Learning Layers

Scaling up Technologies for Informal Learning in SME Clusters

The Influence of Frequency, Recency and Semantic Context on the Reuse of Tags in Social Tagging Systems
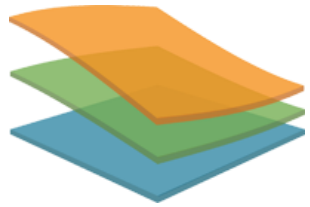
Hypertext 2016

Halifax, Nova Scotia, Canada

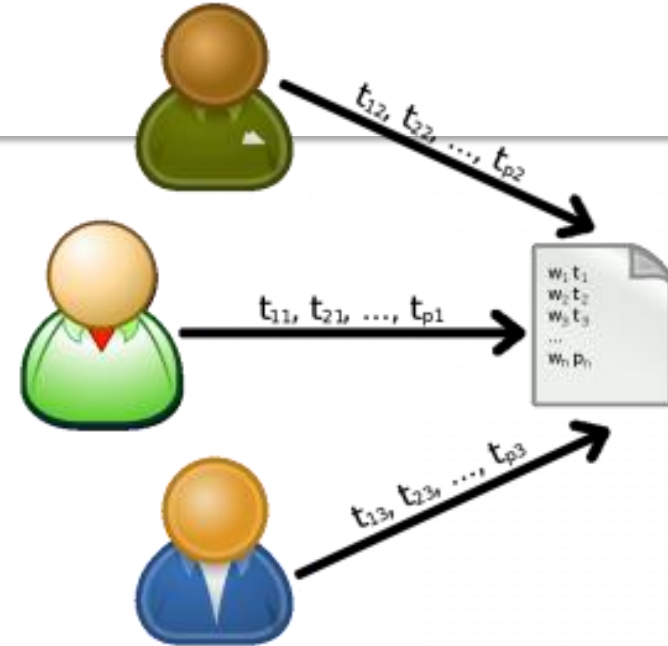**Dominik Kowald**, Elisabeth Lex

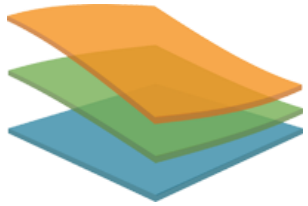Know-Center, Graz University of Technology (Austria)

Social Tagging

- Social tagging is the process of collaboratively **annotating** content
- Essential instrument of Web 2.0 to **structure and search** Web content

- **Issues**
  - No rules for tags → can be **freely chosen**
  - Hard for people to come up with a set of descriptive/relevant tags **by their own**
  - People are **lazy** in applying tags
  - **Language**: Synonyms, spelling errors, singular/plurar …

[http://blog.zubiaga.org/2009/02/what-are-social-tags/]

Solution: Tag Recommendations

[http://de.slideshare.net/idoguy/social-recommender-systems-tutorial-www-2011-7446137]

Previous Work: Cognitive-Inspired Tag Recommendations

- **Activation equation** of the cognitive architecture ACT-R [Anderson et al, 2004]

$$A_i = B_i + \sum_j (W_j \cdot S_{j,i})$$

- Activation of memory unit i (i.e., **tag**) =
  - Base-level activation (**general usefullness**: tag frequency and recency in the past via power function [Anderson et al., 1991])
  - Associative activation (**usefullness** in the current **semantic context**: similarity with resource tags)

- Evaluation results showed that this approach **outperforms other state-of-the-art methods** (e.g., [Kowald et al., 2014] @ WWW; [Trattner et al., 2016] @ Journal of Web Science)
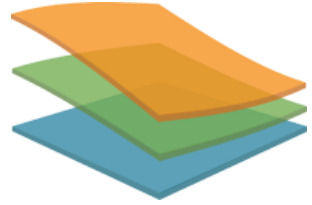
Present Work: Factors that Influence Tag Reuse
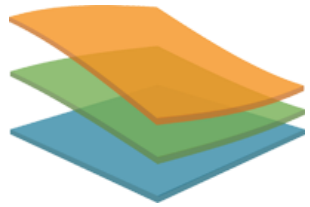
- ## *RQ1*
  - How are the factors of frequency, recency and semantic context **influencing a tag's probability of being reused** in social tagging systems?
    - → Empirical study

