Modeling Activation Processes in Human Memory for Tag Recommendations

Social tagging systems enable users to collaboratively assign freely chosen keywords (i.e., tags) to resources (e.g., Web links). In order to support users in finding descriptive tags, tag recommendation algorithms have been proposed. One issue of current state-of-the-art tag recommendation algorithms is that they are often designed in a purely data-driven way and thus, lack a thorough understanding of the cognitive processes that play a role when people assign tags to resources. A prominent example is the activation equation of the cognitive architecture ACT-R, which formalizes activation processes in human memory to determine if a specific memory unit (e.g., a word or tag) will be needed in a specific context. It is the aim of this thesis to investigate if a cognitive-inspired approach, which models activation processes in human memory, can improve tag recommendations. The findings of this thesis demonstrate that activation processes in human memory can be utilized to improve not only social tag recommendations but also hashtag recommendations. This opens up a number of possible research strands for future work, such as the design of cognitive-inspired recommender systems.

Dr. Dominik Kowald is a senior researcher in the Social Computing division at the Know-Center, Austria's leading research center for data-driven business and big data analytics, where he works on novel algorithms for recommender systems. He has a PhD. (with hons), MSc. (with hons) and BSc. in Computer Science from Graz University of Technology.

978-620-2-32072-6

Graz, TU Graz, Dissertation, 2017

Kowald

Dominik Kowald

Modeling Activation Processes in Human Memory for Tag Recommendations

Using Models from Human Memory Theory to Implement Recommender Systems for Social Tagging and Microblogging Environments

Südwestdeutscher Verlag für Hochschulschriften

Dominik Kowald

Modeling Activation Processes in Human Memory for Tag Recommendations

Dominik Kowald

Modeling Activation Processes in Human Memory for Tag Recommendations

Using Models from Human Memory Theory to Implement Recommender Systems for Social Tagging and Microblogging Environments

Südwestdeutscher Verlag für Hochschulschriften

Imprint

Cover image: www.ingimage.com

Publisher:
Südwestdeutscher Verlag für Hochschulschriften
is a trademark of
International Book Market Service Ltd., member of OmniScriptum Publishing Group
17 Meldrum Street, Beau Bassin 71504, Mauritius

Printed at: see last page
ISBN: 978-620-2-32072-6

Zugl. / Approved by: Graz, TU Graz, Dissertation, 2017

Contents

List of Figures

List of Tables

List of Symbols

text

u	Current user
t	Current tag
r	Current resource
B	Set of bookmarks
B_{train}	Training set
B_{test}	Test set
B_t	Set of bookmarks, in which t was used
U	Set of users
T	Set of tags
R	Set of resources
T_u	T used by user u
T_r	T used for resource r
Y	Set of tag assignments
Y_u	Y of user u
Y_t	Y of tag t
Y_r	Y of resource r

Y_b Y of bookmark b

$Y_{t,r}$ Y of tag t and resource r

$Y_{t,u}$ Y of tag t and user u

β Mixing parameter for linear combinations

c Semantic context cue

d Time-dependent decay parameter

$S(c,t)$ Association strength between cue c and tag t

$A(t,u,r)$ Activation level of tag t for user u and resource r

$B(t,u)$ Base-level activation of tag t for user u

$\widetilde{T}_k(u,r)$ Top-k recommended tags for user u and resource r

$T(u,r)$ Set of relevant tags used by user u for resource r

N_u Neighbors of user u for Collaborative Filtering

Z Number of topics for Latent Dirichlet Allocation

k_U, k_R, k_T Number of factors for Factorization Machines

l Number of iterations for FolkRank and Factorization Machines

α Learning rate for Factorization Machines

λ Regularization constant for Factorization Machines

U_S Set of seed users in Twitter

F Set of followees in Twitter

t Current tweet in Twitter

T Set of tweets

ht Current hashtag

HT Set of hashtags

F_u Set of followees of user u

HT_u HT of user u

HT_{F_u} HT of followees F_u

C_t Terms in tweet t

S_t Similar tweets of tweet t

HT_{S_t} HT of similar tweets S_t

$B_I(ht, u)$ Individual base-level activation for hashtag ht and user u

$B_S(ht, u)$ Social base-level activation for hashtag ht and user u

$CB(ht, t)$ Content-based score for hashtag ht and tweet t

λ Mixing parameter for $\mathrm{BLL}_{I,S,C}$

k Number of recommended tags / hashtags

Abstract

Social tagging systems enable users to collaboratively assign freely chosen keywords (i.e., tags) to resources (e.g., Web links). In order to support users in finding descriptive tags, tag recommendation algorithms have been proposed. One issue of current state-of-the-art tag recommendation algorithms is that they are often designed in a purely data-driven way and thus, lack a thorough understanding of the cognitive processes that play a role when people assign tags to resources. A prominent example is the activation equation of the cognitive architecture ACT-R, which formalizes activation processes in human memory to determine if a specific memory unit (e.g., a word or tag) will be needed in a specific context. It is the aim of this thesis to investigate if a cognitive-inspired approach, which models activation processes in human memory, can improve tag recommendations.

For this, the relation between activation processes in human memory and usage practices of tags is studied, which reveals that (i) past usage frequency, (ii) recency, and (iii) semantic context cues are important factors when people reuse tags. Based on this, a cognitive-inspired tag recommendation approach termed $\mathrm{BLL}_{AC}+\mathrm{MP}_r$ is developed based on the activation equation of ACT-R. An extensive evaluation using six real-world folksonomy datasets shows that $\mathrm{BLL}_{AC}+\mathrm{MP}_r$ outperforms current state-of-the-art tag recommendation algorithms with respect to various evaluation metrics. Finally, $\mathrm{BLL}_{AC}+\mathrm{MP}_r$ is utilized for hashtag recommendations in Twitter to demonstrate its generalizability in related areas of tag-based recommender systems. The findings of this thesis demonstrate that activation processes in human memory can be utilized to improve not only social tag recommendations but also hashtag recommendations. This opens up a number of possible research strands for future work, such as the design of cognitive-inspired resource recommender systems.

