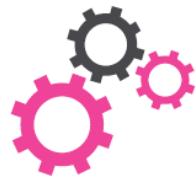


# User-Generated Content for the Semantic Web

Pierre Andrews

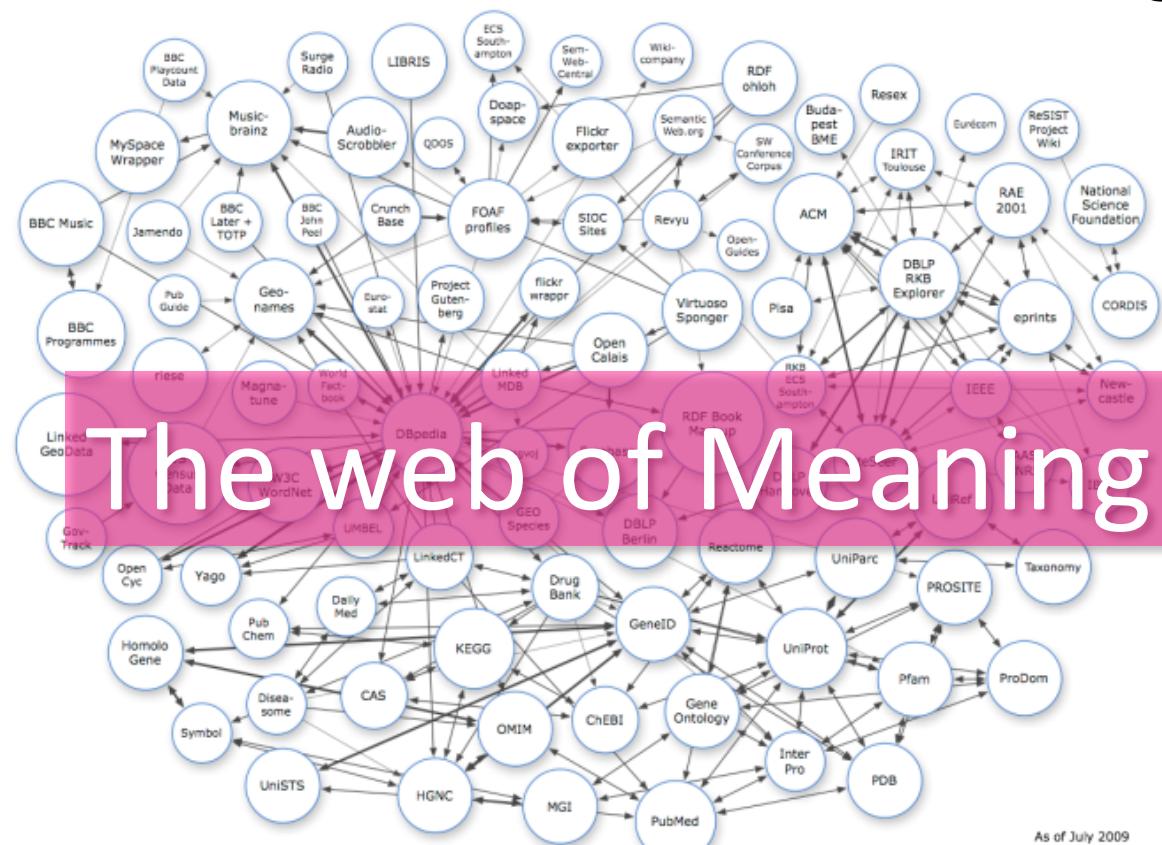
University of Trento, Italy





# The Semantic Web?

# The web of Data? The web of Things?





- Search, Navigation

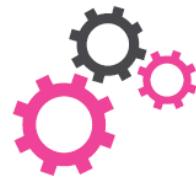


- Question Answering

- “Rich” Data



- To connect heterogeneous systems



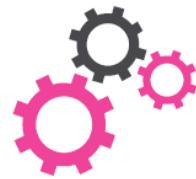
# Users vs. the Semantic Web

## Human Understandable

- **High level representation**
- Images
- Natural Language

## Machine Understandable

- **Low level representation**
- Non ambiguous
- Small context



# Users vs. the Semantic Web





# Insemitives Project



*Realizing the Semantic Web by  
encouraging millions of end-users to  
create semantic content.*



# Insemtives: How?

- Methodologies, methods, tools for the **large-scale** creation of semantic content.
  - Driven by ideas from **incentive theory** and **participatory design**.
  - Optimally combining human input and automatic techniques.
  - For a wide range of content types (**text, images, Web services**).



