

Hidden Universes of Information on the Internet – Part 1



Google™ bing™ Alexa

SearchSystems.net LinkedIn

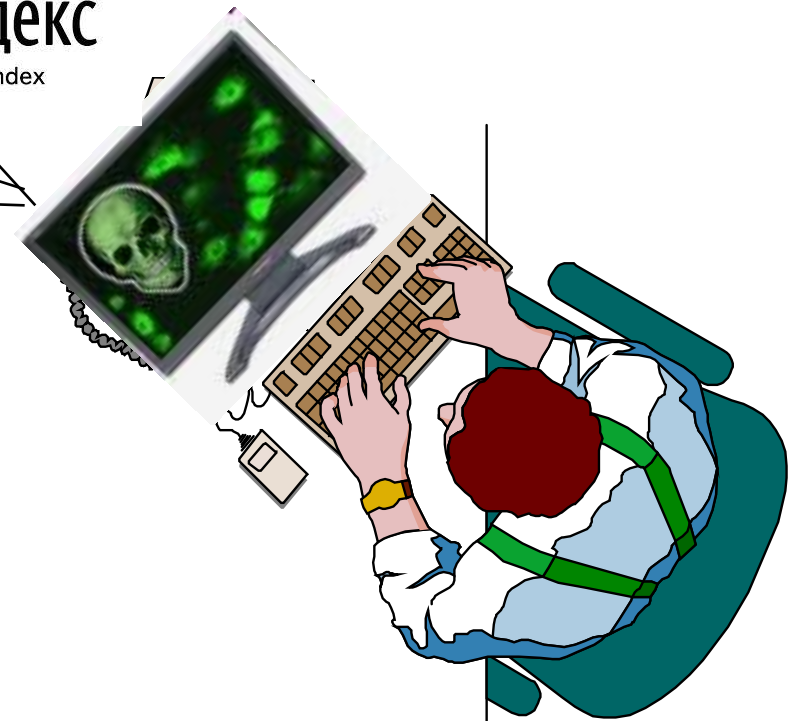
MarineTraffic.com


wikimapia

abyznewslinks.com




ЯНДЕКС
Yandex



 **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 08/2021

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

Copyright © Russ Haynal

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”

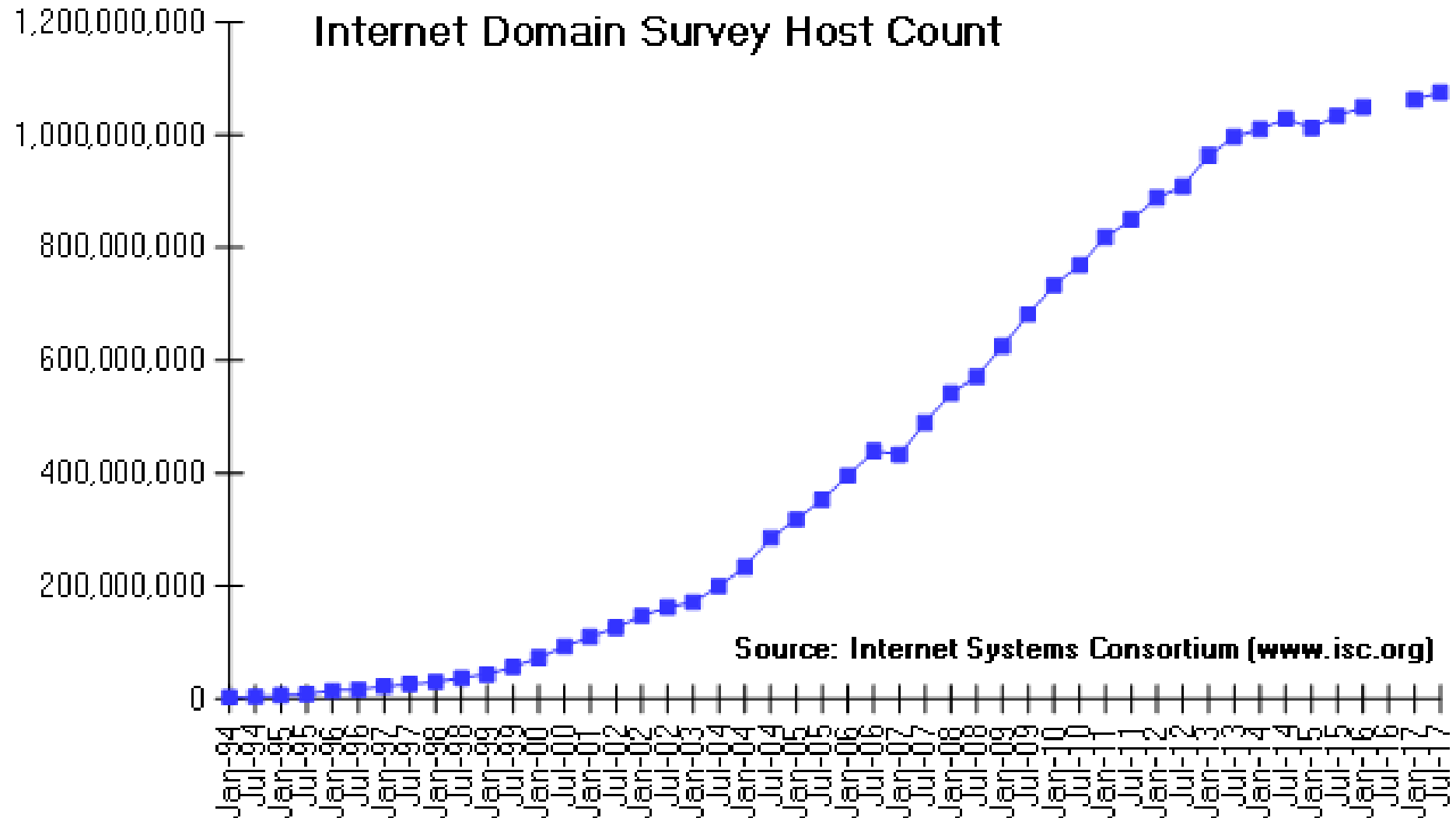
**Internet represents a
once per thousand year event
Last such event = Gutenberg printing press**

Are You Literate in Today's Online World?



Internet's Growth

Internet Domain Survey Host Count



Source: Internet Systems Consortium (www.isc.org)

Number of Hosts in each Domain

Top Level Domains

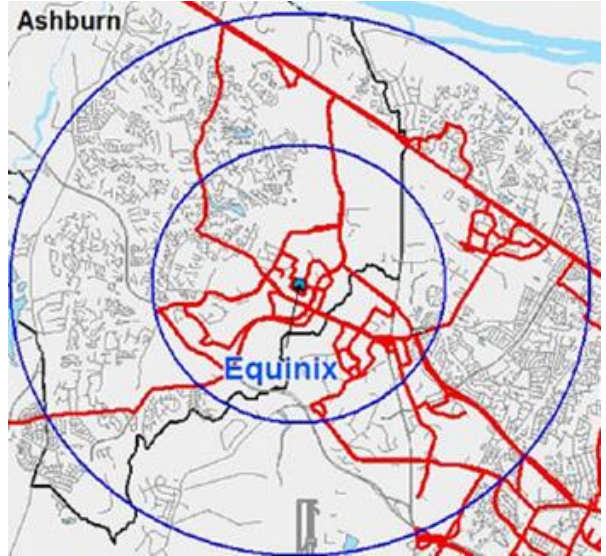
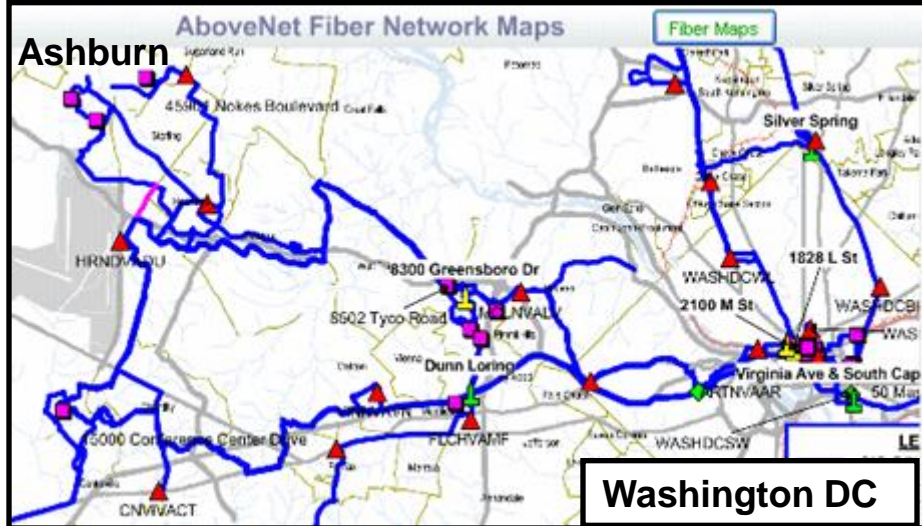
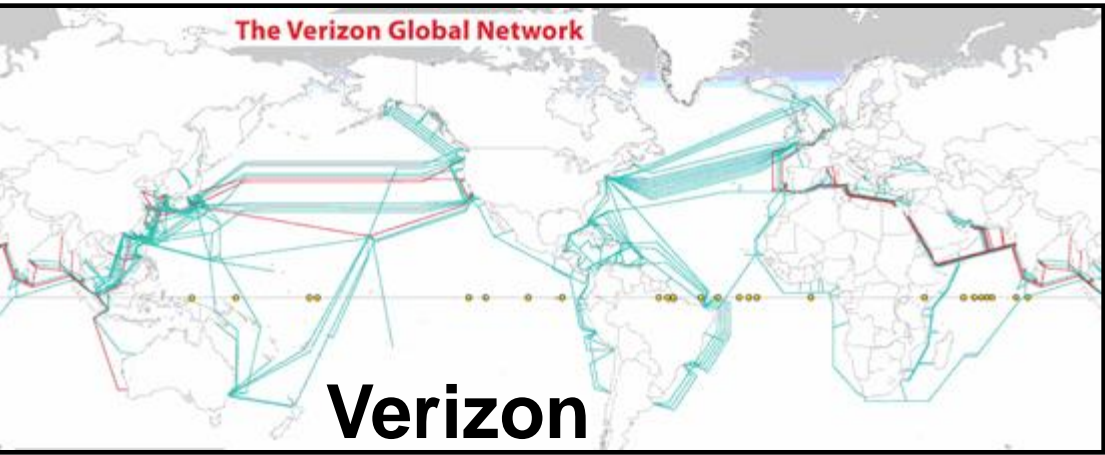
net	386,970,568
com	169,975,462
edu	11,424,990
gov	2,276,632
org	2,161,611
mil	1,443,379

