webserver, without causing a noticeable spike in their traffic?
- What geographic persona and software persona hits are most commonly occurring on target's webserver?
- What 1-click history could be "leaked on purpose" and not raise suspicions?
- Most analytic sites are expensive, but some offer enough free statistics to be very useful for tradecraft purposes:

Alexa
Discontinued May 1, 2022

SimilarWeb
Headquarters in Israel

Cloudflare Radar
Headquarters in San Francisco

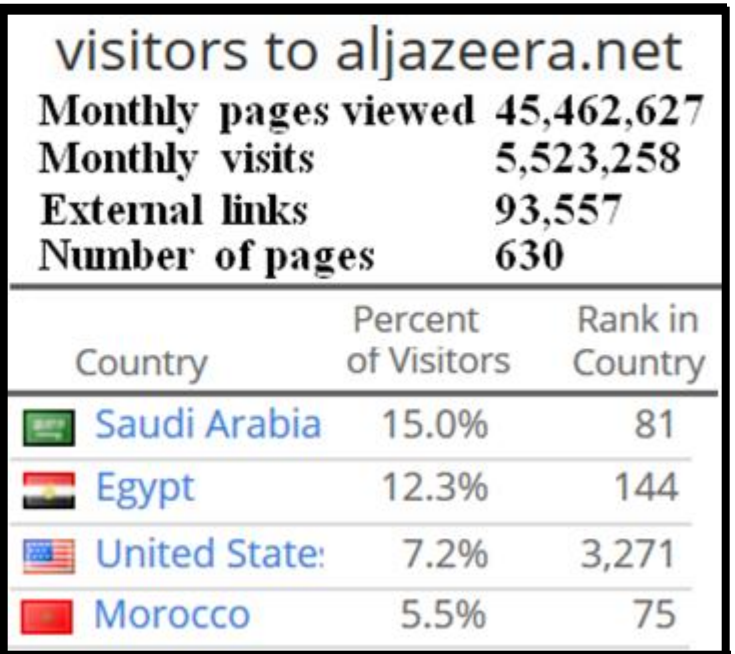
Web Analytics

similarweb.com/website　　**radar.cloudflare.com**

Each of these tools offer a sampling of analytics for free:

- **Popularity of a web site**
- **Audience demographics**
- **Search terms used to find the site**
- **Visitor engagement levels**
- **Traffic history**
- **Related sites = more sites**

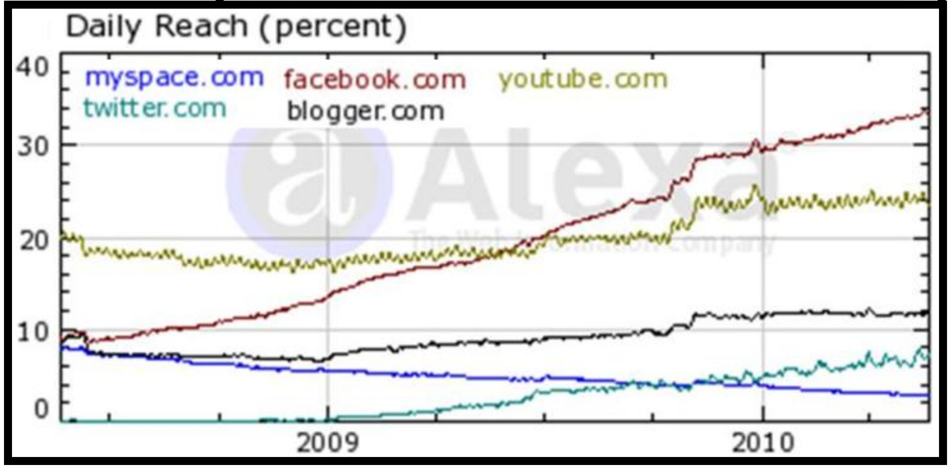
Enter a domain name (not search terms)



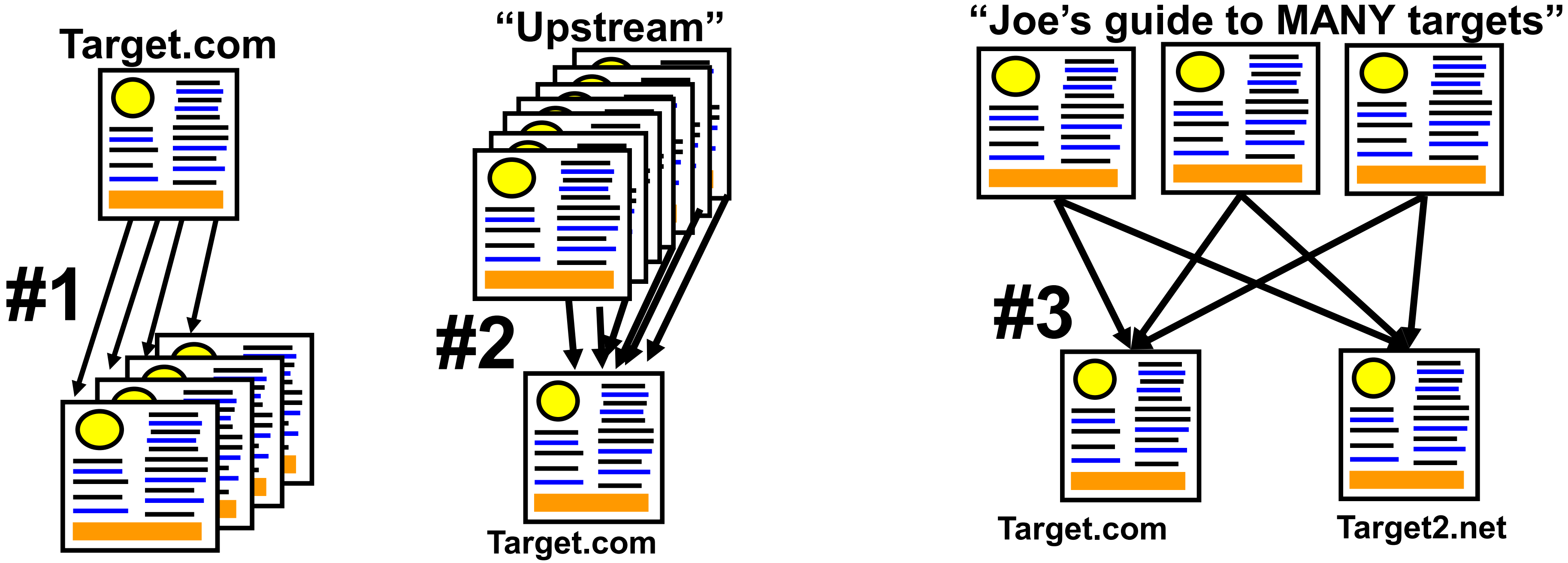
visitors to aljazeera.net

Monthly pages viewed	45,462,627
Monthly visits	5,523,258
External links	93,557
Number of pages	630

Country	Percent of Visitors	Rank in Country
Saudi Arabia	15.0%	81
Egypt	12.3%	144
United State:	7.2%	3,271
Morocco	5.5%	75



Surfing Upstream vs. Downstream



Target.com

#1

“Upstream”

#2

Target.com

“Joe’s guide to MANY targets”

