

The Internet

$$\text{Net Effect on Research Productivity} = \frac{\text{readily available information}}{\text{ways to procrastinate}} = \text{None.}$$

JORGE CHAM © 2007

WWW.PHDCOMICS.COM

using time wisely in research:

computer tools for keeping up + collaborating

jake hofman
physics department
columbia university
2007.09.28

outline

- ▶ rss feeds
- ▶ paper/reference management
- ▶ version control/backup
- ▶ scripting/shell tips

disclaimer

- ▶ mac osx & linux > windows
- ▶ get cygwin for windows: <http://www.cygwin.com/>
- ▶ not an expert, but have some experience
- ▶ please interrupt with comments, suggestions, etc.

rss feeds

- ▶ rss = “really simple syndication”
- ▶ compact format for summarizing internet content
- ▶ published by source on internet
- ▶ user subscribes and reads with client or “feed reader”
- ▶ advantages over email alerts:
 - ▶ inboxes are for mail, not news
 - ▶ centralized
 - ▶ easily flag/tag/share individual items



feed readers

- ▶ stand-alone applications
 - ▶ netnewswire (mac), feed demon (win): <http://www.newsgator.com/Individuals/Default.aspx>
- ▶ browsers w/ integrated readers
 - ▶ safari
 - ▶ firefox
- ▶ website readers
 - ▶ <http://reader.google.com>

google reader

- ▶ accessible from any web browser, e.g. mobile
- ▶ flag items for followup
- ▶ tag interesting items to share
- ▶ integrated email
- ▶ keyboard navigation

The screenshot displays the Google Reader web interface. At the top, the Google Reader logo is visible. Below it, there are navigation links: Home, All items (8), Starred items, Shared items, and Trends. A green bar contains 'Add subscription' and 'Browse' buttons. The main content area is titled 'science-journals' and shows a list of articles with a 'Show: 0 new items - all items' filter and 'Mark all as read' and 'Refresh' buttons. The list includes articles from PubMed and Recent Articles in Phys. Rev. with titles like 'hidden markov', 'cell movement', 'NMR Time Reversal as a Prob', etc. A sidebar on the left shows a tree view of subscriptions, including 'science-blogs' and 'science-journals' with various sub-subscriptions like 'BioCurious', 'bohemian scientist', 'Machine Learning (The...)', 'Machine Learning, etc', 'Not Even Wrong', 'Statistical Modeling...', 'Structure+Strangeness', 'Biophysical Journal', 'Cell', 'Cell Migration Gateway', 'CiteULike: Journal of...', 'PubMed - 0370626[[id]', 'PubMed - 101136452[[id]', 'PubMed - 8409413[[id]', 'Image Processing, IEE...', 'JMLR', 'Journal of Cell Biology', and 'Latest Issue of Machi...'. A 'Manage subscriptions' link is at the bottom of the sidebar.

example feeds: news

- ▶ <http://www.nytimes.com/services/xml/rss/index.html>
- ▶ <http://www.reuters.com/tools/rss>
- ▶ <http://news.google.com>

example feeds: blogs

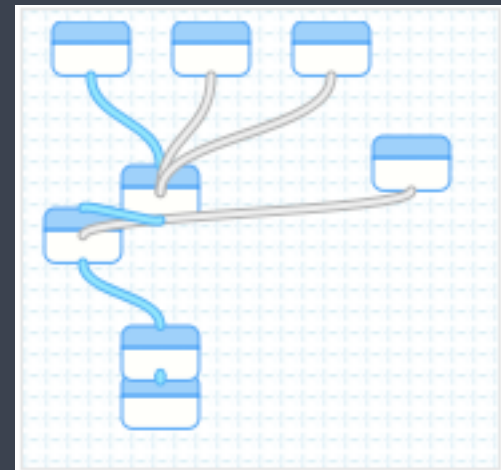
- ▶ gelman: <http://www.stat.columbia.edu/~gelman/blog/>
- ▶ machine learning: <http://hunch.net/>
- ▶ structure + strangeness: <http://www.cs.unm.edu/~aaron/blog/>

example feeds: journals/jobs

- ▶ by publisher:
 - ▶ <http://feeds.aps.org/>
 - ▶ <http://export.arxiv.org/rss/physics.data-an>
- ▶ by third party:
 - ▶ <http://www.citeulike.org/rss/journal/mitpress-jmlr>
- ▶ search over journals
 - ▶ <http://www.pubmed.org>
 - ▶ <http://careers.physicstoday.org/search/results/index.cfm?KEYWORDS=postdoc&ss=1&display=rss>

manipulating feeds: yahoo pipes

- ▶ many feeds = too much info
- ▶ <http://pipes.yahoo.com>
- ▶ filter feeds on keywords
- ▶ combine multiple feeds
- ▶ ... much, much more



example pipes

- ▶ generic feed filter: http://pipes.yahoo.com/pipes/pipe.info?_id=4uqGy9Vt3BGgxxgTdy6ky6g
- ▶ PNAS filtered: http://pipes.yahoo.com/pipes/pipe.info?_id=xqBLH5tG3BG5Pq3HdLq02Q
- ▶ scientific grant search: http://pipes.yahoo.com/pipes/pipe.info?_id=PhKiFckZ3BG2kyovn0artA
- ▶ nytimes through youtube: http://pipes.yahoo.com/pipes/pipe.info?_id=GrnthQ3d2xGYXqOolfXiAA

creating feeds: feed43

- ▶ not all pages offer feeds
- ▶ but many have regular structure
- ▶ create feeds from any page