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 as of July 2017

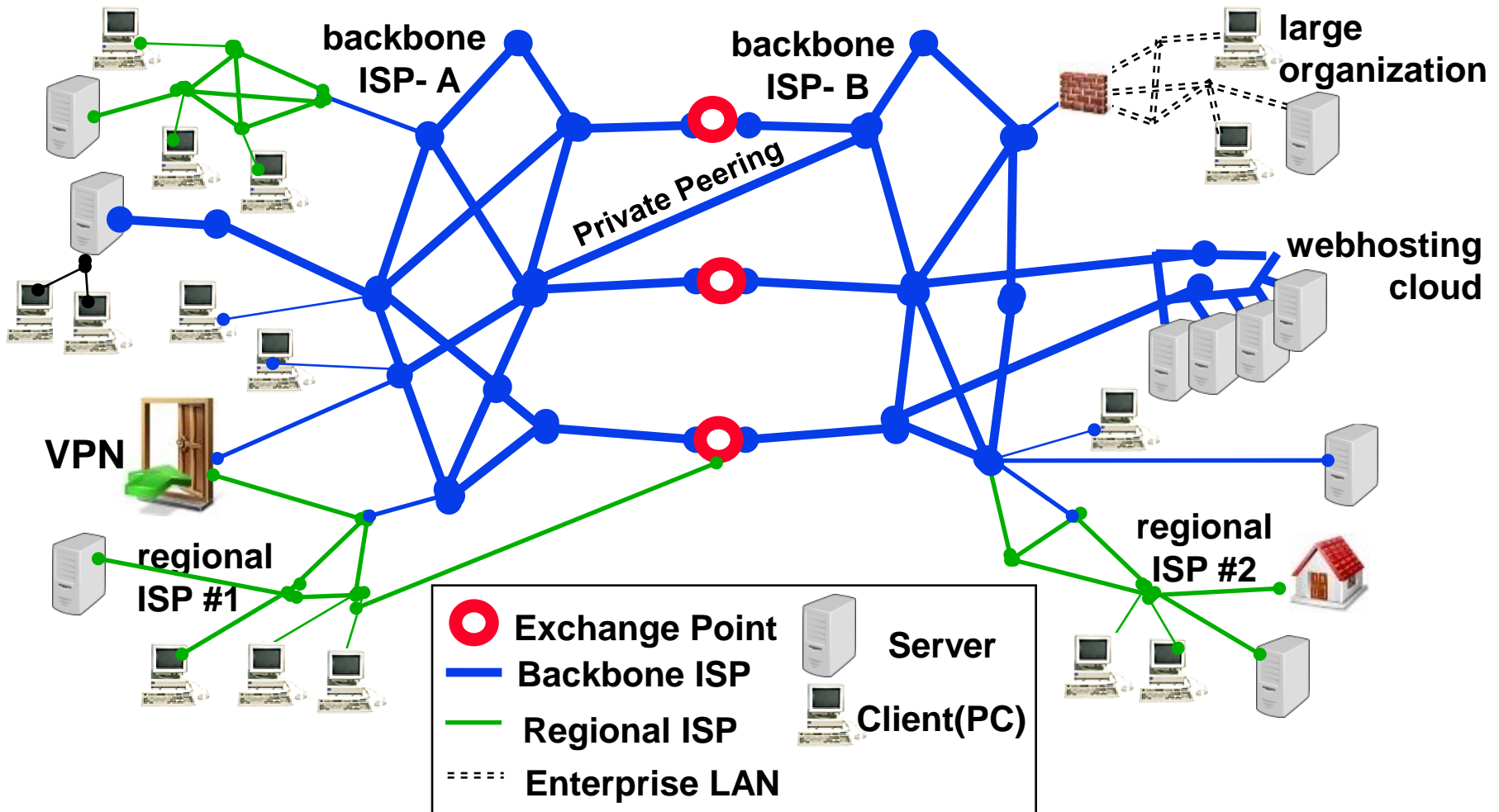
Example Network Maps





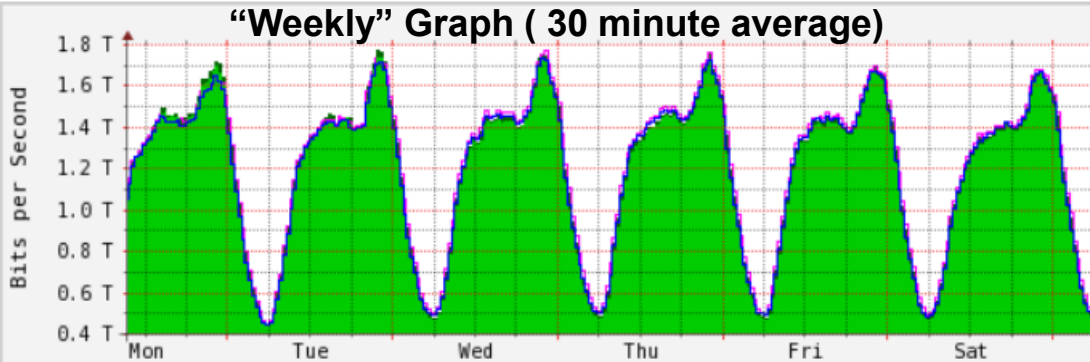
Many People Can Observe Your Internet Usage

Your Internet traffic flows through several providers



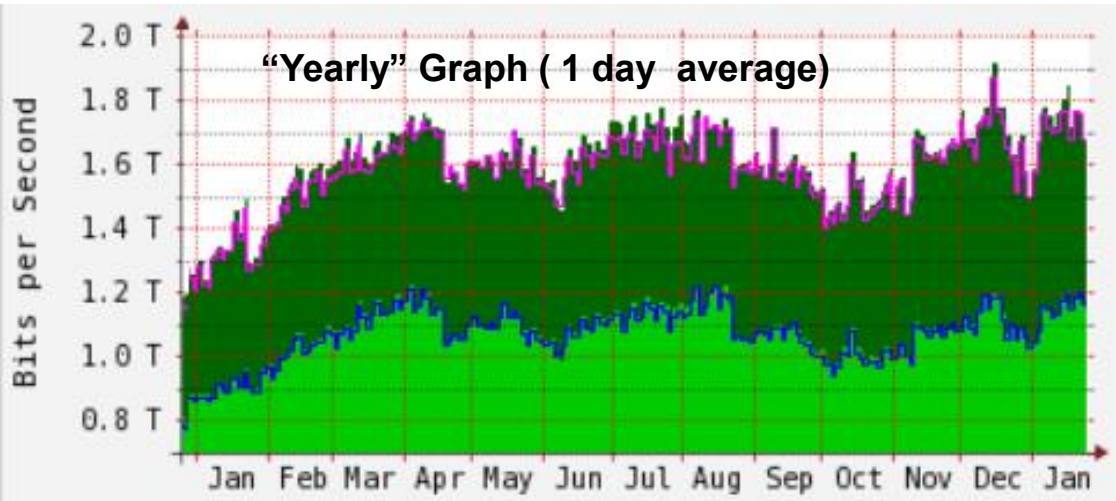
Exchange Point Traffic

“Weekly” Graph (30 minute average)



