ANTaReS – Tag suggestion using multiple sources of information

Ícaro Medeiros
Pável Calado
H. Sofia Pinto

16/11/2009

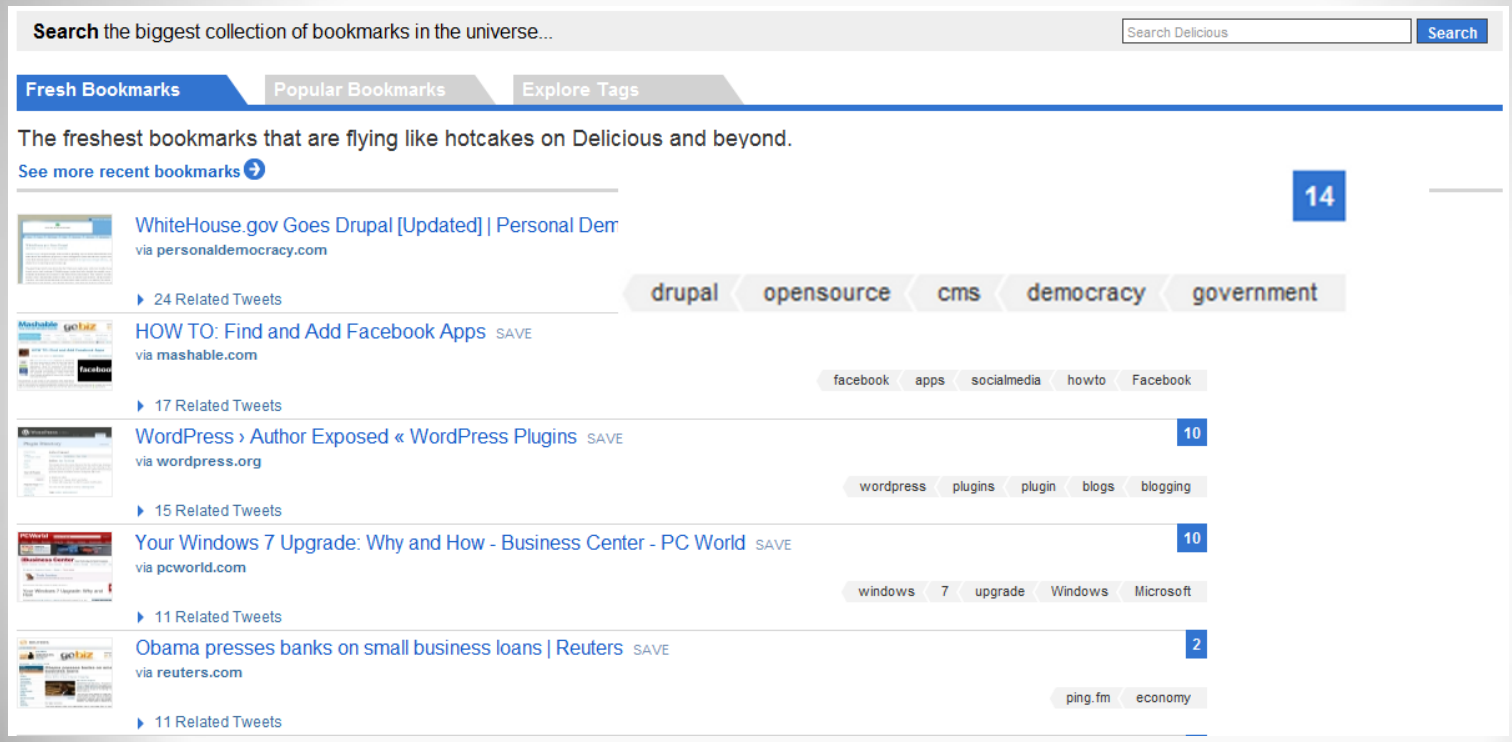
Outline

- 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