# What is here? Two examples

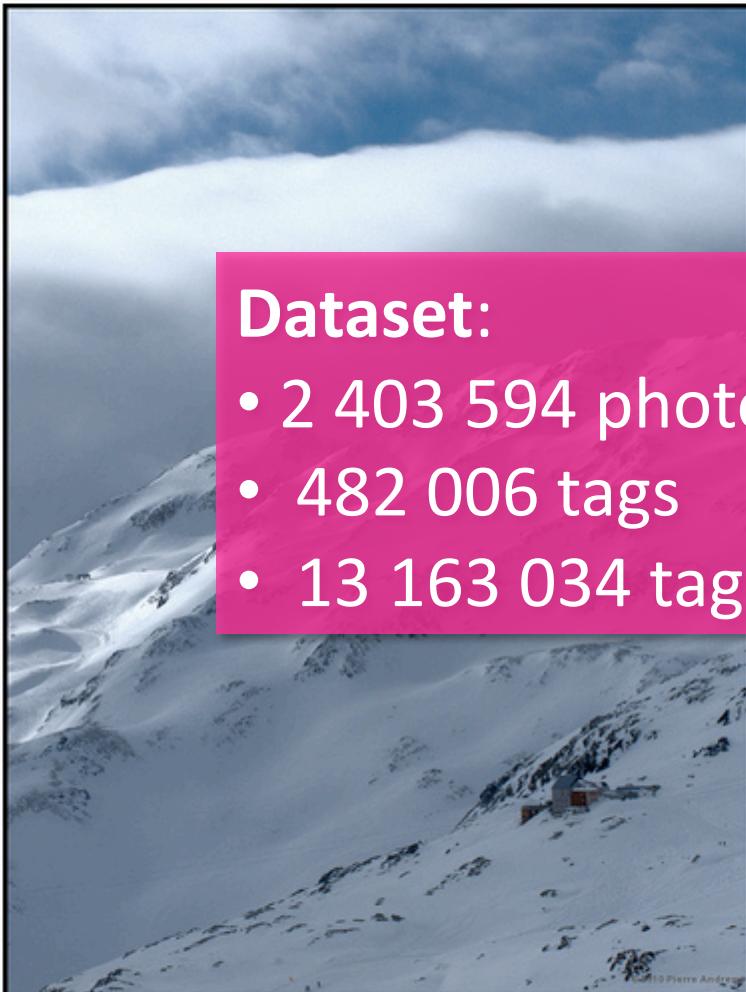
- Web 2.0 Tagging:
  - Flickr Photo Sharing
  - Del.icio.us Social Bookmarking
- Better for, Used for:
  - Search?
  - Sharing?
  - Discovering?
- Machine Understandable?

# The wave

Share This 



ADD NOTE SEND TO GROUP ADD TO SET BLOG THIS ALL SIZES ORDER PRINTS ROTATE EDIT PHOTO DELETE



## Dataset:

- 2 403 594 photos
- 482 006 tags
- 13 163 034 tag-photo pairs

Schöne Aussicht/Bella Vista Mountain Hut (2842m)



Uploaded on March 14, 2010  
by [mortimer?](#)

### [mortimer?'s photostream](#)



200  
uploads

This photo also belongs to:

Vertical Void (Set)

Trento (Set)

◦ 2 people call this photo a favorite

People in this photo

[Add a person](#)