Daily fluctuations - Analysts may want to “schedule” their research

“Yearly” Graph (1 day average)

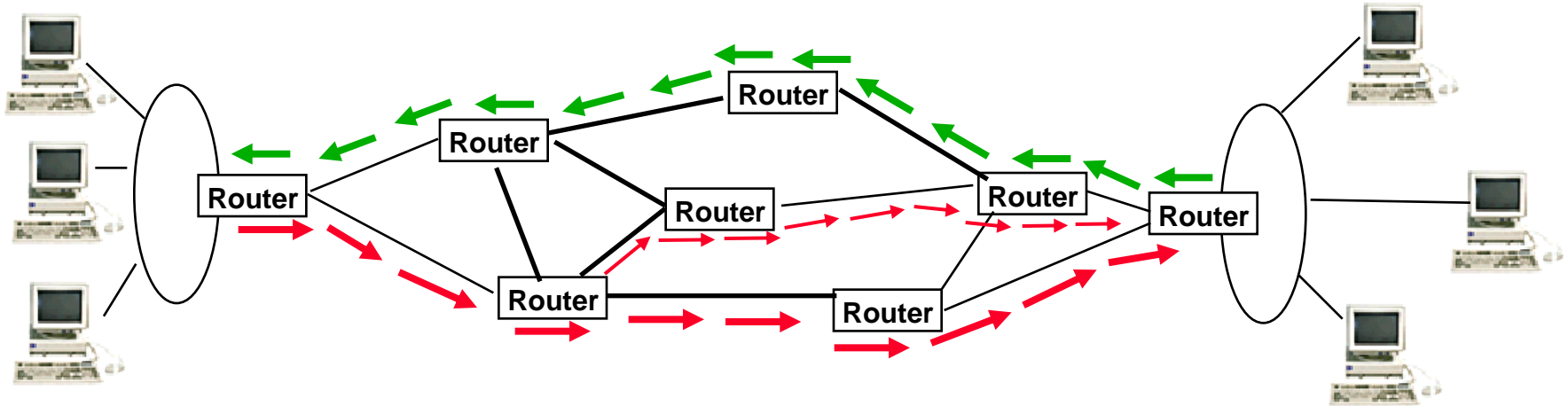


Traffic levels continue to increase

Source: <http://www.hkix.net>

How Does it Work?

- Every Internet connection has a network address consisting of 4 numbers
- Each number has a range of 0-255 (e.g. 198.211.16.134)
- Internet Protocol (IP) numbers are allocated through a hierarchy
 - IANA → ARIN / RIPE / APNIC / LACNIC / AFRINIC → ISP/company/country
- Routers direct your packets of traffic along the “preferred” path



Note: The next version of IP address space (IPV6) is LARGE
3,911,873,538,269,506,102 IP #'s per square meter of the Earth's surface
4,500,000,000,000,000 IP #'s for every observable star in the universe

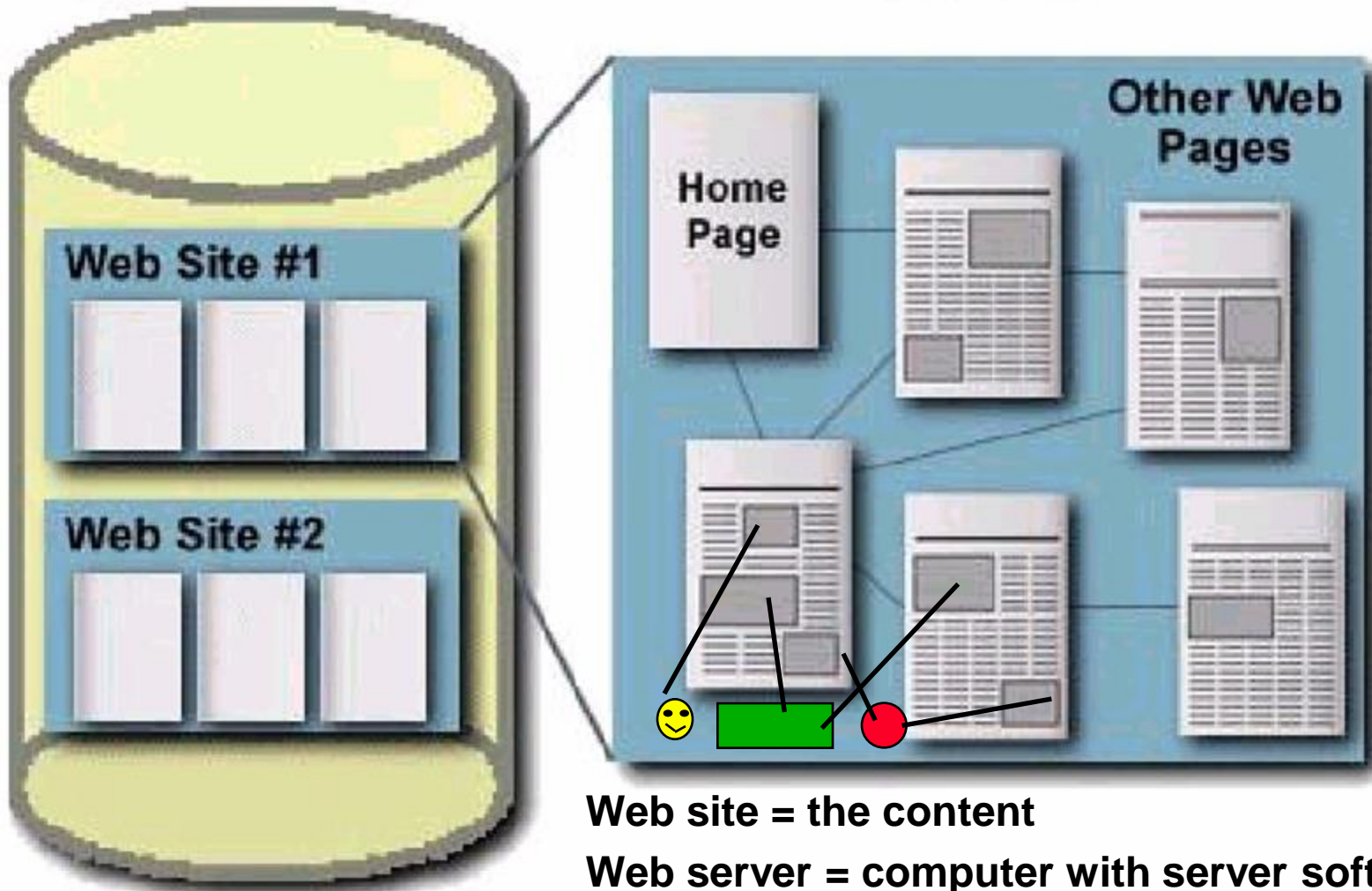
Domain Name System

- **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 act 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 .ru = Russia
- **Reality.... Many country domain names are for sale to ANYONE from ANYWHERE**