#3

Target.com

Target2.net

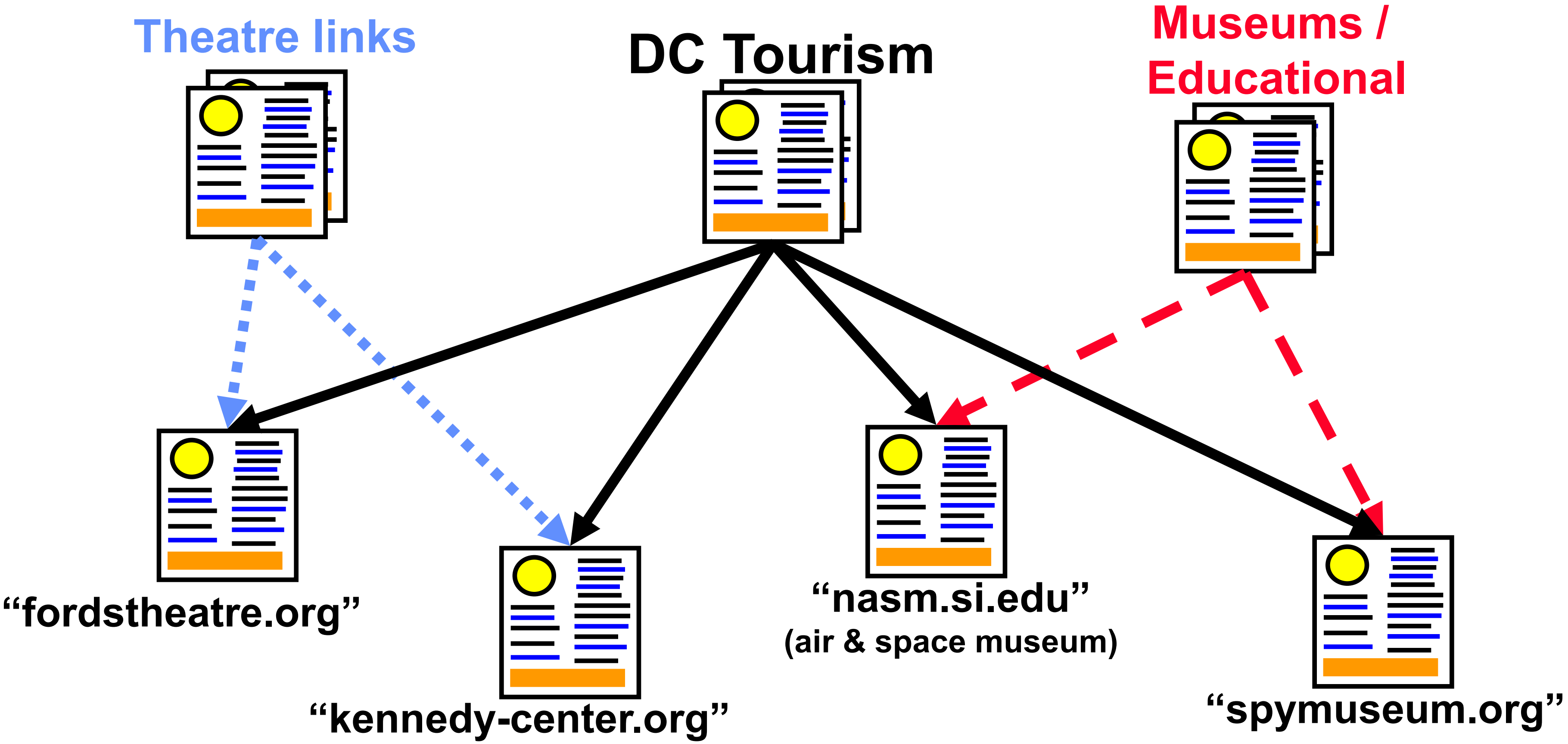
#1 Most researchers follow the links “downstream” from an interesting page

#2 Shows pages that link <u>towards</u> the target (=upstream)
This is an Indication of the page’s “popularity”
= who knows about target.com

#3 Shows pages that link to both target sites …
= “user pages” for that topic

Be Creative When Surfing Upstream

Example: Washington DC Tourist Sites



- **Any combination of these target pages will lead to "DC Tourism" pages, but certain pairings may also lead to subject-specific pages**

Surfing Upstream Details

search format at google or bing	search results
“www.example.com”	contain text: www.example.com
“www.example.com/pageA.html”	contain text of the specific page address
+“www.example1.com” +“www.example2.com”	contain text of both web site addresses This is a great way to discover “user pages” (e.g. Joe's guide to many example-sites)

- Which scenario makes more sense for your scenario; Row #1 or Row #2
  e.g. who links to the home page of the entire site vs
  who links to a specific webpage within the site
- A 3rd and 4th site can be added if they are popular enough
- Note: do not include “http://” (can also omit www)
- Who links to: 2 gov agencies, 2 companies, 2 conferences, 2 technical journals, 2 phone books, 2 hacker sites, 2 social media search tools, etc

Searching Within
a Site or Domain Name

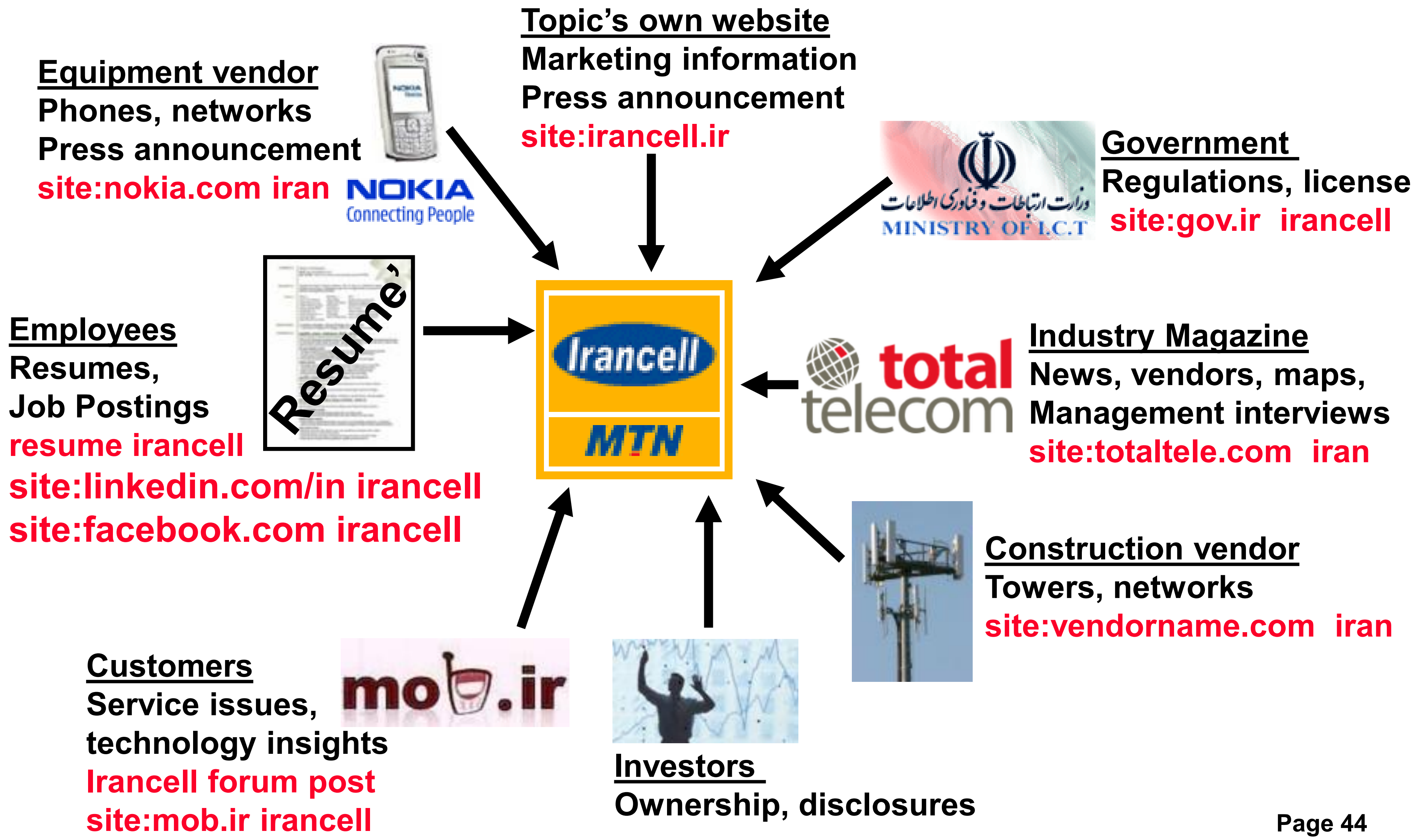
search format at Google	search results
site:example.com	pages hosted on any example.com servers (www.example.com, blog.example.com, etc) = quick way to assess the public size/depth of a domain
site:example.com searchterm	pages hosted at example.com which mention "searchterm"
site:ru searchterm	pages hosted on .ru servers which mention "searchterm"
site:ac.ru nuclear	pages hosted on any academic .Russian servers which mention nuclear
site:iaea.org iran filetype:pdf	PDF documents hosted at iaea web servers which mention iran
site:linkedin.com/in topics	Individual Linkedin profiles that mention your topics

- Faster than reading thousands of pages from a large site
- No space after site:
- Do not include “http://” or “www”
- “use your imagination” to focus these searches

Who Knows About Your Topic?

