



ECML PKDD Discovery Challenge 2009

Tag Recommendations for Social Bookmarking Systems

Folke Eisterlehner, Andreas Hotho, Robert Jäschke

*Knowledge & Data Engineering Group
University of Kassel*

sponsored by



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

viewable for*

relevant for

filter:

relations
(show all | hide all)

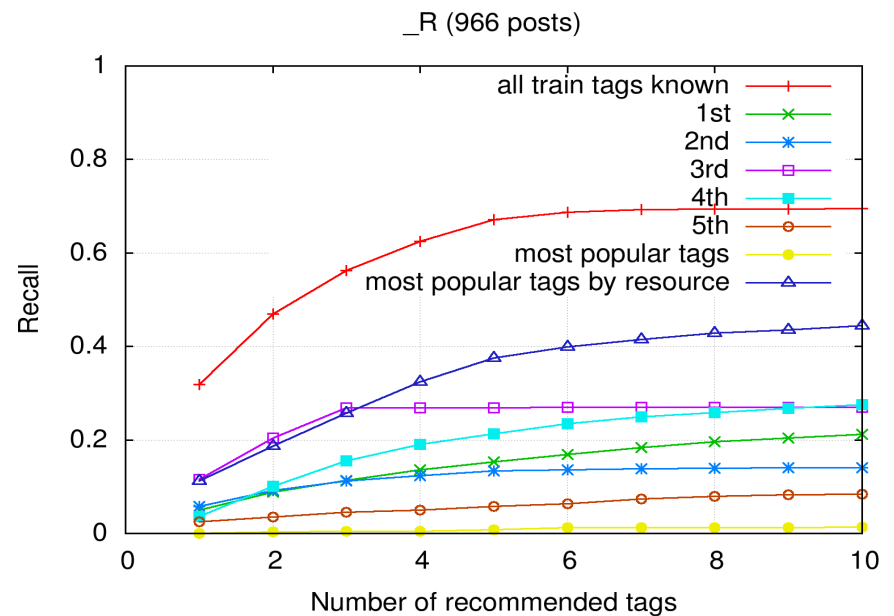
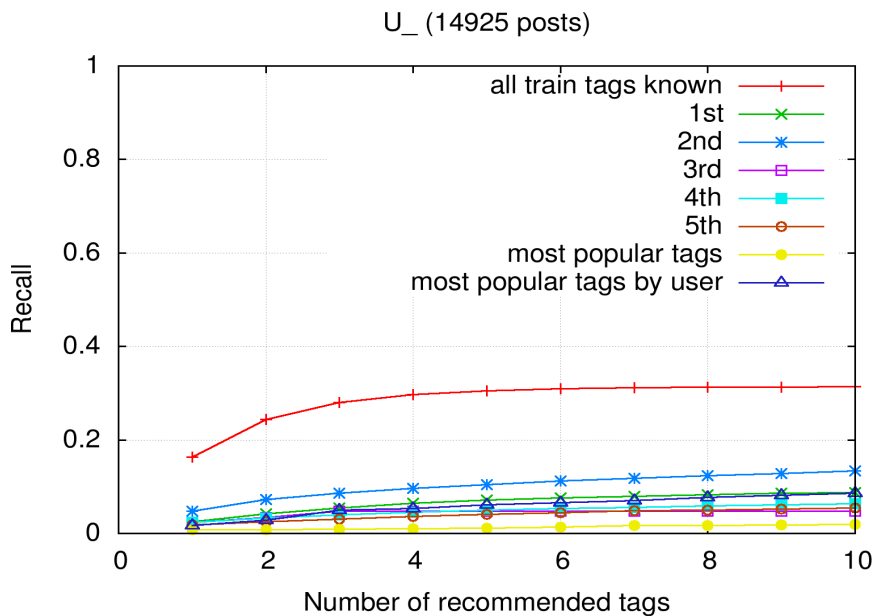
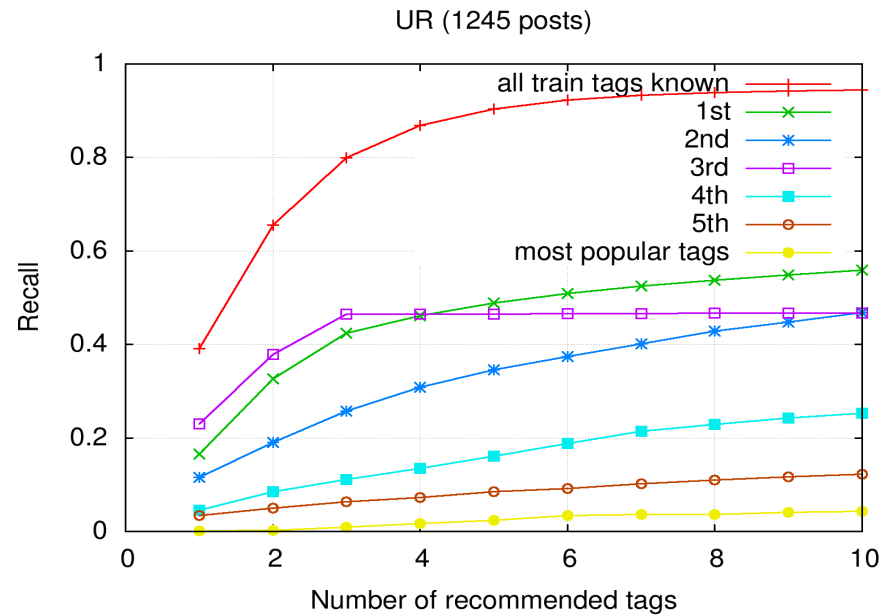