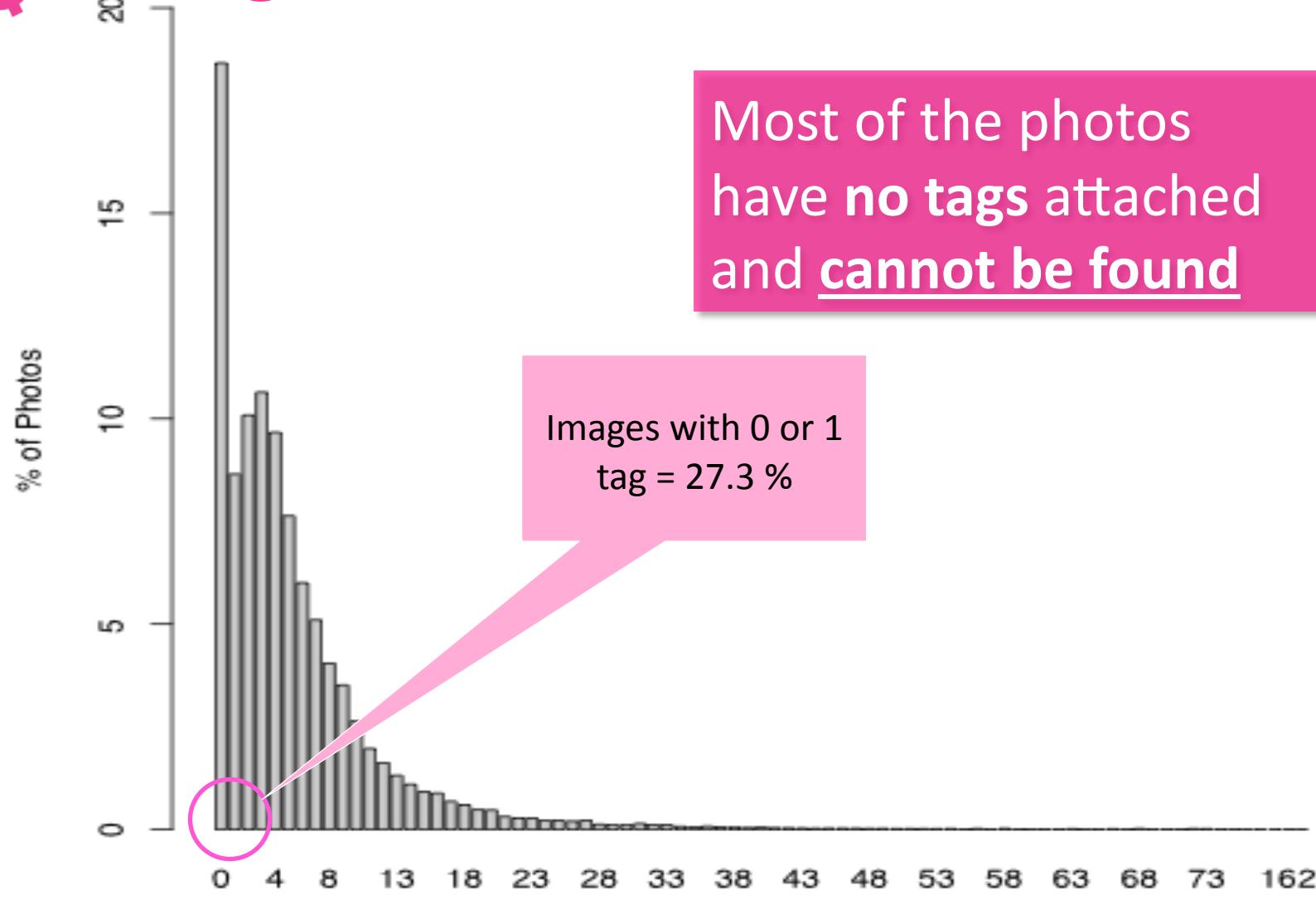
### Tags

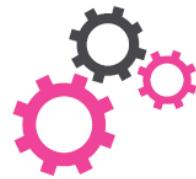
hut ×  
 mountain ×  
 wind ×  
 cloud ×  
 val senales ×  
 mountaineering ×  
 snowboard ×  
 snow ×  
 winter ×

[Add a tag](#)