- ## *RQ2*
  - Can the factors of frequency, recency and semantic context be exploited to **efficiently predict a user's tag reuse given a specific folksonomy type**?
   - → Prediction study

Datasets

- Six social tagging datasets from different domains and of **different folksonomy types**

| Dataset	$|U|$	$|R|$	$|T|$	$|P|$	$|P|/|R|$	
Flickr	9,590	856,755	125,119	856,755	1.000	→ narrow
CiteULike	18,474	811,175	273,883	900,794	1.110	
BibSonomy	10,179	683,478	201,254	772,108	1.129	→ mixed
Delicious	15,980	963,741	184,012	1,447,267	1.501	
LastFM	1,892	12,522	9,748	71,062	5.674	
MovieLens	4,009	7,601	15,238	55,484	7.299	→ broad

- **Train / test sets**
  - For each user → most recent post in test set
    - Rest is used for training (reflecting the **past**)
    - Test set defines tag reuse (reflecting the **future**)
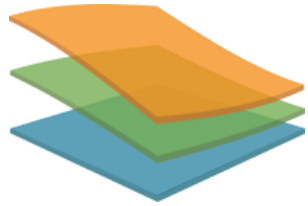
Results (RQ1)

This is for CiteULike but all datasets show similar results
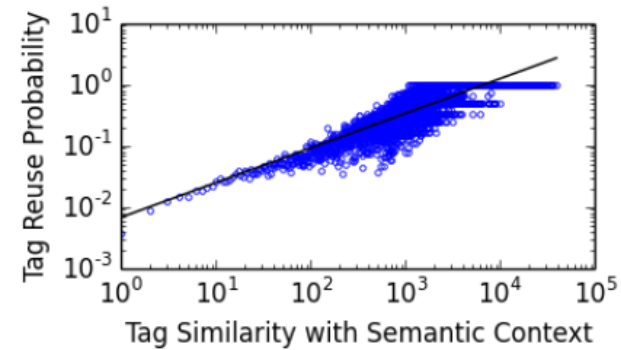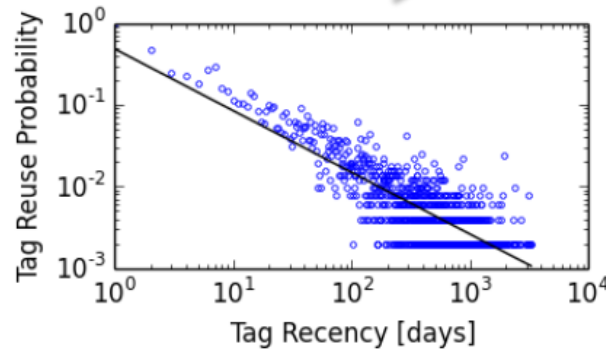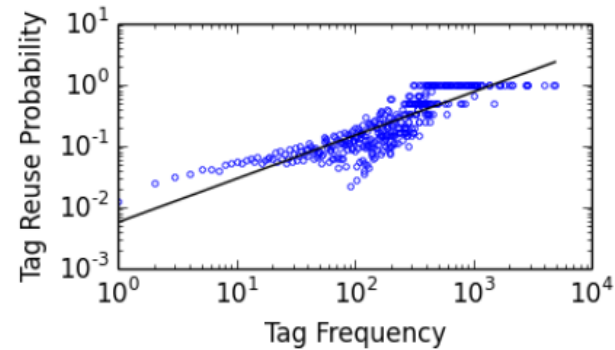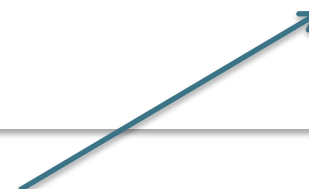
- The **more frequently** a tag was used in the past **(k > 0)**, the higher its reuse probability is.
- The **more recently** a tag was used in the past **(k < 0)**, the higher its reuse probability is.
- The **more similar** a tag is to tags in the **current semantic context (k > 0)**, the higher its reuse probability is.

→ **All three factors are important**

Results (RQ2)

		Most popular	Most recent	Most similar	[Zhang et al., 2012]	[Kowald et al., 2014] Frequ+Rec / all three	FolkRank / Tensor Fact.