Zusammenfassung

Soziale Tagging Systeme ermöglichen das kollaborative Annotieren von Ressourcen (z.B. Web Links) mit Hilfe von frei wählbaren Schlagwörtern (d.h. Tags). Dabei werden die Benutzer dieser Tagging Systeme bei der Findung von passenden Tags von Empfehlungssystemen unterstützt. Eine Schwachstelle gängiger Algorithmen von Tag-Empfehlungssystemen ist es, dass diese oft rein Daten-getrieben arbeiten und somit kognitive Prozesse vernachlässigen, die wichtig für den Auswahlprozess von Tags sind. Ein bekanntes Beispiel stellt die Aktivierungsgleichung der kognitiven Architektur ACT-R dar, welche Aktivierungsprozesse des menschlichen Gedächtnisses formalisiert um den Nutzen einer bestimmten Gedächtniseinheit (z.B. ein Wort oder Tag) in einem bestimmten Kontext zu berechnen. Es ist das Ziel dieser Dissertation festzustellen, ob ein kognitiv-inspirierter Algorithmus, welcher diese Aktivierungsprozesse modelliert, Tag-Empfehlungssysteme verbessern kann.

Dazu wurde die Beziehung zwischen Aktivierungsprozessen und der Verwendung von Tags analysiert, welches zeigte, dass (i) Verwendungshäufigkeit, (ii) Verwendungszeitpunkt, und (iii) der semantische Kontext wichtige Faktoren für die Wahl von Tags darstellen. Darauf aufbauend wurde ein kognitiv-inspirierter Algorithmus, namens BLL_{AC}+MP_r, mit Hilfe der Aktivierungsgleichung von ACT-R entwickelt. Eine umfassende Evaluierung von BLL_{AC}+MP_r zeigte, dass dieser den Stand der Technik von Tag-Empfehlungssystemen gemessen anhand von gängigen Metriken übertrifft. Abschließend wurde BLL_{AC}+MP_r für die Empfehlung von Hashtags in Twitter adaptiert um zu demonstrieren, dass dieser Algorithmus auch für verwandte Arten von Tag-basierenden Empfehlungssystemen generalisiert werden kann. Die Forschungsergebnisse dieser Dissertation demonstrieren, dass Aktivierungsprozesse des menschlichen Gedächtnisses sowohl Tag- als auch Hashtag-Empfehlungen verbessern können. Dies eröffnet zukünftige Forschungsmöglichkeiten, wie z.B. die Entwicklung von weiteren kognitiv-inspirierten Empfehlungssystemen.

Chapter 1

Introduction

"Social tagging systems are excellent examples of distributed cognitive systems" [Fu, 2008]

Social tagging systems enable users to collaboratively assign freely chosen keywords (i.e., *tags*) to resources (e.g., Web links, scientific publications, music, images, movies, etc.). These tags can then be used for not only searching, navigating, organizing and finding content but also serendipitous browsing [Körner et al., 2010a, Rader and Wash, 2008, Chi and Mytkowicz, 2008]. Therefore, social tags have become an essential instrument of Web 2.0 (i.e., the social Web) to assist users during these activities. Another advantage of social tags is that users can freely choose them for annotating their bookmarked resources. However, this also means that these users have to create a set of descriptive tags on their own, which can be a very demanding task [Gemmell et al., 2009, Lipczak, 2012].

As a solution, tag recommendation algorithms have been proposed, which suggest a set of tags for a given user and a given resource. These tag suggestions are typically calculated based on previously used tags and/or the content of resources [Kowald, 2015]. Thus, tag recommendation algorithms aim to help not only the individual to find appropriate tags [Jäschke et al., 2008] but also the collective to consolidate the shared tag vocabulary with the aim to reach semantic stability and implicit consensus [Wagner et al., 2014, Kopeinik et al., 2017]. Furthermore, it was shown that personalized tag recommendations can increase the indexing quality of resources, which makes it easier for users to understand the information content of an indexed resource solely based on its assigned tags [Dellschaft and Staab, 2012].

Another strand of research on the underlying cognitive mechanisms of social tagging has shown that the way users choose tags for annotating resources corresponds to processes and structures in human memory [Cress et al., 2013, Held et al., 2012, Ley and Seitlinger, 2010, Fu, 2008]. In this respect, a prominent example is the activation equation of the cognitive architecture ACT-R, which formalizes activation processes in human memory. Specifically, the activation equation determines the activation level (i.e., probability) that a specific memory unit (e.g., a word or tag) will be needed (i.e., activated) in a specific context [Anderson et al., 2004].

However, while current state-of-the-art tag recommendation approaches perform reasonably well in terms of recommendation accuracy, most of them are designed in a purely data-driven way. Consequently, they are based on either simply counting tag frequencies [Jäschke et al., 2007] or computationally expensive calculation steps (e.g., calculating user similarities [Marinho and Schmidt-Thieme, 2008], modeling topics [Krestel et al., 2009] or factorizing the features of resources [Rendle, 2010]). Hence, these approaches typically ignore the above mentioned insights originating from cognitive research on how humans access information, such as words or tags, in their memory. This is contrary to the assumption that tag recommendation algorithms should attempt to mimic the user's tagging behavior [Kowald, 2015].