Web Server / Web Site

Web Server

Web Site



Web pages
= htm, html

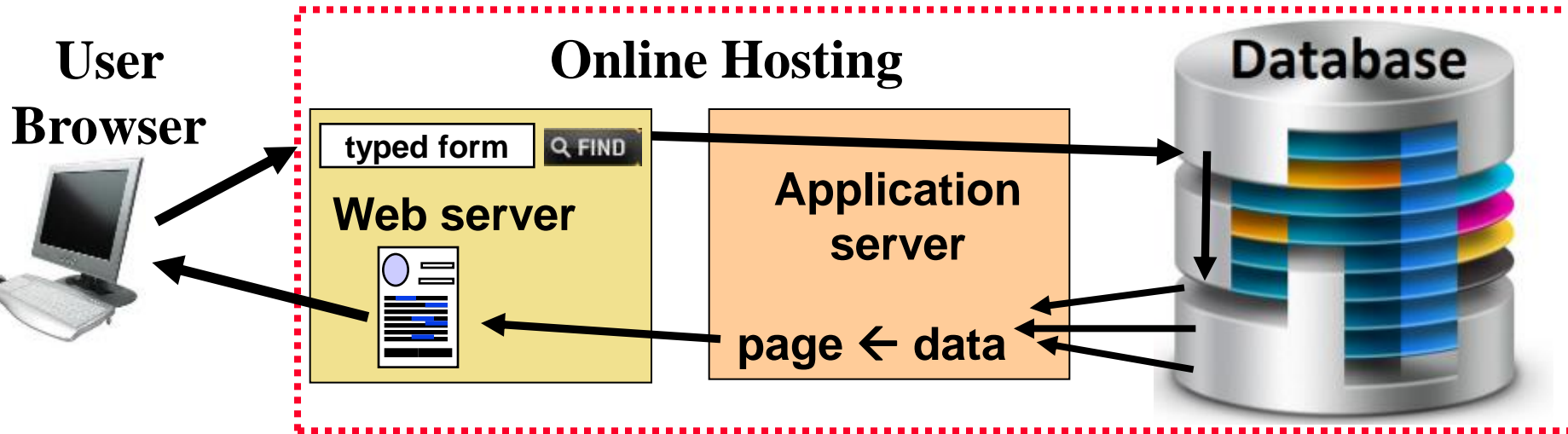
Graphics
= gif, jpg

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

Web site = the content

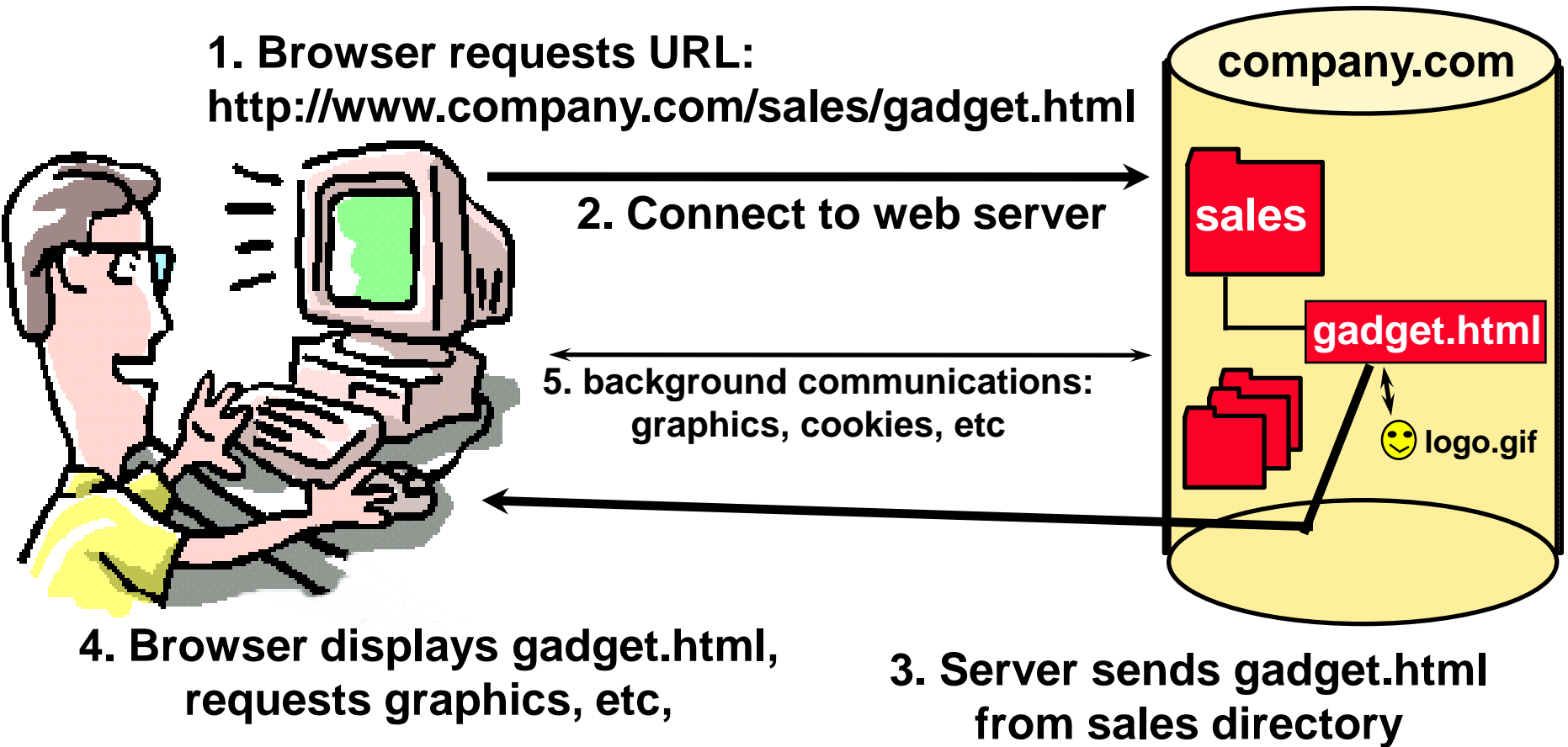
**Web server = computer with server software and
reliable Internet connection**

A More Complex Environment



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

Accessing a Web Page



“Document not found”? - Try shortening the URL!

Course Outline

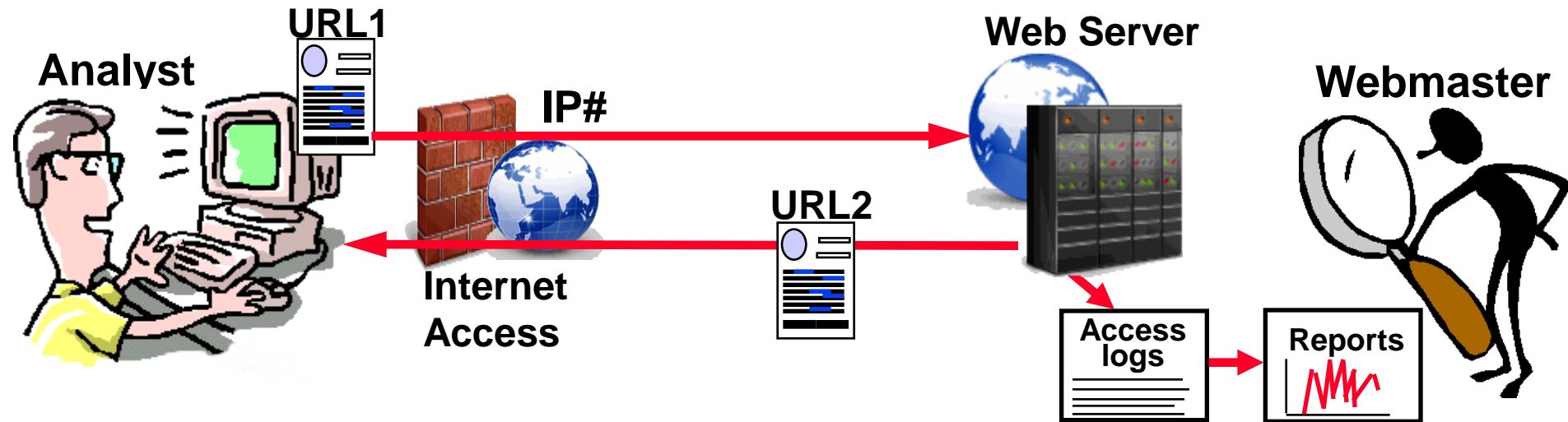
- Introduction to Internet Architecture
- • “Persona” issues
- Search: Search Engines
- Search: “User pages”
- Search: Specialized Tools
- Source Evaluation
- Review / Summary

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



Introduction to “Persona”