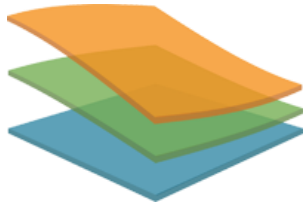|		Individual factors			Combination			Social	
Dataset	Metric	Frequency	Recency	SemCon	GIRP	BLL	BLL_{AC}	FR	PITF
Flickr	F_1@5	.371	.464	-	.455	**.470**	.470	.365	.350
	nDCG@10	.569	.702	-	.686	**.711**	.711	.561	.535
CiteULike	F_1@5	.231	.236	.041	.243	.254	**.259**	.250	.178
	nDCG@10	.367	.385	.069	.394	.413	**.422**	.392	.294
BibSonomy	F_1@5	.253	.252	.063	.262	.269	**.280**	.279	.215
	nDCG@10	.371	.368	.090	.386	.396	**.409**	.408	.327
Delicious	F_1@5	.173	.179	.108	.190	.203	**.243**	.196	.199
	nDCG@10	.267	.287	.158	.298	.318	**.374**	.292	.302
LastFM	F_1@5	.193	.189	.202	.198	.202	.251	.270	**.276**
	nDCG@10	.292	.293	.302	.303	.313	.375	.399	**.414**
MovieLens	F_1@5	.077	.076	.077	.077	.079	.086	.153	**.156**
	nDCG@10	.177	.183	.176	.177	.187	.203	.319	**.324**

Folksonomy type	Frequency	Recency	SemCon	Comb	Social
Narrow	+/-	+	-	+/-	-
Mixed	+	+	+/-	+	+/-
Broad	+/-	+/-	+	+/-	+
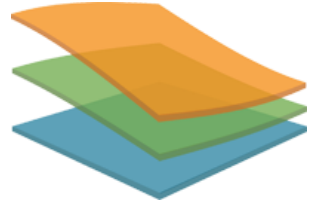
Conclusion & Future Work

- *RQ1*: All three factors influence tag reuse
  - $k > 0$ for **frequency and semantic context**
  - $k < 0$ for **recency**
- *RQ2*: Prediction accuracy depends on folksonomy type
 - **Recency** is most important in the **narrow** case
  - **Combination** of the factors works best in the **mixed** case
  - **Social** influence become better, the **broader** the folksonomy is

- **Future Work**
  - Analyze **social** influence (tag „imitation")
  - Extent **semantic context** (e.g., resource title or content)
  - Apply findings to **hashtags** (e.g., Twitter, Facebook, Instagram)
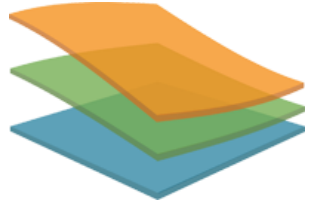
References

- J. R. Anderson and L. J. Schooler. Refections of the environment in memory. Psychological science, 2(6):396-408, 1991.

- J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An integrated theory of the mind. Psychological review, 111(4):1036, 2004.

- D. Kowald, P. Seitlinger, C. Trattner, and T. Ley. Long time no see: The probability of reusing tags as a function of frequency and recency. In Proc. Of WWW'14 companion, pages 463-468. International World Wide Web Conferences Steering Committee, 2014.

- D. Kowald, E. Lacic, and C. Trattner. Tagrec: towards a standardized tag recommender benchmarking framework. In Proc. of Hypertext'14, pages 305-307. ACM, 2014.

- D. Kowald and E. Lex. Evaluating tag recommender algorithms in real-world folksonomies: A comparative study. In Proc. of RecSys'15, pages 265-268, New York, NY, USA, 2015. ACM.

- C. Trattner, D. Kowald, P. Seitlinger, T. Ley, and S. Kopeinik. Modeling activation processes in human memory to predict the use of tags in social bookmarking systems. The Journal of Web Science, 2(1):1-18, 2016.
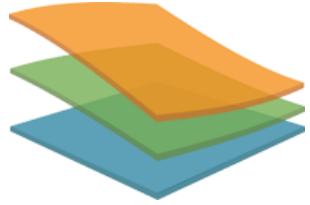
Thank you for your attention! Questions?

TagRec: open-source recommender evaluation framework

https://github.com/learning-layers/TagRec/

Dominik Kowald

dkowald@know-center.at

Know-Center, Graz University of Technology (Austria)