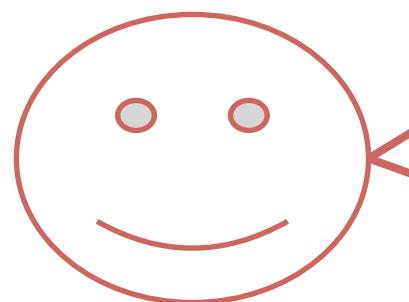


# Tag Use in Flickr





# Tag reuse



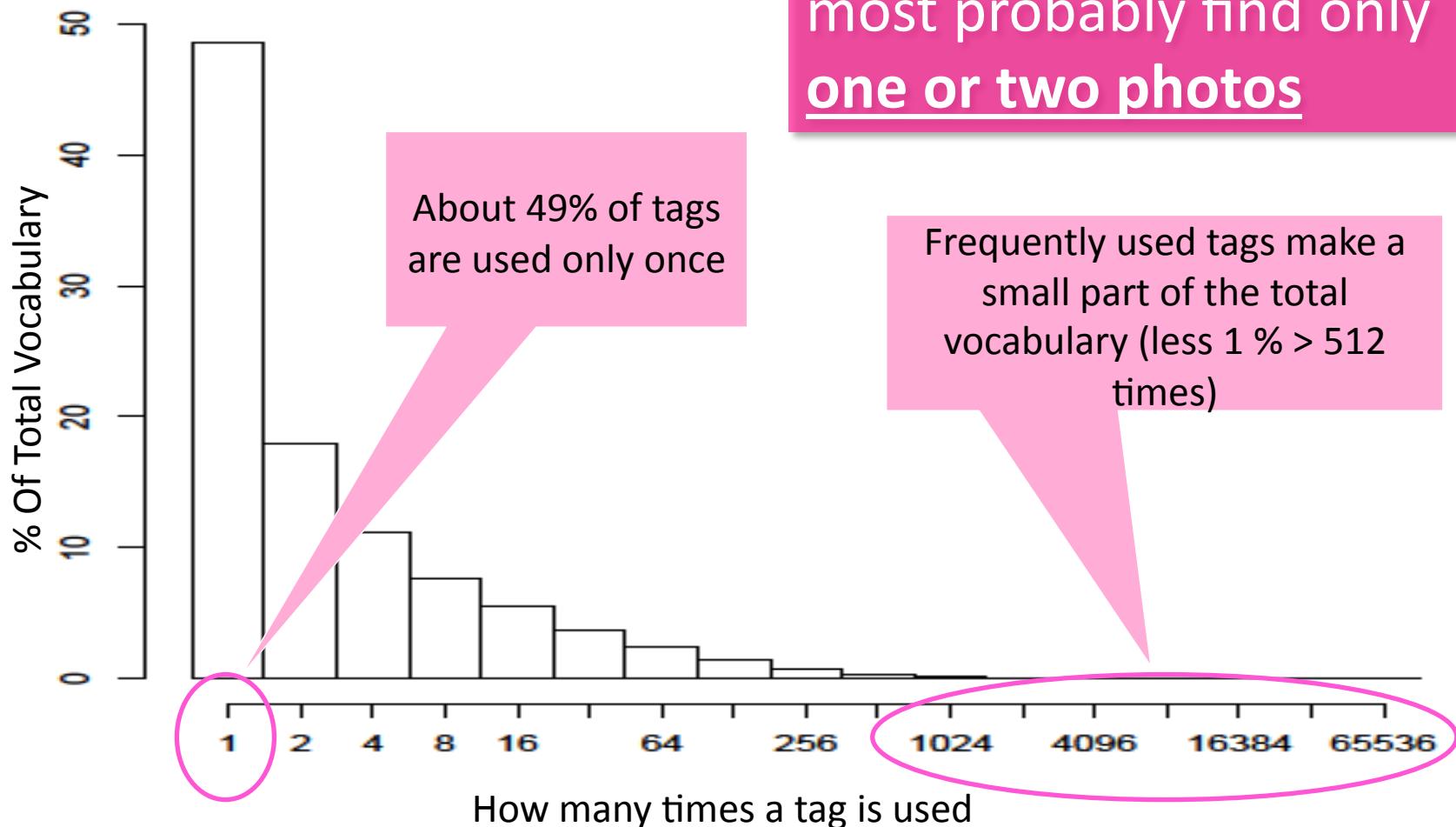
Balloon

Hot air balloon





# Tag reuse in Flickr





# More Complex Annotations?

Drag and drop your photos or videos on to the map!

You're on page 1 of 1  
127 items taken here  
Explore this map

NEW Save map view Find a location GO Map Hybrid Satellite

Map

Hybrid Satellite

120km 15mi

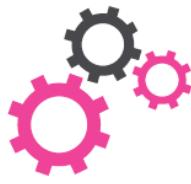
100 items :: 1 selected | Select all | Clear selection

SEARCH More options Jump to date Need help?

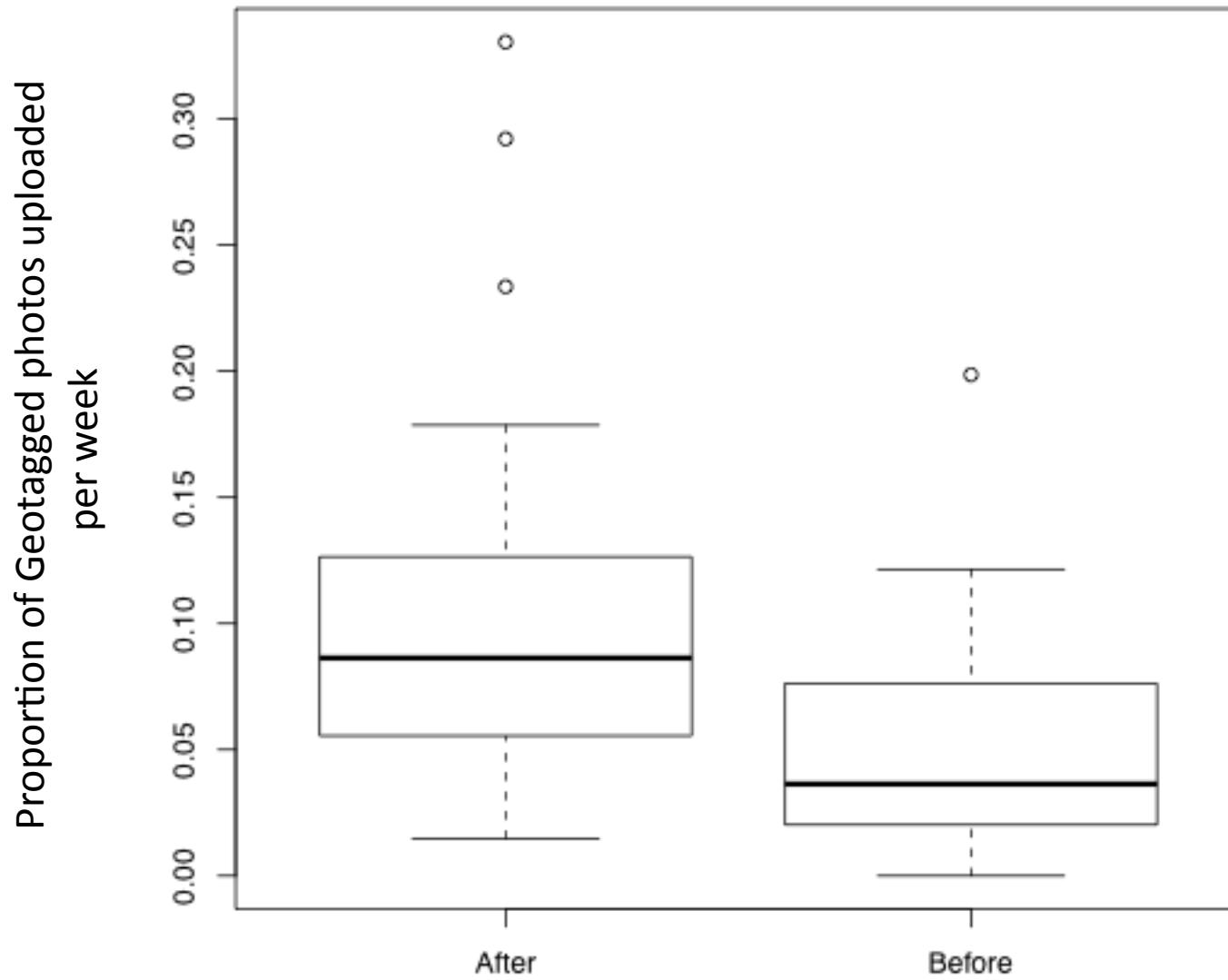
10/21/10

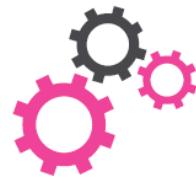
www.insemitives.eu

13



# Flickr GeoTags





## Maybe because...

- Not made/used for discovering through tags
- Not used for “archival”
- Missing a “helpful” tagging interface
- More idea?

