Discover the Invisible Web

Tom Keefe
Reference Librarian
Loyola University Chicago School of Law
Law Library

April 28, 2015
Topics Covered

• **Topic 1**: Invisible Web: The Traditional Definition
• **Topic 2**: Now you see it, now you don’t: superseded information, reference rot, and invisible information due to poor information architecture
• **Topic 3**: Present but not accounted for: information that searchers miss due to search engine tricks and poor research skills
• **Topic 4**: Tips for success: how to ensure that you’re not missing something
• **Topic 5**: Invisible web search tools
What is the ‘Invisible’ Web

• Definition (traditional)
• Size?
• Contents (traditional)
  o Info in databases/catalogs
  o Password protected/subscription info
  o Info that people/organizations don’t want you to see

• Illustration
The ‘Invisible’ Web: Deceased or Non-digital Info

- Info never existed in digital form
- Info removed/superseded because no longer current

“Reference Rot”
- Webpage restructuring
- Domain changes
- Shuttered organizations
- Defunct servers

Poor Information Architecture
- Example: National Bar Association evaluation
  - content is difficult to locate online
  - not constructed with a clear directory
  - multiple domains/subdomains and redirects/rewrites that prevent good crawling and browsing
Other Reasons for Invisibility

• Search engine optimization (SEO) and paid links
  o How do search engines work?
  o The first word in any search algorithm is??

• “personalized” searching
  o Custom search results
  o The “filter bubble
    • A filter bubble is a result of a personalized search in which a website algorithm selectively guesses what information a user would like to see based on information about the user (such as location, past click behavior and search history) and, as a result, users become separated from information that disagrees with their viewpoints, effectively isolating them in their own cultural or ideological bubbles.

• Unrefined research techniques
  o Exclusive use of one search engine
  o Failure to go beyond first page of results
  o Inability to navigate to information contained within a relevant website
Tips for Using Search Engines Effectively

- If you’re trying to be thorough, use multiple search engines
- Advanced search features are your friend
- Internet research often requires a two-step process; if you know that information resides on a website, go there
- Internet subject directories help overcome the invisible web
A Few Helpful Sites

- Google Advanced Search
  - [http://www.google.com/advanced_search](http://www.google.com/advanced_search)
- The Wayback Machine
  - [http://www.archive.org](http://www.archive.org)
- Internet Public Library Resources by Subject
  - [http://www.ipl.org/div/subject/](http://www.ipl.org/div/subject/)
- DuckDuckGo
  - [http://www.duckduckgo.com](http://www.duckduckgo.com)
Google Advanced Search

http://www.google.com/advanced_search

Advanced word search features

Filter by site/domain

Filter by file type
The Wayback Machine

https://www.archive.org
What is the CyberCemetery?

The CyberCemetery is an archive of government websites that have ceased operation (usually websites of defunct government agencies and commissions that have issued a final report). This collection features a variety of topics indicative of the broad nature of government information. In particular, this collection features websites that cover topics supporting the university's curriculum and particular program strengths.

Who created and maintains the CyberCemetery?

The University of North Texas Libraries and the U.S. Government Printing Office, as part of the Federal Depository Library Program, created a partnership to provide permanent public access to the Web sites and publications of defunct U.S. government agencies and commissions. This collection was named the “CyberCemetery” by early users of the site.

This collection supports the mission of the UNT Libraries by acquiring, preserving, and providing access to recorded knowledge in the form of government websites.
DMOZ
http://www.dmoz.org

Arts
Movies, Television, Music...

Games
Video Games, RPGs, Gambling...

Kids and Teens
Arts, School Time, Teen Life...

Reference
Maps, Education, Libraries...

Shopping
Clothing, Food, Gifts...

World
Català, Dansk, Deutsch, Español, Français, Italiano, 日本語, Nederlands, Polski, Русский, Svenska...

Business
Jobs, Real Estate, Investing...

Health
Fitness, Medicine, Alternative...

Regional
US, Canada, UK, Europe...

Society
People, Religion, Issues...

Computers
Internet, Software, Hardware...

Home
Family, Consumers, Cooking...

Recreation
Travel, Food, Outdoors, Humor...

Science
Biology, Psychology, Physics...

Sports
Baseball, Soccer, Basketball...