(Google search terms are in red)

Example: Iranian cell phone company (Irancell-MTN)



Topic's own website
Marketing information
Press announcement
site:irancell.ir

Equipment vendor
Phones, networks
Press announcement
site:nokia.com iran

Government
Regulations, license
site:gov.ir irancell

Employees
Resumes,
Job Postings
resume irancell
site:linkedin.com/in irancell
site:facebook.com irancell

Industry Magazine
News, vendors, maps,
Management interviews
site:totaltele.com iran

Construction vendor
Towers, networks
site:vendorname.com iran

Customers
Service issues,
technology insights
Irancell forum post
site:mob.ir irancell

Investors
Ownership, disclosures

Cautions about Social Media

- Confirm policies for viewing, joining, or interacting on social media
- Understand each site's different levels of interactions:
  - viewing, following, group member, connecting, friend, messaging
- What information is shared to the other end user?
- What information is shared with 3rd party advertisers / data brokers?
- ALL interactions are known to the owner of the social media site
  --> learn who owns the site
- Who has "jurisdiction" over the site? (VK → Russia, QQ → China)

PREMIUM

Russ Haynal

Instructor -> Internet: OSINT / Tradecraft / Cyber Security Awareness

92
Who's viewed your profile

365
Views of your post

- Linkedin example:
- Different membership levels have various capabilities
- free ($0/month), premium, premium personal, premium career, sales navigator, recruiter lite, recruiter ($900/month)
- "recruiter" has unlimited access to everyone's full profiles, and leaves no "hits" on the people they view

Free account = YOU are the "product" being sold!

Alternative Techniques

Web analytics: Assess Popularity, choose "access point", 1-click history

- = your browser → google.com or bing.com
- "target.com" –site:target.com = # of outside pages" pointing to target.com
- "target.com" site:de = # of .de pages pointing to target.com
- "target.com" site:fr = # of .fr pages pointing to target.com

OPSEC Issues…

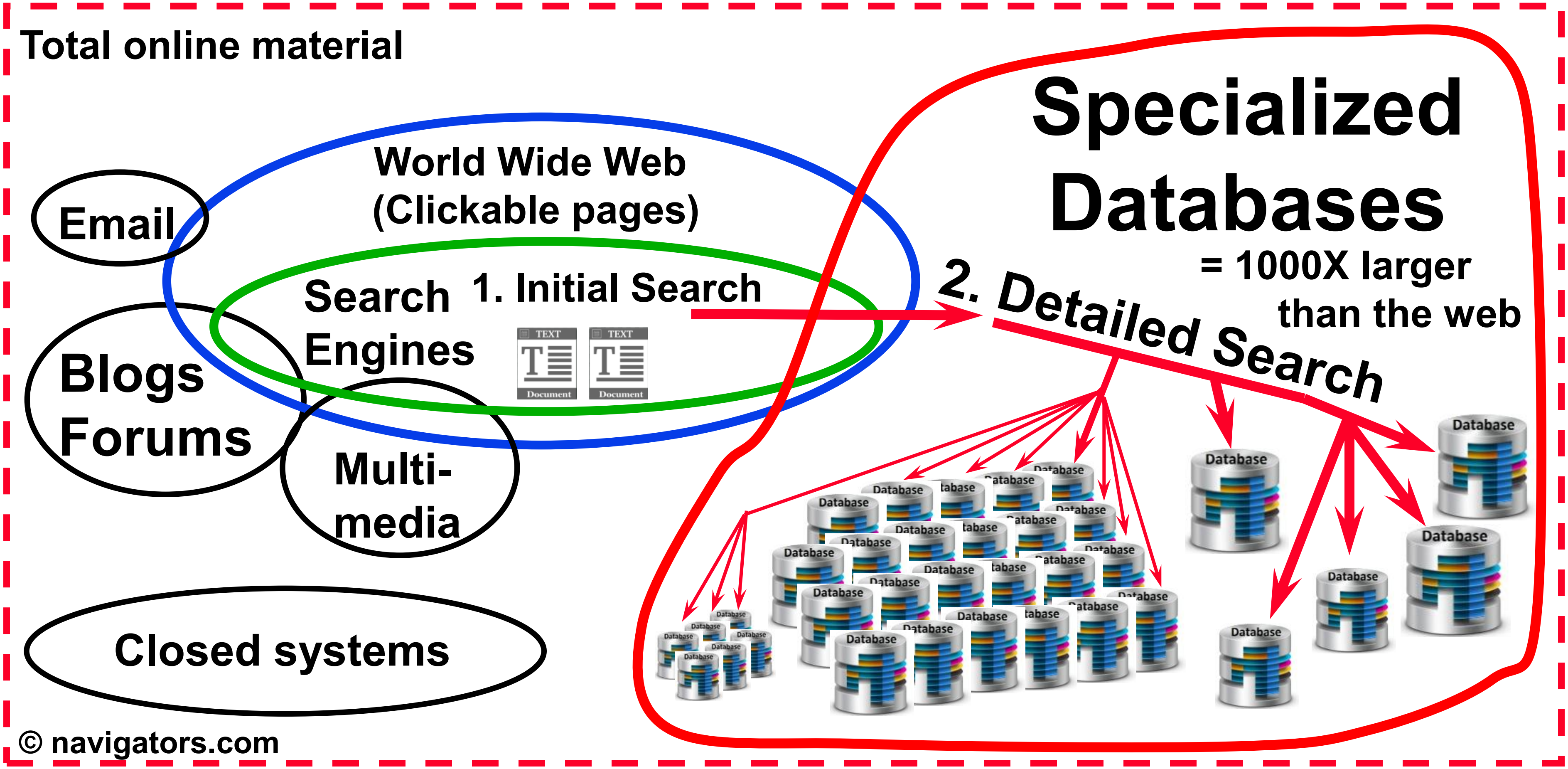
- Mobile Friendly test (https://search.google.com/test/mobile-friendly)
  - "Testing Live URL" = Google mobile tester→ target.com
- HTML viewer (https://htmledit.squarefree.com)
  - Cut and paste HTML from another source (e.g. Mobile friendly test)
  - = your browser → target.com hits for graphics, etc!!!
- https://pdfmyurl.com/ = PDF My URL → target.com
- https://archive.org = way back machine → target.com/robots.txt

Course Outline

- Introduction to Internet Architecture
- “Persona” issues
- Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

Online Web page = http://navigators.com/opensource.html

The "clickable web" is TINY



- Many detailed searches are a two-step process
  - find the specialized database
  - then type appropriate query into that database

Lists of Databases

- For specific info, use a specialized database
- Over 100,000 specialty databases

SearchSystems.net

70,000 public record databases

- Search for the organization that would host the specialized database
- Try searching: your_topic database
  - patent database → patft.uspto.gov
  - arms transfer database → sipri.org/databases/armstransfers
  - fish database → fishbase.org

Specialized databases contain content that search engines can't reach

Business databases can be quite useful

U.S. SECURITIES AND EXCHANGE COMMISSION • MCMXXXIV •

Home | Previous Page

U.S. Securities and Exchange Commission

dun & bradstreet

KOMPASS
Connects business to business

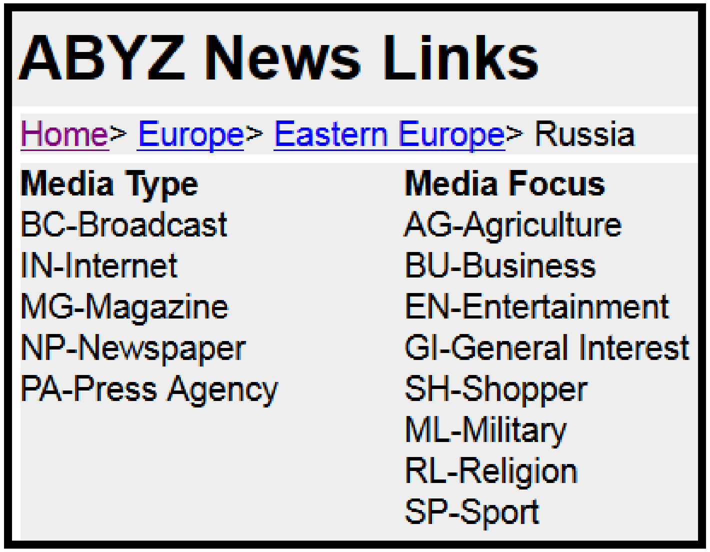
- Most publicly held companies are required to file financial statements with the Securities Exchange Commission
- These filings are accessible to the public through the SEC's EDGAR database
- READ forms 10-Q and 10-K (quarterly and annual report) Detailed reports about the company's activities, plans, sales, etc
- Seek out other business databases: financial, investment, government regulatory, etc
- Databases may be available at your library (internal or public)

