


Core univ. program seminar, Aug. 20-21, 2008.
Busan, Korea.

Speculative information discovery from social bookmarks

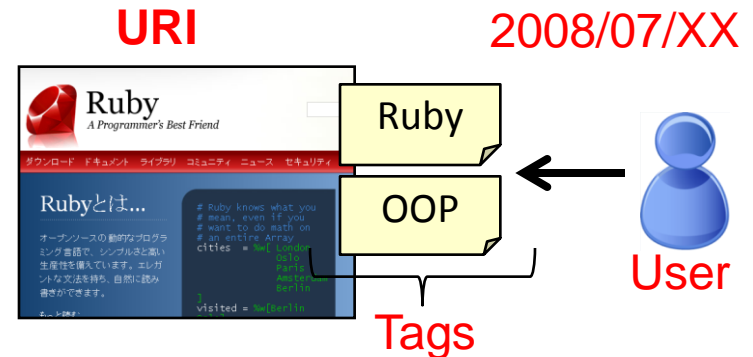
Eisuke ITO and Makoto MOMOTA
*Research Institute for Information Technology,
Dept. of Informatics,
Kyushu University*
itou@cc.kyushu-u.ac.jp

1. Introduction

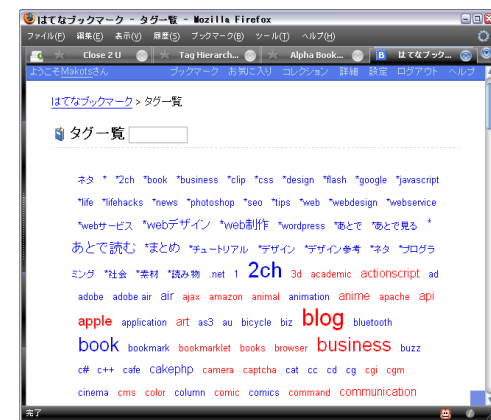
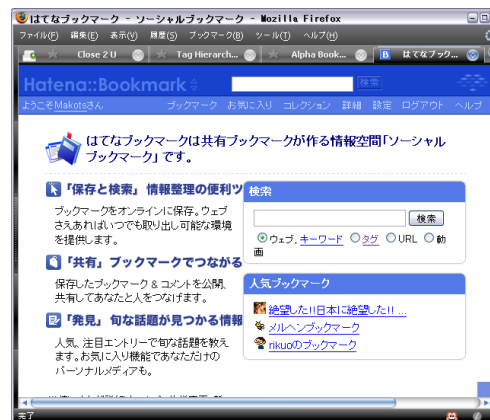
- 
1. Introduction
 2. What do you want?
 3. The alpha-bookmarker method
 4. Experiments and Results
 5. Conclusion

1. Introduction

- Social Bookmark
 - Online bookmark & bookmark sharing
 - Tagging web pages
- Elements of Social-bookmark
 - $\langle \text{URI, User, Tag(s), Date} \rangle$
- Good features
 - Location/machine free
 - Refer Tags by others
 - Classification using tags
 - Tag Cloud
 - Popular tags
 - Trend analysis



Hatena Bookmark (in Japan)



Feature of Social Bookmark

- Folksonomy
 - Folks + Taxonomy
 - Folks are tagging to contents
 - Web: Social Bookmark
 - Photo
 - Movie
 - Contents filtering by folks
 - User only tags interest contents.
- Merits
 - Low-cost (No expert)
 - Cover every category
- Demerits
 - Uncertain (not strict)
 - No Hierarchy (not classify, not categorize)

Related works

- Provided function
 - Tag cloud Not Personalized
 - Search by tags Not clear because tags are uncertain.
- Previous researches
 - Apply PageRank to gain precision
 - Yamaie et.al.(2007), Hotho et.al.(2007)
 - Reduce uncertainty by tag clustering
 - Rui et al. (2007), Niwa et al
- These researches try to reduce demerit.