Delicious: 150,000 bookmarks/day

Youtube

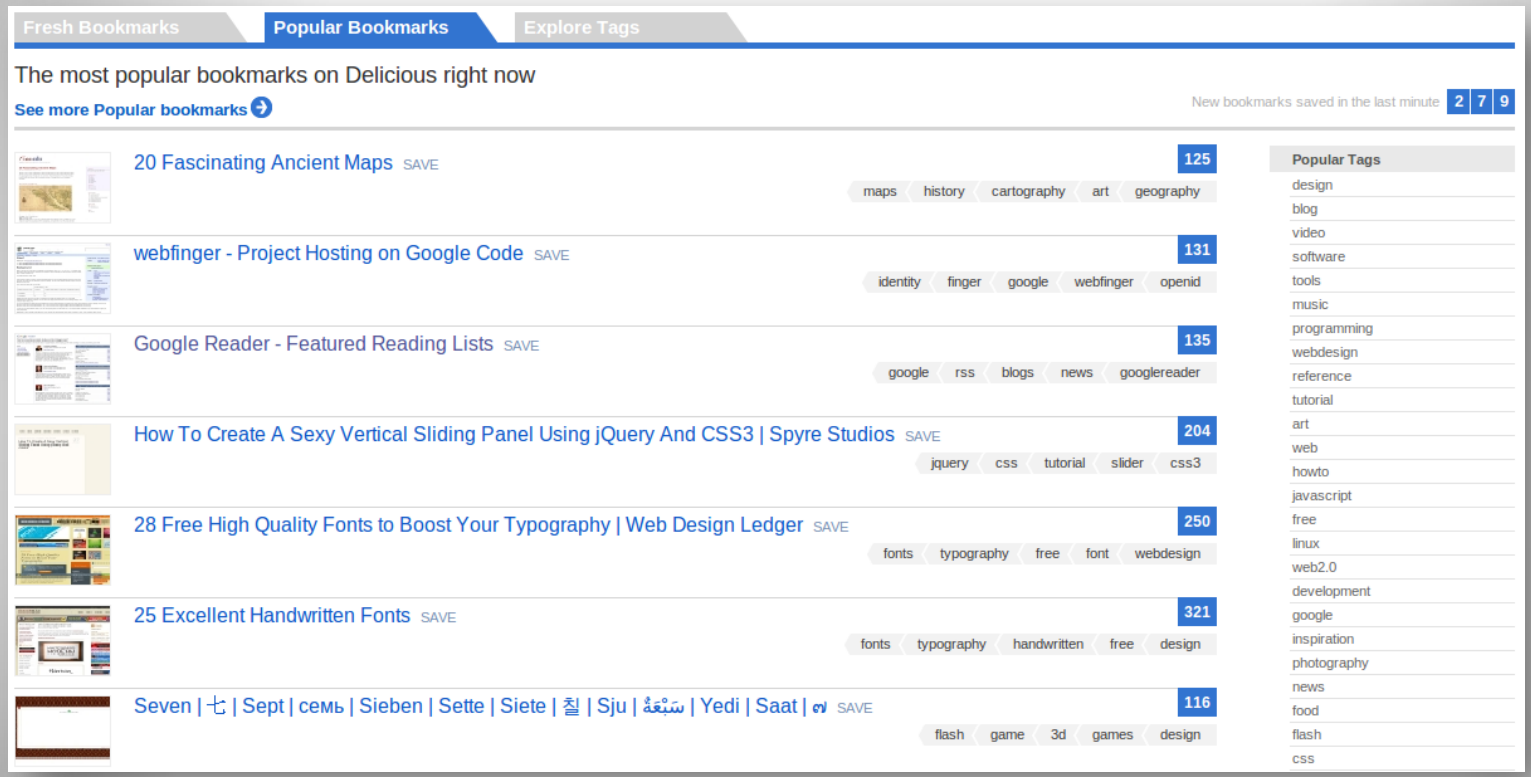
52 million visitors/month

Social Tagging Systems: features

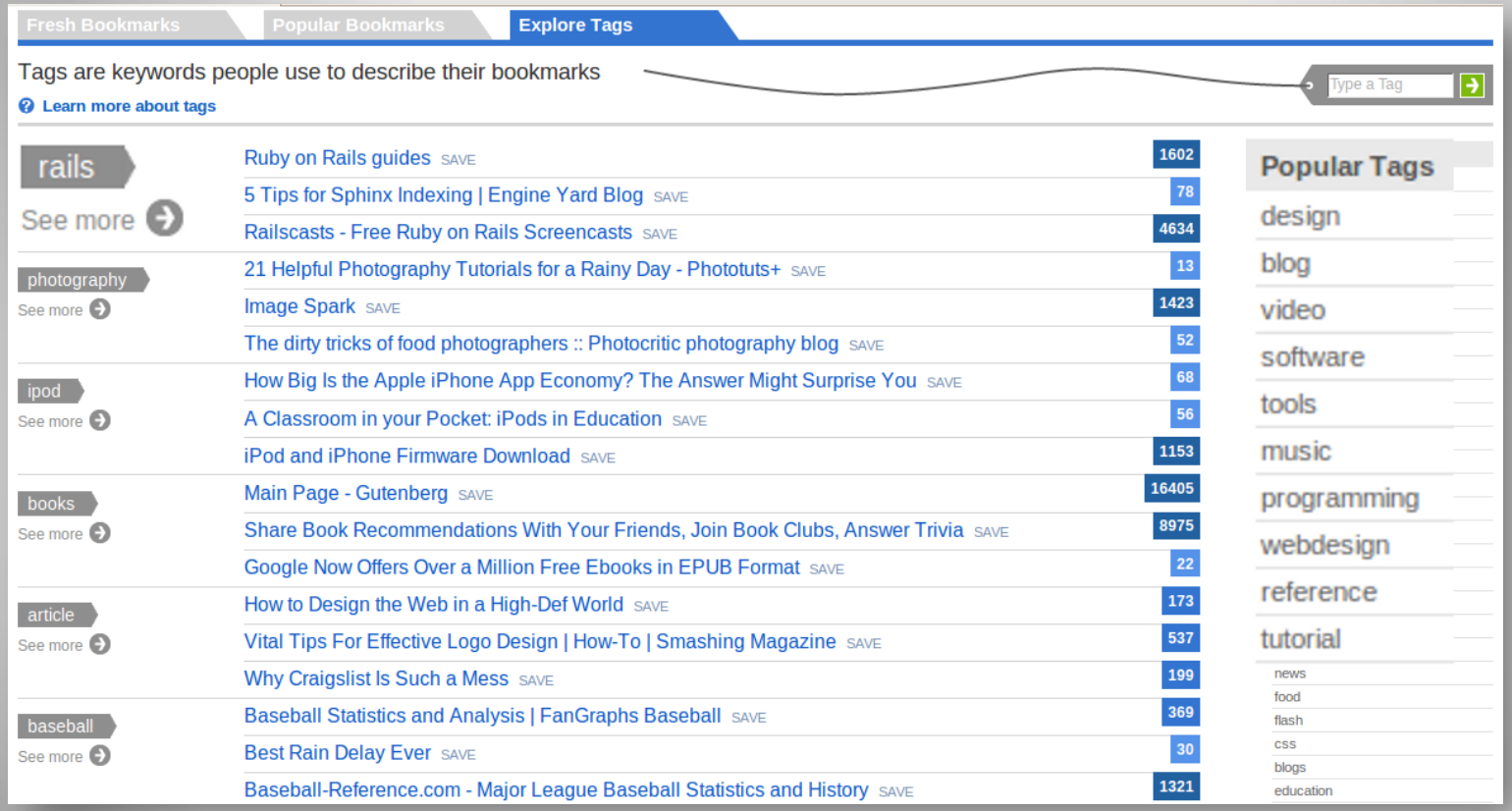
- Tags are freely chosen keywords
  - NO predefined vocabulary (Mathes, 2004)
  - Encourage tag assignment
- Tag assignment is collaborative
  - Generally these systems are in social networks and resource sharing systems
  - Content organization increases STS's popularity

