



# Hacking Google

How to Find and Create Free Quality  
Research Databases on the Web

R Philip Reynolds

Stephen F. Austin State University

# What do I Mean by Hack?



- Traditionally hacking has been associated with illegal, or at least ethically questionable. However, a "hack" can also refer a clever workaround or solution to a particularly vexing programming glitch. Hacks often extend the functionality of an application or program beyond its original intent or expected capabilities.

# What you will learn

- We will cover some Google Hacks
- Have a brief look at Google's Custom Search Engines (CSE)
- These hacks will help you to discover and create new research resources

Bausch, Paul, Tara Calishain, and Rael Dornfest. Google Hacks. 3rd. Sebastopol, CA: O'Reilly Media, Inc., 2006. Safari Books Online, Ralph W. Steen Library, Nacogdoches, TX. 06 Nov 2006  
<http://proquest.safaribooksonline.com/home>.

johnny. "Welcome to the Google Hacking Database (GHDB)!."  
[johnny.ihackstuff.com](http://johnny.ihackstuff.com). 2006. IHS. 6 Nov 2006  
<http://johnny.ihackstuff.com/index.php?module=prodreviews>.

# Cataloging the Web



- Over the years great effort has gone into implementing metadata like Dublin Core, SGML, XML and other information into web page's. While there have been a few niche uses found for some of these, to our great exasperation, none have turned into that which deep down inside, every librarian really longs for; a card catalog for the Internet.
- Despite the failure of these schemes to be widely adopted, the web has developed some conventions and even standards for meta data that are used ubiquitously and can be searched with search engines.

# What can Google Really Do?



- If you count all Google services you end up with over seventy different products and services, fifty-two of them are research related.
- Just by studying the documentation and tools provided by Google we can become expert searchers compared to a majority of Internet users
- Being familiar with common meta data on the web even though it lacks a controlled vocabulary is still useful.

# Common Meta Data



- Currently most web pages contain at least the following metadata
- Domain, <Title>, URL, File Name
- site:gov (intitle:database | inurl:database)
- site:edu (intitle:database | inurl:database)
- site:il.us (intitle:database | inurl:database)
- You can of course repeat this search with any domain (.org .net .int .uk etc.) and refine the results to specific types of information by adding keywords for the subject you are interested in like crime, plants, medicine, etc..

# Goggle Hacks and Beyond

- Custom Search Engines (CSE)
- True Virtual Collections
- WWI Diaries and Memoirs
- (WWI OR “The Great War” OR “World War One”) (intitle:diary OR intitle:diaries OR inurl:diary OR inurl:diaries)
- [http://google.com/coop/cse?cx=014997348189176913774%3A\\_b4nqqxhk2k](http://google.com/coop/cse?cx=014997348189176913774%3A_b4nqqxhk2k)
- <http://protoknowledge.com/iwasthere/war/worldwarone/index.html>

# The Three P's Of Search Engine Mastery

- **Pick** two or three search engines.
  - **Print** out and study the help pages
  - **Practice** using them and their special features in your daily work
- 
- Google help Pages A-Z  
[http://www.google.com/support/bin/static.py?page=alpha\\_index.html](http://www.google.com/support/bin/static.py?page=alpha_index.html)  
(Incomplete, Changes Frequently)
  - Google Hacks, Third Edition  
<http://www.oreilly.com/catalog/googlehks3/>
  - Welcome to the Google Hacking Database (GHDB)!  
<http://johnny.ihackstuff.com/ghdb.php>