Objectives

- We focus on merits of folksonomy
 - Use merits, and don't care demerits
- Speculative discovery of information from social bookmarks
 - Speculative
 - Recommend brand-new web pages which may become popular
 - Personalized service
 - May fit for one's interests.
 - Serendipity

2. What you want?

1. Introduction
- 2. What do you want?
3. The alpha-bookmarker method
4. Experiments and Results
5. Conclusion

2. What do you want?

- Search system in library
 - Well categorized
 - Well controlled vocabulary (metadata)
- Web search engine
 - Authorized Web pages
 - PageRank by Google
 - HITS algorithm
 - Correct/Qualified/Accurate pages
 - Past information
- Social Bookmark system
 - Up-to-date information
 - Tagging to News/blogs
 - Filtering by personal interest
 - Not categorized

What do you want from SBM?


- Speculative Information discovery
 - Recommend brand-new web pages which may become popular
- Personalized service
 - May fit for one's interests.

Strategy

- In each topic field, there are few leaders and a lot of followers
 - Leaders are good trend watchers of the topic.
 - Early detector of valuable news.
 - Followers catch up later.
 - They mentioned those news later
- Definition of alpha-bookmarker
 - Good trend watchers of the topic.
 - He/She has a lot of follower.

Find alpha-bookmarker for given topic,
Recommend alpha-bookmarker's watching web pages.

3. The alpha-bookmarker method

1. Introduction
2. What do you want?
-  3. The alpha-bookmarker method
4. Experiments and Results
5. Conclusion

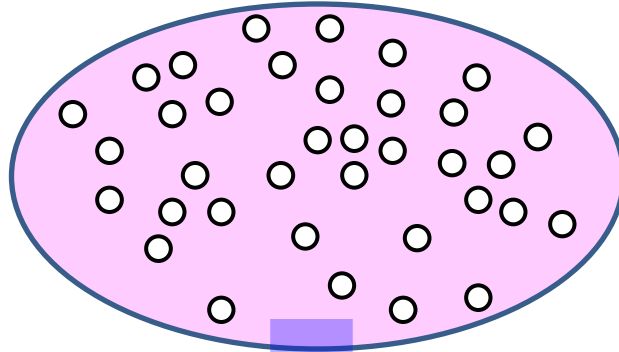
3. The Alpha-Bookmarker method

Find alpha-bookmarker for given topic,
Recommend alpha-bookmarker's watching web pages.

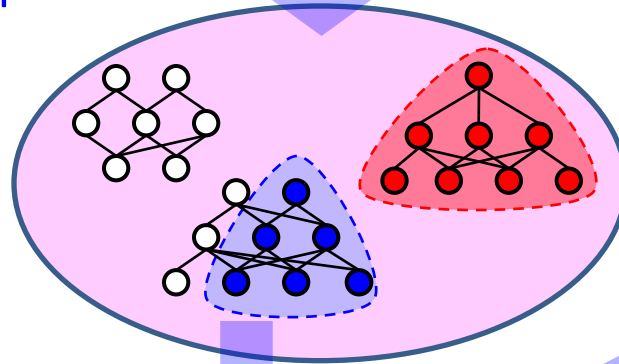
- Problems.
 - How to extract topics?
 - How to find alpha-bookmarker from SBM?
 - How to rank pages for recommendation?

Outline of the alpha-bookmarker method

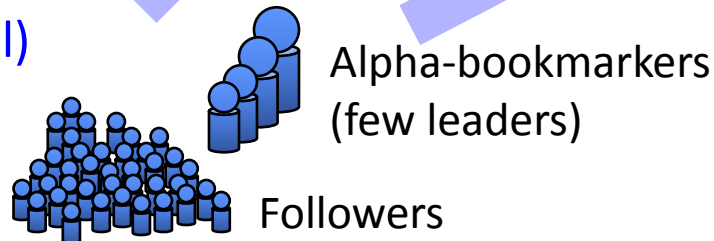
Collect massive
bookmarks
<URI, date, tags, user>



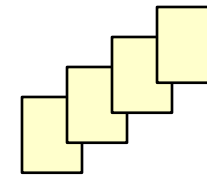
Tag Hierarchization
(Topic extraction)



