

ANTaReS – Tag suggestion using multiple sources of information



Ícaro Medeiros
Pável Calado
H. Sofia Pinto

16/11/2009

Outline

- Introduction
 - Social Tagging Systems and Folksonomies
 - Motivation
- Related Work
 - Tag Suggestion
- ANTaReS
 - Implementation
 - Results
- Conclusion

Introduction

Social tagging systems

- Users can assign **tags** (keywords) to **resources**
 - Webpages, photos, videos

The screenshot shows the Delicious website interface. At the top, there is a search bar with the text "Search the biggest collection of bookmarks in the universe..." and a "Search Delicious" button. Below the search bar, there are three tabs: "Fresh Bookmarks" (selected), "Popular Bookmarks", and "Explore Tags". The main content area displays a list of bookmarks under the heading "The freshest bookmarks that are flying like hotcakes on Delicious and beyond." Each bookmark entry includes a thumbnail, the title, the source, a "SAVE" button, and a list of tags. The first bookmark is "WhiteHouse.gov Goes Drupal [Updated] | Personal Derr" via personaldemocracy.com, with tags: drupal, opensource, cms, democracy, government, and 24 Related Tweets. The second is "HOW TO: Find and Add Facebook Apps" via mashable.com, with tags: facebook, apps, socialmedia, howto, Facebook, and 17 Related Tweets. The third is "WordPress > Author Exposed « WordPress Plugins" via wordpress.org, with tags: wordpress, plugins, plugin, blogs, blogging, and 15 Related Tweets. The fourth is "Your Windows 7 Upgrade: Why and How - Business Center - PC World" via pcworld.com, with tags: windows, 7, upgrade, Windows, Microsoft, and 11 Related Tweets. The fifth is "Obama presses banks on small business loans | Reuters" via reuters.com, with tags: ping.fm, economy, and 11 Related Tweets. A blue box with the number "14" is visible on the right side of the page.

Delicious: 150,000 bookmarks/day

Youtube

The screenshot shows the YouTube interface for a video titled "Wave - Tom Jobim & Toquinho". The video player shows a man playing a flute. The video has 1475 ratings (5 stars) and 752481 views. The video is at 0:18 / 2:28. The video is categorized as "Música" with keywords: "wave tom jobim toquinho bossa nova jazz joao gilberto miucha vinicius de Moraes mpb brazil brasil". The video is by EdgarRodrigues, who joined on July 7, 2006. There are related video suggestions below, including "WAVE - Tom Jobim" (171317 views) and "Tom Jobim+Vinicius de Moraes+Toquinho+Miúcha" (742610 views).

You Tube Broadcast Yourself™ [icarow](#) [Sair](#)

[Página inicial](#) [Vídeos](#) [Canais](#) [Inscrições](#) [Histórico](#)

Wave - Tom Jobim & Toquinho

EdgarRodrigues
7 de julho de 2006
(menos informações)

Categoria: **Música**
Palavras-chave:
wave tom jobim toquinho bossa nova
jazz joao gilberto miucha vinicius de
moraes mpb brazil brasil

Mais de: EdgarRodrigues

Vídeos relacionados

- WAVE - Tom Jobim**
171317 exibições
nidiagutierrez07
- Tom Jobim+Vinicius de Moraes+Toquinho+Miúcha**
742610 exibições
EdgarRodrigues
- Tom Jobim - One Note Samba**

1475 avaliações **752481** exibições

[MySpace](#) [Facebook](#) [Enviar vídeo](#) (mais opções de compartilhamento)

52 million visitors/month

Social Tagging Systems: features

- Tags are freely chosen keywords
 - **NO** predefined vocabulary (Mathes, 2004)
 - Encourage tag assignment
- Tag assignment is **collaborative**