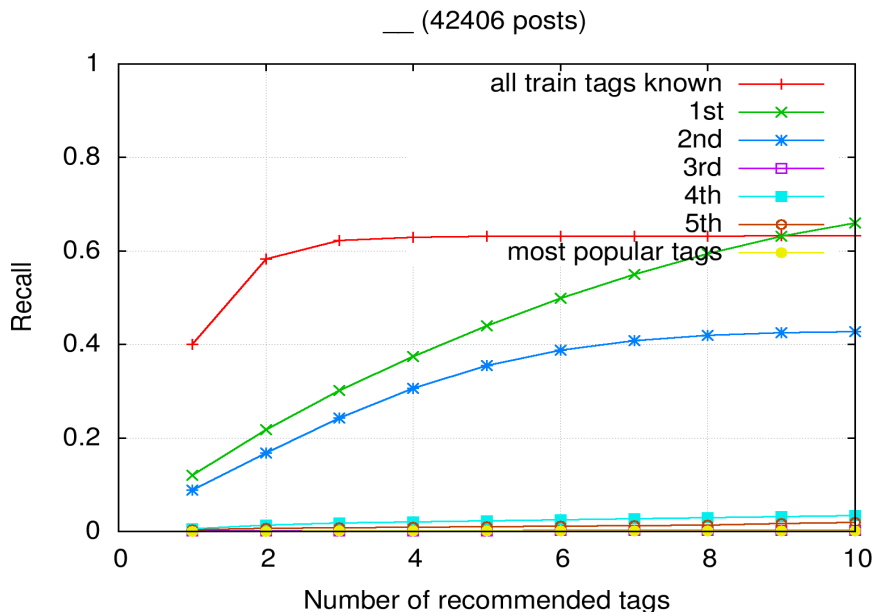
- ↓ author ← newman
- ↓ conference ← ecai ecmipkdd gvd iccs icdm icfca recsys
- ↓ folksonomy ← bookmarking tagging
- ↓ geo ← gps map utm
- ↓ howto ← manual reference tutorial
- ↓ location ← anhalt berlin bitterfeld bled celle dresden europe frankfurt hannover kassel london magdeburg ort sachsen sachsen_anhalt 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 weka wine word x11 zope
- ↓ sport ← basketball football volleyball