As you surf the Internet, you give-off a certain persona



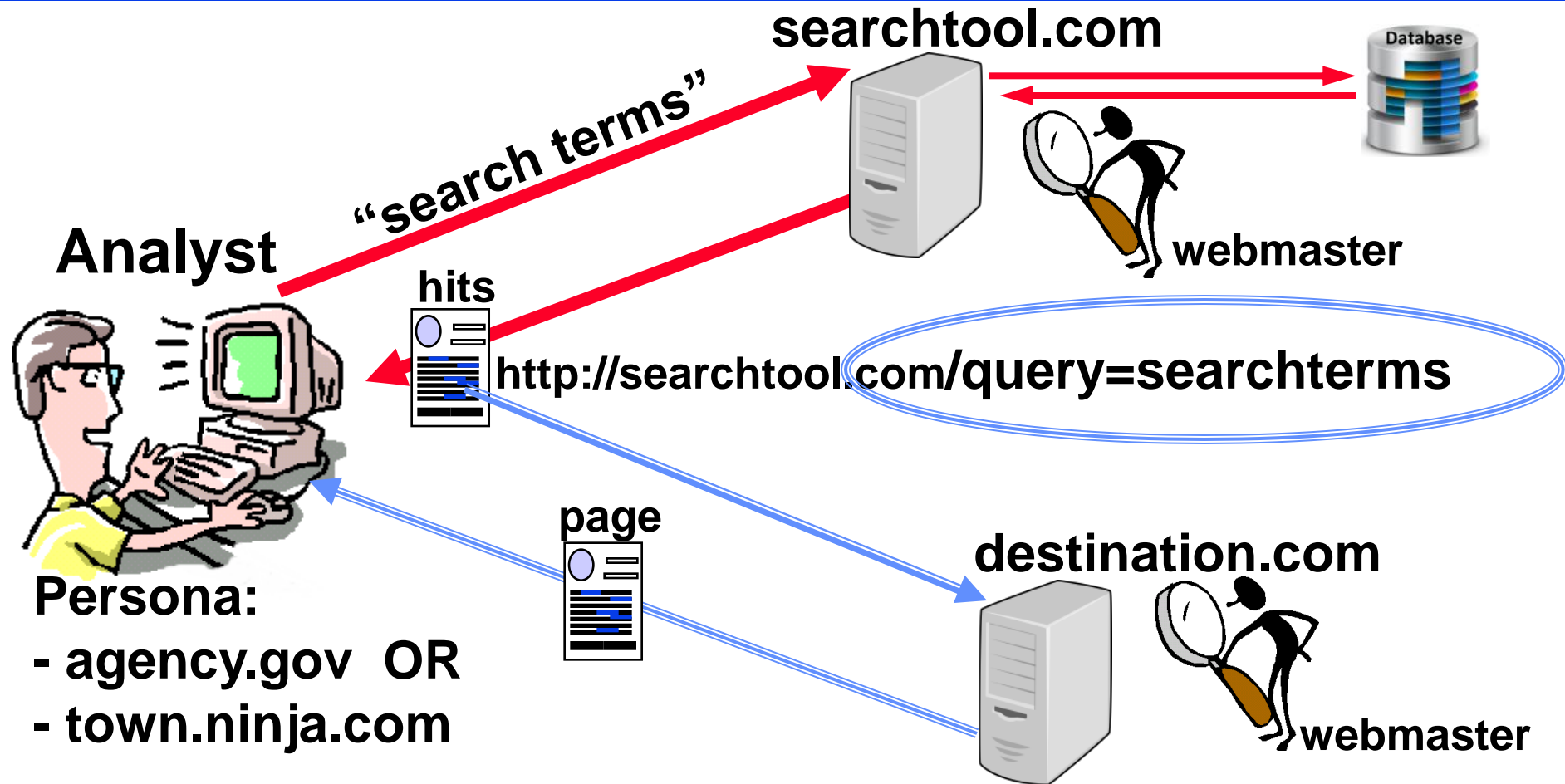
- While viewing a page (URL1), you click on link to another page (URL2)
- Your browser sends “environment variables” to the web server
- Webmasters use this information to learn about you and your organization (physical location, your interests, software)

You should understand what websites know about you

Persona Details

- You should know your persona before 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 previous page that you clicked on
- 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
- Your persona will also be transmitted via Java applets such as ga.js (google analytics) and Adobe flash

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 technique you 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 you click 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**

Think Before You Click...

- Does your connection transmit a referring URL?
- IF IT DOES... do **NOT** “Click” on your search results

The screenshot shows a Mozilla Firefox browser window titled "terrorist - Google Search - Mozilla Firefox". The address bar contains the URL "http://www.google.com/query=terrorist_&start=110", which is circled in red. A red box labeled "Referring URL" points to this address bar. Below the search results, a link to "http://www.orgnet.com/hijackers.html" is circled in red. A red box labeled "Hover over the link to see its URL" points to this link.

- Clicking on this link will tell orgnet.com’s webmaster that you found them while searching for “terrorist”

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

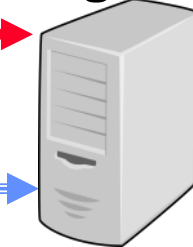
The “parallel visit” Problem...

Analyst #1



agency.gov

target.com



Analyst #2



ninja.com

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

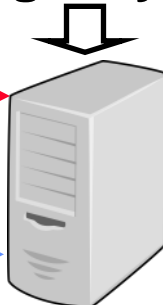
The “portal” Problem...

Analyst #1



agency.gov

agency_portal.com/page_names



Persona=agency.gov + referrer = portal

target.com



Analyst #2



ninja.com

Persona=ninja.com + referrer = portal

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

You will search for the same topic throughout the course. This enables you to compare results among the various search tools / techniques.

Pick a topic you can stay with for 2 days

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

**Lecture - I will introduce the search tool/technique
(Please refrain from using your computer)**

**Demo - I will 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

- **S**pell it Out - Define the topic, key words, acronyms, “what” and “who”
- **S**trategize - Choose your approach, online resources, specific search tools
- **S**earch - Get online, stay focused, use advanced search features
- **S**ift - Filter the results, follow the leads
- **S**ave – Make bookmarks, take notes, organize results, share with co-workers



Spell out the topic...

1. Name of topic, and what do you want to learn

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

- **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...

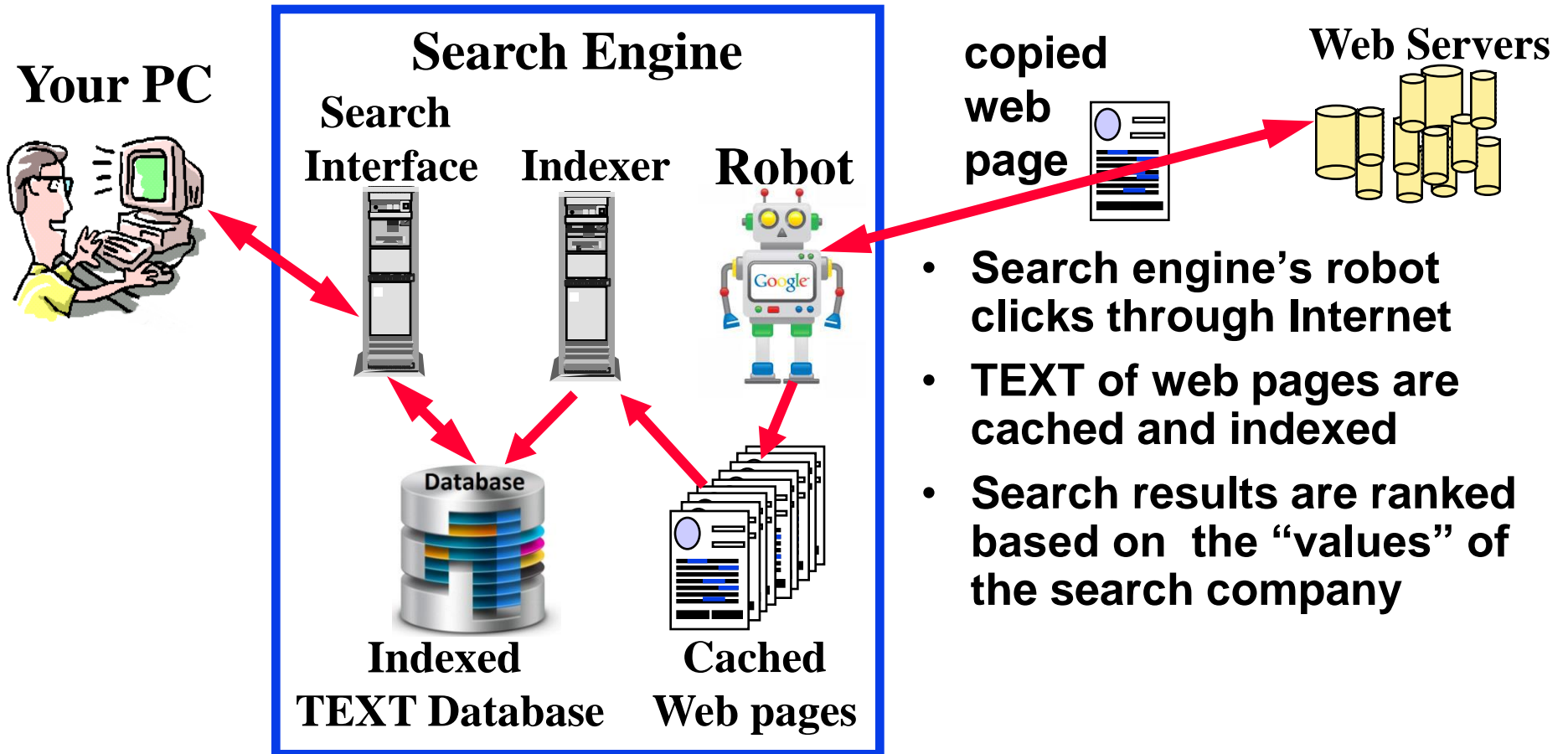


Each tool has strengths and weaknesses



Search Engines

(google.com , bing.com)