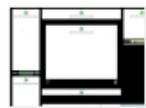


Design\*Sponge » Blog Archive » sweet idea: diy catalog of love [SAVE](#)

via [www.designspongeonline.com](http://www.designspongeonline.com)

76

diy crafts inspiration love design



DOMMUNE [SAVE](#)  
via [www.dommune.com](http://www.dommune.com)

203

music ustream streaming dj culture



Leadership Lessons from Dancing Guy | Derek Sivers [SAVE](#)

via [sivers.org](http://sivers.org)

200

leadership video inspiration psychology culture



Obituary For IE6 [SAVE](#)  
via [www.webresourcesdepot.com](http://www.webresourcesdepot.com)

77

ie6 browser ie hacks webdesign



Cowboy Steak with Chimichurri Sauce [SAVE](#)

via [www.elise.com](http://www.elise.com)

74

steak beef recipes chimichurri recipe



Seth's Blog: Yeah, but he really knows his stuff... [SAVE](#)

via [sethgodin.typepad.com](http://sethgodin.typepad.com)

75

sethgodin management business leadership advice



Zoofs - the video conversation right now [SAVE](#)

via [www.zoofs.com](http://www.zoofs.com)

203

video twitter youtube videos social



iPhone開発向けTips追加分 (20090217) - もとまかのiPhone・iPod touch戯れ日記 [SAVE](#)

via [d.hatena.ne.jp](http://d.hatena.ne.jp)

75

iphone development tips cocoa links



Starting from an external USB storage device (Intel-based Macs) [SAVE](#)

via [support.apple.com](http://support.apple.com)

79

mac osx usb boot apple

## Popular Tags

design  
blog  
video  
software  
tools  
music  
programming  
webdesign  
reference  
tutorial  
art  
web  
howto  
javascript  
free  
linux  
web2.0  
development  
google  
inspiration  
photography  
news  
food  
flash  
css  
blogs  
education  
business  
technology  
travel



Save a Bookmark on Delicious

delicious.com/save?url=http://www.flickr.com/photos/mortimer/4433480528/in/photostream/&title=

**delicious Save Bookmark**

Hi, mortimer ▾

URL  Required

TITLE  Required

NOTES

1000 characters left

**TAGS** te

Space separated, 128 characters per tag

Space separated

**Dataset:**

- 6 200 544 Tags
- 50 185 343 Bookmarks
- 947 729 Users
- 404 940 010 Annotations

Save Cancel

Sort: Alpha | Frequency

metal punk rock

▼ language

ajax cocoa css hcal html java javascript latex lisp perl php prolog python rss  
rubyonrails xhtml xml xsit xul