Delicious: features

- Popular tags -> Trends



Tag browsing



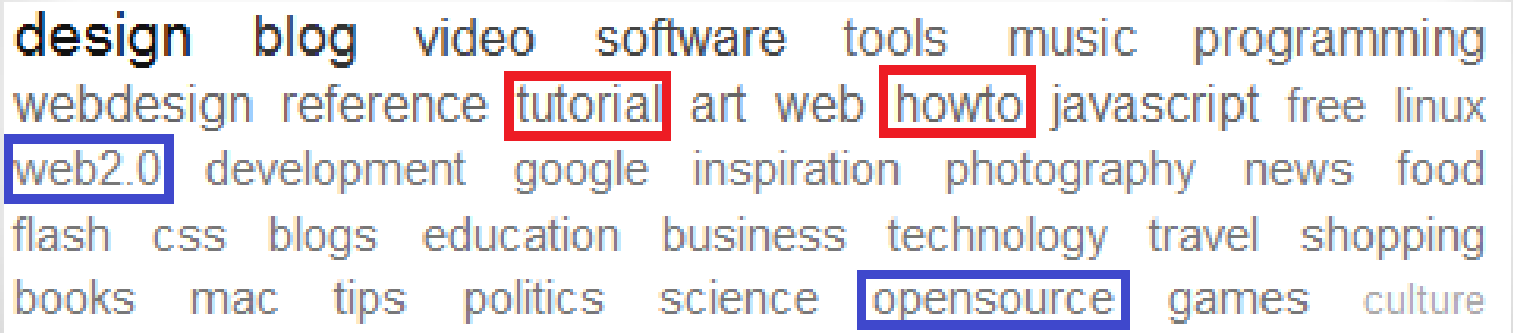
Folksonomies

- 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

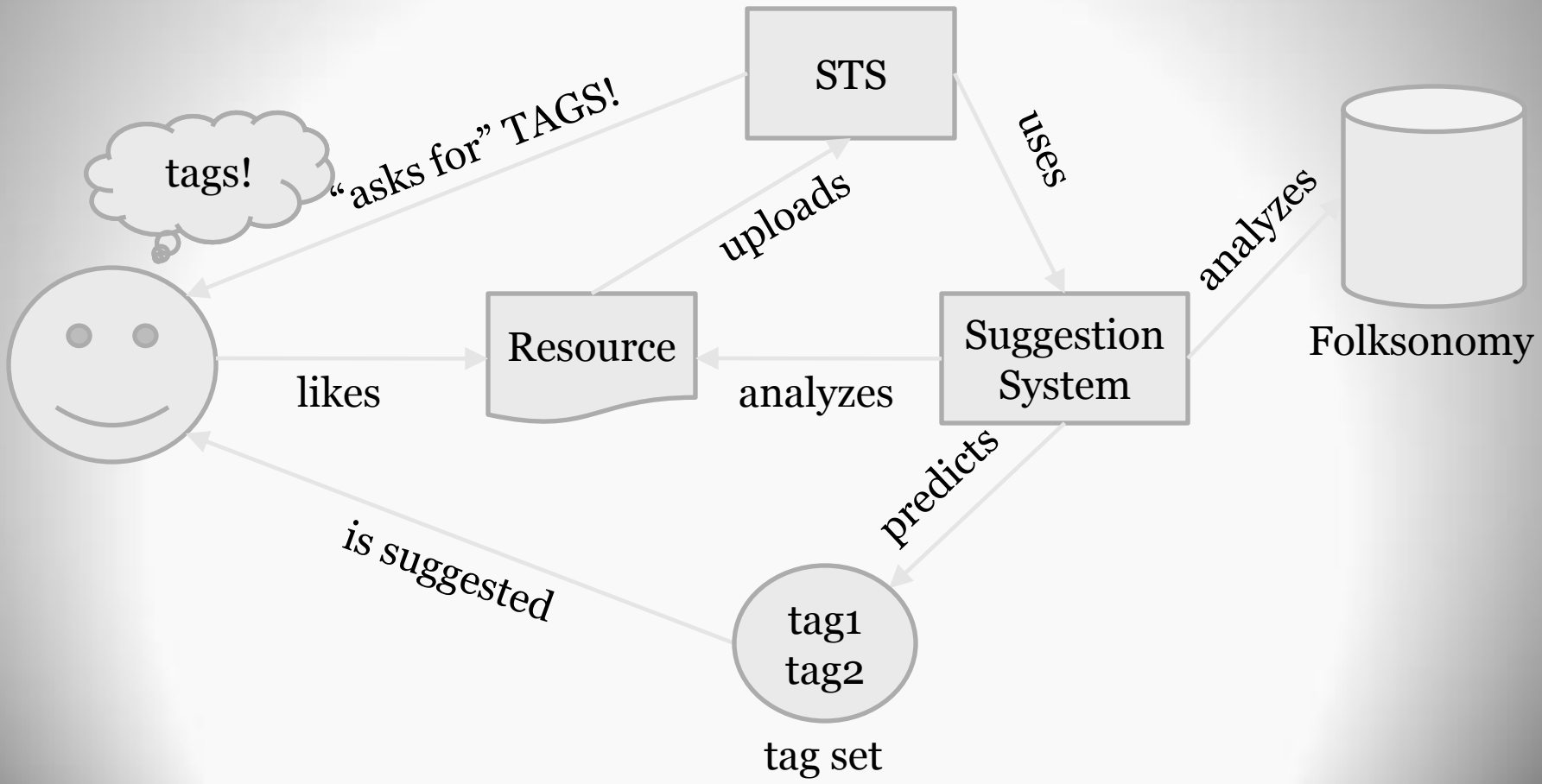
- Free vocabulary = ambiguity
  - Polysemy: *whistle* (device) e *whistle* (sound)
  - Homonym: Java (language) e Java (coffee)
  - Synonym: Buy/purchase
  - Binded Words: webdev, semweb, semantic_web