Thus, it is the aim of this thesis to show that a cognitive-inspired approach, which is build upon activation processes in human memory, can improve the state-of-the-art of tag recommendations. Furthermore, such an approach should help to better understand the underlying cognitive processes of social tagging and tag recommendations. Taken together, the problem this thesis aims to tackle is the following:

"There is a lack of knowledge about (i) how activation processes in human memory can be modeled for the task of predicting and recommending tags, and (ii) if this could lead to improvements in real-world tag recommendation settings."

1.1 Research Questions

Based on the aforementioned problem statement, four research questions guide this thesis, which are visualized in Figure 1.1.

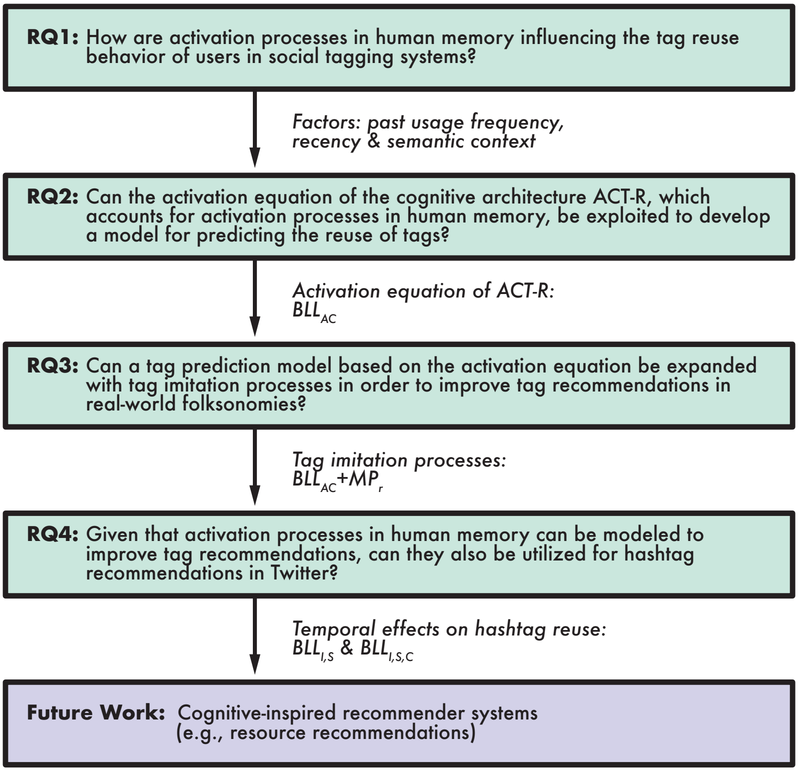


Figure 1.1: The four research questions of this thesis. While Research Question 1 deals with the identification of the relevant activation processes in human memory that influence the reuse of tags in social tagging systems (i.e, past usage frequency, recency and the current semantic context), Research Questions 2, 3 and 4 utilize the activation equation of the cognitive architecture ACT-R to design, implement and evaluate the tag recommendation algorithms BLL_{AC} and $\text{BLL}_{AC}+\text{MP}_r$ as well as the hashtag recommendation algorithms $\text{BLL}_{I,S}$ and $\text{BLL}_{I,S,C}$. The findings of this thesis open up a possible strand for future research, which is the design of cognitive-inspired recommender systems (e.g., for resource recommendations).

1.1.1 Research Question 1: The Influence of Activation Processes in Human Memory on Tag Reuse

"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"

The first research question of this thesis deals with the relation of activation processes in human memory and the tag reuse behavior of users in social tagging systems. According to [Anderson et al., 2004], the activation of a memory unit (e.g., a tag) should depend on at least two variables: (i) the general usefulness of this memory unit given by past usage frequency and recency, and (ii) its usefulness in the current semantic context. Therefore, it is the aim of Research Question 1 to test if this is also applicable for social tagging settings.

This research question is addressed in Chapter 4, in which a study is presented that sheds light on the influence of frequency, recency and semantic context on the reuse of tags. The results of this study show that there is a strong relation between activation processes in human memory and the use of tags in social tagging systems. Additionally, these findings act as a prerequisite for designing a cognitive-inspired approach for tag reuse prediction, which is investigated by Research Question 2.

1.1.2 Research Question 2: Designing a Cognitive-Inspired Algorithm for Tag Reuse Prediction

"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"

The aim of Research Question 2 is to identify the usefulness of the activation equation of the cognitive architecture ACT-R [Anderson et al., 2004] for the design of a tag reuse prediction algorithm termed BLL_{AC}. The activation equation quantifies the activation level of a piece of information in human memory by integrating three factors: (i) past usage frequency, (ii) past usage recency, and (iii) similarity with the current semantic context. Since Research Question 1 showed that these three factors influence the reuse of tags, Research Question 2 validates if this finding can be utilized for the prediction of tags.

To do so, in Chapter 5, the cognitive-inspired tag reuse prediction approach

BLL_{AC}, which is based on the activation equation of ACT-R, is presented and evaluated. The evaluation results show that BLL_{AC} provides higher accuracy and ranking estimates than algorithms reflecting its individual components and combinations of its components. Furthermore, these results show that social influences, by means of recommending popular tags of other users, have an impact on a user's choice of tags as well. This finding suggests that combining tag imitation processes with BLL_{AC} should lead to further improvements, which is addressed by Research Question 3.

1.1.3 Research Question 3: Implementing a Hybrid Approach for Tag Recommendations in Real-World Folksonomies

"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"

Research Question 3 deals with the extension of BLL_{AC} by tag imitation processes to realize a hybrid tag recommendation approach for real-world folksonomies called BLL_{AC}+MP_r. This question is addressed in Chapter 6. Specifically, it is shown that popular tags used by other users to annotate the current resource can be used to enhance the recommendation accuracy of BLL_{AC}. Another focus of this research question lies on the evaluation of BLL_{AC}+MP_r in a real-world folksonomy setting.