USA Weekend Box-Office Summary

30 December 2005

Rank	Title	Weekend Gross
1.	The Chronicles of Narnia: The Lion, the Witch and the Wardrobe (2005)	\$33.7M \$224.0K
2.	King Kong (2005)	\$31.8M \$174.0K
3.	Fun with Dick and Jane (2005)	\$21M \$44.4K
4.	Cheaper by the Dozen 2 (2005)	\$18.9M \$36.7K
5.	Rumor Has It... (2005)	\$11.8M \$26.9K
6.	The Family Stone (2005)	\$10.3M \$46.3K
7.	Memoirs of a Geisha (2005)	\$10.2M \$30.4K
8.	The Ring 2 (2005)	\$8.55M \$21.7K
9.	Harry Potter and the Goblet of Fire (2005)	\$7.44M \$277K
10.	Murich (2005)	\$6.44M \$15.9K

R.S.S. FEED

Feed: USA Weekend Box-Office Summary
Title: 30 December 2005

Rank	Title	Weekend Gross
1.	The Chronicles of Narnia: The Lion, the Witch and the Wardrobe (2005)	\$33.7M \$224.0K
2.	King Kong (2005)	\$31.8M \$174.0K
3.	Fun with Dick and Jane (2005)	\$21M \$44.4K
4.	Cheaper by the Dozen 2 (2005)	\$18.9M \$36.7K
5.	Rumor Has It... (2005)	\$11.8M \$26.9K
6.	The Family Stone (2005)	\$10.3M \$46.3K
7.	Memoirs of a Geisha (2005)	\$10.2M \$30.4K
8.	The Ring 2 (2005)	\$8.55M \$21.7K
9.	Harry Potter and the Goblet of Fire (2005)	\$7.44M \$277K
10.	Murich (2005)	\$6.44M \$15.9K

example: ego feed

- ▶ original url: http://www.google.com/search?as_qdr=d&as_q=jake+hofman
- ▶ feed43 edit: http://feed43.com/feed.html?name=google_jake_hofman_24hr
- ▶ resulting feed: http://feed43.com/google_jake_hofman_24hr.xml

example: bishop's publications

- ▶ original url: <http://research.microsoft.com/users/cmbishop/publications.htm>
- ▶ feed43 edit: http://feed43.com/feed.html?name=cmbishop_publications
- ▶ resulting feed: http://feed43.com/cmbishop_publications.xml

example: sciencemag postdoc

- ▶ original url: <http://aaas.sciencecareers.org/js.php?view=1&lookid=aaas&tmpld=&q=postdoc&qMatch=all&qField=All&qSort=date&qCity=&qCountry=&qMiles=&qRegion=aaasregionnortheast&qInd=categoryphysics&qDate=&pp=20&fromSearch.x=57&fromSearch.y=5>
- ▶ feed43 edit: http://feed43.com/feed.html?name=sciencemag_physics_postdoc
- ▶ resulting feed: http://feed43.com/sciencemag_physics_postdoc.xml

example: columbia career events

- ▶ original url: <http://www.cce.columbia.edu/events/index.php>
- ▶ feed43 edit: <http://feed43.com/feed.html?name=cu-cce-events>
- ▶ resulting feed: <http://feed43.com/cu-cce-events.xml>

sharing via feeds

- ▶ <http://www.google.com/reader/shared/{user-id}>
 - ▶ tags (e.g. wigginslab)
 - ▶ shared
 - ▶ starred
- ▶ <http://del.icio.us>
 - ▶ tags
 - ▶ public
 - ▶ popular

organizing papers/references

- ▶ papers (mac): <http://mekentosj.com/papers>
- ▶ bibdesk (mac): <http://bibdesk.sourceforge.net>
- ▶ zotero (firefox plugin): <http://www.zotero.org>

version control for writing/coding

- ▶ keeps and organizes all paper/code revisions
- ▶ automatic merging of multiple users' simultaneous changes
- ▶ store repository locally (network-free access)
- ▶ store repository remotely (collaborate, backup)

version control software

- ▶ concurrent version system (cvs): <http://www.nongnu.org/cvs/>
- ▶ subversion (svn): <http://subversion.tigris.org/>
- ▶ comparisons:
 - ▶ http://www.pushok.com/soft_svn_vscvs.php
 - ▶ <http://users.footprints.net/~kaz/mcvs-svn-comparison.html>

cvcs demo

- ▶ import existing project on laptop to repository on cunix
- ▶ checkout copy of module to laptop
- ▶ make changes on laptop
- ▶ view log, diff (xxdiff?), etc.
- ▶ merge changes

scripting shell tips

- ▶ *bash* aliases (e.g. ls, cd, emacs, pdflatex + bibtex, ssh, cvs, apps, proxy)
- ▶ sshkey (e.g. cvs)
- ▶ sed/awk/perl foo (e.g. gml, svmlight)
- ▶ wget/imagemagick (scripts for figures, movies)
- ▶ books/links

thanks.