



ECML PKDD Discovery Challenge 2009

Tag Recommendations in BibSonomy

*Folke Eisterlehner, Andreas Hotho, Robert Jäschke
Knowledge & Data Engineering Group, University of Kassel*

Sponsored by:





- Task 1 = mostly unknown data
 - Keep it simple
 - Gather tags from various sources (title, content, external services, different spellings)
 - Re-order tags for personalization
 - Clever selection and combination strategies win
- Task 2 = user, resource, tags known
 - Sophisticated methods are useable and useful
 - Factor models, classification, graph-based, CF, etc.
 - Simple methods work, too





Since Task 1 is the >90% case, use simple methods (they're quick, too!)

Online Tag Recommendations for Social Bookmarking Systems



BibSonomy :: edit bookmark
A blue social bookmark and publication sharing system.

logged in as jaeschke · settings · logout · help · blog · about ·  

4 picked in basket · edit tags

Home ▾ myBibSonomy post bookmark post publication tags authors relations ▾ groups popular

Feel free to edit your bookmark

general information

url*

title*

description, comment

tags* (space separated)

recommendation **linux software opensource gnu**

viewable for*

relevant for*

filter:

relations
(show all | hide all)

- ↓ author ← newman
- ↓ conference ← ecai ecmipkdd gvd iccs icdm icfca recsys
- ↓ folksonomy ← bookmarking tagging
- ↓ geo ← gps map utm
- ↓ howto ← manual reference tutorial
- ↓ location ← anhalt berlin bitterfeld bled celle dresden europe frankfurt hannover kassel london magdeburg ort sachsen sachsen_anhalt toulouse tübingen wittenberg
- ↓ ort ← location
- ↓ programming ← ada c fortran java lisp perl ruby
- ↓ protocol ← ftp http smtp
- ↓ researcher ← devadze shannon turing
- ↓ science ← chemistry math
- ↓ software ← apache beagle cocoa debian eclipse firefox haystack nextstep photoshop protege thunderbird wine word x11

tags
(alpha | freq) (cloud | list) (minfreq 1 | 2 | 5)

- Provide tag recommendations for the user in a live setting
 - Respect time constraints (!)
- ➔ Try it out yourself: <http://www.bibsonomy.org>



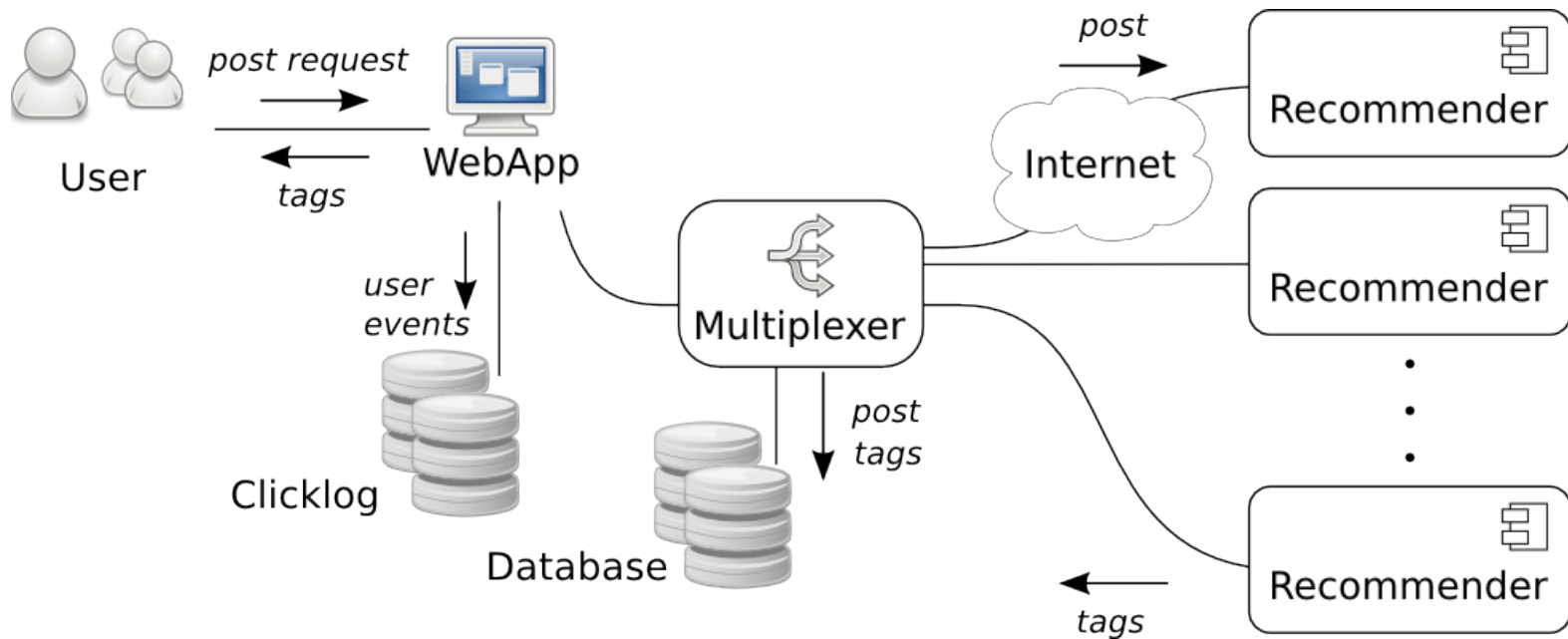
- Participants:
 - 13 Different Recommender Systems
 - 10 Participants
 - 7 Countries



