Making Effective Use of The Internet: Tips and techniques

Brant Houston, Knight Chair, U. of Illinois and IRE Staff
Facts of online life

- You want: Fast, Cheap, Easy
- You usually get: Two out of three
- You want: One-stop shopping
- You usually have to: Bargain hunt at several stores
Overview

- Information strategy
- What search engines are really for
- Beyond Google
- A look at the Invisible (Deep) ‘Net
- Pearls and perils
HOW REPORTERS START THEIR DAY AT WORK...

TODAY I AM AN EXPERT IN:

- Politics
- Economy
- Car Repair
- Health Care
- Television
- Aerospace
- Biotechnology
- Music
- Foreign Affairs
- County Fairs
- Oil
- Vinegar
- Sports
- Stocks
- Bondage
- OBO
- Explosives
- Computers
- Plumbing
- Food
- Septic Tanks
- Fusion
- Liposuction
- Cartoons
- World Peace
Strategy

- Will something other than the Internet have what I need?
- Is there a Web site or tip sheet already there?
- What are the key words I’ll need if I search?
- Where are the credible sources?
Know the source

- Individuals (social media) *Informal sources*
- Organizations *Institutional sources*
  - Businesses (com, biz)
  - Governments (gov)
  - Associations (org, net)
- Universities (edu) *Scholarly sources*
- News organizations (com, org) *Journalistic sources*
Examples of places to go

- **Institutional Sources**
  - Consumer Product Safety Commission
  - Company Website: products, corporate

- **Informal Sources**
  - Google Groups
  - Video: Expo TV
  - Facebook, etc

- **Journalistic Sources**
  - Google News

- **Scholarly Sources**
  - Google Scholar

- **Aggregators**
  - Google Finance
What do I need?

- What is the key topic?
- What kind of information do you need?
- Why do you need the information?
- How much information do you need?
- How far back do you need to go?
- Who do you need to talk to?
- Where have you already looked?
What do I need?

- Documents and data to “interview” for story – Using Google advanced search.
- Reference – Get fast facts
  - Wikipedia (for tips on the primary documents)
  - Internet Public Library “Reference” page
  - Google Scholar, Infomine (scholarly internet resources)
  - Find the ones most useful for your beat or area of expertise and bookmark them
Beat tasks

Keep up with a beat with Alerts

- Alerts are customized – you plug in the terms
- General / news: Google Alerts ("federal stimulus package")

Get notified about new material with RSS and email lists

- RSS feeds / email lists are general – everyone gets the same
- Google Desktop or Yahoo!
- Feedreader
Beat tasks

- Find people
  - If you don’t have a specific name:
    - Subject-specific directories
    - Google Scholar
  - If you have a name:
    - Telephone/e-mail directories
    - Public records databases

- Hang out in virtual communities
  - Social networking sites
Use proven journalist’s list

- Compiled by journalists for journalists. The best:
  - NICAR ‘Net Tour

http://www.ire.org/training/nettour/
Which search engine to use

- Google is No. 1.
- It has the highest number of pages it “searches,” including different file types, although Yahoo! is now challenging this claim and Microsoft is shifting to that model too.
- Can search Groups, images, blogs, maps, academic journals, etc.
But it has its limits

- Library of Congress: 11 terabytes
- Surface web: 167 terabytes  
  (70 billion web pages)
- Deep web: 91,000 terabytes  
  (38,000 billion web pages)
- Estimate: Google indexes about 10% of  
  surface web (8 billion pages)

http://websearch.about.com/od/invisibleweb/a/invisible_web.htm
Quicken your searches

- Use Advanced Search in Google
- Search for phrases
- Use Boolean logic – and, or, not
- By domain or site
- By file type
- By date
Who is behind the site?

Domain searches reveal all.

- Domain Tools
  http://www.domaintools.com

- Allwhois http://www.allwhois.com
Go back in time

Many Web pages are stored:
- The Way Back Machine
  http://www.archive.org/index.php
- Remember there is text, images and video
Multiple sources

- Search two or three search engines besides Google – including their advanced functions.
- Each search engine has its own strengths and weaknesses.
  - Total shared results of any two search engines: 8.9 percent.
  - Any three search engines: 2.2 percent
Multiple sources

- Alltheweb.com (advanced search options)
- Ask.com (Subject-Specific Popularity™)
- Clusty.com (clustered results)
- Yahoo! (directory)
- Jux2 (Multiple search site results returned)
- Bing
The Deep Web

- The Visible Web is made up of the typical results you see in Google, etc.
- The Invisible (Deep) Web includes searchable databases and other dynamic Web content.
How deep you can go

What search engines (Google, et. al.) can’t find

- Data-based content
- Content behind registration screens
- Content behind firewalls
- ASP / dynamically generated pages
- Robot.txt excluded pages
Invisible Web sites

- Databases: CompletePlanet
- Business: Explorit Now!
- Science: Science.gov
- Video: Blinkx
Pearls

- Public records: Searchsystems.net
- People search: Zabasearch.com
- Web owners: GoDaddy.com
- Tips: IRE resource center
- Book: Computer-Assisted Research (Paul and Hansen, IRE Beat Books)
More Pearls

- Public records: Portico
  http://indorgs.virginia.edu/portico/
- People profiles: Pipl
  http://pipl.com/
- Tips: IRE resource center
What do Yogi Berra, Niels Bohr, Samuel Goldwyn, Dan Quayle and Mark Twain have in common?

According to Google results, they all are responsible for the saying, “Prediction is difficult, especially about the future.”

Thanks to John Martin
So remember ...

- For the most part, the Internet doesn’t have an editor.
- Misinformation replicates exponentially.
- GIGO – Garbage in, garbage out.
- Multi-source and cross-reference