Generally these systems are in social networks and resource sharing systems

 - Content organization increases STS's popularity


Delicious: features


- Popular tags -> Trends







The screenshot displays the 'Popular Bookmarks' section of the Delicious website. At the top, there are three tabs: 'Fresh Bookmarks', 'Popular Bookmarks' (which is active), and 'Explore Tags'. Below the tabs, the text reads 'The most popular bookmarks on Delicious right now' and 'See more Popular bookmarks'. On the right side, there is a notification: 'New bookmarks saved in the last minute 2 7 9'. The main content area lists several bookmarks, each with a thumbnail, title, 'SAVE' button, and a count of bookmarks. The first bookmark is '20 Fascinating Ancient Maps' with a count of 125 and tags: maps, history, cartography, art, geography. The second is 'webfinger - Project Hosting on Google Code' with a count of 131 and tags: identity, finger, google, webfinger, openid. The third is 'Google Reader - Featured Reading Lists' with a count of 135 and tags: google, rss, blogs, news, googlereader. The fourth is 'How To Create A Sexy Vertical Sliding Panel Using jQuery And CSS3 | Spyre Studios' with a count of 204 and tags: jquery, css, tutorial, slider, css3. The fifth is '28 Free High Quality Fonts to Boost Your Typography | Web Design Ledger' with a count of 250 and tags: fonts, typography, free, font, webdesign. The sixth is '25 Excellent Handwritten Fonts' with a count of 321 and tags: fonts, typography, handwritten, free, design. The seventh is 'Seven | 七 | Sept | семь | Sieben | Sette | Siete | 칠 | Sju | سبعة | Yedi | Saat | 7' with a count of 116 and tags: flash, game, 3d, games, design. On the right side, there is a 'Popular Tags' section with a list of 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.

Tag browsing

Fresh Bookmarks Popular Bookmarks **Explore Tags**

Tags are keywords people use to describe their bookmarks 

 [Learn more about tags](#)

rails See more 	Ruby on Rails guides <small>SAVE</small> 1602
	5 Tips for Sphinx Indexing Engine Yard Blog <small>SAVE</small> 78
photography See more 	Railscasts - Free Ruby on Rails Screencasts <small>SAVE</small> 4634
	21 Helpful Photography Tutorials for a Rainy Day - Phototuts+ <small>SAVE</small> 13
	Image Spark <small>SAVE</small> 1423
ipod See more 	The dirty tricks of food photographers :: Photocritic photography blog <small>SAVE</small> 52
	How Big Is the Apple iPhone App Economy? The Answer Might Surprise You <small>SAVE</small> 68
	A Classroom in your Pocket: iPods in Education <small>SAVE</small> 56
books See more 	iPod and iPhone Firmware Download <small>SAVE</small> 1153
	Main Page - Gutenberg <small>SAVE</small> 16405
	Share Book Recommendations With Your Friends, Join Book Clubs, Answer Trivia <small>SAVE</small> 8975
article See more 	Google Now Offers Over a Million Free Ebooks in EPUB Format <small>SAVE</small> 22
	How to Design the Web in a High-Def World <small>SAVE</small> 173
	Vital Tips For Effective Logo Design How-To Smashing Magazine <small>SAVE</small> 537
baseball See more 	Why Craigslist Is Such a Mess <small>SAVE</small> 199
	Baseball Statistics and Analysis FanGraphs Baseball <small>SAVE</small> 369
	Best Rain Delay Ever <small>SAVE</small> 30
	Baseball-Reference.com - Major League Baseball Statistics and History <small>SAVE</small> 1321

Popular Tags

- design
- blog
- video
- software
- tools
- music
- programming
- webdesign
- reference
- tutorial
- news
- food
- flash
- css
- blogs
- education

Folksonomies

- The result of users interacting with the STS tagging resources
- Collaborative tags +
- Semantics emerging from the community =
- Bottom-up Classification known as Folkosonomy (Mathes, 2004)

STS Drawbacks

- Free vocabulary = ambiguity
 - Polysemy: *whistle* (device) e *whistle* (sound)
 - Homonym: **Java** (language) e **Java** (coffee)
 - Synonym: Buy/purchase
 - Binded Words: **webdev**, **semweb**, **semantic_web**