- BibSonomy:
 - 1500 active users
 - >28.000 posting processes, >2000 per recommender system
 - For evaluation only public non-spam posts are considered
 - ~410 relevant posting processes per recommender system

Website: <http://www.kde.cs.uni-kassel.de/ws/dc09/>

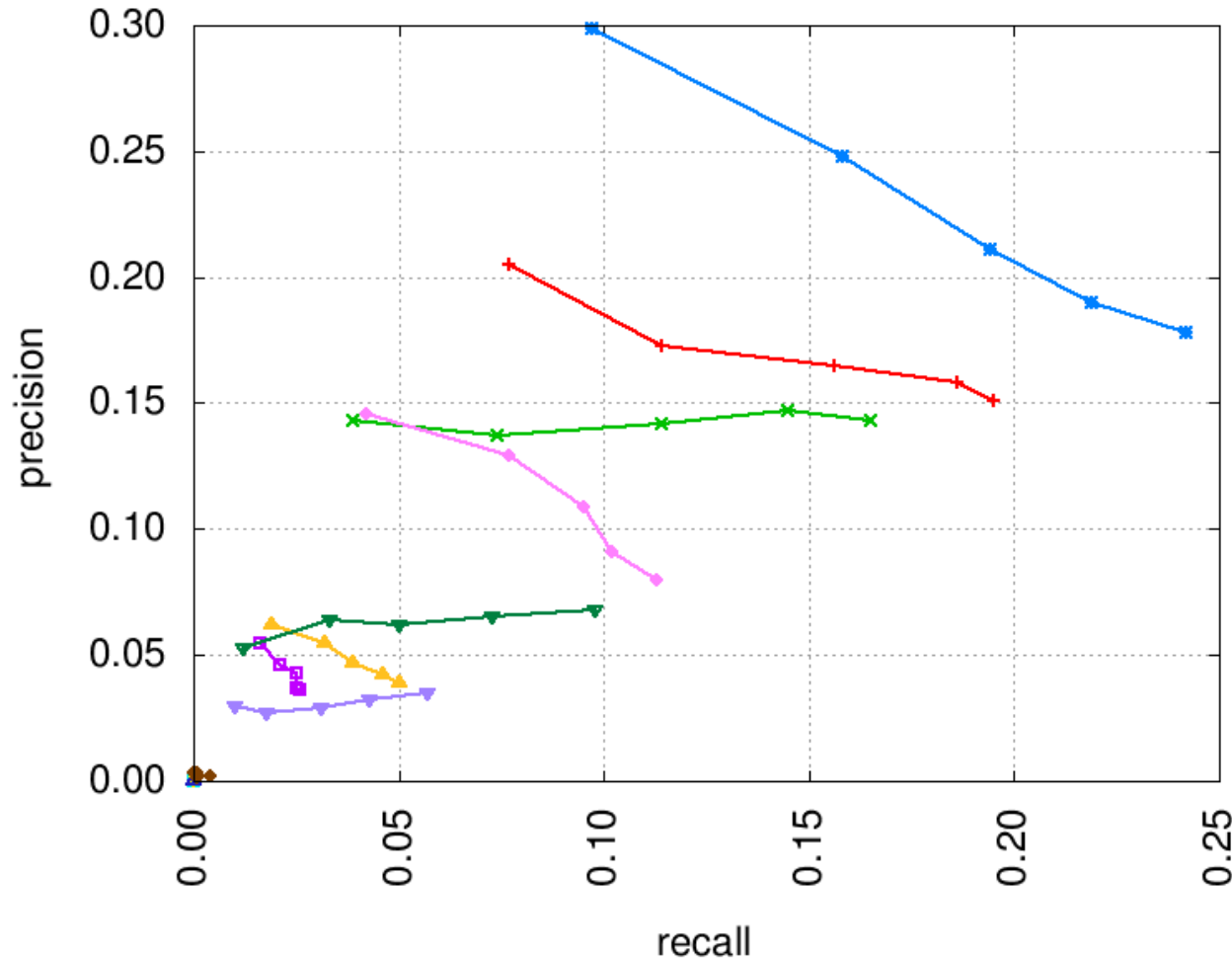
Tag Recommendation Framework



- Each *post request* is sent to every recommender system
- One of which is *randomly selected* (without replacement)
