Advanced Online Media
Dr. Cindy Royal
Texas State University - San Marcos
School of Journalism and Mass Communication

Web Scraping and Extracting

Download Them All – Firefox addon http://www.downthemall.net/ - Extracts links and images on a page. Install the plugin, then find it on context menu or under tools. It's free, but there's a $10 suggested contribution.

Chrome Scraper Extension – https://chrome.google.com/extensions/detail/mbigbapnjcgaffohmbkdlecaccepngjd free extension for Chrome. Select content on a page, use the context menue to Scrape Similar. Can export results to a Google Doc.

iWeb Tools Link Extractor - http://www.iwebtool.com/link_extractor Put in a url and it extracts the links. Limited to 10 requests per hour in free version. Web-based.

OutWit Hub - Firebug Addon - Limited in free version, but only $34.90 for pro version http://www.outwit.com/products/hub/ Has quick start tutorials. Install the plugin, use the icon in the toolbar to extract when you are on a page. Click Data, Tables or Lists. You can also look at links and images on a page. Under automators, scapers – you can have a little more flexibility to extract certain items, customize.

ScraperWiki.com - for developers, journos and researchers. Full support and access to data and tools for scraping. Start by modifying Hello World file in your language of choice. Lots of examples and tutorials.

Yahoo Pipes – feed aggregator and manipulator; scraping RSS feeds. Tutorial on how to build a pipe - http://www.youtube.com/watch?v=J3tS_DkmbVA
More complex, but visual and easy to use. Consider your data sources and how you'd like to merge or filter. Then you can publish and use in a variety of ways, including a badge (embed code).

Use programming for utmost flexibility - Ruby, PHP, Python, etc. Make your own scripts to extract the data

Scraping Resources

Scraping for Journalists - Dan Nguyen
http://danwin.com/2010/04/coding-for-journalists-101-a-four-part-series/

How to Scrape Data from Websites Without Programming Skills - Michelle Minkoff - http://www.poynter.org/how-tos/digital-strategies/e-media-tidbits/102589/how-to-scrape-websites-for-data-without-programming-skills/

Almost Scraping Presentation from IRE - Michelle Minkoff - http://www.slideshare.net/michelleminkoff/almost-scraping-web-scraping-without-programming

An Introduction to Compassionate Screen Scraping – Will Larson - http://dev.lethain.com/an-introduction-to-compassionate-screenscraping/

Use programming for utmost flexibility - Ruby, PHP, Python, etc.

Other Development Tools – Firefox Extensions
Firebug
FireFTP
Firefox Web Developers Tools