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 shopping
books mac tips politics science **opensource** games culture

Incorrect tag assignment

Everyone's Bookmarks for:

Java Coffee Break - your free guide to the world of Java programmin...

www.javacoffeebreak.com/index.html

History | Notes

Saved 222 times, first saved by [Nick Sieger](#) on 20 Mar 02. [View Chart](#)

14 OCT 09 [hellosrini](#) Java **Coffee** Break - your free guide to the world of **Java programming** packed full of free articles, tutorials, book reviews, and FAQs

17 SEP 09 [lakshmiteja](#) Java, education

06 SEP 09 [cakturk](#) java programming tutorial

31 AUG 09 [juliosergio](#) La secc. de FAQs de este sitio, trae muchos ejemplos de como programar en Java

24 AUG 09 [Deyi](#) java tutorial

16 AUG 09 [cmetzger4](#) java programming tutorial

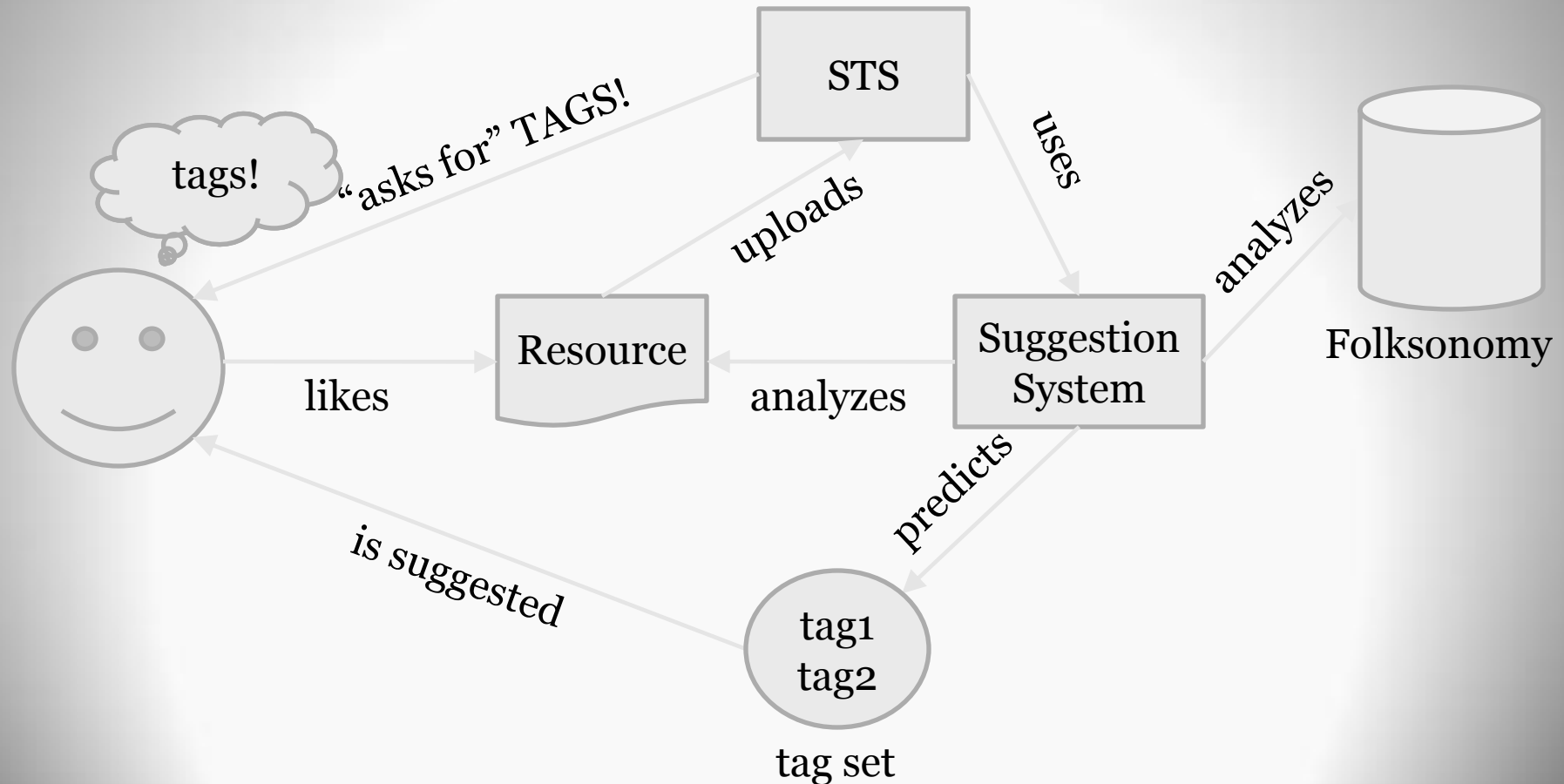
Save this bookmark
Look up another URL

Tags

Top Tags

java	183
programming	95
tutorial	71
tutorials	38
development	36