Find alpha-bookmarker
for each topic (tag level)



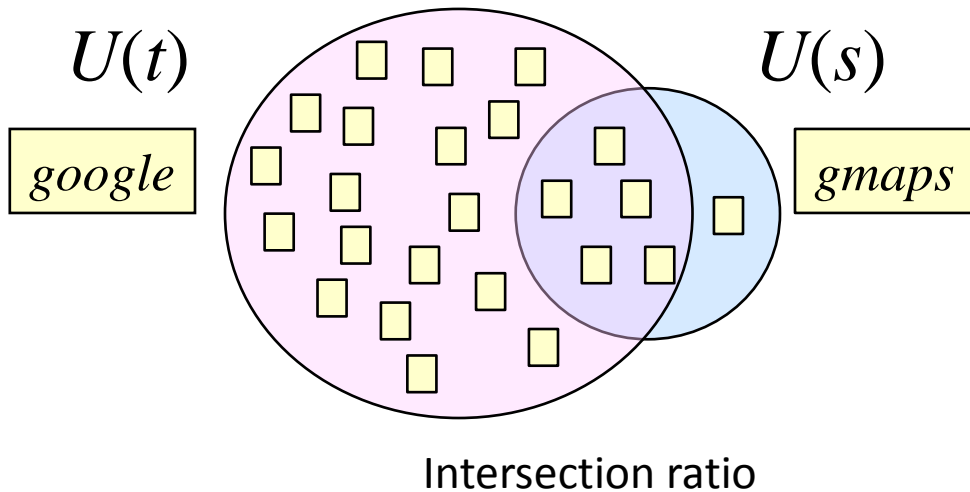
Recommended pages



URI (web page)
recommendation
based on ABM.

3.1 Tag Hierarchization (1/2)

$U(t)$: set of URIs (pages) which are tagged t .



- 'Google' is more popular than 'gmaps'.
- 'Google' tagged pages are rarely tagged 'gmaps'.
- 'gmaps' tagged pages are frequently tagged 'google'.

'google' is higher than 'gmaps'.

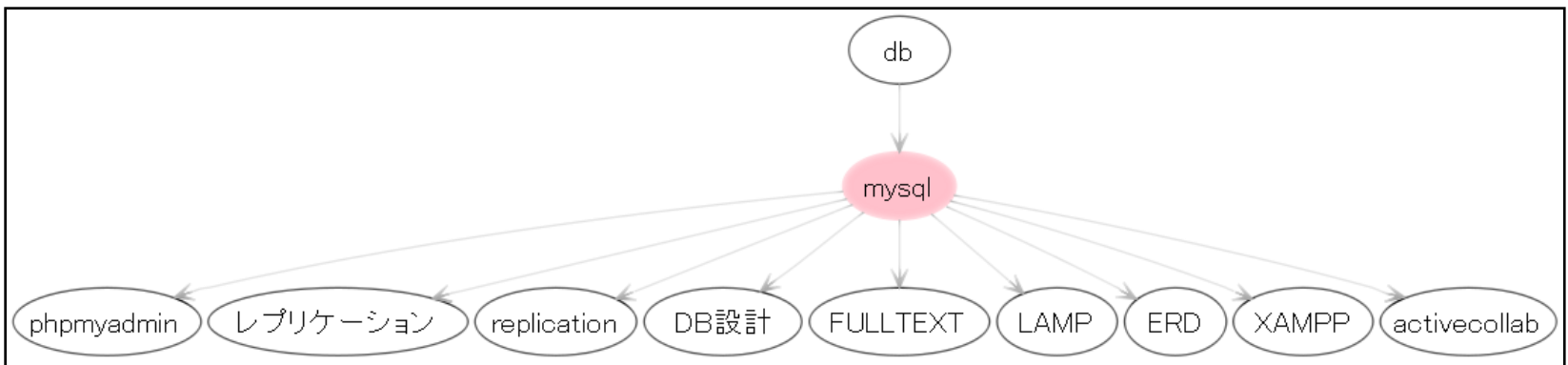
Definition

$$|U(t)| > |U(s)| \text{ and } \frac{|U(t) \cap U(s)|}{|U(s)|} > a \text{ then } t > s$$

(We used $a=50\%$.)