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

tags: bibsonomy clustering community detection diploma tagging ubuntu windows x11 zope

- Support users during the tagging process
- Recommend tags on the posting page
- <http://www.kde.cs.uni-kassel.de/ws/dc09/>

Last Year's Tag Recommendation Results





- Given a user and a resource, recommend up to five tags.
- Task 1: *Content-Based Tag Recommendations*
 - User, resource, or tags might be unknown
- Task 2: *Graph-Based Recommendations*
 - User, resource, and tags contained in training post-core
- Task 3: *Online Tag Recommendations*
 - Deliver live recommendations respecting timeouts, etc.
- 48h time for Tasks 1 & 2, Task 3 running for five weeks



Social bookmarking data from BibSonomy

<http://www.bibsonomy.org/>

- Training data
 - Released on March 25th, 2009
 - Posts from BibSonomy up to 31st Dec. 2009
 - Complete snapshot: 3,617 users, 93,756 tags, 378,378 resources
 - Post-core at level 2: 1,185 users, 13,276 tags, 22,389 resources
- Test data
 - Released on July 6th, 2009
 - 6 months snapshot (2009-01-01 - 2009-06-30)
 - Task 1: 1,591 users, 34,045 tags, 40,729 resources, 43,002 posts
 - Task 2: 136 users, 862 tags, 667 resources, 778 posts



- Preprocessing
 - Removed spam, posts from user *dblp*
 - Tag cleansing: only letters+numbers, common system tags removed
- Identity of resources
 - Intrahash + user name uniquely identifies a *post*
 - *Overlap* between resources by interhash
- Post-core at level 2
 - Each user, tag, resource appears in at least two posts
 - Iterative pruning process
 - Interhash to identify resources



- Test data = posts without tags
- Submitted data = recommended tags for the posts
- First 5 recommended tags compared against true tags from user (again, cleaning tags - only letters+numbers, no system tags)

- Precision and recall per post:

$$precision = \frac{|recommended \cap true|}{|recommended|} \quad recall = \frac{|recommended \cap true|}{|true|}$$

- Averaged over all posts

- Final criterion: F1-Measure = $\frac{2 \cdot precision \cdot recall}{precision + recall}$

Participants & Submissions

- 150 registered mailing list users (= access to training data)
- 21 result submissions for each of the Tasks 1 & 2
- 27 paper submissions - 24 accepted

