

WEBSCRAPING GOVERNMENT INFORMATION

Michael Samson, Public Service Librarian, Wayne State University, Arthur Neef Law Library
2008 Depository Library Council Spring Meeting - March 31, 2008, Kansas City, MO

This presentation introduces government documents librarians to several exciting, easy-to-use and free web 2.0 information harvesting tools. These tools provide automated 'webscraping' of Internet government information and archival harvesting procedures so that librarians can customize the delivery of depository library information to their patrons via library web pages, course web systems (e.g.: Blackboard, Pipeline), and online catalogs (e.g.: Innovative).

TABLE OF CONTENTS

WebScraping Government Information


1. Innovative catalog & Law Library RSS feeds
2. Innovative catalog & CLJC RSS feeds
3. PACER feeds via Justia
4. RSS 2 HTML tools
5. Step-by-step capture screens (if NO Internet access@FDLP conference!)
 - a. USA.gov (feeds used for demo)
 - b. RSS2HTML (step-by-step demo)
 - c. RSS2JS (step-by-step demo)
 - d. HTML code
6. FREE & Open Source applications
7. More Web tools


WEBSCRAPING GOVERNMENT INFORMATION


1. INNOVATIVE CATALOG & LAW LIBRARY RSS FEEDS

 **Innovative Interfaces - A Neef News Feeds (RSS)**
<http://elibrary.wayne.edu/screens/feeds_law.html>

2. INNOVATIVE CATALOG & CLJC RSS FEEDS


 **Current Law Journal Content (WSU catalog)**
<<http://elibrary.wayne.edu/search/?searchtype=t&searcharg=current+law+j&searchscope=47&SORT=R&exte...>>


 **Adelaide Law Review (non WSU)**
<<http://elibrary.wayne.edu/search?tAdelaide+Law+Review/tadelaide+law+review/1%2C1%2C2%2CB/frame...>>

 **British Tax Review (WSU contribution)**
<<http://elibrary.wayne.edu/search~/?searchtype=t&searcharg=british+tax+review&SORT=R&searchscope=47>>

3. PACER FEEDS VIA JUSTIA

 **Justia and RSS feeds 2 PACER** <<http://onward.justia.com/justia-news-149-federal-court-case-filings-docketsjustiacom.html>>






 **Justia Federal District Court Filings and Dockets** <<http://dockets.justia.com/>>
Justia Federal Filings

 **Docket Search: Filed in Michigan - Justia Federal District Court Filings and Dockets & RSS feed** <<http://dockets.justia.com/search?query=&search=Search&stateorcourt=state-mi&lawsuittype=&document...>>

Justia Federal Filings - Docket Search: Filed in Michigan



 **Justia RSS feed 2 PACER Michigan dockets**

4. RSS 2 HTML TOOLS

-  [mySyndicaat <http://www.mysyndicaat.com/home>](http://www.mysyndicaat.com/home)
-  [RSS2HTML \(no account needed; use 4 demo\) <http://www.rss2html.com/>](http://www.rss2html.com/)
-  [RSSinclude <http://www.rss-info.com/en_rssinclude.html>](http://www.rss-info.com/en_rssinclude.html)
-  [RSS Feed Reader <http://www.rssfeedreader.com/>](http://www.rssfeedreader.com/)
-  [RSS-to-JavaScript \(no account needed; use 4 demo\) <http://www.rss-to-javascript.com/>](http://www.rss-to-javascript.com/)

5. STEP-BY-STEP CAPTURE SCREENS (IF NO INTERNET ACCESS@FDLP CONFERENCE!)

A. USA.GOV (FEEDS USED FOR DEMO)

-  [USA.gov RSS Library <http://www.usa.gov/Topics/Reference_Shelf/Libraries/RSS_Library.shtml>](http://www.usa.gov/Topics/Reference_Shelf/Libraries/RSS_Library.shtml)
-  [USA.gov-Reference and General Government RSS Feeds <http://www.usa.gov/Topics/Reference_Shelf/Libraries/RSS_Library/Reference.shtml>](http://www.usa.gov/Topics/Reference_Shelf/Libraries/RSS_Library/Reference.shtml)
-  [Library of Congress RSS Feeds <http://www.loc.gov/rss/>](http://www.loc.gov/rss/)


B. RSS2HTML (STEP-BY-STEP DEMO)

-  [1. RSS2HTML <http://www.rss2html.com/>](http://www.rss2html.com/)
-  [2. RSS2HTML <http://www.rss2html.com/step-1.php>](http://www.rss2html.com/step-1.php)
-  [3. RSS2HTML <http://www.rss2html.com/step-2.php>](http://www.rss2html.com/step-2.php)
-  [4. RSS2HTML <http://www.rss2html.com/step-3.php>](http://www.rss2html.com/step-3.php)
-  [5. RSS2HTML <http://www.rss2html.com/step-4.php>](http://www.rss2html.com/step-4.php)



C. RSS2JS (STEP-BY-STEP DEMO)

-  [1. RSS2JS <http://www.rss-to-javascript.com/p/138.html>](http://www.rss-to-javascript.com/p/138.html)
-  [2. RSS2JS -PREVIEW <http://www.rss-to-javascript.com/?p=151%2C381&src=http%3A%2F%2Fwww.loc.gov%2Frss%2Flaw%2Fnews.xml...>](http://www.rss-to-javascript.com/?p=151%2C381&src=http%3A%2F%2Fwww.loc.gov%2Frss%2Flaw%2Fnews.xml...>)
-  [3. RSS2JS -GENERATE JS <http://www.rss-to-javascript.com/?p=151%2C381&src=http%3A%2F%2Fwww.loc.gov%2Frss%2Flaw%2Fnews.xml...>](http://www.rss-to-javascript.com/?p=151%2C381&src=http%3A%2F%2Fwww.loc.gov%2Frss%2Flaw%2Fnews.xml...>)






D. HTML CODE

-  Web page HTML code (PASTED link & JS code)

6. FREE & OPEN SOURCE APPLICATIONS

-  [The Best Free Software \(PCMAG.COM\) <http://www.pcmag.com/print_article2/0,1217,a=200444,00.asp>](http://www.pcmag.com/print_article2/0,1217,a=200444,00.asp)
-  [The Top 50 Proprietary Programs that Drive You Crazy and Their Open Source Alternatives \(WHDb.COM\) <http://whdb.com/2008/the-top-50-proprietary-programs-that-drive-you-crazy-and-their-open-source-a...>](http://whdb.com/2008/the-top-50-proprietary-programs-that-drive-you-crazy-and-their-open-source-a...>)

7. MORE WEB TOOLS

-  [Browsershots <http://browsershots.org/>](http://browsershots.org/)
-  [dapper <http://www.dapper.net/>](http://www.dapper.net/)
-  [METAfy \(MacOS X\) <http://www.metafy.com/>](http://www.metafy.com/)
-  [openkapow <http://openkapow.com/Default.aspx>](http://openkapow.com/Default.aspx)
-  [screen-scraper <http://www.screen-scraper.com/>](http://www.screen-scraper.com/)