Many country resources are online

Assess popularity of resources using web analytics

If necessary use site:

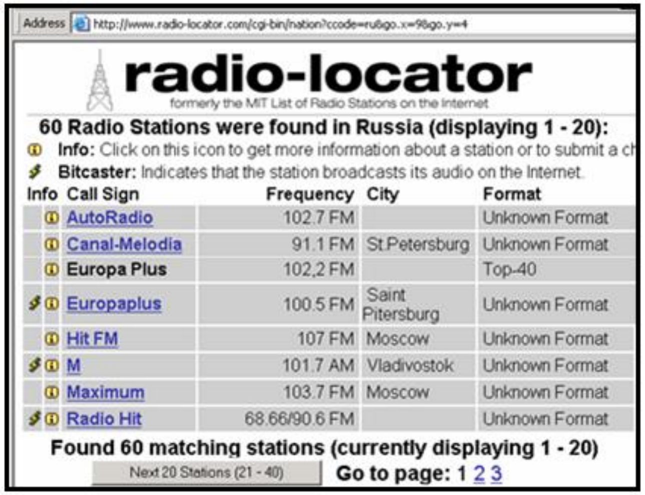
wayp.com



ABYZ News Links

Home> Europe> Eastern Europe> Russia

Media Type	Media Focus
BC-Broadcast	AG-Agriculture
IN-Internet	BU-Business
MG-Magazine	EN-Entertainment
NP-Newspaper	GI-General Interest
PA-Press Agency	SH-Shopper
	ML-Military
	RL-Religion
	SP-Sport


Address
http://www.radio-locator.com/cgi-bin/nation?ccode=ru&go.x=9&go.y=4

radio-locator

formerly the MIT List of Radio Stations on the Internet

60 Radio Stations were found in Russia (displaying 1 - 20):

ⓘ **Info:** Click on this icon to get more information about a station or to submit a c

⚡ **Bitcaster:** Indicates that the station broadcasts its audio on the Internet.

Info	Call Sign	Frequency	City	Format
ⓘ	AutoRadio	102.7 FM		Unknown Format
ⓘ	Canal-Melodia	91.1 FM	St.Petersburg	Unknown Format
ⓘ	Europa Plus	102,2 FM		Top-40
⚡ ⓘ	Europaplus	100.5 FM	Saint Pitersburg	Unknown Format
ⓘ	Hit FM	107 FM	Moscow	Unknown Format
⚡ ⓘ	M	101.7 AM	Vladivostok	Unknown Format
ⓘ	Maximum	103.7 FM	Moscow	Unknown Format
⚡ ⓘ	Radio Hit	68.66/90.6 FM		Unknown Format

Found 60 matching stations (currently displaying 1 - 20)

Go to page: 1 2 3

Lists of OSINT Resources

“OSINT Resources” can be found using Hidden Universes techniques: (eg. surfing upstream, filetype:pdf, OSINT guide, OSINT Handbook, toolkit, conference, journal)

https://i-intelligence.eu/uploads/public-documents/OSINT_Handbook_2020.pdf
= 500 pages of OSINT hyperlinks

I-INTELLIGENCE

OPEN SOURCE INTELLIGENCE
TOOLS AND RESOURCES HANDBOOK
2020

https://metaosint.github.io = Directory of 4,000 OSINT resources
https://metaosint.github.io/chart ← explore visually
https://metaosint.github.io/table ← searchable text listings

https://bit.ly/bcattools = Bellingcat’s online investigation toolkit
(click on tabs along bottom of spreadsheet)

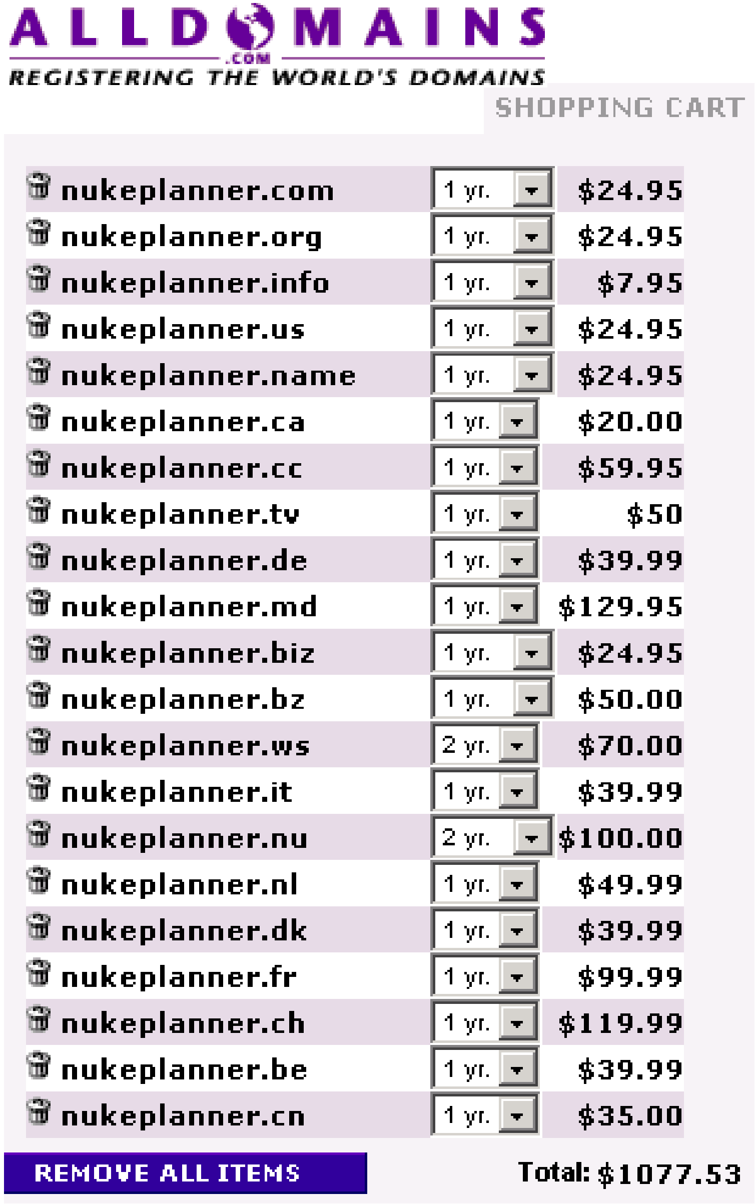
https://inteltechniques.com/tools/index.html = A supplement to book:
OSINT Techniques by Michael Bazzell

Course Outline

- Introduction to Internet Architecture
- “Persona” issues
- Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

Online Web page = http://navigators.com/opensource.html

Most Countries Sell Their Domains



ALLDOMAINS .COM
REGISTERING THE WORLD'S DOMAINS

SHOPPING CART

nukeplanner.com	1 yr.	$24.95
nukeplanner.org	1 yr.	$24.95
nukeplanner.info	1 yr.	$7.95
nukeplanner.us	1 yr.	$24.95
nukeplanner.name	1 yr.	$24.95
nukeplanner.ca	1 yr.	$20.00
nukeplanner.cc	1 yr.	$59.95
nukeplanner.tv	1 yr.	$50
nukeplanner.de	1 yr.	$39.99
nukeplanner.md	1 yr.	$129.95
nukeplanner.biz	1 yr.	$24.95
nukeplanner.bz	1 yr.	$50.00
nukeplanner.ws	2 yr.	$70.00
nukeplanner.it	1 yr.	$39.99
nukeplanner.nu	2 yr.	$100.00
nukeplanner.nl	1 yr.	$49.99
nukeplanner.dk	1 yr.	$39.99
nukeplanner.fr	1 yr.	$99.99
nukeplanner.ch	1 yr.	$119.99
nukeplanner.be	1 yr.	$39.99
nukeplanner.cn	1 yr.	$35.00

