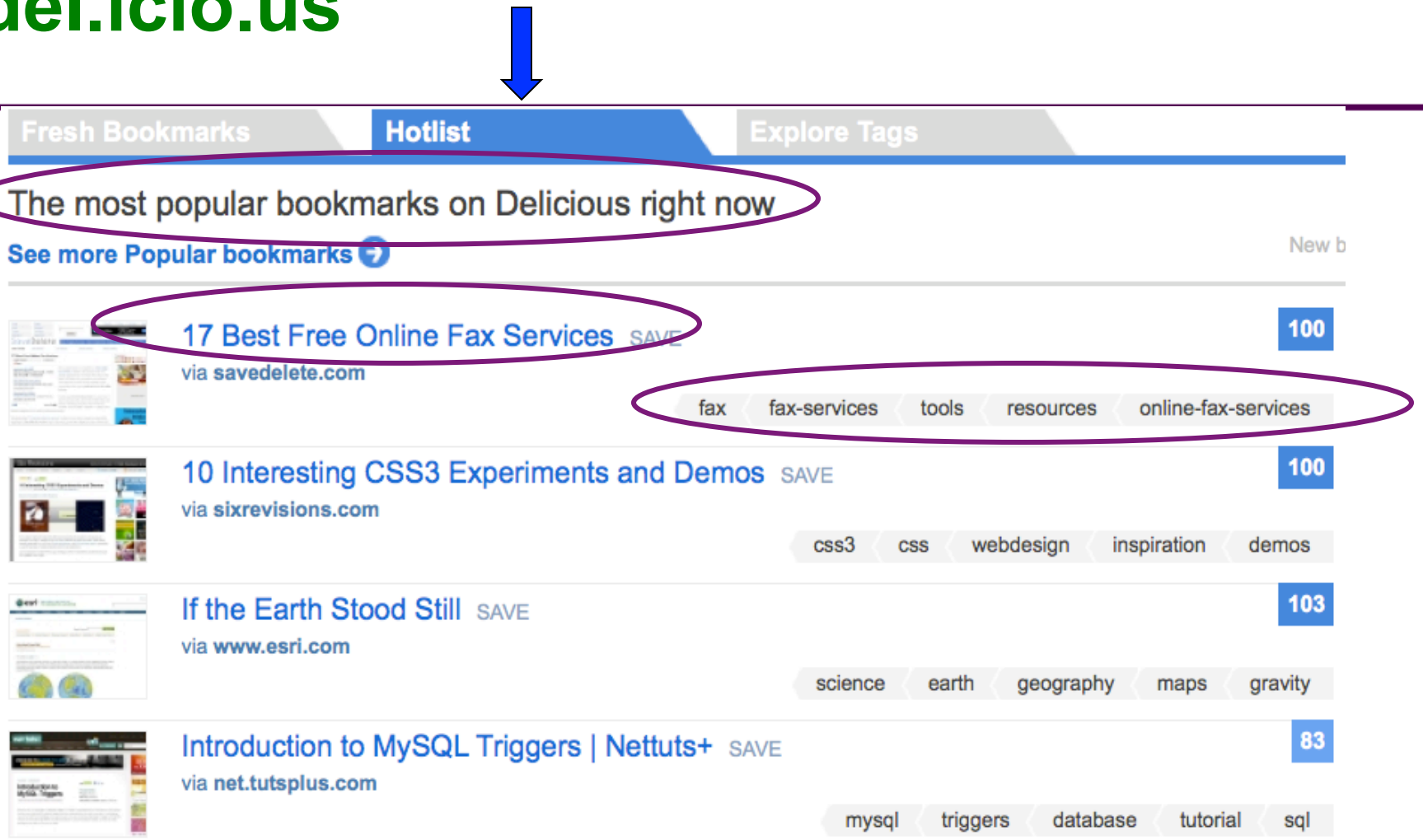
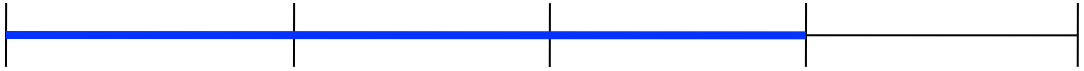
del.icio.us


Fresh Bookmarks
Hotlist
Explore Tags
The most popular bookmarks on Delicious right now
See more Popular bookmarks
New b
17 Best Free Online Fax Services SAVE
via savedelete.com
100
fax
fax-services
tools
resources
online-fax-services
10 Interesting CSS3 Experiments and Demos SAVE
via sixrevisions.com
100
css3
css
webdesign
inspiration
demos
If the Earth Stood Still SAVE
via www.esri.com
103
science
earth
geography
maps
gravity
Introduction to MySQL Triggers | Nettuts+ SAVE
via net.tutsplus.com
83
mysql
triggers
database
tutorial
sql

del.icio.us Hotlists Experiment

- **116,177 del.icio.us users**
  - who tagged 175,691 distinct URLs
  - using 903 tags
  - for a total of 2,322,458 tagging actions
  - for 1 month in 2006
- **Evaluate how networks predict user's interest**
  - *J. Stoyanovich, S. Amer-Yahia, C. Yu, C. Marlow: Leveraging Tagging Behavior to Model Users' Interest in del.icio.us (AAAI Workshop on Social Information Processing 2008)*



