

Information Integration

Exercise 2 – Social News Popularity

Daniel Rinser

Ingmar Rötzer

Starting Point – Digg Stats Mashup

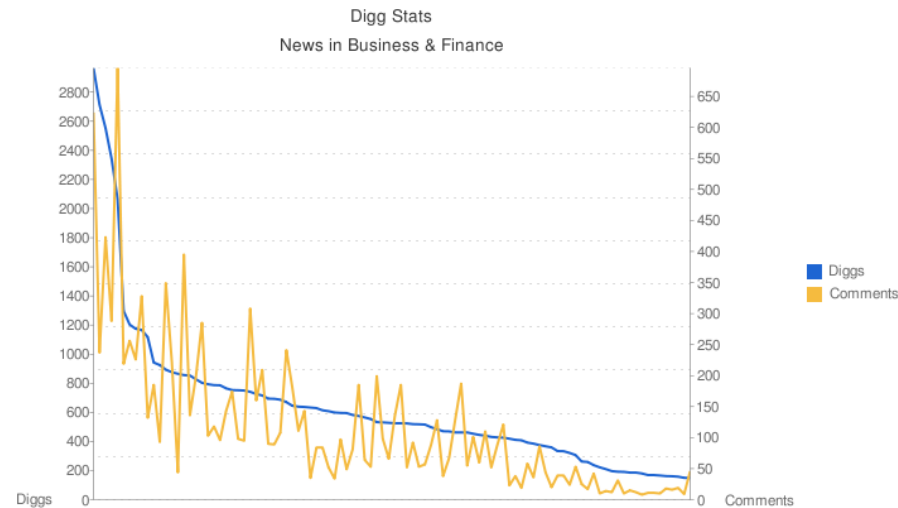
2

Mashup - Digg and Google Chart API

Topic:

Media Type:

Results:



[How YouPorn is ruining Porn DVD Sales](#)

2967 diggs, 624 comments

[Why Zappos Pays New Employees to Quit—And You Should Too](#)

2714 diggs, 237 comments

[Perhaps 60% of today's oil price is pure speculation](#)

2553 diggs, 423 comments

[150 Funniest Resume Mistakes, Bloopers and Blunders Ever](#)

2343 diggs, 288 comments

[Starbucks logo under fire: shows partially covered breasts](#)

2065 diggs, 696 comments

- Integration of further social news/bookmarking services
 - Yigg (API)
 - del.icio.us (API)
 - Mister Wong, StumbleUpon, LinkArena, ...

- Visualisation of "Popularity" in a chart

- Problem: Categories vs. Tags
 - Digg and Yigg arrange the items in categories
 - Categories differ very much from each other (not mergeable)
 - Bookmarking sites mostly rely on social tagging
- Solution: Keyword search

Keyword:

Domain:

Results: ▼

■ SocialItem

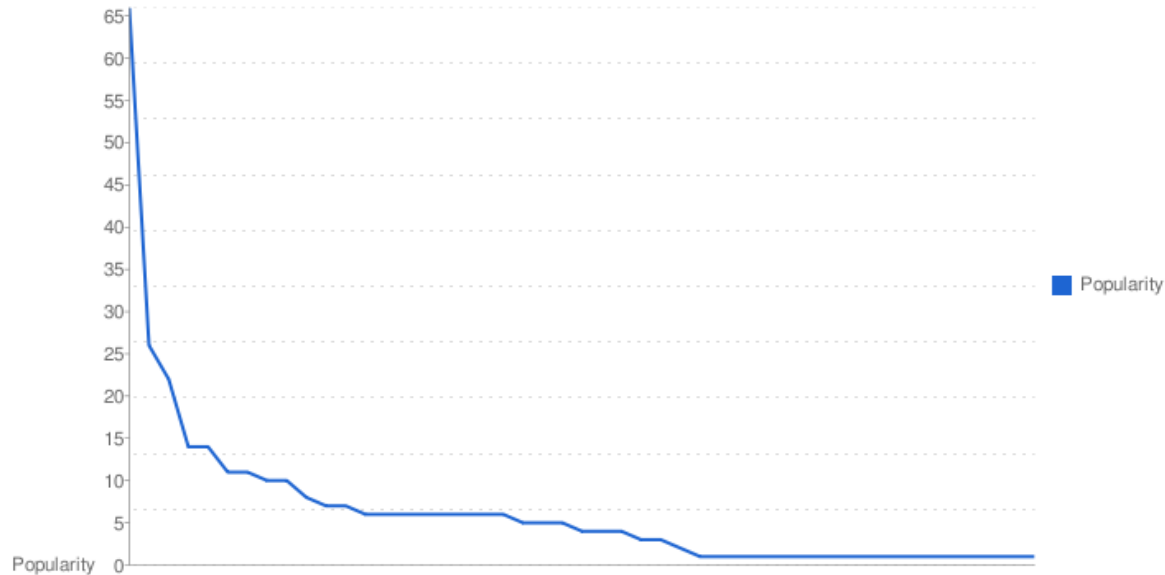
- Title – Name of story or bookmark
- URL
- Date
- Popularity – Diggs or number of bookmarking users
- Tags – if available
- Origin


- **del.icio.us:** API only for personal bookmarks -> kicked!
- **Yigg:** API no free text search, google as search engine -> kicked!
- **Digg:** nice API, but does not allow free text search! -> web scraping
- **Diigo:** no API -> web scraping
- **Mister Wong:** same

- Popularity
 - Social news services: votes/diggs
 - Social bookmarking services: number of users
 - Solution: scale factor
- Date
- HTML structure

- Merge identical items from different sources (duplicate detection)
 - Identified by same URL
 - Popularity added together
 - Tags merged
- Domain filtering
- Sort by popularity
- Reduce result set

Social News Stats
Keyword: gentoo



 **Gentoo Linux -- Gentoo Linux News**

Tags: logiciel_libre source gentoo linux sysadmin distribution distro, Popularity: 66, 2008-05-13

 **Gentoo Linux Wiki**

Tags: gentoo wiki, Popularity: 26, 2006-06-03

 **Free Linux Disks (not Ubuntu)**

Popularity: 22, 2006-07-13

 **3D Desktop! TouchScreen and XGL on Linux!**

Popularity: 14, 2006-08-18