Popular Delicious tags 25-10-09

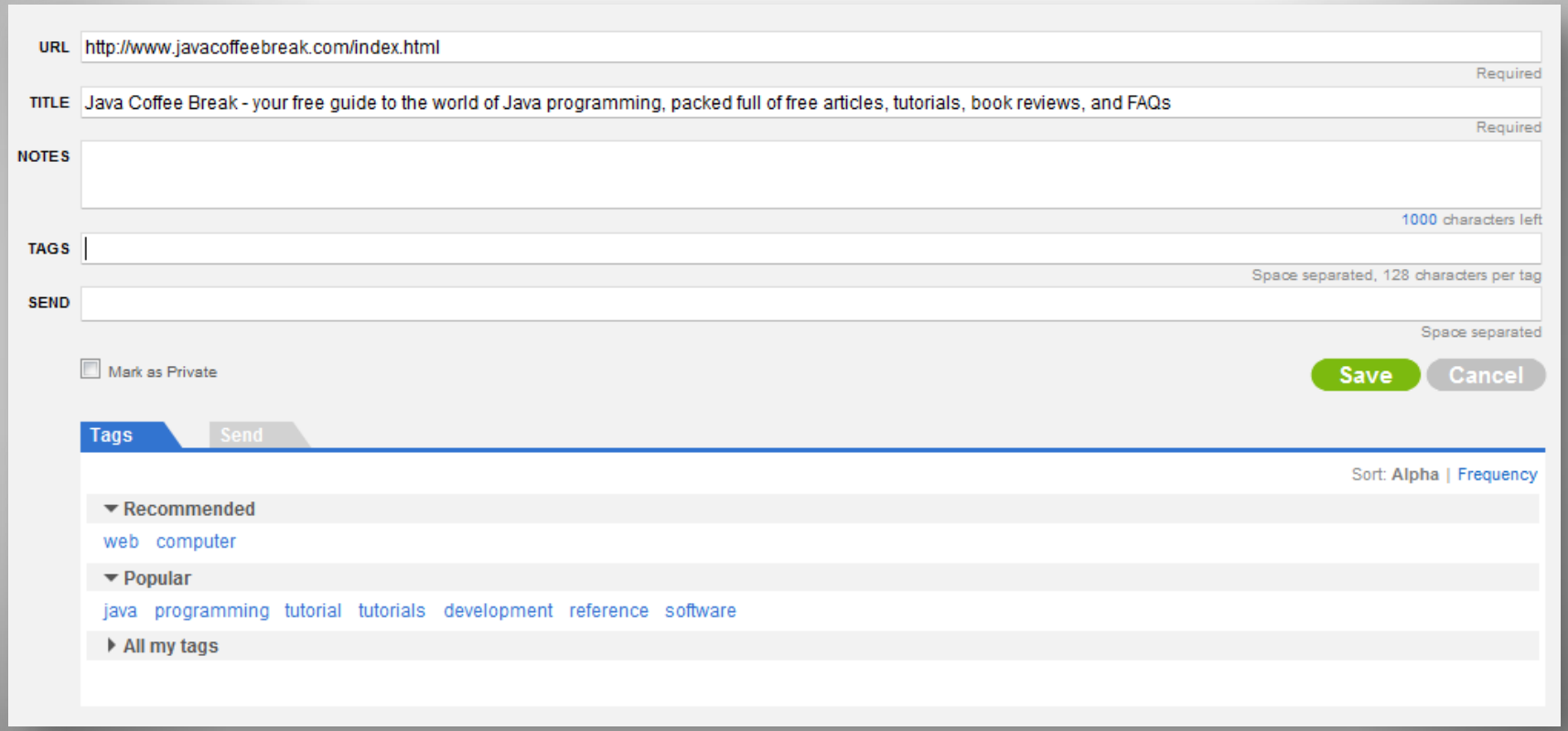
Incorrect tag assignment

Tag suggestion process



Suggestion on Delicious

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



Advantages of suggestion
(Heymann et al.,2008b)

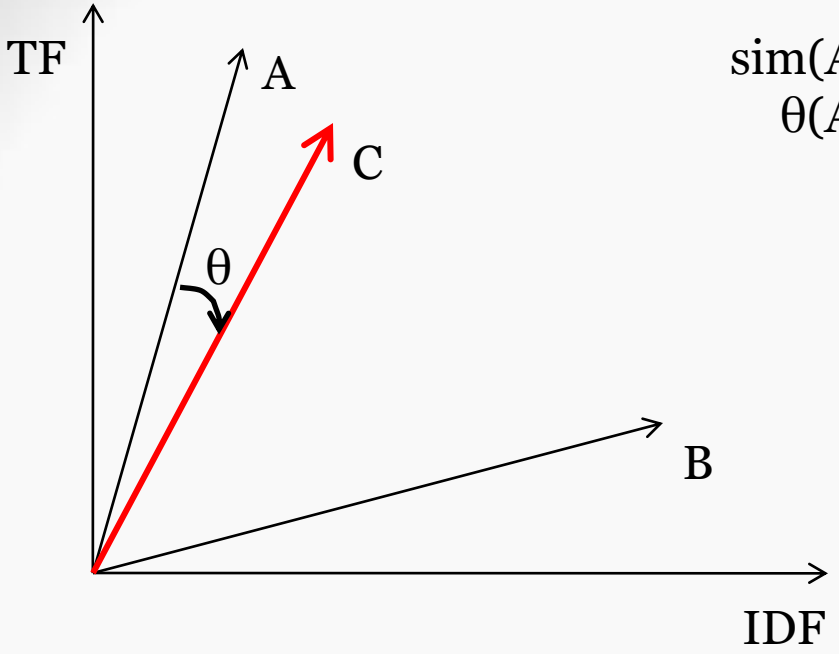
- Vocabulary convergence
- Increase of recall for tag queries
- Inter-user agreement
- Tag desambiguation
- Bootstrapping

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)