reference	35
software	22
howto	11
resources	6
learning	6
guide	5
books	5
community	5
webdev	4
faq	4
articles	4
computer	3

Tag suggestion process



Suggestion on Delicious

- Tags mostly used for a resource
- Intersection with personal folksonomy

URL Required

TITLE Required

NOTES

TAGS 1000 characters left

SEND Space separated, 128 characters per tag

Mark as Private

Tags

Sort: Alpha | Frequency

▼ Recommended

[web](#) [computer](#)

▼ Popular

[java](#) [programming](#) [tutorial](#) [tutorials](#) [development](#) [reference](#) [software](#)

▶ All my tags

Advantages of suggestion

(Heymann et al.,2008b)

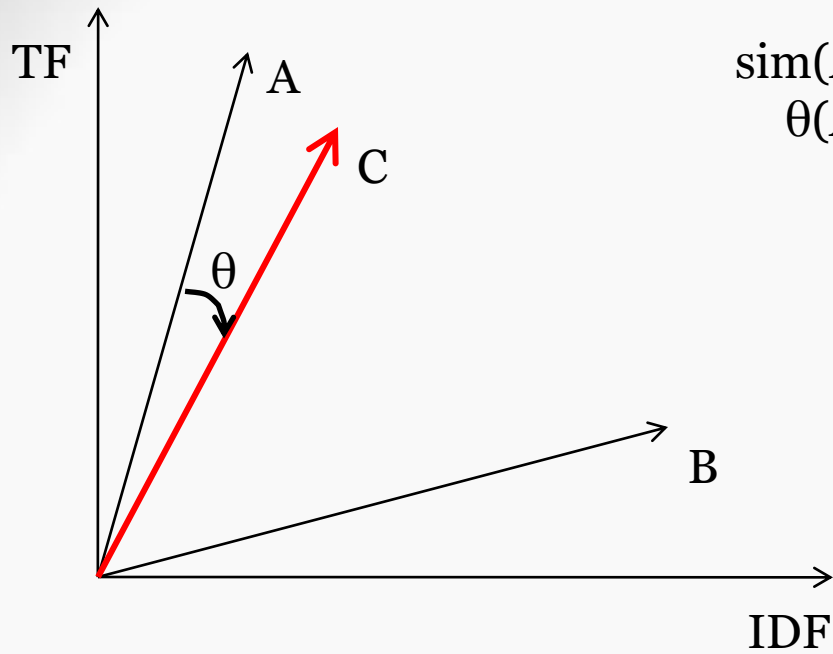
- Vocabulary convergence
- Increase of recall for tag queries
- Inter-user agreement
- Tag desambiguation
- Bootstrapping

Related Work

Collaborative Filtering

- Given the resources
 - A with tags {x,y}
 - B with tags {z,w}
 - And a new document C
- Suggest to resource C tag of the most similar documents (Jäschke et al., 2007)