▼ langue

français italiano

▼ phd

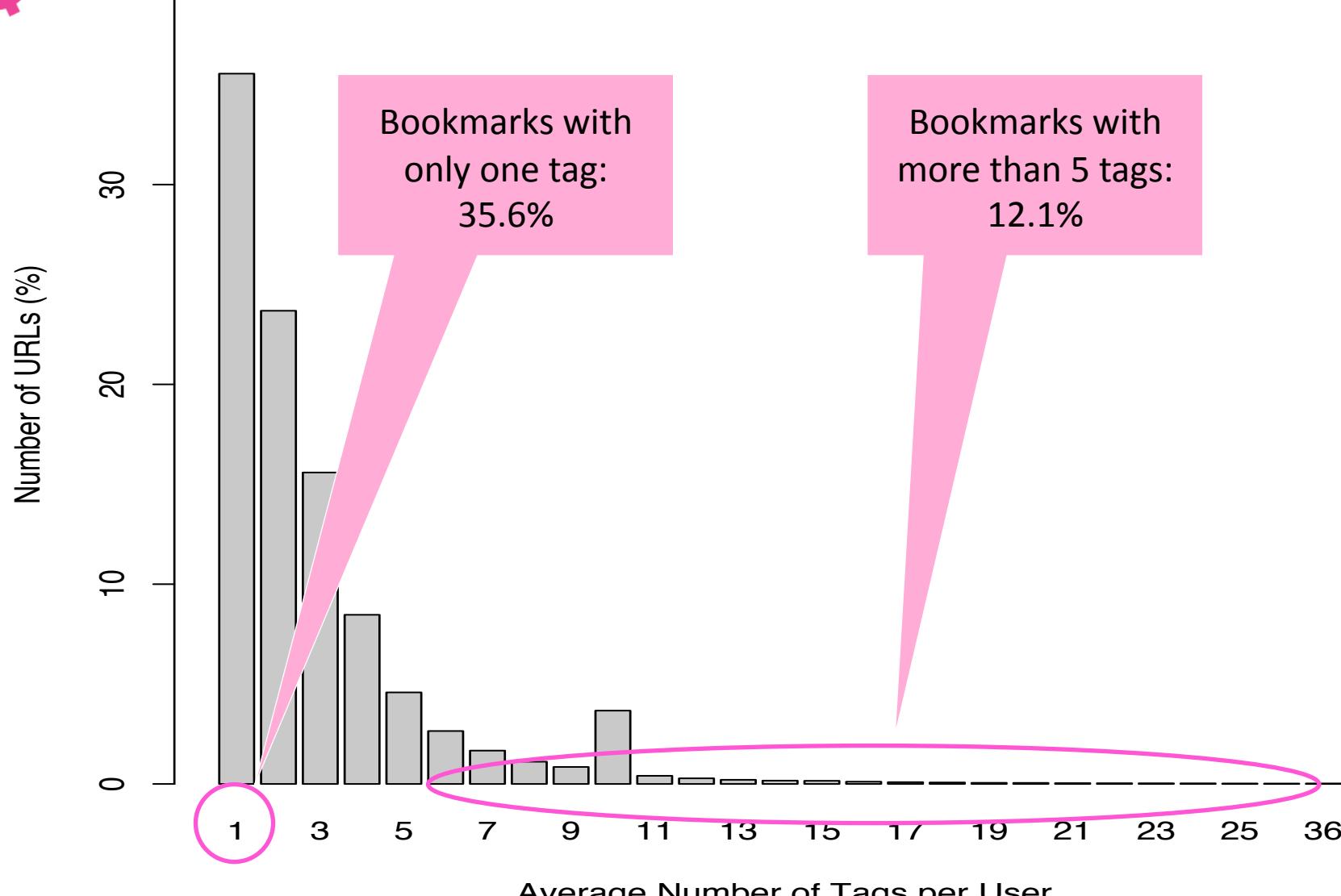
AI algorithmic alice argument aristotle bibliography classification conference corpus  
dialogue discourse done evaluation feature-extraction going graph language  
learning linguistics missed must-submit NLP phd rhetoric rst submitted suresh  
text-processing thesis visualisation

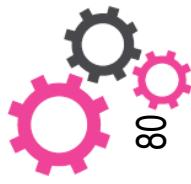
▼ subject

AI argument aristotle evaluation feature-extraction graph linguistics rst text-processing