Therefore, unfiltered datasets from six social tagging environments (i.e, Flickr, CiteULike, BibSonomy, Delicious, MovieLens and LastFM, see Section 3.2) are used to compare BLL_{AC}+MP_r to a rich set of state-of-the-art tag recommendation algorithms. This is done using various evaluation metrics that validate not only the accuracy and ranking quality of the recommendations but also the diversity, novelty and computational costs. The results of this study show that BLL_{AC}+MP_r provides the most robust results over all datasets with respect to various evaluation metrics and folksonomy settings. These strong results raise the question if activation processes in human memory can also be utilized for related recommendation tasks, which is the aim of Research Question 4.

1.1.4 Research Question 4: Utilizing the Approach for Hashtag Recommendations in Twitter

"Given that activation processes in human memory can be modeled to improve tag recommendations, can they also be utilized for hashtag recommendations in Twitter?"

The fourth and last research question of this thesis addresses the generalizability of BLL_{AC}+MP_r by utilizing it for related use cases in the area of recommender systems, such as hashtag recommendations in Twitter. Therefore, in Chapter 7, two data collections are crawled from Twitter and temporal dynamics are studied in these data collections in order to propose two cognitive-inspired hashtag recommendation approaches called $\text{BLL}_{I,S}$ and $\text{BLL}_{I,S,C}$.

The evaluation of $\text{BLL}_{I,S}$ and $\text{BLL}_{I,S,C}$ shows that both approaches outperform related algorithms in the field of hashtag recommendations. These findings demonstrate that activation processes in human memory cannot only be utilized for social tag recommendations but also for hashtag recommendations. Furthermore, these findings open up a number of possible research strands for future work, such as the design of cognitive-inspired recommender systems (e.g., for resource recommendations, see Section 8.3).

1.2 Scientific Contributions

With relation to the research questions and methodology of this thesis, the five main contributions are as follows:

1. It is shown that activation processes in human memory impact tag reuse practices in social tagging systems. Specifically, the factors of (i) past usage frequency, (ii) recency, and (iii) semantic context highly correlate with the reuse probability of social tags (Research Question 1).

2. A cognitive-inspired algorithm for tag reuse predictions termed BLL_{AC} is presented, which utilizes the activation equation of the cognitive architecture ACT-R. This algorithm provides higher accuracy and ranking estimates than related tag prediction approaches (Research Question 2).

3. BLL_{AC} is combined with tag imitation processes by means of popular tags used by other users to realize a hybrid algorithm for tag recommendations in real-world folksonomies termed BLL_{AC}+MP_r. BLL_{AC}+MP_r outperforms state-of-the-art tag recommendation algorithms with respect to various evaluation metrics (Research Question 3).

4. It is demonstrated that activation processes in human memory can also be utilized for related use cases in the research area of tag-based recommender systems, such as hashtag recommendations in Twitter. This is done by analyzing temporal effects on hashtag reuse and by proposing two cognitive-inspired hashtag recommendation approaches called $\mathrm{BLL}_{I,S}$ and $\mathrm{BLL}_{I,S,C}$ (Research Question 4).

5. Finally, the *TagRec* framework is presented as an open-source tag recommendation benchmarking toolkit. *TagRec* is used to address all four research questions of this thesis and thus, fosters the reproducibility of the presented results and findings.

These contributions have been published in 13 scientific publications:

(P1) **Kowald, D.** (2015). Modeling cognitive processes in social tagging to improve tag recommendations. In *Proceedings of the 24th International Conference on World Wide Web, WWW '15 Companion*, pages 505-509, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

This work presents a general idea of how cognitive processes can be utilized to improve tag recommendations in social tagging systems. Therefore, three types of cognitive processes are studied, which are (i) categorization processes, (ii) temporal dynamics, and (iii) tag imitation processes, to derive a tag recommendation algorithm termed 3LT+MP_r [Kowald, 2015]. One weakness of this approach is that it depends on sparsely available category information of resources, which is the reason why the algorithm described in this thesis (i.e., BLL_{AC}+MP_r) neglects categorization processes. Instead, BLL_{AC}+MP_r takes the semantic context of social tagging into account via implementing the full activation equation of the cognitive architecture ACT-R.

However, the presentation of this work in the Doctoral Symposium track of the 24th International Conference on World Wide Web in Florence, Italy was an important step for the realization of this thesis and for improving the concept of a tag recommendation algorithm based on cognitive processes.

(P2) **Kowald, D.** and Lex, E. (2016). The influence of frequency, recency and semantic context on the reuse of tags in social tagging systems. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media, HT '16*, pages 237-242, New York, NY, USA. ACM.

This paper analyzes the influence of frequency, recency and semantic context on the reuse of tags in social tagging systems [Kowald and Lex, 2016]. It is shown that all three factors play a role when people reuse tags and that the intensity of this influence greatly depends on the given social tagging system (i.e., narrow versus broad folksonomies). This analysis contributes to Research Question 1 and is used as a prerequisite for designing BLL_{AC} using the activation equation of ACT-R.

Furthermore, this paper contains a prediction study, which validates the tag prediction quality of BLL_{AC} and its components, which contributes to Research Question 2. The author of this thesis has received a student travel grant in order to be able to present this paper at the 27th ACM Conference on Hypertext and Social Media in Halifax, Canada.

(P3) **Kowald, D.**, Seitlinger, P., Trattner, C., and Ley, T. (2014). Long time no see: The probability of reusing tags as a function of frequency and recency. In *Proceedings of the 23rd International Conference on World Wide Web, WWW '14 Companion*, pages 463-468, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

In this work, the first version of the cognitive-inspired tag recommendation algorithm BLL_{AC} is proposed, which contributes to Research Question 2. This first version of the algorithm solely utilizes the Base-Level Learning (BLL) equation of ACT-R and thus, neglects the semantic context of tag assignments [Kowald et al., 2014b].