3.1 Tag Hierarchization (2/2)

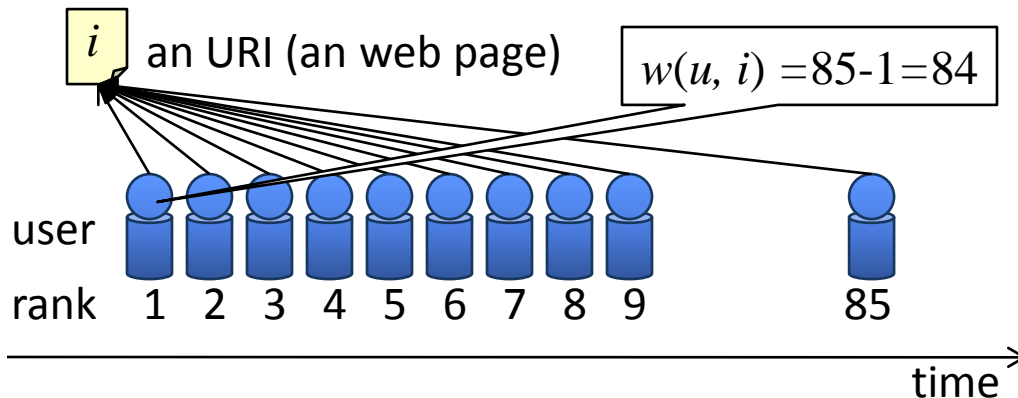
- Lower tags belongs to the higher tag.
- Def.
 - $\text{TOPIC}(t) = \{ t, \{\text{lower tags of } t\} \}$
- Ex.
 - $\text{TOPIC}(\text{mysql}) = \{ \text{mysql}, \text{phpmyadmin}, \text{replication}, \text{FULLTEXT}, \text{LAMP}, \text{ERD}, \text{XAMPP}, \text{activecollab}, \dots \}$



3.2 Find alpha-bookmarker (1/2)

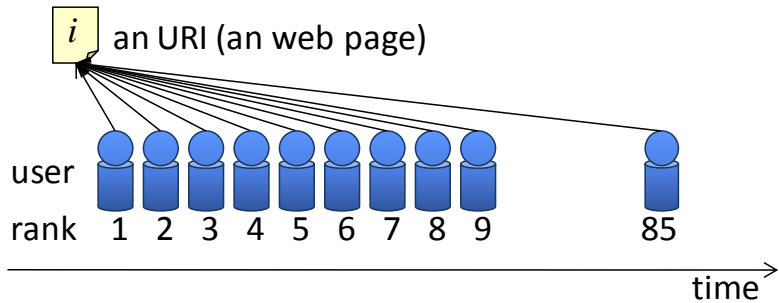
- Def. of alpha-bookmarker degree for a TOPIC(t).
 - t : a tag , u : a user, i : an URI .
 - $U(t) = \text{URI}(\text{TOPIC}(t))$: Set of URIs which is tagged one of tag $t \in \text{TOPIC}(t)$.
 - $\text{alpha}(u, t)$. : alpha-bookmarker degree of user u for a TOPIC(t).

$$\text{alpha}(u) = \sum_{i \in U(t)} w(i, u) \quad w(i, u) \text{ is weight of a user } u \text{ for the URI } i ,$$
$$w(i, u) = (\# \text{ of bookmarked users } \} - \text{rank}(u) .$$



3.3 Find alpha-bookmarker (2/2)

$$\alpha(u) = \sum_{i \in U(t)} w(i, u)$$



- Early bookmarking and many follower, then weight becomes heavy.

- Def. of alpha-bookmarkers for a TOPIC(t).