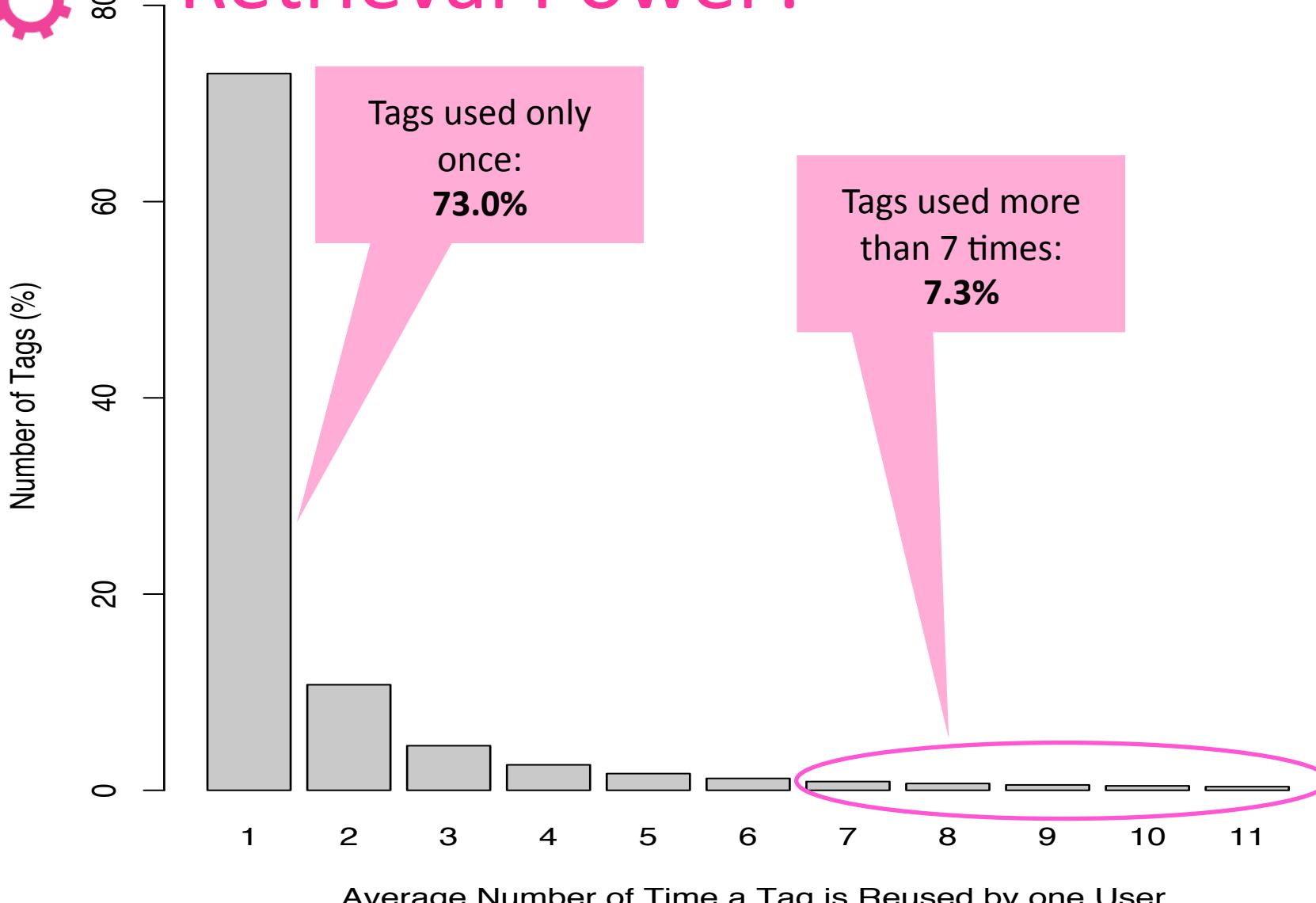


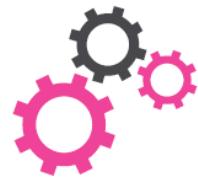
# Retrieval Power?



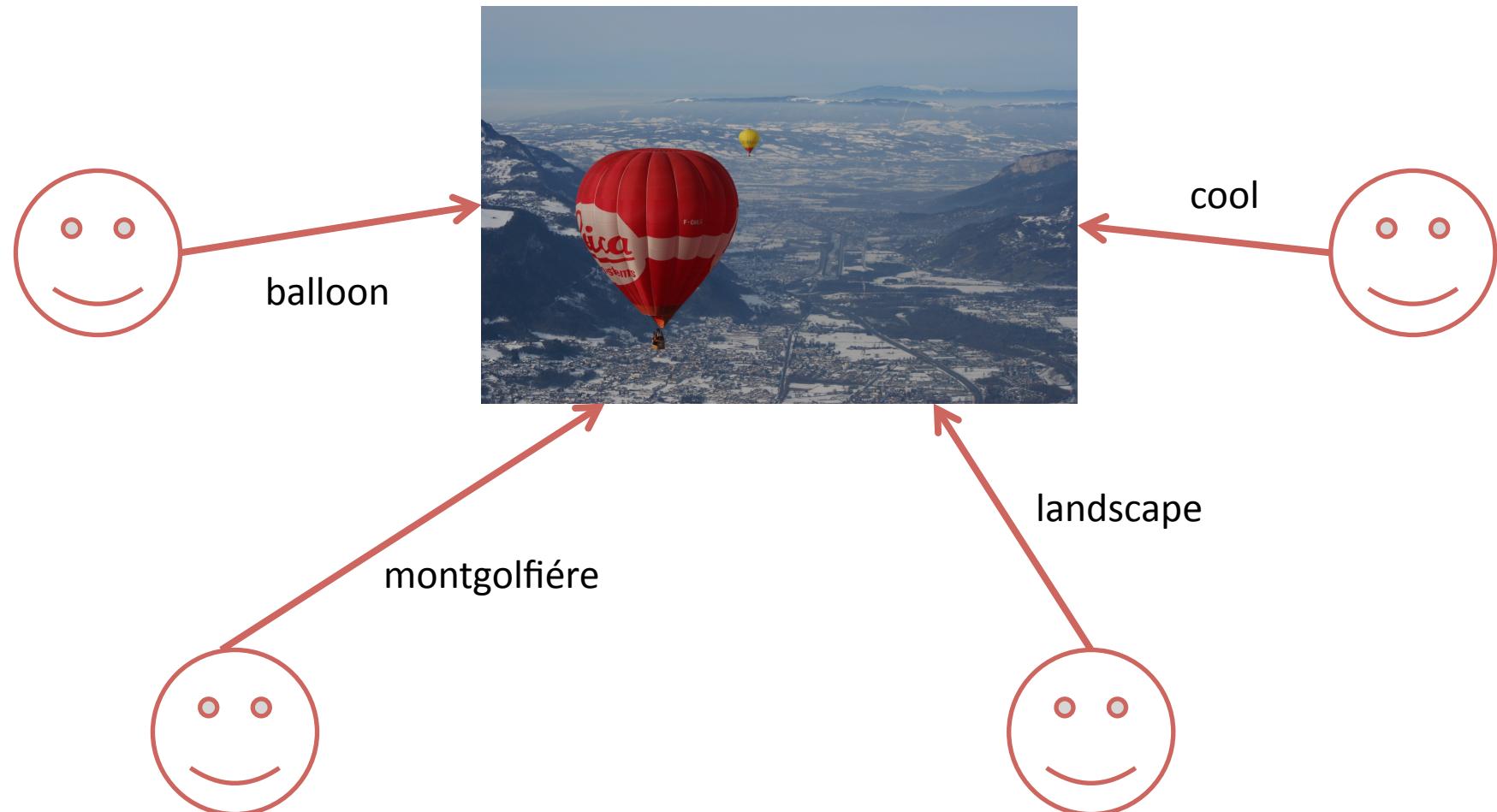


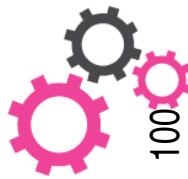
# Retrieval Power?