- Only responses within a *1000 ms timeout* are considered
- Recommended tags are presented to the user



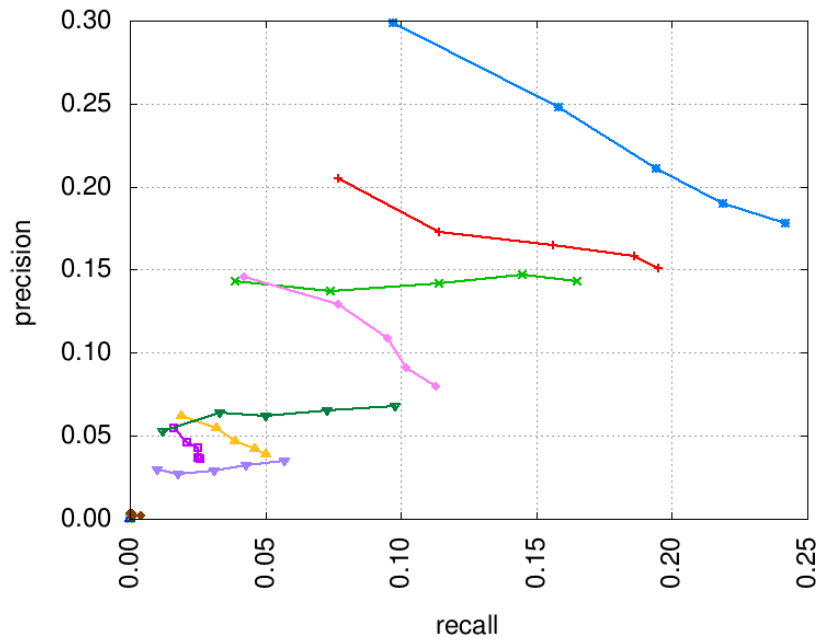
- **Precision:** „How many recommended tags where adequate“?
- **Recall:** „How many of the true tags where recommended?“



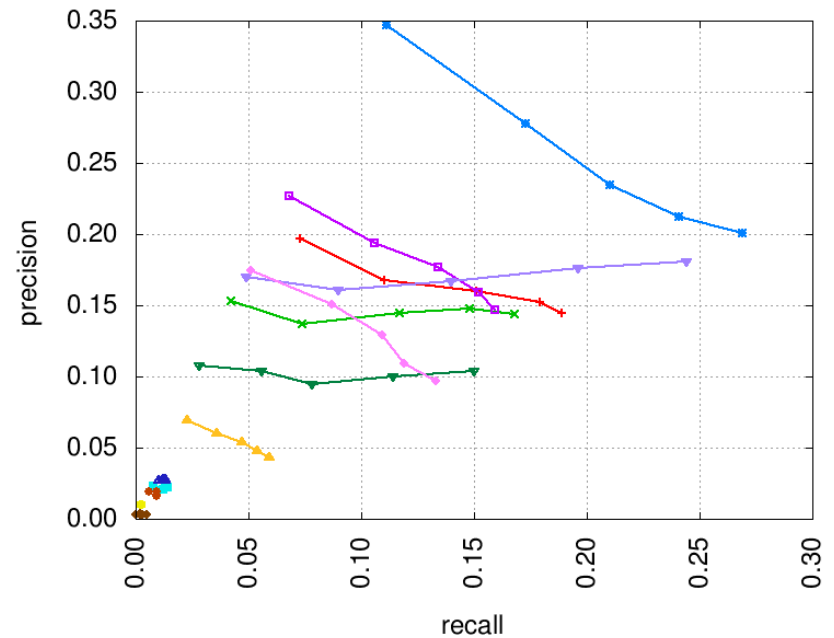
Some Results



- **Precision:** „How many recommended tags where adequate“?
- **Recall:** „How many of the true tags where recommended?“