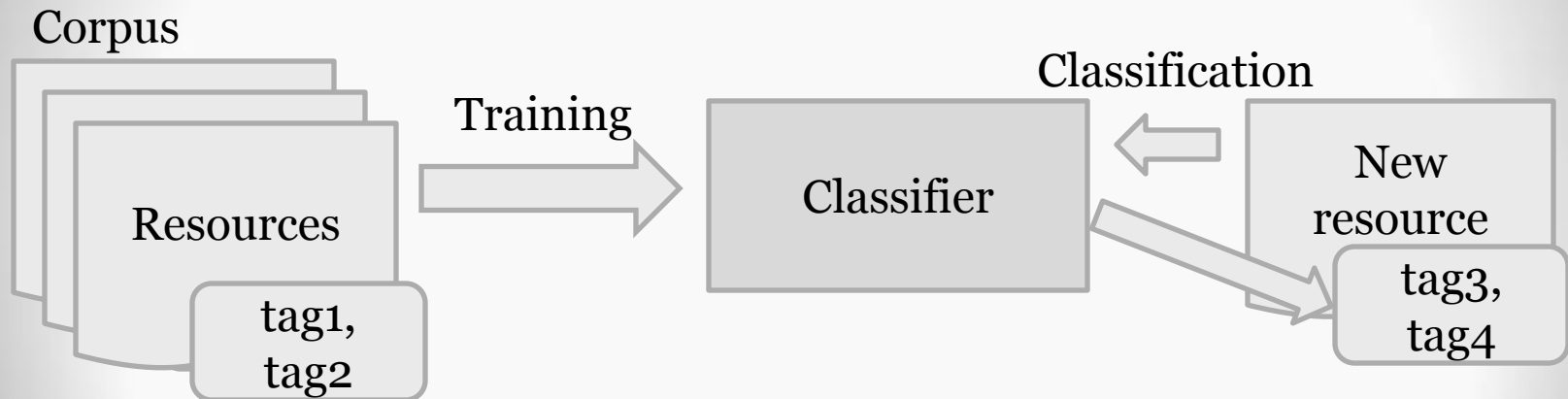
Example



$$\text{sim}(A,C) > \text{sim}(B,C)$$
$$\theta(A,C) > \theta(B,C)$$

Machine learning

- **Learn patterns** to “know” when a term is suitable to be a tag



Tag suggestion with ML

- Resources are represented as a **set of features**
- A classifier is built using patterns from these features
 - Ex: A term in the **title** might indicate a **good tag**
- Sparse gaussian (Song et al., 2008)
- Support Vector Machine (Wang and Davison, 2008)

ANTaReS

A Novel TAg REcommendation System

Goals

- Suggest tags to Web pages
 - Specially (novel) keywords **not present** in the folksonomy
 - Users might “discover” **new** interesting tags
- The method must be able to **adapt** to **vocabulary dynamics**