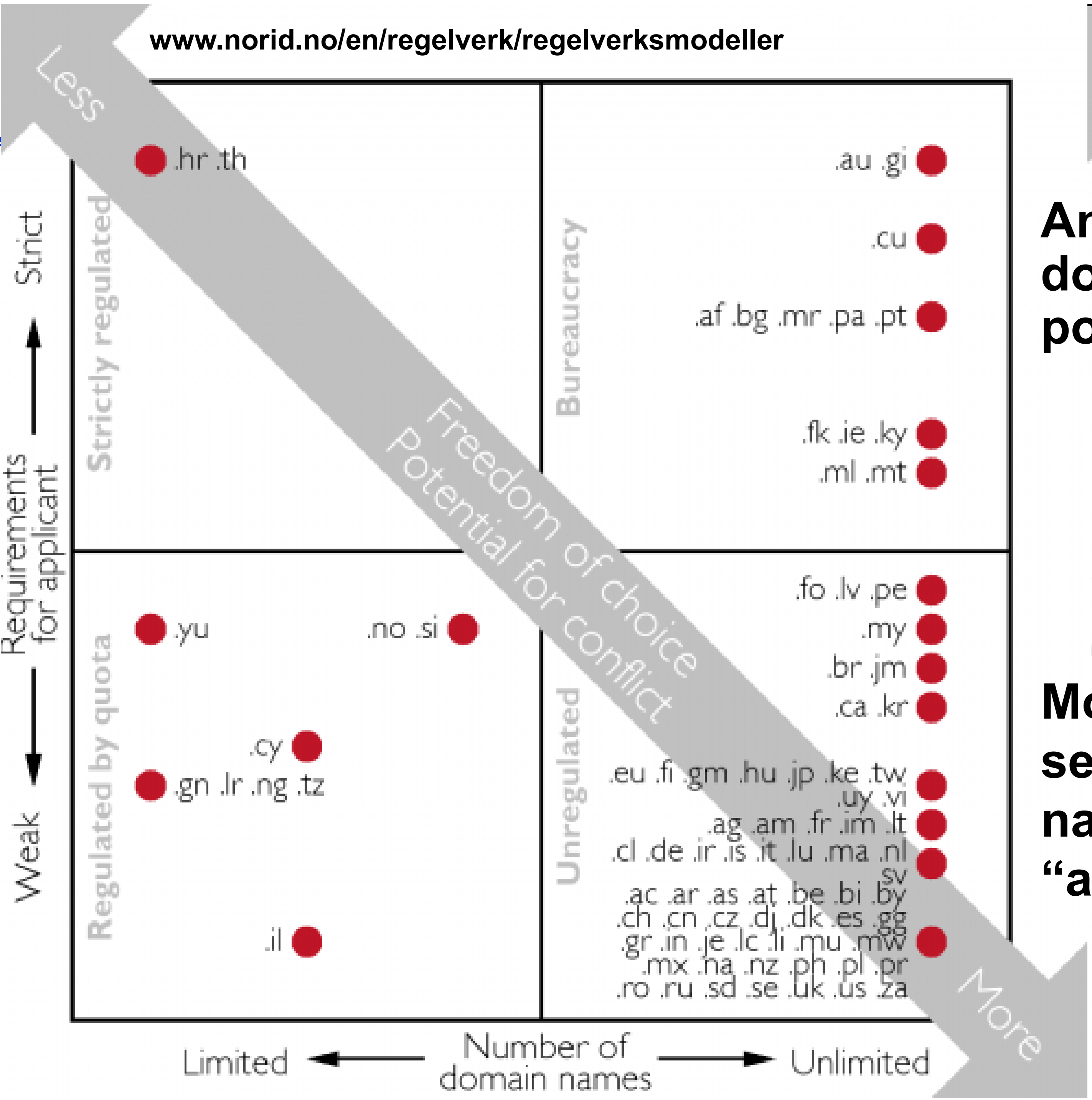
Total: $1077.53

- These were just some of the country domains available for sale
- “All Domains” happened to be a licensed “registrar” for these countries
- Most countries sell their domain names to “anyone”

Learn About the 2-letter code

- Visit your county's domain name registrar
  - iana.org/domains/root/db
- What is the policy for getting a domain name? (citizenship, trademark, local presence, money)
  - What is the cost to register a domain name?
  - Are there any censorship clauses?
- Does the registrar require any proof of identity? (drivers license, passport, business license)
- Is there a whois service? (make a bookmark)

An analysis of domain name policies



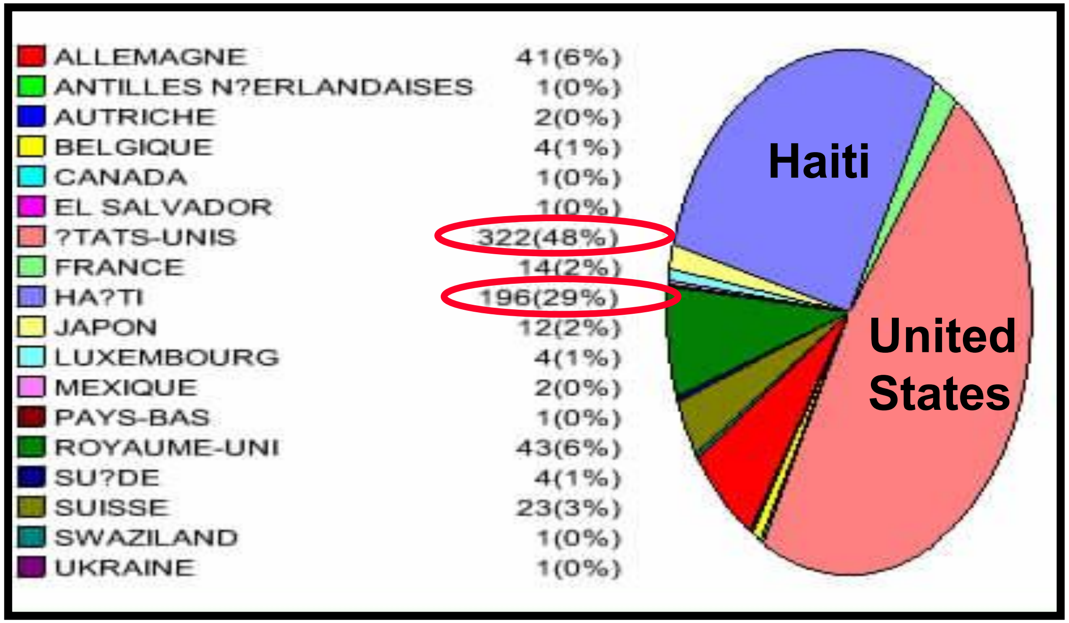
www.norid.no/en/regelverk/regelverksmodeller

Most countries sell their domain names to "anybody"

Domain Names for Sale

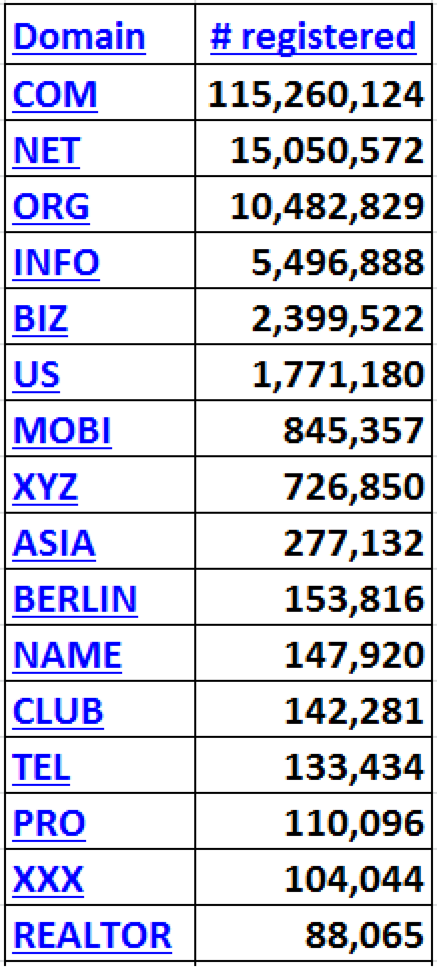
- Only 29% .HT domain names were registered to people with a Haitian address
- 48% of Haiti’s Domain names were registered to U.S addresses
- When you see a .ht website… is it necessarily foreign?