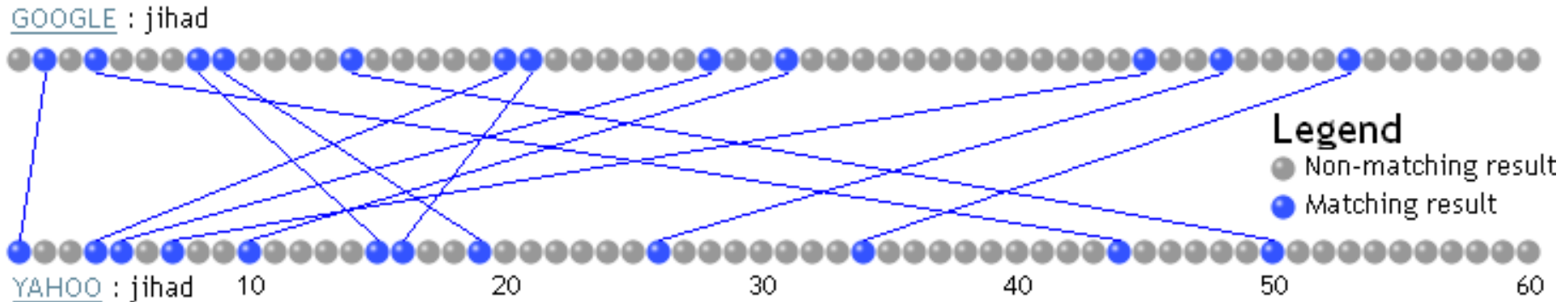
You must envision the target page
"Use your imagination"

- Set "results per page" = 100 : Settings → Search Settings

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 any good sites**

Search Engine Comparison



Search engine market share



- <http://ranking.thumbshots.com> – Compared the first sixty hits from two search engines
- Example search “jihad”, Only 12 out of 60 hits appeared in both Google and Yahoo... Most hits are unique to each search engine



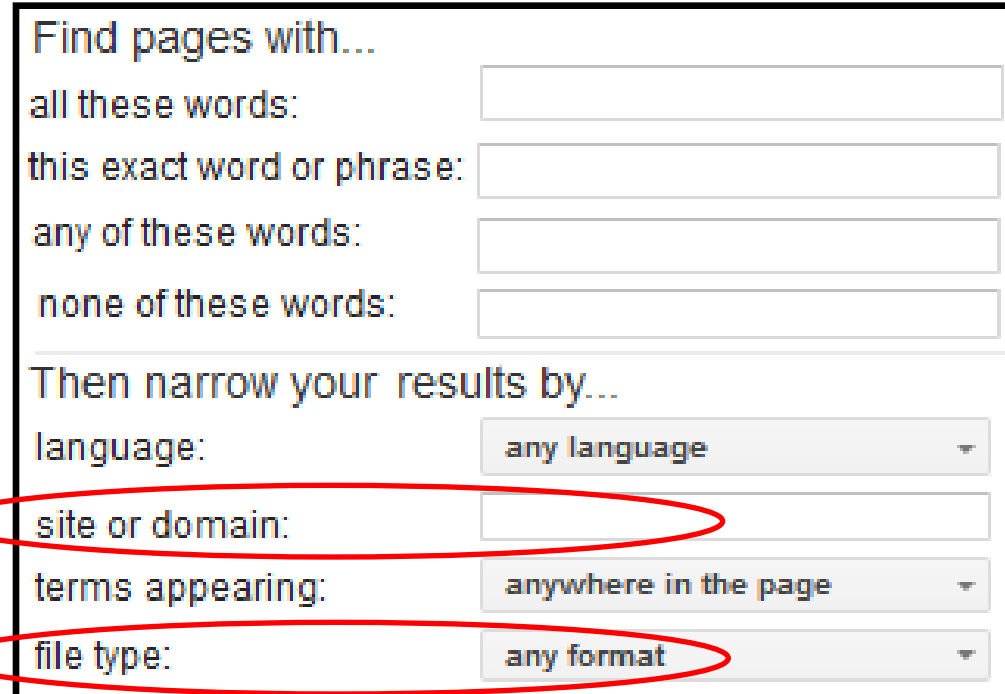
- News, forums and analysis of search engines

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:

site or domain:

terms appearing:

file type:

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: Settings → Advanced Search



Viewing Google's Cache Causes Hits to Target Webserver!

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

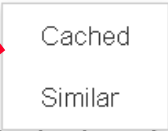
Try it: Firefox → Tools → Web Developer → Network



How to View html Pages as Cached “text only” in Google

[Russ Haynal's ISP Page](#) → link to live website

navigators.com/isp.html



Major Internet Back
... The inet-access
Includes Searchable Index; Network ...

Link to Google’s text of a cached webpage. The graphics, videos, etc. are then downloaded by your browser from the target’s webserver.

To view a “text only” version of Google’s cache...

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

<http://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
no space

Note: cached documents (e.g. pdf, ppt, etc) are already “text only”

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





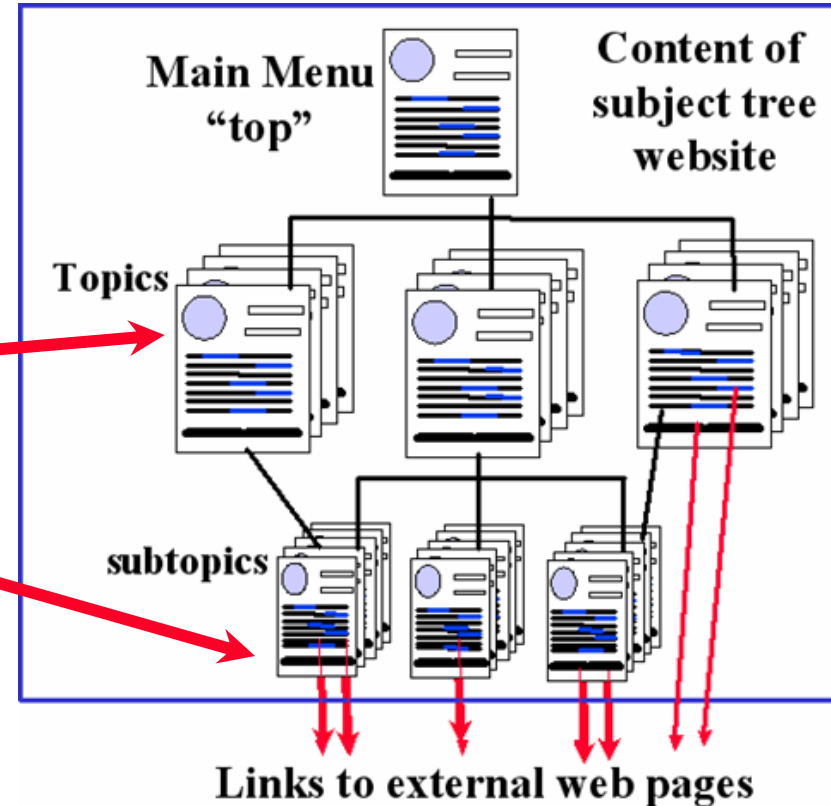
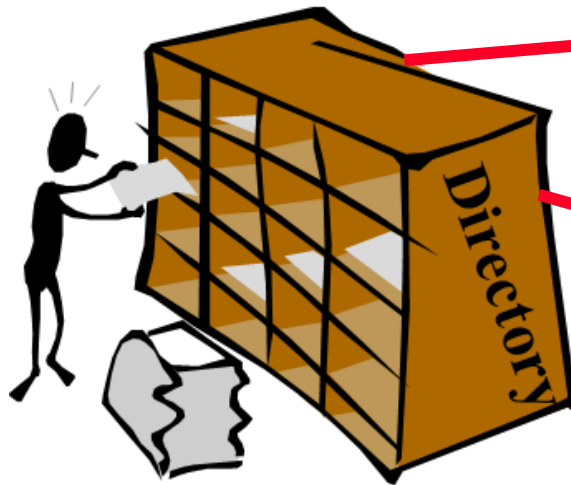
Finding “User pages”

- Subject directory
- Groups of users in a forum, conference, journal, club
- Contribute to wikipedia, wikimapia
- Watch for sites labeled:
 “Joe’s ultimate guide to widgets”
- “User pages” often point to other “user pages”
- “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



Finding Groups of Users

- Forum – discussion focused on a particular topic
 - Many users can participate by posting messages
 - **Search : your_topic forum post replies**
= threads and posts that discuss your topic
 - Membership requirements are a barrier to search engine robots
-
-
- 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, society, club, league, “user group”, alumni,
 - **Search: your_topic conference , your_topic festival, etc.**
 - Individual: resume, Curriculum Vitae, CV, biography