However, by combining this approach with a frequency-based analysis of the most popular tags assigned to the bookmarked resource, this paper already shows that the approach is able to outperform related tag recommendation methods.

(P4) **Kowald, D.**, Kopeinik, S., Seitlinger, P., Ley, T., Albert, D., and Trattner, C. (2015). Refining frequency-based tag reuse predictions by means of time and semantic context. In *Mining, Modeling, and Recommending 'Things' in Social Media*, pages 55-74. Springer.

The complete version of BLL_{AC}+MP_r is presented in this book chapter. Hence, the full activation equation of ACT-R is implemented via combining the base-level activation by means of the BLL equation with an associative activation by means of normalized tag co-occurrences [Kowald et al., 2015a].

Furthermore, BLL_{AC}+MP_r is evaluated against state-of-the-art tag recommendation methods (e.g., Collaborative Filtering, FolkRank and Tensor Factorization, see Section 2.2) and it is shown that BLL_{AC}+MP_r outperforms these methods in terms of recommendation accuracy and ranking (Research Question 3).

(P5) Trattner, C., **Kowald, D.**, Seitlinger, P., Ley, T., and Kopeinik, S. (2016). Modeling activation processes in human memory to predict the use of tags in social bookmarking systems. *The Journal of Web Science*, 2(1), pages 1-18.

This journal article gives a detailed description of the development process of BLL_{AC}+MP_r and thus, contributes to Research Questions 1, 2 and 3. With respect to Research Question 1, the time-dependent decay of tag reuse is investigated and it is shown that a power function is better suited to model these temporal effects. With respect to Research Questions 2 and 3, a detailed evaluation is presented, which also discusses the computational complexity of the algorithms [Trattner et al., 2016].

Apart from that, the evaluation results are validated using the ECML PKDD Discovery Challenge 2009 dataset in order to increase the reproducibility of the evaluation results.

(P6) **Kowald, D.**, Lacic, E., and Trattner, C. (2014). TagRec: Towards a standardized tag recommender benchmarking framework. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media, HT '14*, pages 305-307, New York, NY, USA. ACM.

This work describes the initial version of the *TagRec* framework, which is developed as part of this thesis in order to conduct the tag recommendation evaluation procedures addressing Research Questions 1, 2 and 3. Apart from that, this work won the best poster award at the 25th ACM Conference on Hypertext and Social Media in Santiago, Chile.

TagRec is an open-source tag recommender benchmarking framework with the aim to support developers and researchers of tag-based recommender systems in all steps of the development and evaluation process of novel algorithms [Kowald et al., 2014a].

(*P7*) Kowald, D., Kopeinik, S., and Lex, E. (2017). The TagRec framework as a toolkit for the development of tag-based recommender systems. In *Adjunct Publication of the 25th Conference on User Modeling Adaptation and Personalization, UMAP '17*, pages 23-28, New York, NY, USA. ACM.

This paper describes the final version of *TagRec*, which extends the framework with algorithms and evaluation methods for other types of recommender systems, such as resource and hashtag recommendations [Kowald et al., 2017a].

Specifically, this final version of the framework enables to address the hashtag recommendation evaluation procedure with respect to Research Question 4.

(*P8*) Kowald, D. and Lex, E. (2015). Evaluating tag recommender algorithms in real-world folksonomies: A comparative study. In *Proceedings of the 9th ACM Conference on Recommender Systems, RecSys '15*, pages 265-268, New York, NY, USA. ACM.

The evaluation of tag recommendation algorithms in real-world folksonomy settings is not a trivial task and greatly depends on the given user needs. Therefore, this paper compares a rich set of state-of-the-art tag recommendation algorithms in six unfiltered social tagging datasets via various evaluation metrics, measuring prediction accuracy, diversity, novelty and computational costs of the algorithms [Kowald and Lex, 2015].

This contributes not only to Research Question 3 but also to the general methodology used in this thesis (see Chapter 3).

(*P9*) Kopeinik, S., **Kowald, D.**, and Lex, E. (2016). Which algorithms suit which learning environments? A comparative study of recommender systems in TEL. In *Proceedings of the 11th European Conference on Technology Enhanced Learning, ECTEL'16*, pages 124-138. Springer.

This work presents an extensive study comparing a variety of recommendation algorithms in Technology Enhanced Learning (TEL) settings. These TEL datasets are typically small and sparse and thus, are not suitable for classic tag recommendation approaches such as MostPopular, Latent Dirichlet Allocation, FolkRank or

Collaborative Filtering [Kopeinik et al., 2016b].

The evaluation results described in this paper show that BLL_{AC}+MP_r outperforms these classic approaches not only in social bookmarking environments but also in TEL settings (Research Question 3).

(P10) **Kowald, D.**, Pujari, S., and Lex, E. (2017). Temporal effects on hashtag reuse in Twitter: A cognitive-inspired hashtag recommendation approach. In *Proceedings of the 26th International Conference on World Wide Web, WWW'17*, pages 1401-1410, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

This paper describes how activation processes in human memory can be used for designing a hashtag recommendation algorithm [Kowald et al., 2017b]. To that end, the effect of time on individual and social hashtag reuse is studied to derive a predictive model using the BLL equation of ACT-R. Furthermore, this model is combined with a content-based tweet analysis in order to further increase the recommendation quality.

Specifically, it is the aim of this work to show that BLL_{AC}+MP_r can be generalized for related use cases in the research area of recommender systems, such as hashtag recommendations in Twitter (Research Question 4).