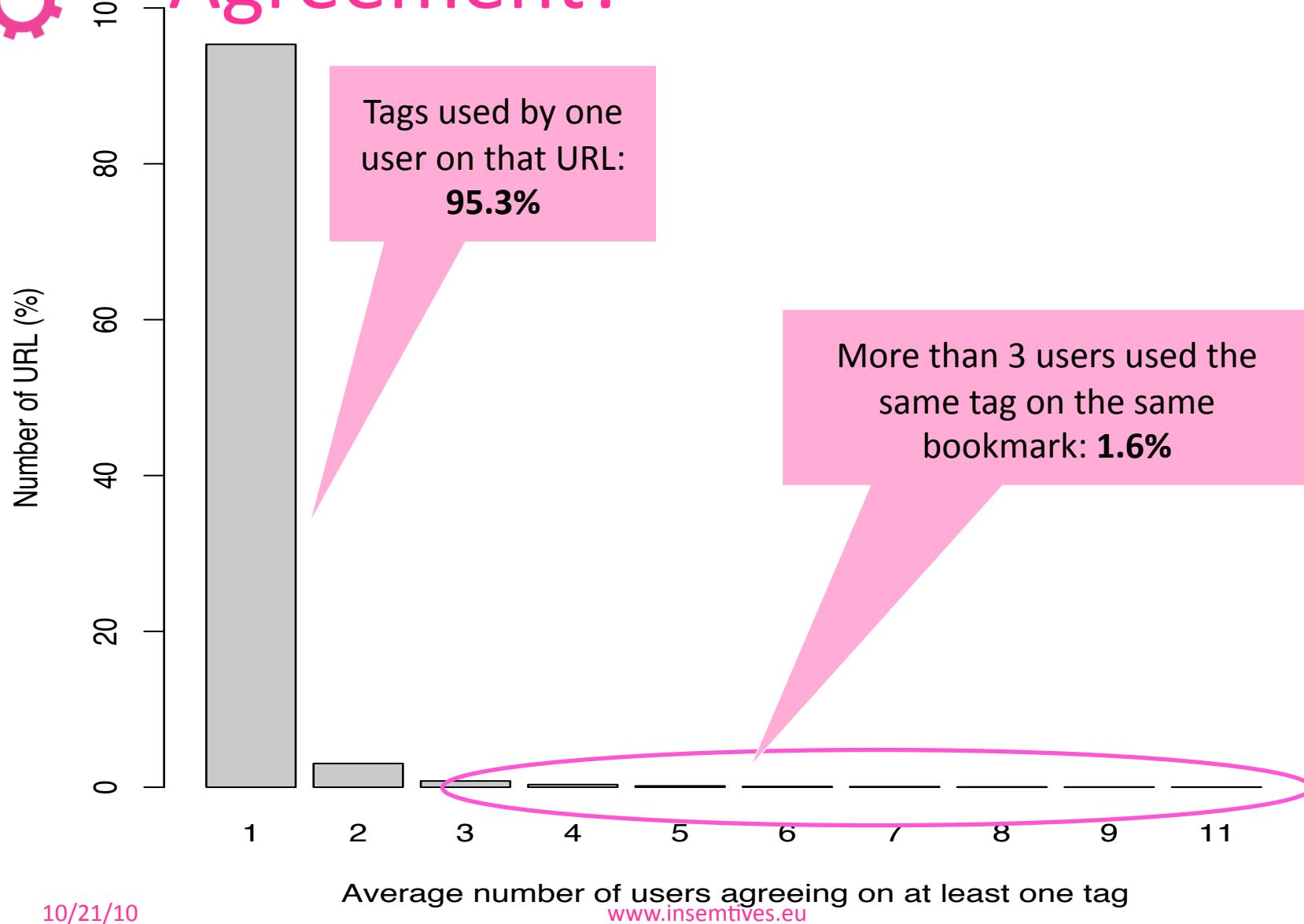


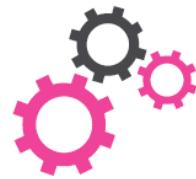
# Agreement vs. Diversity





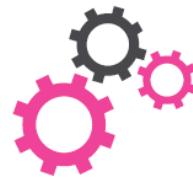
# Agreement?





## Maybe because...

- Not made/used for discovering through tags
- ~~Not used for “archival”~~
- ~~Missing a “helpful” tagging interface~~
- More idea?



# So What?

- Data is in Natural Language
  - Synonymy problem: “image” vs. “picture”
  - Polysemy problem: “java island” vs. “java programming language”
  - Specificity gap problem: “dog” vs. “animal”
- Data is not actually provided by most users
- There is no agreement on the meaning of things



# Insemtives Project



Create incentives, good practice,  
integration to automatic processing  
to “solve” this issues.

*Realizing the Semantic Web by  
encouraging millions of end-users to  
create semantic content.*

# Discussion