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**



- **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**

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:



owned by Amazon



owner based in New Zealand



owner based in Israel

Web Analytics

alexa.com/siteinfo

urlm.co

similarweb.com

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

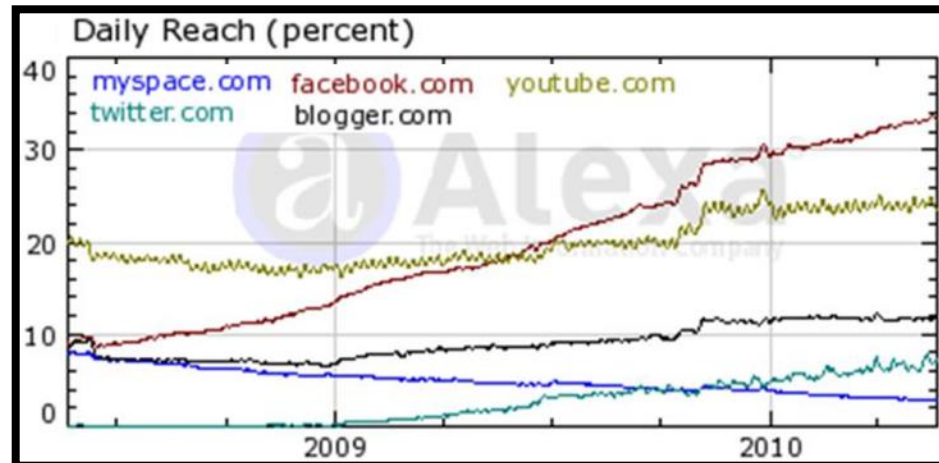
- Popularity of a web site
- Audience demographics
- Search terms used to find the site
- Visitor engagement levels
- Traffic history
- Related sites = more sites

**Enter a domain name
(not search terms)**

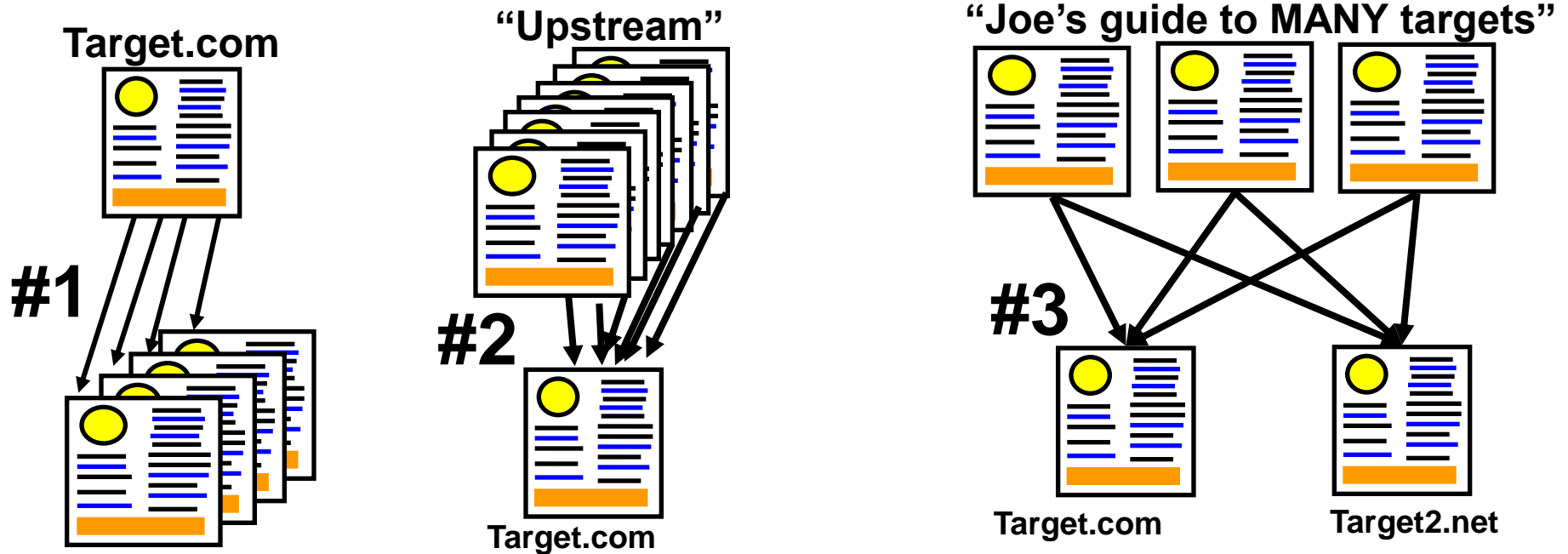
visitors to aljazeera.net

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
 Algeria	5.2%	123



Surfing Upstream vs. Downstream



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

#2 Shows pages that link towards 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

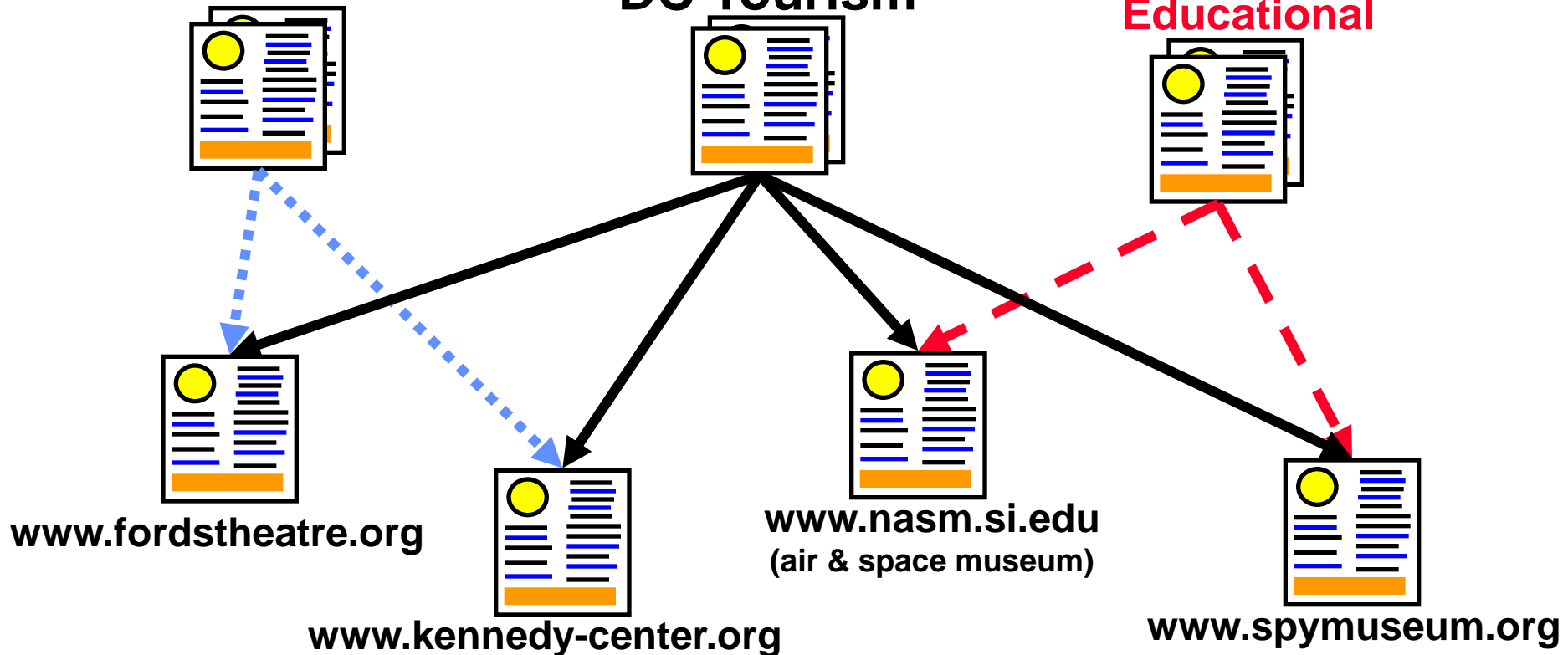
Be Creative When Surfing Upstream

Example: Washington DC Tourist Sites

Theatre links

DC Tourism

Museums /
Educational



- Any combination of these target pages will lead you to “DC Tourism” pages, but certain pairings may also lead you 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 <u>both</u> web site addresses This is a great way to discover “user pages” (e.g. Joe's guide to <u>many</u> example-sites)

- **You need to decide which scenario makes more sense; 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