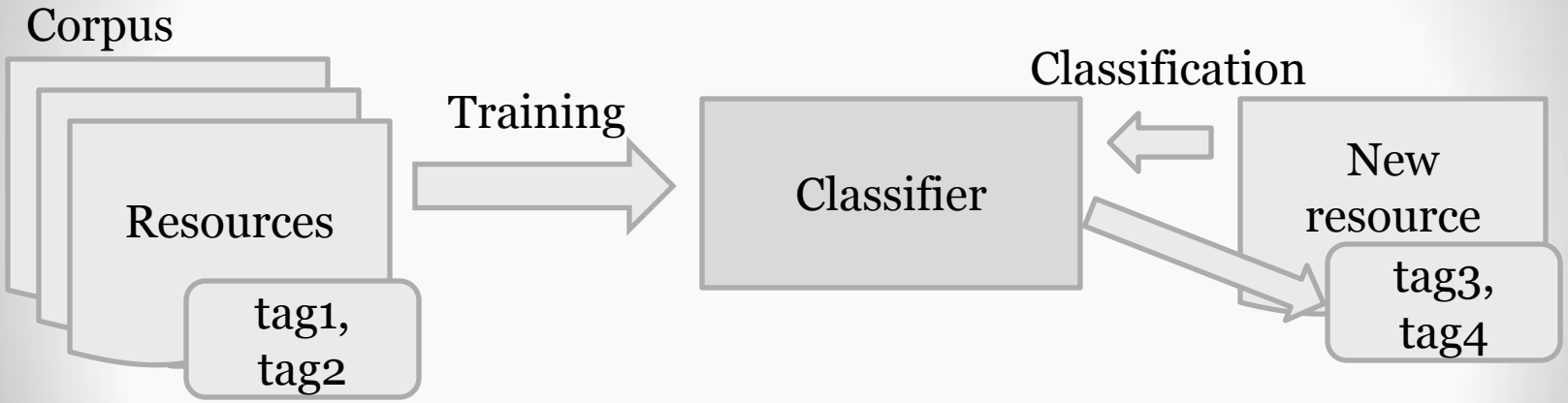
Example



sim(A,C) > sim(B,C)
θ(A,C) > θ(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

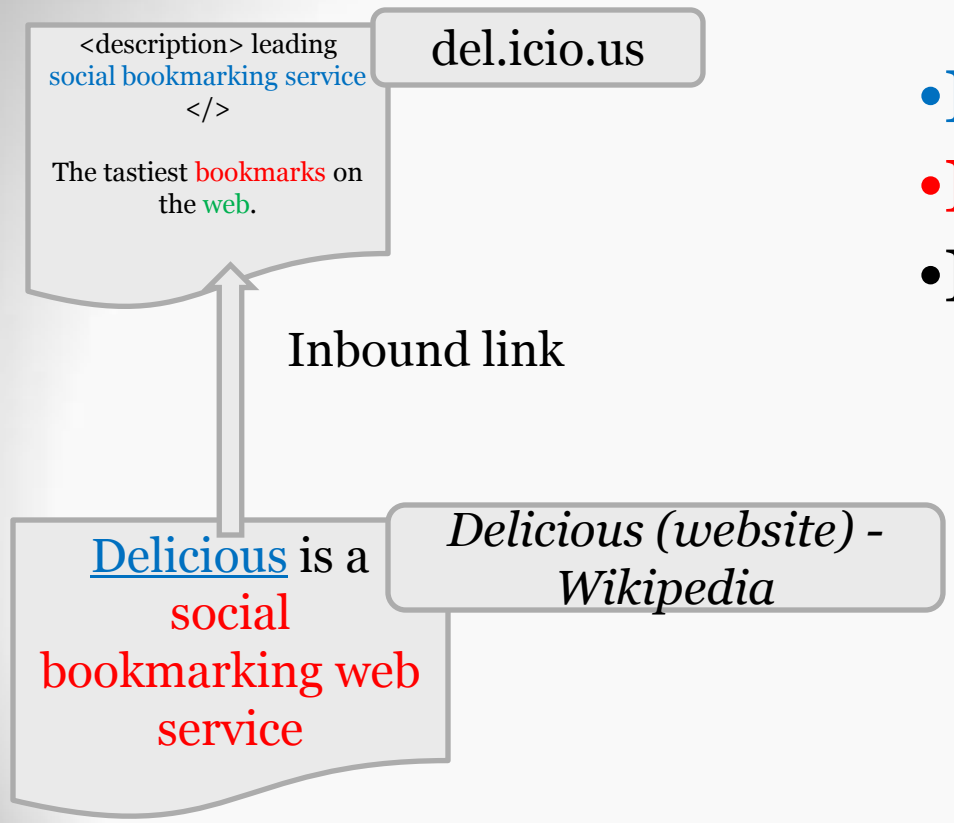
<description> leading social bookmarking service </>