Postal address for .HT Domain Owners



New Generic Top-Level Domains

ICANN



Domain	# registered
COM	115,260,124
NET	15,050,572
ORG	10,482,829
INFO	5,496,888
BIZ	2,399,522
US	1,771,180
MOBI	845,357
XYZ	726,850
ASIA	277,132
BERLIN	153,816
NAME	147,920
CLUB	142,281
TEL	133,434
PRO	110,096
XXX	104,044
REALTOR	88,065

Source Evaluation

- Pick apart the URL: protocol://computer.domain.name/pathname/filename.ext
- Determine where "ownership" of the web page begins
  - www.facebook.com/joesmith/info.html
  - www.joesmith.com/stuff/info.html
- Browse the directories (shorten URL if necessary)
- Look at domain's home page - Is it a web hosting site? Is "pathname" a user account?
- IF the domain home page looks like the "owner" of the content, then move forward with whois and traceroute

Source Evaluation - Using WHOIS

- Domain names are “registered” at Internet registrars (global, country-specific)
- Each registrar develops its own policies
  - may sell to anyone/anywhere (.com, .org, .net, .tv, .pk)
  - may have strict qualification requirements (.gov, .mil, .au)
- Registrants provide “point of contact” information, for at least invoicing purposes
- Domain “point of contact” information is often available from the registrars’ database via a “WHOIS” query
- WHOIS contents may be inaccurate, although usually the email, or postal address will be correct to receive renewal invoice

Performing a “Whois” Query

- “whois” reveals the “owner” of a domain (searchenginewatch.com)

```
Administrative contact: Ron Doobay
HAYMARKET HOUSE
28-29 HAYMARKET
LONDON SW1Y 4RX UK
+44.2074849700        +44.2079302238
dns@incisivemedia.com

Technical contact: Domain Administrator
3rd Floor Prospero House
241 Borough High Street
Borough London SE1 1GA UK
+44.2070159370       +44.2070159375
corporate-services@netnames.com

Created on: 1998-03-20
Expires on: 2026-03-19

Domain name servers:
NS3.INCBASE.NET 85.133.68.200
NS2.INCBASE.NET 62.140.213.136
NS1.INCBASE.NET 62.140.213.135
```

- Spam concerns caused many domain names being registered via “privacy enhanced” options
- EU GDPR Law in 2018 impacts WHOIS records for Europeans (General Data Protection Regulation)

Traceroute

- Shows a network path between 2 machines
- Traceroute designed to help de-bug network connections
- Can initiate traceroute from your workstation, or from public “traceroute servers” located throughout the Internet
- Each Internet provider has their own naming convention for their infrastructure
  - Location labels: City names or 3-letter airport codes
  - Exchange points (LINX, HKIX, AMS-IX)
  - Infrastructure Topology (T3, FDDI, GE, SMW3)
- A website can be hosted anywhere
  - Could be at organizations’ building,
    or more likely at a well-connected hosting facility

Results of Traceroute

```
traceroute output from WWW.Telcom.Arizona.EDU to www.nsa.gov:
1 128.196.128.253 (128.196.128.253) 1 ms
2 192.80.43.25 (192.80.43.25) 1 ms
3 192.80.43.58 (192.80.43.58) 1 ms
4 207.250.65.133 (207.250.65.133) 5 ms
5 core-01-ge.phnx.twtelecom.net (209.234.146.45) 5 ms
6 core-02-so.lsag.twtelecom.net (168.215.53.73) 17 ms
7 tran-01-ge.lsag.twtelecom.net (168.215.54.98) 17 ms
8 POS1-1.GW3.LAX1.ALTER.NET (208.222.8.245)  17 ms
9 CL2.LAX4.ALTER.NET (152.63.52.246) 18 ms
10 TL2.LAX9.ALTER.NET (152.63.115.146) 18 ms
11 so.TL2.DCA8.ALTER.NET (152.63.3.193)  74 ms
12 so.XL2.DCA8.ALTER.NET (152.63.35.250) 74 ms
13 ATM6-0.GW3.BWI1.ALTER.NET (152.63.39.41) 76 ms
14 * * *
15 * * *
```

Time-Warner and Alternet swap traffic at Los Angeles

Baltimore airport code

Traceroute helps reveal the dynamic architecture of the Internet

Try different starting points for Traceroutes

Starting from Arizona University

```
 1 128.196.128.253  0 ms
2 192.80.43.25 0 ms
3 192.80.43.58 1 ms
4 207.250.65.133 5 ms
5 core-02-ge.phnx.twtelecom.net 5 ms
 6 core-02-so.chcg.twtelecom.net 46 ms
7 peer-01-ge.chcg.twtelecom.net 46 ms
8 aads.verio.net 47 ms
9  chcgil01.us.bb.verio.net 47 ms
10 chcgil06.us.bb.verio.net  47 ms
11 dllstx01.us.bb.verio.net 47 ms
13 stngva01.us.bb.verio.net 82 ms
17 navigators.com  82 ms
```

Times are real-time round trip measurements from step 1 to step #_

Starting From University of Maryland

```
1 Vlan5.css-core-r1.net.umd.edu 0.53 ms
2 128.8.1.222 0.43 ms
3 qwest-bdr.net.umd.edu 1.49 ms
4 63-237-64-1.cust.qwest.net 1.38 ms
6 dca-brdr.inet.qwest.net 1.48 ms
7 qwest.stngva01.us.bb.verio.net 2.45 ms
9 ge.stngva01.us.verio.net 3.09 ms
10 stngva01.us.verio.net  2.75 ms
11 navigators.com 2.48 ms
```

Navigators.com "must" be near University of Maryland's server
2.48 x 120 / 2 = ~150 miles

The speed of light can serve as a yardstick in traceroutes

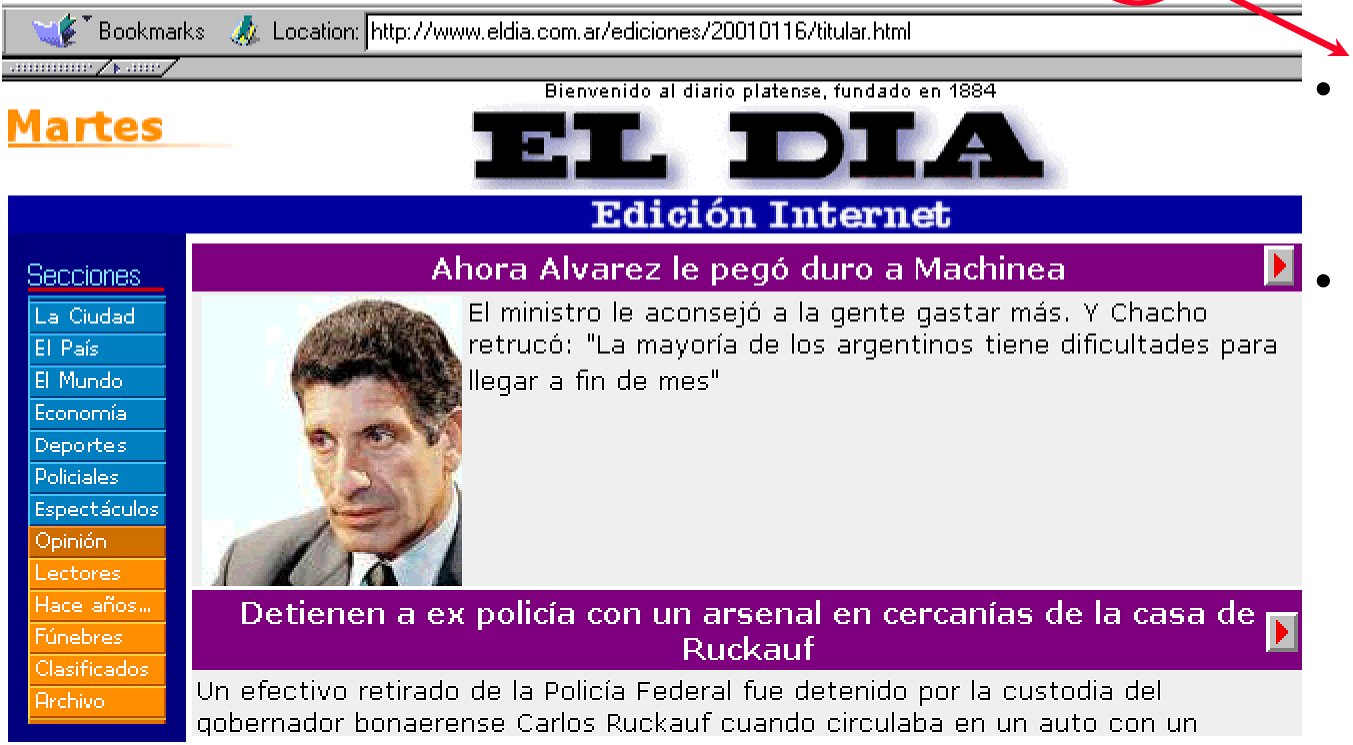
Speed of light:

186,000 miles/sec (in vacuum)
120,000 miles/sec (in glass fiber)
= 120 miles/ms (in glass fiber)

Note: Each hop via geostationary satellite must take at least 240 ms
Low-Earth satellites can have lower latencies than terrestrial networks

A Foreign Newspaper ???

URL = http://www.eldia.com.ar



http://www.eldia.com.ar/ediciones/20010116/titular.html

Bienvenido al diario platense, fundado en 1884

Martes

EL DIA

Edición Internet

Ahora Alvarez le pegó duro a Machinea

El ministro le aconsejó a la gente gastar más. Y Chacho retrucó: "La mayoría de los argentinos tiene dificultades para llegar a fin de mes"

Detienen a ex policía con un arsenal en cercanías de la casa de Ruckauf

Un efectivo retirado de la Policía Federal fue detenido por la custodia del gobernador bonaerense Carlos Ruckauf cuando circulaba en un auto con un

- ".ar" implies the site is from Argentina?
- Traceroute reveals this website is physically hosted in the U.S.

```
traceroute  from WWW.Telcom.Arizona.EDU to www.eldia.com.ar:
1 woody-netops.telcom.Arizona.EDU (128.196.128.1) 1 ms
…..
8 peer-01-ge.chcg.twtelecom.net (168.215.53.194) 46  ms
….
10 r01.chcgil01.us.bb.verio.net (129.250.2.254) 48 ms
11 r02.stngva01.us.bb.verio.net (129.250.5.103) 83 ms
12 ge.r0728.stngva01.us.wh.verio.net (129.250.27.219) 81 ms
13 ge.stngva01.us.wh.verio.net (161.58.129.13) 81 ms
14 noticiasargentinas.com (161.58.165.155) 80 ms 80 ms 81 ms
```

- Chicago, Illinois
- Sterling, Virginia
- wh = web hosting

Course Outline

- Introduction to Internet Architecture
- “Persona” issues
- Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

Online Web page = http://navigators.com/opensource.html

Each Search Tool is Different

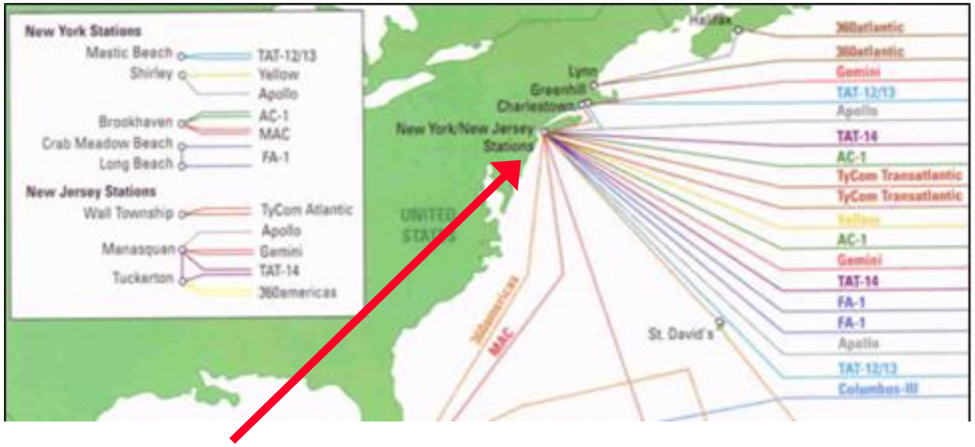
Google
bing
dmoz open directory project
AnyWho
Fairfax County Property Search
Wikipedia The Free Encyclopedia
Internet Archive WayBackMachine
АПОРТ ищет что надо
IRAN YELLOW PAGES iranian information at your fingertips

- Each search tool has it’s own unique set of defaults and options
- Take the time to learn the options of each tool
    - Don’t assume anything
- These tools are competing, trying to be unique
- **Read the help**

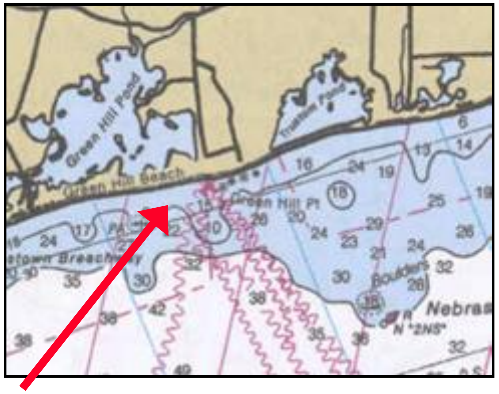
Search - Review

- Stay organized in your search
  - (spell, strategize, search , sift, save)
- Be conscious of the type of tool you are using (and read its help)
- The “right” search terms,
  placed correctly into the “right” search tool,
  should quickly yield “good” results
- Discover the best “user pages” and online communities for your topic - follow their leads
  (They have already weeded through the junk)
- OSINT handbooks, genealogy search sites,

Several Open Sources can be Combined to Build a Complete Picture



Start with a simple cable map



Nautical charts show exact cable locations

Satellite imagery follows cable

FCC Filings, Building Permits, etc. provide additional details:

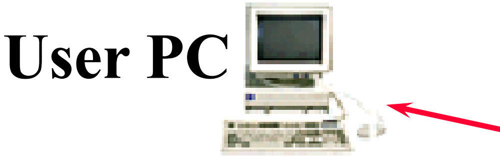
fcc.gov filings: “12. C&W USA states that the Apollo Cable landing stations in the United States will be located in New York and New Jersey. In New York, the cable landing station will be located in Tritec Park, Brookhaven Technology Center, Shirley, New York, at coordinates 40º 50 minutes 30 seconds north and 72º 53 minutes 4 seconds west.”