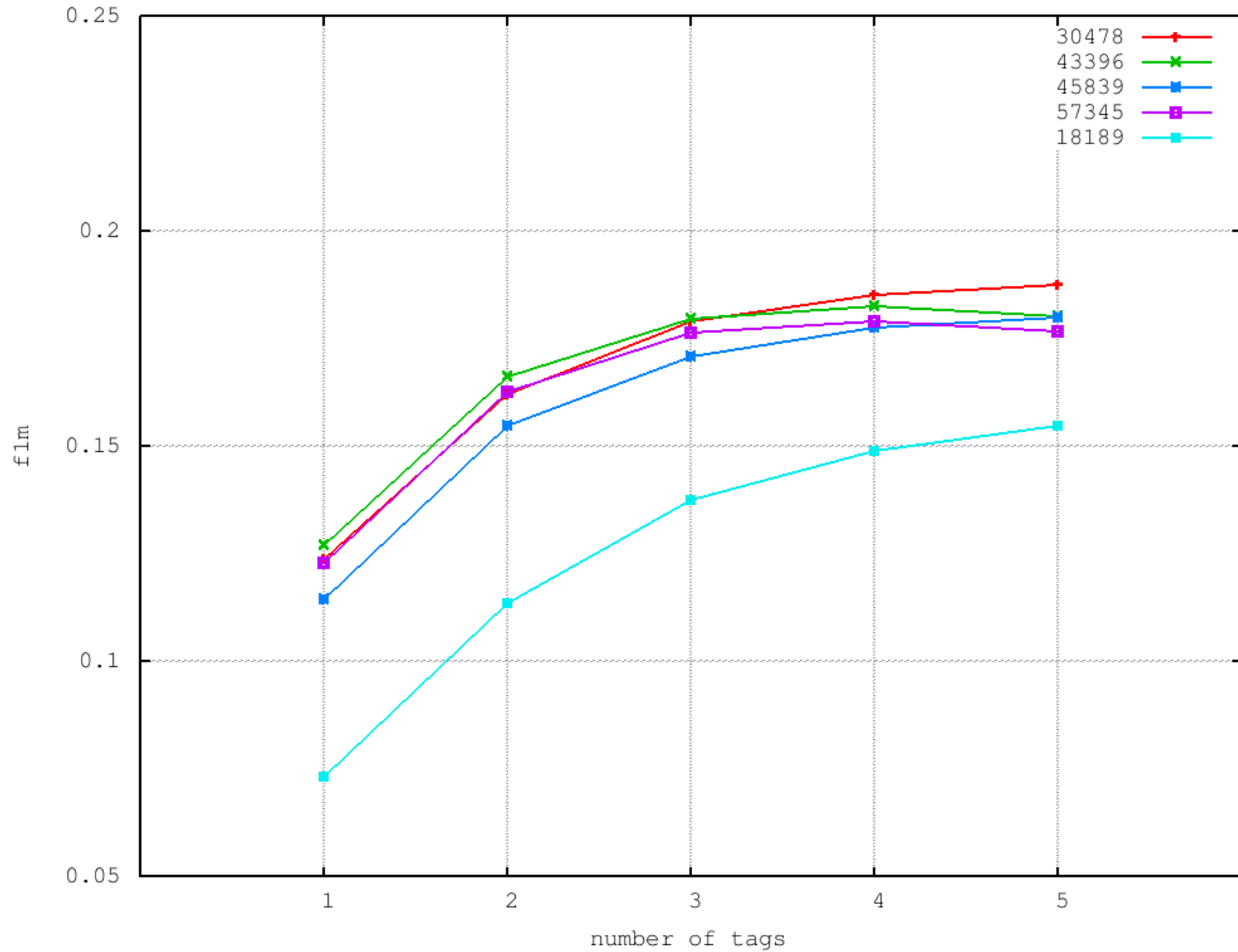


Results Task 1



submission	f1m	team	
1	30478	0.18740	Tag Sources for Recommendation in Collaborative Tagging Systems Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios
2	43396	0.18001	Content- and Graph-based Tag Recommendation: Two Variations Johannes Mrosek, Stefan Bussmann, Hendrik Albers, Kai Posdziech, Benedikt Hengefeld, Nils Opperman, Stefan Robert, and Gerrit Spira
3	45839	0.17975	A Weighting Scheme for Tag Recommendation in Social Bookmarking Systems Sanghun Ju and Kyu-Baek Hwang
4	57345	0.17661	
5	18189	0.15464	
6	67797	0.15412	
7	13482	0.14398	
8	28068	0.14199	
9	76565	0.14151	
10	20418	0.14018	