A/B testing: user behavior in first 3 weeks to predict 4th week

Data Model

- users $u \in U$, tags $t \in T$, items $i \in I$
- *friends(u)* directional

- *tags(u)*
- *items(u)* & *items(u,t)*

- *taggers(i)* & *taggers(i,t)*

Tagging data has a long tail

- we have to clean it for efficiency (relational processing)
- we removed unpopular tags (< 4 uses) & URLs (< 10 uses), reduced to 27% of original size

Global

10 URLs that are tagged most often over-all

Performance

coverage (`global`) = 3%

scope (`global`) = 100%

Global Top-10

Rank	URL	Votes
1	google.com	980
2	facebook.com	820
3	iTunes.com	729
4	twitter.com	720
5	jonasbrothers.com	680
6	cnn.com	678
7	amazon.com	620
8	yahoo.com	525
9	youtube.com	524
10	**techcrunch.com**	492

Items(Chris)

URL	Tag
jars.com	java
java.sun.com	java
techcrunch.com	news
devshed.com	tutorial

Items(Ben)

URL	Tag
bbc.co.uk	news
pbs.org	news
tomwaits.com	music
nick-cave.com	music
loureed.com	music

Tag-based

- **If a user tags with *sports*, he is interested in sports-related content**
  - interest(u,t) = |items(u,t)| / |items(u)|

Top-10 for "news"

Rank	URL	Votes
1	cnn.com	610
2	bbc.co.uk	503
3	npr.org	427
4	nytimes.com	414
5	slashdot.org	392
6	reuters.com	330
7	news.cnet.com	290
8	msnbc.msn.com	250
9	news.yahoo.com	180
10	digg.com	149

Top-10 for "music"

Rank	URL	Votes
1	iTunes.com	542
2	eMusic.com	420
3	pandora.com	350
4	thebeatles.com	330
5	jonasbrothers.com	215
6	madonna.com	175
7	rhapsody.com	148
8	rollingstones.com	133
9	lastfm.com	120
10	beyonce.com	107

Items(Ben)

URL	Tag
bbc.co.uk	news
pbs.org	news
tomwaits.com	music
nick-cave.com	music
rollingstones.com	music

Build one global hotlist per tag, use in one of two ways

- `best_tag`
  hotlist = top-10 for tag for which user has highest interest

- `dominant_tags`
  hotlist is a combination of up to 3 top-10 lists s.t. interest(u,t) ≥ 0.3 (user has *strong interest* for these tags)

Performance of Tag-based

`best_tag`

coverage = 9%

***scope* = 100%**

`dominant_tags`

1 tag	coverage = 10%	scope = 32%
2 tags	coverage = 14%	scope = 14%
3 tags	coverage = 18%	scope = 6%

Network-based

Choose 10 most popular URLs from those tagged by a user's friends.

coverage (`friends`) = 43%

scope (`friends`) = 31%

Common Interest Networks: URL-interest

Identify the seed -- a set of users who tag many of the same URLs as the user u ("agree with u"). Hotlist = 10 most popular URLs tagged by users in seed.

agr (u,*f*) = |items(*u*) ∩ items(*f*)| / |items(*u*)|

U_{scope}= {u ∈ U | ∃ f ∈ U, agr (*u*, *f*) > threshold}

U_{seed}= {f ∈ U | agr (*u*, *f*) > threshold}

thresh = 0.3	coverage = 61%	scope = 1.2%
thresh = 0.5	coverage = 71%	scope = 0.7%

Common Interest Networks: Tag-URL-Interest

Agreement across the board is rare, let's look at agreement per-tag: may agree with adviser on research, but with mom on cooking.

agr (*u*,*f*,*t*)=|items(*u*,*t*) ∩ items(*f*,*t*)| / |items(*u*,*t*)|

U_{scope}, U_{scope}defined as for `url-interest`, combined as in `dominant-tags`.

scope (`tag-url-interest`) = 7%

Tag/Interest-based Methods: a Comparison

Users in the intersection of `dominant-tags`, `url-interest` and `tag-url-interest`, with a strong interest in 2 tags, all thresholds = 0.3

	$\|U_{scope}\|$	avg ($\|U_{seed}\|$)	coverage
`dominant-tags`	1235	26,856	17%
`tag-url-interest`	1235	227	82%
`url-interest`	205	203	85%