(P11) Lacic, E., **Kowald, D.**, Seitlinger, P., Trattner, C., and Parra, D. (2014). Recommending items in social tagging systems using tag and time information. In *Proceedings of the 1st International Workshop on Social Personalization co-located with the 25th ACM Conference on Hypertext and Social Media, HT '14*. CEUR-WS.

In this work, tag and time information is used to improve resource recommendations based on Collaborative Filtering. Therefore, the BLL equation of the cognitive architecture ACT-R is used to identify the importance of a resource for a user based on temporal tag usage patterns [Lacic et al., 2014b].

This demonstrates that activation processes in human memory are also useful for resource recommendations and thus, opens up this research strand as potential future work (see Section 8.3).

(P12) Seitlinger, P., **Kowald, D.**, Trattner, C., and Ley, T. (2013). Recommending tags with a model of human categorization. In *Proceedings of the 22nd ACM international conference on Conference on information and knowledge management,* *CIKM '13*, pages 2381-2386, New York, NY, USA. ACM.

This paper describes the *3Layers* tag recommendation approach, which is based on a model of human categorization. This approach uses categories assigned to resources in order to simulate a categorization process, which matches the currently bookmarked resource against already bookmarked resources. Then, the tags of the most semantically similar resources are suggested to the user [Seitlinger et al., 2013].

While categorization processes are not directly reflected in $\mathrm{BLL}_{AC}+\mathrm{MP}_r$, this paper provides important insights for the integration of cognitive processes into the tag recommendation process (see also 2.1.2).

(*P13*) Kowald, D., Seitlinger, P., Kopeinik, S., Ley, T., and Trattner, C. (2015). Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic tag recommender. In *Mining, Modeling, and Recommending 'Things' in Social Media*, pages 75-95. Springer.

In this paper, the *3Layers* tag recommendation algorithm is extended by incorporating the time-dependent decay of tag reuse. This is realized by integrating the BLL equation of the cognitive architecture ACT-R [Kowald et al., 2015b].

Thus, this work shows that the BLL equation cannot only be used to design a novel tag recommendation approach but also to improve existing algorithms (see also Section 8.3).

1.3 Structure of this Thesis

The remainder of this thesis is structured as follows: In Chapter 2, related work is presented, which includes the motivation and cognitive processes behind social tagging as well as the most important approaches in the area of tag-based recommender systems. This is followed by Chapter 3, in which a description of the methodology used in this thesis is given. This includes the conducted experiments, the collected data, the implemented baseline algorithms, the evaluation method and the *TagRec* framework.

In Chapter 4, the influence of activation processes in human memory on tag reuse is studied, which contributes to Research Question 1. Next, Research Question 2 is investigated in Chapter 5 by presenting the design of a cognitive-inspired algorithm for tag reuse prediction using the activation equation of the cognitive architecture

ACT-R. This algorithm is extended in Chapter 6 with tag imitation processes to implement a hybrid approach for tag recommendations in real-world folksonomies in order to contribute to Research Question 3.

Chapter 7 utilizes this approach for hashtag recommendations in Twitter, which shows that activation processes in human memory can be generalized for related use cases in the area of recommender systems (Research Question 4). Finally, a summary of the contributions of this thesis along with possibilities for future work are given in Chapter 8.

1.4 Terms and Definitions

This section lists and describes the most important terms and definitions used in this thesis in alphabetical order.

Activation processes. Activation processes describe how human memory tunes the activation of its memory units to statistical regularities of the environment [Anderson and Schooler, 1991]. The activation equation of the cognitive architecture ACT-R formalizes these activation processes.

ACT-R. ACT-R, which is short for "Adaptive Control of Thought – Rational", is a cognitive architecture developed by John Robert Anderson [Anderson, 1996, Anderson et al., 2004, Anderson et al., 1997]. ACT-R deals with defining and formalizing the basic cognitive operations of the human mind (e.g., the access on information in human memory).

Bookmark. With respect to social tagging, a bookmark (or sometimes also called "post") is a set of tags used by a given user to annotate a given resource (e.g., a Web link) [Chi and Mytkowicz, 2008]. Bookmarks are used to store favorite resources for later retrieval.

Data sparsity. Data sparsity describes to which extent a specific measure contains empty values. In social tagging systems, this refers to the degree of narrowness of a folksonomy and thus, indicates how often a resource was bookmarked [Helic et al., 2012]. p-core pruning is one way to unnaturally create denser (or broader) folksonomies [Doerfel et al., 2016].

Evaluation metric. An evaluation metric measures the performance of an algorithm with respect to a certain dimension and is typically used to compare algorithms with each other. In the area of recommender systems, mainly metrics for measuring the accuracy, ranking quality, diversity, novelty and computational costs of recommendations are used [Gunawardana and Shani, 2009, Konstan, 2004, Baeza-Yates et al., 1999].

Folksonomy. Folksonomies are hierarchical structures obtained from social tagging data. Thus, a folksonomy is the set of all bookmarks in a social tagging system. The literature distinguishes between "narrow" (i.e., only one user tags a resource) and "broad" (i.e., multiple users tag a resource) folksonomies [Helic et al., 2012]. Folksonomies that cannot explicitly be categorized into these two types are referred as "mixed" folksonomies (i.e., only a few users tag the same resource) [Kowald and Lex, 2016].

Need probability. The need probability of a memory unit (e.g., a word or tag) is the probability (or activation level) that this unit will be needed in a specific situation. It depends on at least two variables: (i) the general usefulness of this memory unit, and (ii) its associations to the current semantic context [Anderson and Schooler, 1991].

Recommendation accuracy. The recommendation accuracy of an algorithm is given by the number of correctly predicted recommendations. Thus, it is given by the intersection of the set of recommended tags with the set of relevant tags. Precision and Recall are two of the most popular metrics for measuring recommendation accuracy [Van Rijsbergen, 1974].