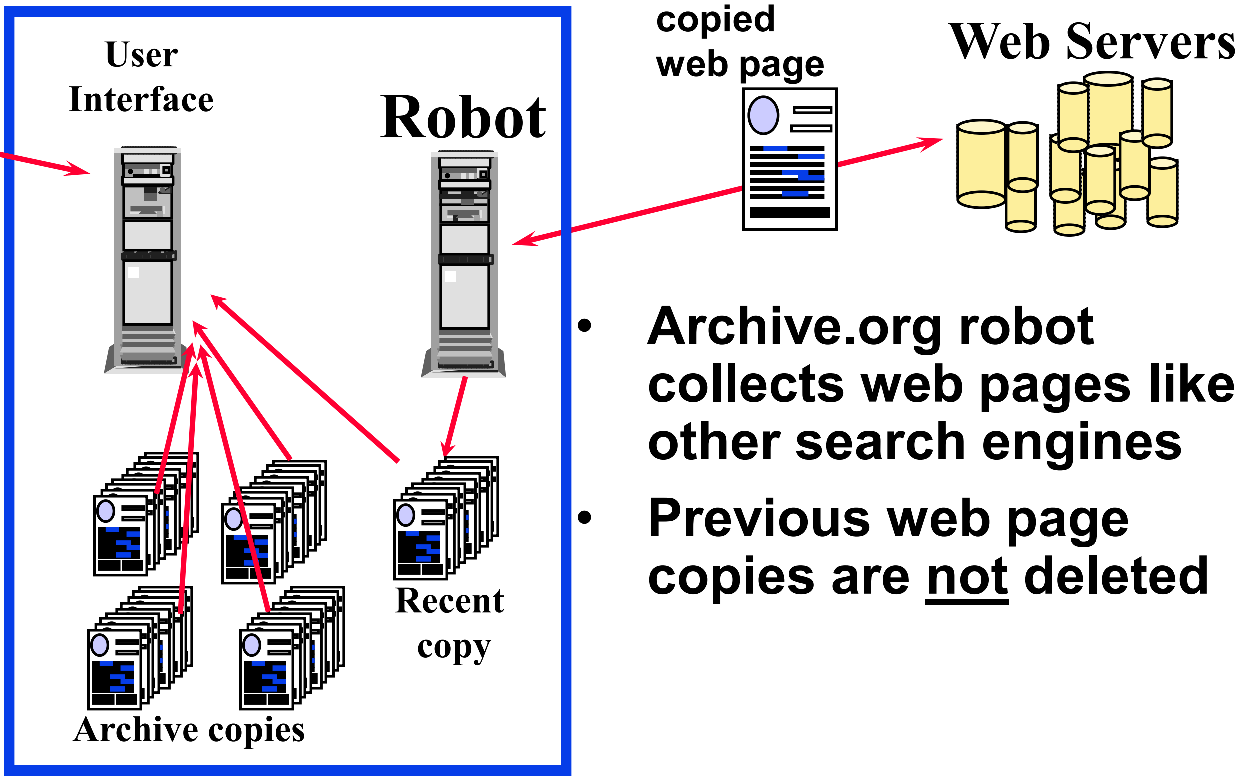
Newspaper / Building Permit Section: “USA Apollo Cable Landing Station, Ramsay Rd. and Precision Dr., site plan-land division station, construct 25,573-square-foot one-story building to house computer equipment for a fiber optic cable landing station on one lot of a two-lot land division in Phase 1. External generators and associated above-ground vaulted diesel fuel tanks to be installed in Phase II. Cable & Wireless USA, Shirley.”

Here is the cable landing station

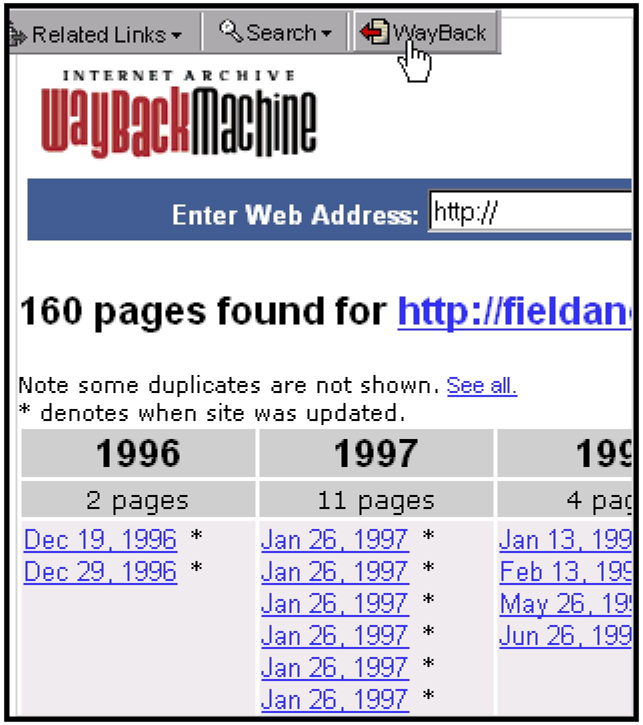
Reference: http://cryptome.org/eyeball/cable/cable-eyeball.htm

web.archive .org





- Archive.org robot collects web pages like other search engines
- Previous web page copies are not deleted



Enter Web Address: http://

160 pages found for http://fieldan

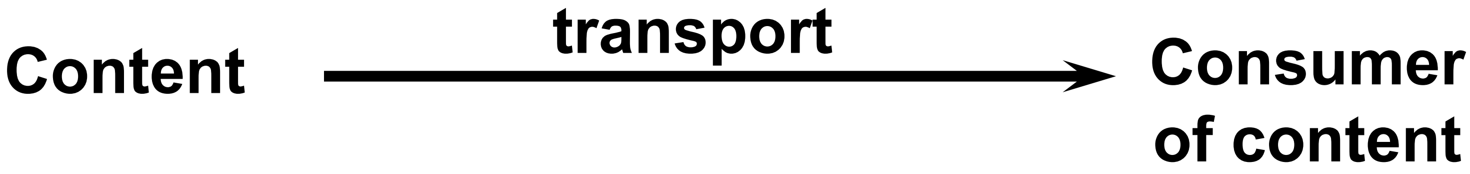
Note some duplicates are not shown. See all.
* denotes when site was updated.

1996	1997	199
2 pages	11 pages	4 pag
Dec 19, 1996 *	Jan 26, 1997 *	Jan 13, 199
Dec 29, 1996 *	Jan 26, 1997 *	Feb 13, 199
	Jan 26, 1997 *	May 26, 19
	Jan 26, 1997 *	Jun 26, 199
	Jan 26, 1997 *	
	Jan 26, 1997 *	

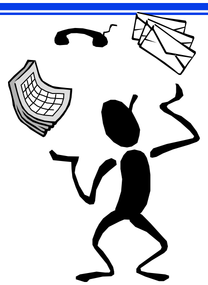
- Surf through previous copies of a web site
- Deleting sensitive information from today's web server does not remove it from archive.org

- "document not found" – Paste the address into archive.org
- Viewing archived web pages will cause hits to live target website

The Future of the Internet



Content → transport → Consumer of content



- Types of content
  - Information, entertainment, business, leisure
- Content origins
  - corporations, hollywood, other people
- Content formats
  - text, audio, video, interactive reality
- Transport mechanism
  - Phone line (copper/fiber), coaxial cable, wireless, direct satellite, electric lines

Mergers and acquisitions are occurring horizontally and vertically

Summary

- Internet contains a large, fragmented information space
- Search engines are limited to billions of “clickable” pages
- The best content is organized by “people without lives”
- The Internet will transcend all other communication technologies
- Change is the only constant

The Future is Clear...
Master the Information Superhighway
or
Become Roadkill

Hidden Universes Links

- Persona: ipleak.com/full-report whoer.net browserleaks.com
  coveryourtracks.eff.org amiunique.org/fingerprint
- Search tools: google.com google.com/advanced_search bing.com
  chatpgt.com wikipedia.org old.wikimapia.org
  unclaimed.org Searchsystems.net
- International searchenginecolossus.com abyznewslinks.com/allco.htm
  radio-locator.com wayp.com
- OSINT tools: metaosint.github.io bit.ly/bcattools
  inteltechniques.com/tools
  i-intelligence.eu/uploads/public-documents/OSINT_Handbook_2020.pdf
- Analytics: radar.cloudflare.com similarweb.com/website
- WayBack machine: web.archive.org
- Source Evaluation: iana.org/domains/root/db traceroute.org who.is
  search.arin.net db.ripe.net/whois
  afrinic.net/whois lacnic.net/cgi-bin/lacnic/whois
  wq.apnic.net/static/search.html

Security and Privacy Links

- Security and Privacy Issues: annualcreditreport.com
- Download your data:

  facebook.com/settings?tab=your_facebook_information

  www.linkedin.com/mypreferences/d/download-my-data

  takeout.google.com
- Detailed Privacy Guides:

  - odni.gov/files/NCSC/documents/campaign/DoD_IAPM_Guide_March_2021.pdf
- - www.socom.mil/Documents/SOCOMSmartcards.pdf

  - www.soc.mil/IdM/publications/docs/general/Id_Privacy_Full.pdf

  - www.pa.gov/content/dam/copapwp-pagov/en/pccd/documents/victim-services/documents/2023-stop-conference/digital_exhaust_guide-law_enforcement_partners_version_2.0_final.pdf
- Trackers: whotracks.me
- Watch: “The Social Dilemma” on Netflix