Sources of information

Web page features

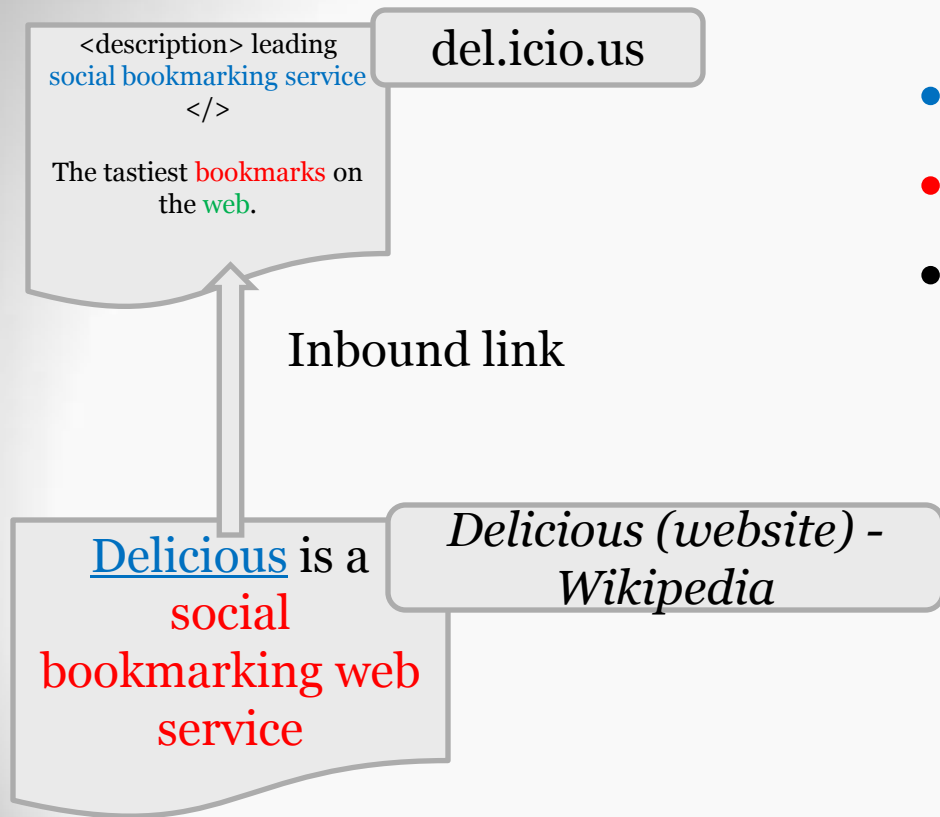
<description> leading social
bookmarking service </>

The tastiest bookmarks
on the web.

del.icio.us

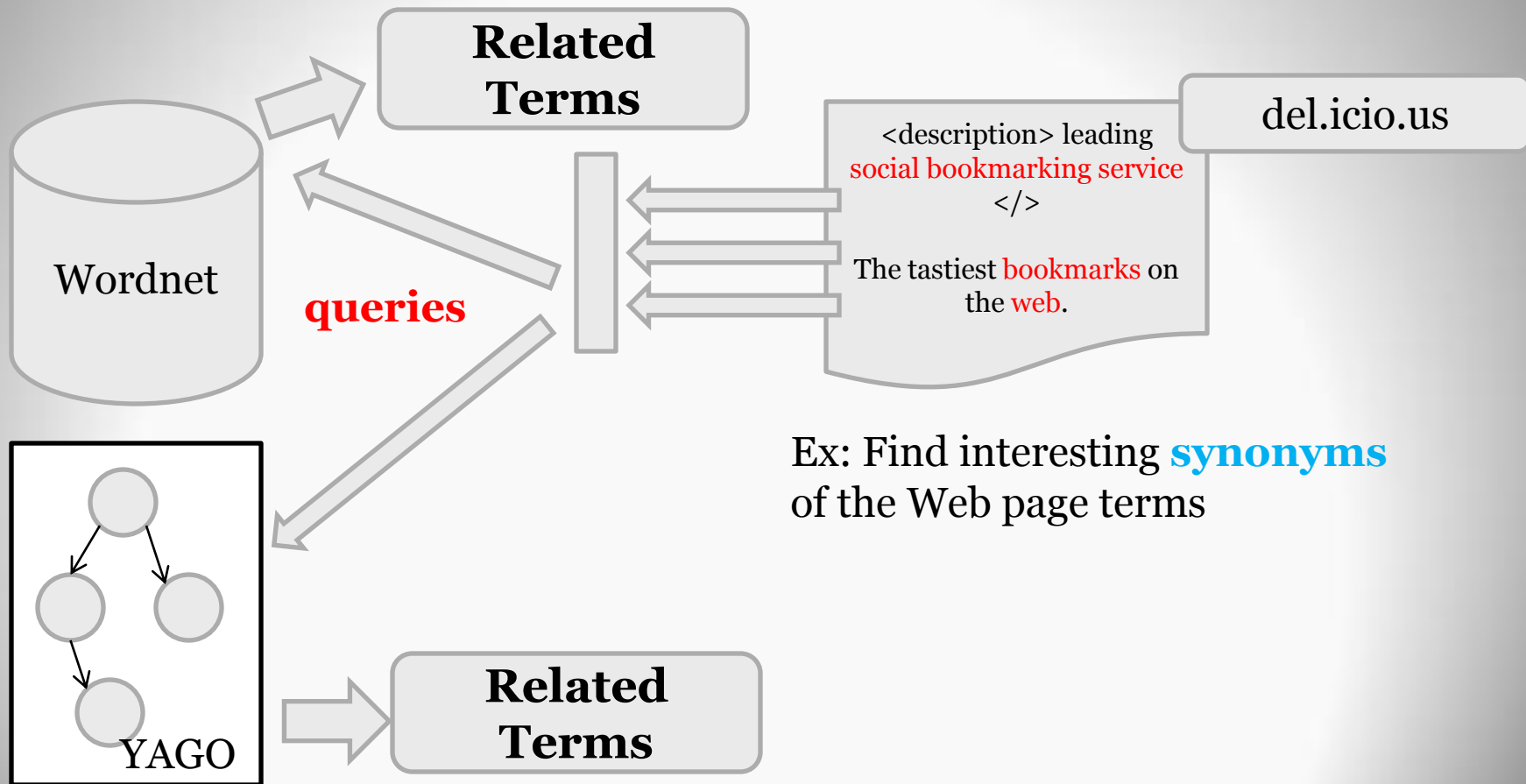
- HTML features (title, keywords, anchors)
- Information retrieval (TF, IDF)
- URL information
- Linguistic features