Semantic context. When a user tags a resource on the Web, the tag choices of this user are influenced by the semantic context of the currently bookmarked resource. This current semantic context typically consists of information related to the resource such as the title, description text and tags. Because of the lack of content data in publicly available social tagging datasets, this thesis focuses on the resource's tags to model the semantic context [Kowald and Lex, 2016].

Social tagging system. Social tagging systems are important instruments, which enable their users to collaboratively bookmark and annotate resources with freely-chosen keywords (i.e., tags) [Körner et al., 2010a]. In this thesis, six social tagging systems are investigated: Flickr (images), CiteULike (scientific references), Bib-

Sonomy (Web bookmarks and scientific references), Delicious (Web bookmarks), LastFM (music) and MovieLens (movies).

Tag assignment. A tag assignment is a triple consisting of a user, a resource and a tag [Jäschke et al., 2008]. Thus, the set of tag assignments of a given user and a given resource is denoted as a bookmark.

Tag cloud. A tag cloud visualizes the most popular tags in a social tagging system by using the font size of the tags as a proxy for their popularity [Helic et al., 2011, Sinclair and Cardew-Hall, 2008]. Such a visualization can be generated based on the tags of a user, the tags of a resource or all tags in a folksonomy.

Tag recommendations. Tag recommendations support users in finding descriptive tags for bookmarking resources. Thus, a tag recommendation system is a type of recommender system, which suggests a set of tag for a given user and a given resource [Jäschke et al., 2008].

Time-dependent decay. The time-dependent decay of memory access describes the effect of recency on the reuse probability of a memory unit (e.g., a word or tag). Thus, the more recent a memory unit was used, the higher the probability that it will be reused. It was shown that this time-dependent decay can be modeled using a power-law distribution [Anderson and Schooler, 1991].

Web 2.0. The Web 2.0 (also referred to as the social Web or the participatory Web) describes the part of the Web that focuses on user-generated content. Important examples of Web 2.0 systems are wikis (e.g., Wikipedia), social networking sites (e.g., Facebook) and social tagging systems (e.g., BibSonomy) [O'reilly, 2005].

Chapter 2

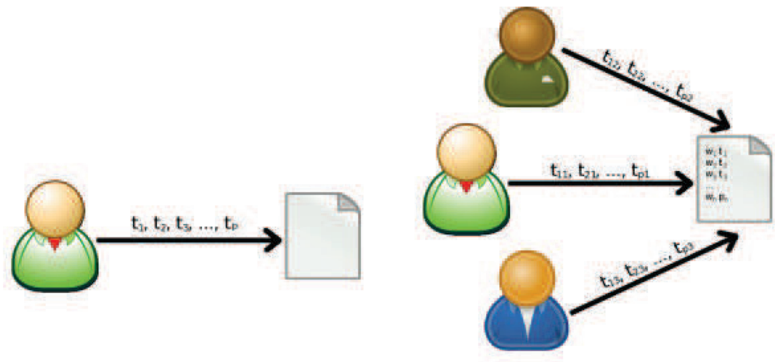
Related Work

"Tag recommendation reduces the cognitive effort from generation to recognition." [Gemmell et al., 2009]

In this chapter, the state-of-the-art research related to this thesis is presented in order to describe the preliminaries for Research Questions 1 to 4 outlined in Chapter 1. Thus, this chapter covers the motivation and cognitive processes behind social tagging (Section 2.1) as well as the most important approaches in the area of tag-based recommender systems (Section 2.2). This includes not only tag recommendation methods but also algorithms for recommending hashtags. This chapter is based on P1 [Kowald, 2015], P5 [Trattner et al., 2016], P12 [Seitlinger et al., 2013] (Sections 2.1 and 2.2.1) and P10 [Kowald et al., 2017b] (Section 2.2.2).

2.1 Social Tagging

Social tagging is the process of collaboratively assigning freely-chosen keywords (i.e., tags) to resources such as Web bookmarks, academic references, images, music or videos [Zubiaga, 2009] (see Figure 2.1). Recent years have shown that social tagging is an important feature of the social Web, supporting users with a simple mechanism to collaboratively organize and find content [Körner et al., 2010a]. Thus, in this section, the motivation behind the use of social tagging systems is discussed. Furthermore, a literature overview describing the underlying cognitive processes of social tagging is given. This section is based on P1 [Kowald, 2015] and P12 [Seitlinger et al., 2013].



(a) Narrow folksonomy (b) Broad folksonomy

Figure 2.1: Illustration of the social tagging process, where users assign freely chosen keywords (i.e., tags) to resources. These tripartite structures are also referred as "folksonomies". In narrow folksonomies (a), only the user who uploads the resource is allowed to tag it, whereas in broad folksonomies (b), all users are allowed to tag the resources. These figures were used with permission from Arkaitz Zubiaga [Zubiaga, 2009].

2.1.1 Motivation of Social Tagging

In order to understand the requirements for a successful tag recommendation algorithm, a general understanding of the motivation behind the use of social tagging systems is needed [Kowald, 2015]. To that end, two main types of tagging models have been defined in the literature: (i) the personal tagging model, and (ii) the collaborative tagging model [Lipczak, 2012, Lipczak et al., 2009].

Personal Tagging Model

The *personal tagging model* focuses on the individual user and assumes that she mainly uses the social tagging system as an own repository for storing and organizing bookmarks. For this reason, the user mainly draws on her own tag vocabulary, thus reuses tags she has used before [Rader and Wash, 2008, Sen et al., 2006]. It has been shown that these individual tags significantly improve the search process for bookmarked resources [Dellschaft and Staab, 2012, Trattner et al., 2012].