With timeout constraints

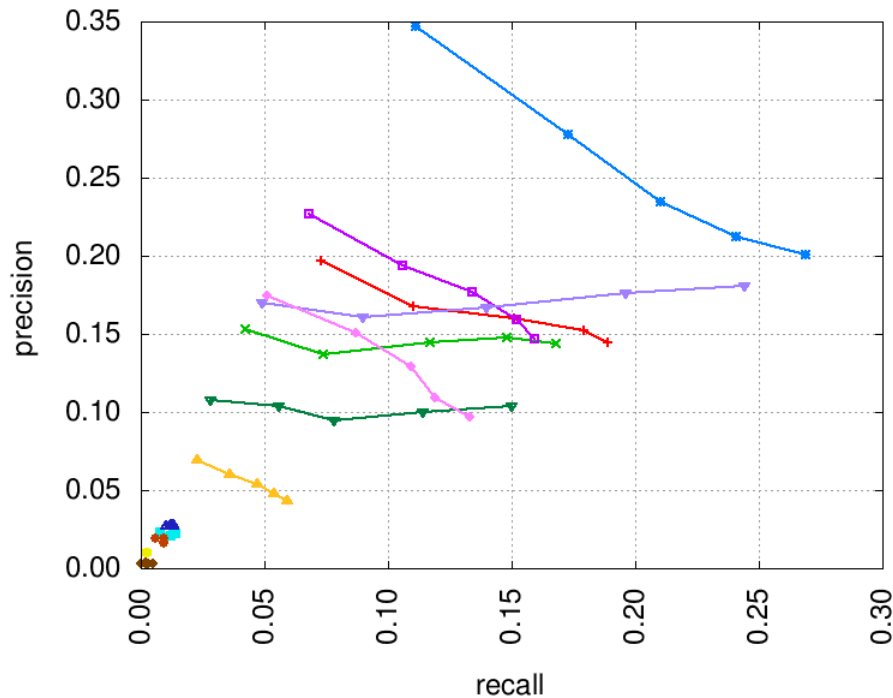


Without timeout constraints

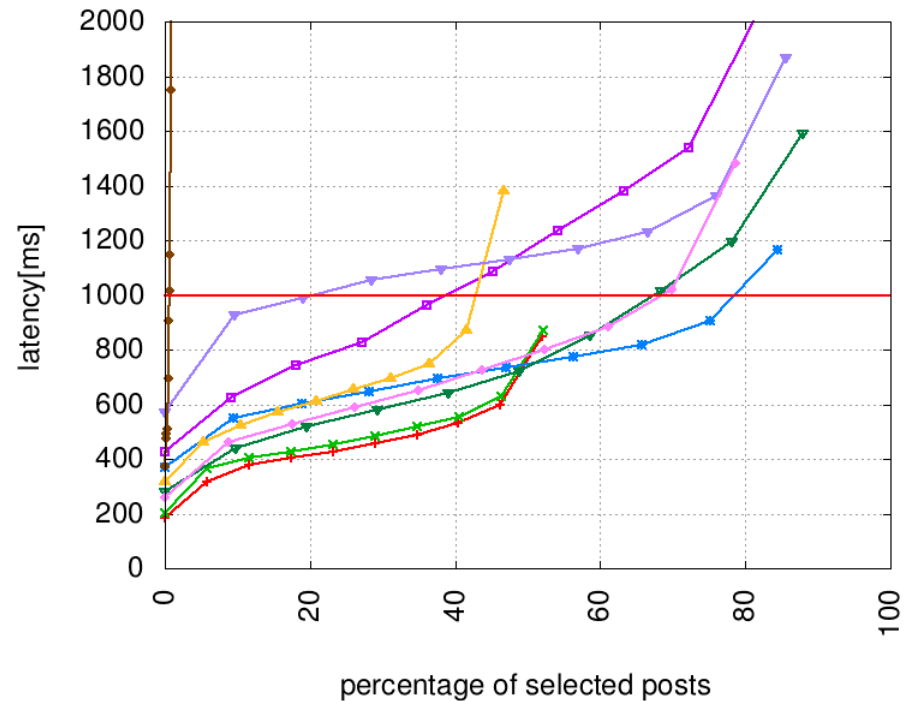
Some Results



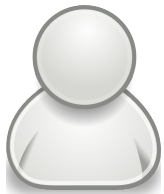
- **Precision:** „How many recommended tags where adequate“?
- **Recall:** „How many of the true tags where recommended“?



P/R with timeout constraints



Latency, ordered ascending



(will be announced this evening)

Precision	Recall	F1-Measure
0.178	0.242	0.205
0.151	0.195	0.171
0.143	0.165	0.154



CONGRATULATIONS !!!