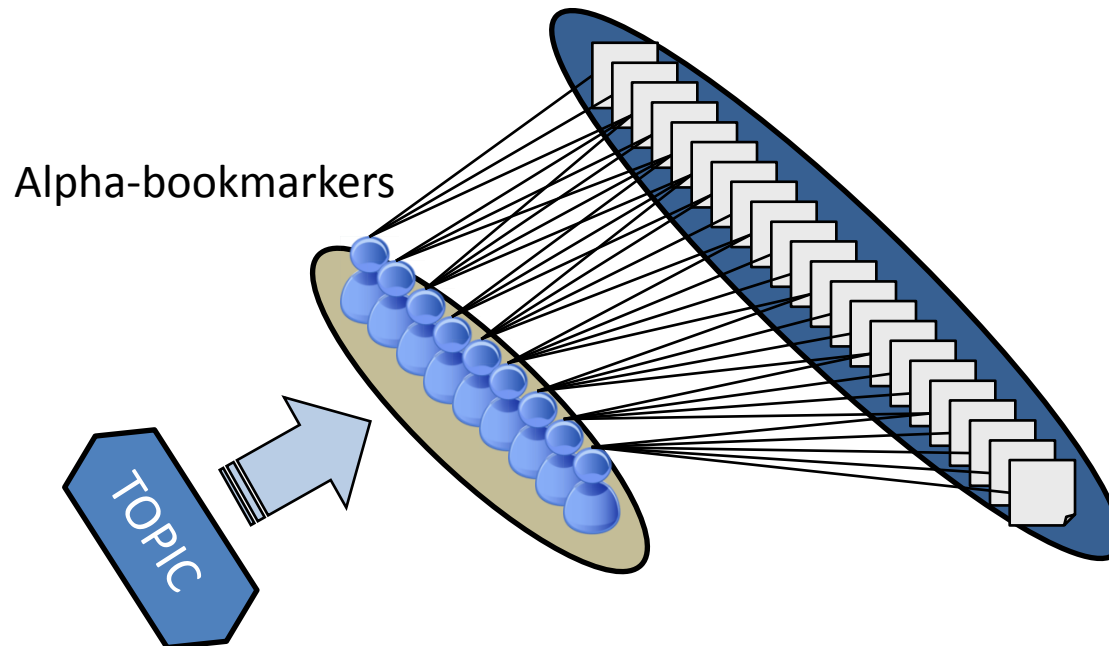
Top N th users by $\alpha(u, t)$ are alpha-bookmarkers for the TOPIC(t) .

(We used $N=10$.)


3.3 URI recommendation based on ABM

- Recommend high rank
- Def. : $\text{alpharank}(i,t)$, where i is an URI.

$$\text{alpharank}(i,t) = \sum_{i \in \text{bookmarkedURI}} \text{alpha}(u,t)$$



4. Experiments and Results

1. Introduction
2. What do you want?
3. The alpha-bookmarker method
-  4. Experiments and Results
5. Conclusion

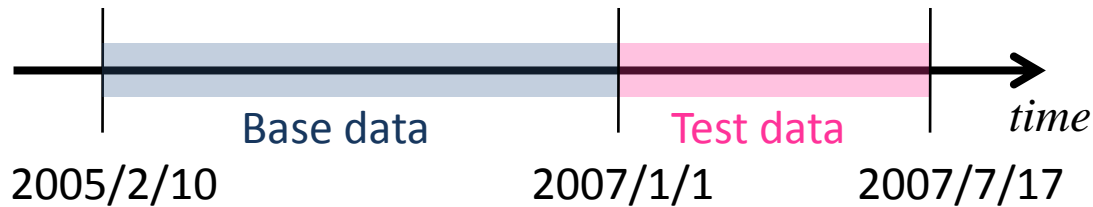
Experiments and Results

- Apply our method to real data (*Hatena Bookmark*)
 - *Hatena Bookmark* is the most popular SBM in Japan
 - Over 100,000 users
- Crawled data

Period	Feb. 10 2005 – Jul. 17, 2007
# of SBM	6,648,994
# of unique Tags	108,753
# of unique URI	350,279
# of unique Users	42,753