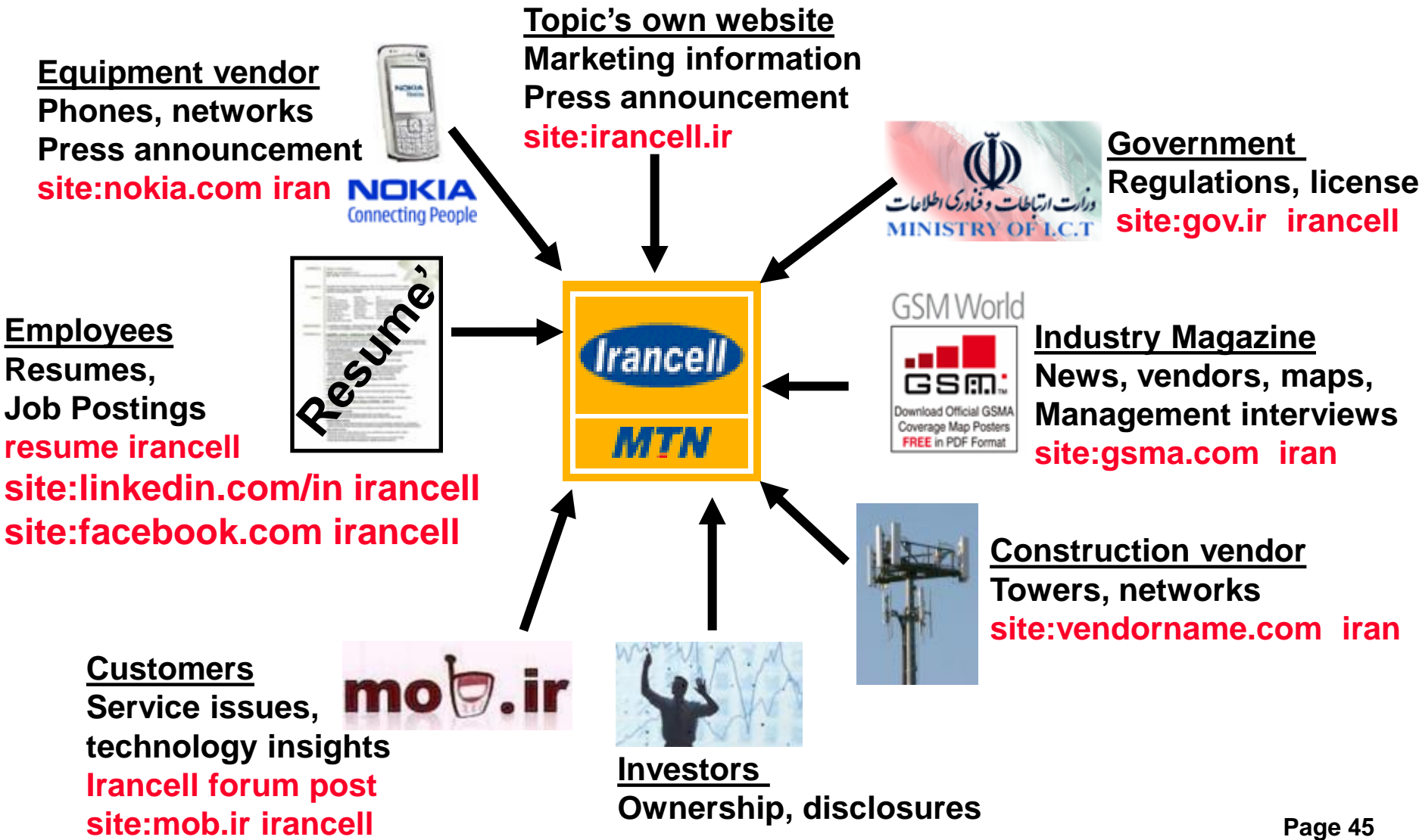
search format at Google	search results
site:example.com	pages hosted on any example.com servers (www.example.com, blog.example.com, etc) = quick way to assess the public size/depth of a domain
site:example.com searchterm	pages hosted at example.com which mention "searchterm"
site:ru searchterm	pages 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 topic

- **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)



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)



- 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!

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>



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

AnyWho

- A phone book for the entire U.S. Includes reverse look-ups

KOMPASS
Connects business to business

- list of 35 million companies

 **MarineTraffic.com**

- Real-time tracking of ships from around the world

govinfo

- Federal Register and much more

Specialized databases contain content that search engines can't reach

Business databases can be quite useful

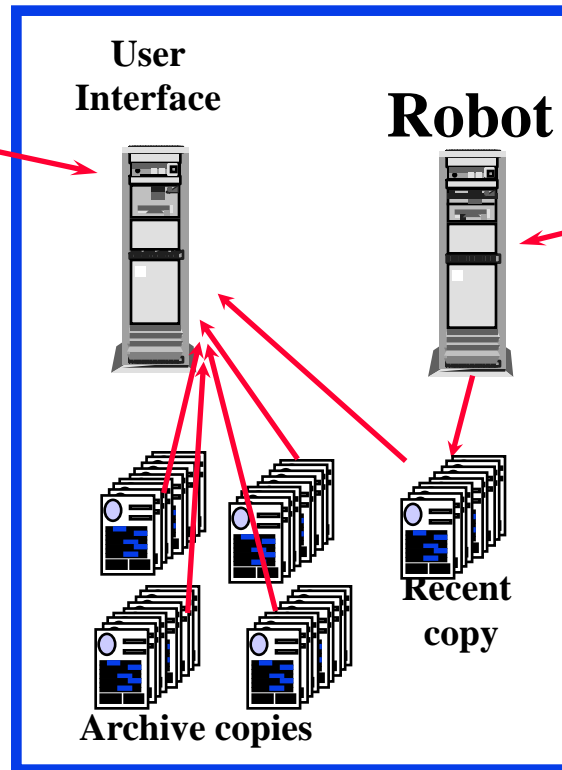


[Home](#) | [Previous Page](#)

U.S. Securities and Exchange Commission

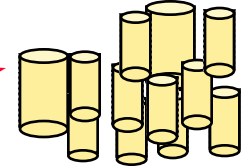


- **Most publicly held companies are required to file financial statements with the Securities Exchange Commission**
- **These filings are accessible to the public through the SEC's EDGAR database**
- **READ forms 10-Q and 10-K (quarterly and annual report)**
These are very detailed reports about the 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)**



copied
web page

Web Servers



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

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

Related Links Search WayBack

INTERNET ARCHIVE
WayBackMachine

Enter Web Address:

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 *	

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

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>

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 you can 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: 2017-03-19

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 has lead to many domain names being registered via “privacy enhanced” options
- EU GDPR Law in 2018 also impacts WHOIS records for European (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, but most likely hosted 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 Altnet swap traffic at Los Angeles

Baltimore airport code

Traceroute helps reveal the dynamic architecture of the Internet

A Foreign Newspaper ???

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



- “.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

Each Search Tool is Different

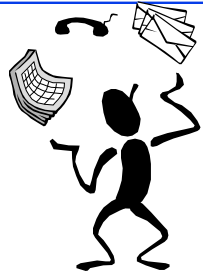


- Each search tool has it's own unique set of defaults and options
- Take the time to learn the options of each tool
 - Don't assume anything
- These tools are competing, trying to be unique
- Read the help

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)**

The Future of the Internet



- **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 tests:** whoer.net amiunique.org/fp browserspy.dk/headers.php
ipleak.com/full-report samy.pl/evercookie panopticlick.eff.org
- **Search tools:** google.com google.com/advanced_search bing.com
wikipedia.org old.wikimapia.org
- Searchsystems.net unclaimed.org marinetraffic.com
- searchenginecolossus.com abyznewslinks.com/allco.htm radio-locator.com
infobel.com/en/world/index.aspx wayp.com web.archive.org
- **Analytics:** alexa.com/siteinfo urlm.co similarweb.com
- **Cached "text only":** webcache.googleusercontent.com/search?strip=1&q=cache:
- **Source Evaluation:** iana.org/domains/root/db norid.no/en/om-domenenavn/domreg
traceroute.org who.is search.arin.net db.ripe.net/whois
wq.apnic.net/static/search.html afrinic.net/whois lacnic.net/cgi-bin/lacnic/whois
- **Exchange points:** en.wikipedia.org/wiki/List_of_Internet_exchange_points
peeringdb.com
- **Cables maps:** submarinecablemap.com live.infrapedia.com subtelforum.com
en.wikipedia.org/wiki/List_of_international_submarine_communications_cables
- **Satellites:** lyngsat.com
<https://ucsusa.org/resources/satellite-database>
- **Datacenters:** datacentermap.com datacenterknowledge.com
- https://www.arcyber.army.mil/Portals/34/Fact%20Sheets/DOD%20Identity%20Protection%20Guide/DoD_Identity_Awareness_Protection_Management_Guide_March2019.pdf