Results Task 1

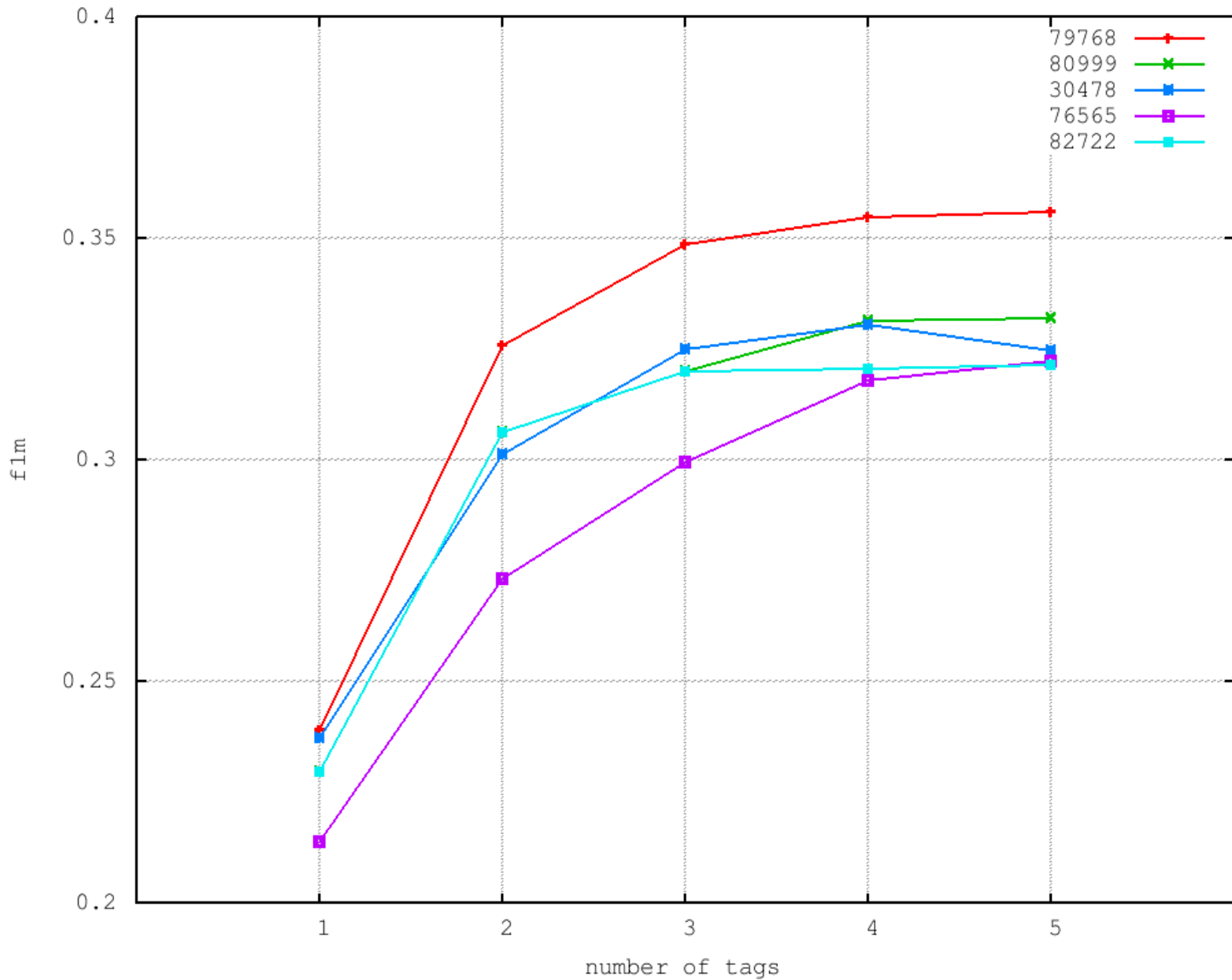


Results Task 2



submission	f1m	team	
1	79768	0.35594	Factor Models for Tag Recommendation in BibSonomy Steffen Rendle and Lars Schmidt-Thieme
2	80999	0.33185	Relational Classification for Personalized Tag Recommendation Leandro Balby Marinho, Christine Preisach, and Lars Schmidt-Thieme
3	30478	0.32461	Tag Sources for Recommendation in Collaborative Tagging Systems Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios
4	76565	0.32230	A Collaborative Filtering Tag Recommendation System based on Graph Yuan Zhang, Ning Zhang, and Jie Tang
5	82722	0.32134	
6	61431	0.32124	
7	45839	0.32039	
8	56390	0.31396	
9	16379	0.31368	
10	11472	0.30751	

Results Task 2





- 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!)



- Conference organizers, in particular Tina Anzic
- Our sponsors  
NOKIA
Connecting People
- YOU - for your vivid participance



9:00 - 10:30

- Welcome & Introduction
- **Tag Sources for Recommendation in Collaborative Tagging Systems** - Marek Lipczak, Yeming Hu, Yael Kollet, and Evangelos Milios
- **Factor Models for Tag Recommendation in BibSonomy** - Steffen Rendle and Lars Schmidt-Thieme

10:30 - 11:00

- Coffee Break (+poster preparation)



11:00 - 12:30

- **Content- and Graph-based Tag Recommendation: Two Variations** - Johannes Mrosek, Stefan Bussmann, Hendrik Albers, Kai Posdziech, Benedikt Hengefeld, Nils Opperman, Stefan Robert, and Gerrit Spira PDF
- **Relational Classification for Personalized Tag Recommendation** - Leandro Balby Marinho, Christine Preisach, and Lars Schmidt-Thieme PDF
- **A Weighting Scheme for Tag Recommendation in Social Bookmarking Systems** - Sanghun Ju and Kyu-Baek Hwang PDF

12:30 - 14:00

- Lunch



14:00 - 15:20

- **A Collaborative Filtering Tag Recommendation System based on Graph** - Yuan Zhang, Ning Zhang, and Jie Tang PDF
- **Minute Madness:** Every poster presenter has one minute to convince everybody to look at his/her poster.
- Closing Session

15:20 - 15:40

- Coffee Break (+ poster preparation)

15:40 - 17:00

- Poster Session

17:20 - 18:00

- Conference Opening & Awards