The tastiest bookmarks on the web.

del.icio.us

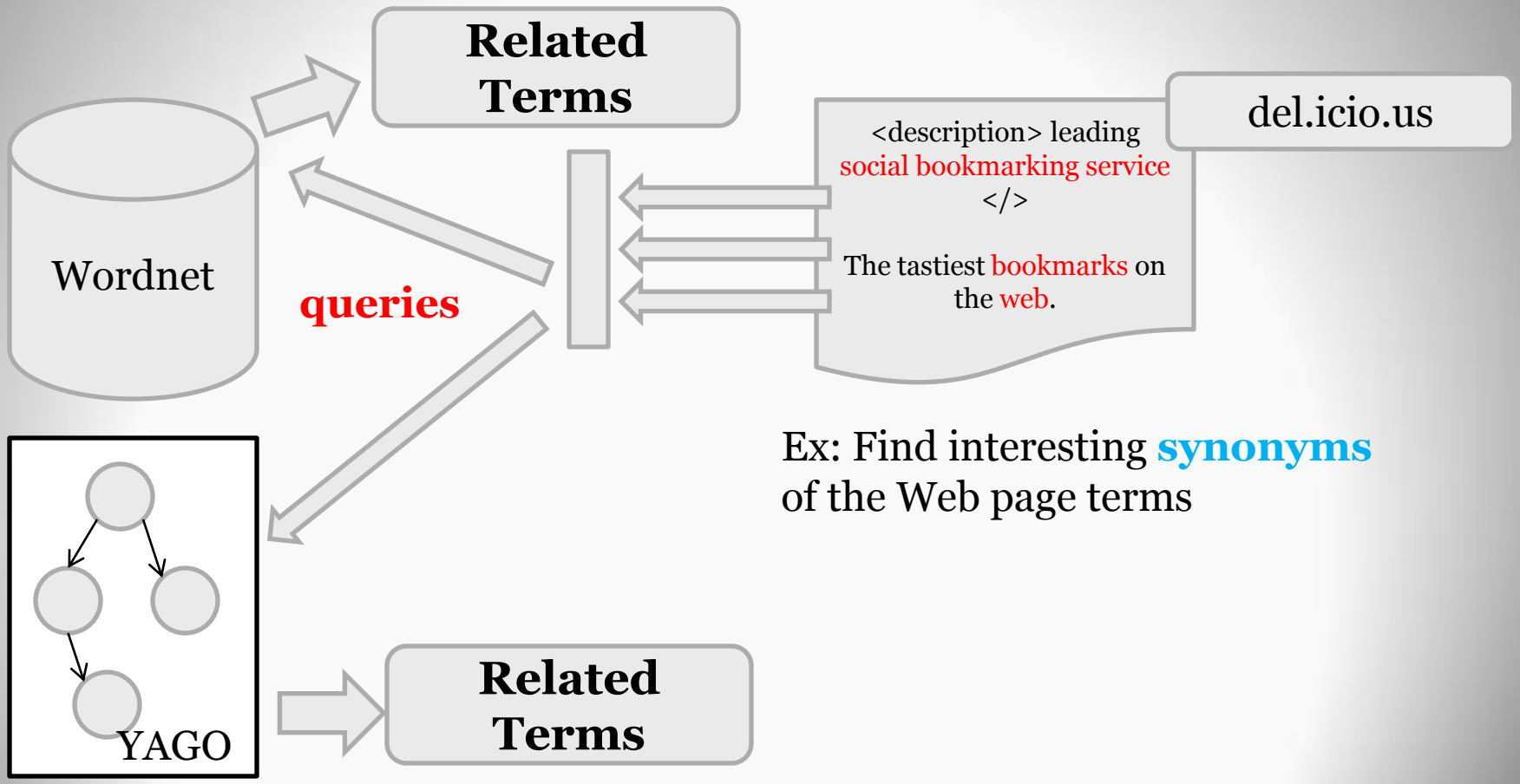
- HTML features (title, keywords, anchors)
- Information retrieval (TF,IDF)
- URL information
- Linguistic features

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 wheter a term is a **tag** ou **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 classication 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