Experiments

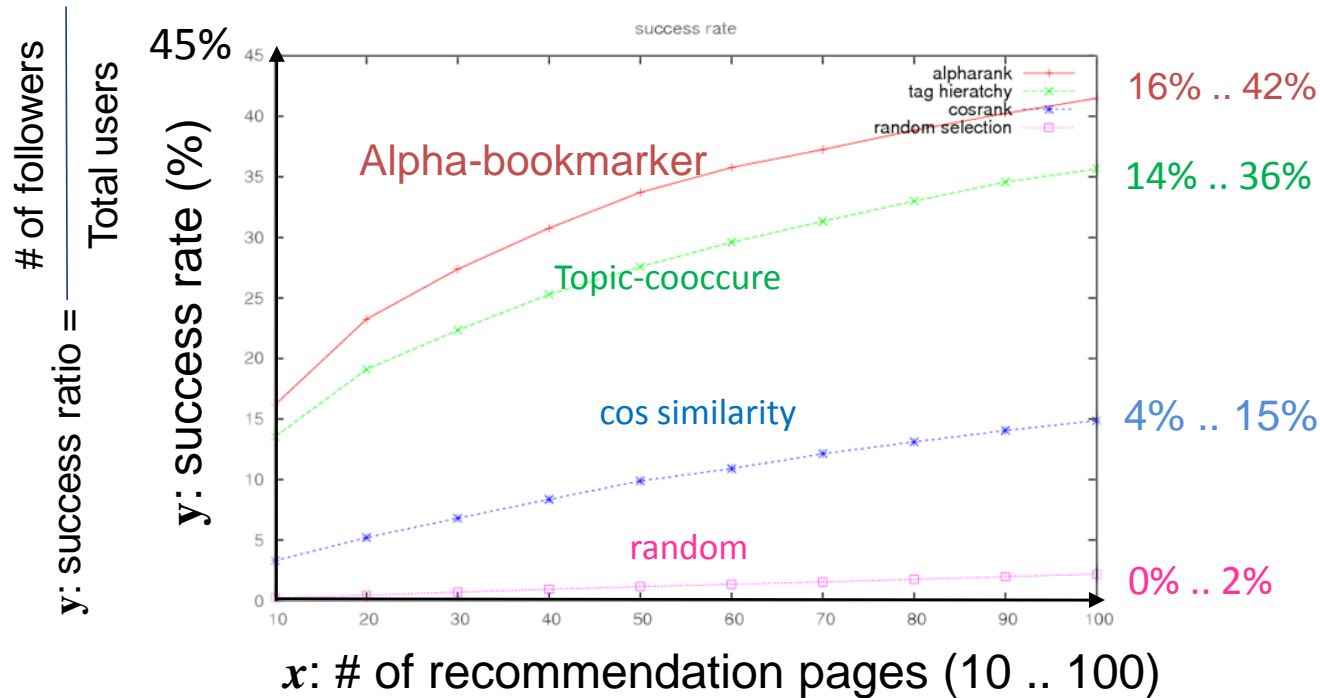
- Separate SBM data into 2 parts



- Assumption
 - If alpha-bookmarker's bookmarked pages are also bookmarked by follower.
- Definition of success
 - If recommended URI (from base data, before period) is bookmarked by someone at later period, then that is success.
- Test data
 - Tags: Top 20 TF(term frequency) tags
 - Users: Random sampling from users who tagged related tags.
- Compare 4 methods
 - Alpha-bookmarker
 - Topic co-occurrence
 - Cosine similarity
 - Random


Results

- Random Sampling 1,000 uses , 100 times, and calculate average



- Results
 - Alpha-bookmark method is most useful

5. Conclusion

1. Introduction
2. What do you want?
3. The alpha-bookmarker method
4. Experiments and Results
-  5. Conclusion

Search 'Alpha Bookmarker'

6. Conclusion

- Summary
 - Speculative information discovery from social bookmarks.
 - We developed *Alpha-Bookmarker* method.
 - Apply the method to *Hatena bookmark* data.
 - *Hatena bookmark* is the most popular SBM in Japan.
- Future works
 - Check adaptability for long period SBM data
 - Apply to other folksonomy system
 - del.icio.us (SBM), Flickr(photo), ...