Inbound pages features



- Inbound anchor
- Inbound link context
- Inbound page text

Knowledge databases features



Ex: Find interesting **synonyms** of the Web page terms

Methodology

- A classifier (SVM) was built using these features to predict whether a term is a **tag** or **not**
- Analysis
 - Source x Tags
 - Comparison with Delicious tags
 - User's evaluation

Dataset extraction

- **Popular tags** in Delicious crawled (199)
- For each top tag, **popular pages using it** were retrieved (~ 7 pages/tag)
- For each page, the **top tags** used to describe it were extracted (~ 9.8 tags/page)
- Result: **1394** Web pages
 - 13667 tags (2201 **unique**)

Delicious comparison (early) results

- Using **only Web page features**:
 - Precision 83.65%
 - Recall 43.28%

Conclusion

Actual work

- The recall is low – features from other sources are needed
- But these features “confuse” the classifier if mixed with Web page ones
- Solution: different classifiers combined using methods like majority vote or product rule

Conclusion

- Tag suggestion using **a lot of** features from **multiple** sources of information
 - Suggest interesting and new **unforeseen** tags
- Early results show that our method can perform well
- **I hope I have a better conclusion when I finish the work**

References

- P. Heymann, D. Ramage, and H. Garcia-Molina. **Social tag prediction**. In Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval(SIGIR), pages 531-538. ACM, 2008.
- R. Jäschke, L. Marinho, A. Hotho, L. Schmidt-Thieme, and G. Stumme. **Tag recommendations in folksonomies**. In PKDD 2007: Proceedings of the European Conference on Principles and Practice of Knowledge Discovery in Databases, pages 506-514. Springer-Verlag, 2007.
- A. Mathes. **Folksonomies cooperative classification and communication through shared metadata**.
<http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.pdf>, 2004.

References II

- Y. Song, L. Zhang, and C. L. Giles. **A sparse gaussian processes classification framework for fast tag suggestions.** In Proceeding of the ACM conference on Information and Knowledge Management (CIKM), pages 93-102. ACM, 2008.
- J. Wang and B. D. Davison. **Explorations in tag suggestion and query expansion.** In SSM '08: Proceeding of the 2008 ACM Workshop on Search in Social Media, pages 43-50. ACM, 2008.



icaro.medeiros@gmail.com