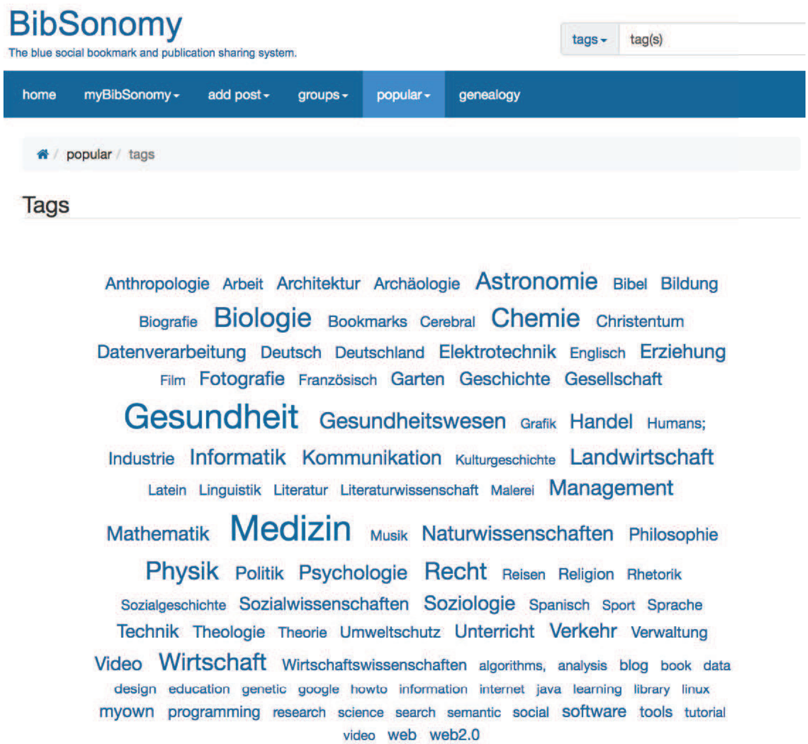


Figure 2.2: Example of a tag cloud in BibSonomy (retrieved 07-July-2017). The font size reflects the popularity of the tags.

According to [Körner et al., 2010b], two types of users can be defined based on the personal tagging motivation: (i) categorizers, and (ii) describers. While categorizers use tags for *categorizing* resources, describers use tags for *describing* resources. Simplified, this can be defined via the tag/resource ratio (*trr*) metric of the user. The *trr* is given by the vocabulary size of the user (i.e., the number of used tags) divided by the number of resources bookmarked by this user. Since categorizers use a limited vocabulary for tagging resources, they should achieve a lower *trr* score than describers, who use a variety of tags for this purpose.

Collaborative Tagging Model

The *collaborative tagging model* assumes that a user does not only reuses her own tags but also that a user is influenced by tags used by other users [Golder and Huberman, 2006, Dellschaft and Staab, 2008]. This in turn leads to a shared knowledge base / vocabulary of the community [Robu et al., 2009], which is often visualized as a tag cloud in the social tagging system [Helic et al., 2011, Sinclair and Cardew-Hall, 2008]. In Figure 2.2, an example of a tag cloud in the social bookmarking system BibSonomy (see Section 3.2.1) is illustrated. In this respect, [Ley and Seitlinger, 2015] showed that shared tag collections have an impact on learning with respect to collaborative knowledge building.

This is in line with the work of [Wagner et al., 2014], in which the semantic stability of social tagging systems is explored. Here, semantic stability is defined as the consensus reached on the tags used for the annotation of a resource among various users. The authors found that a combination of tag imitation processes and shared background knowledge of the bookmarked resource's topics leads to the fastest and highest semantic stability. Apart from that, related research showed that tags can be used to collaboratively classify Web content [Zubiaga et al., 2013] and to enhance the navigation in large knowledge repositories [Helic et al., 2012, Helic et al., 2010].

The tag recommendation algorithm BLL_{AC}+MP_r developed in course of this thesis builds on a combination of the personal and the collaborative tagging model. While the BLL_{AC} component reflects the personal tagging motivation, the MP_r component reflects the collaborative one.

2.1.2 Cognitive Processes in Social Tagging

Related literature has shown that an understanding of the cognitive processes involved in social tagging can help not only predicting the individual tagging behavior of users [Seitlinger et al., 2015b] but also modeling phenomena on the collective level, such as the emergence of stable tag distributions [Fu, 2008]. Specifically, when users categorize and tag resources on the Web (e.g., images), they draw on their semantic-lexical memory to retrieve corresponding memory units [Kowald, 2015]. For instance, a user might add the tag "banana" as the image shows a fruit she has recently eaten [Lindstaedt et al., 2009].

When looking closer at the underlying processes of social tagging, the way users choose tags for annotating resources strongly corresponds to processes in human memory and its cognitive structures [Cress et al., 2013, Held et al., 2012, Ley and Seitlinger, 2010]. As a prominent example in this respect, the work of [Fu, 2008] discusses an interplay between individual micro-level processes (e.g., associating the currently bookmarked resource with tags stored in memory) and collective macro-level processes (e.g., imitating other users' tags) in social tagging systems.

To that end, generative models of social tagging have been developed based on models of information theory [Halpin et al., 2007] and human memory theory [Fu et al., 2010] to provide insights into the emerged data in social tagging systems [Seitlinger et al., 2013]. These generative models implement assumptions about human information processing in order to derive computational models for predicting tag distributions. When comparing the predicted to the empirical tag distributions, claims about the validity of the underlying cognitive assumptions can be made [Seitlinger et al., 2013].

In order to provide a stricter test of these theoretical claims, controlled experiments can be conducted as these experiments allow for testing causal relationships more directly. Examples of such user studies have been conducted in [Cress et al., 2013] to test a social variant of information foraging theory, in [Fu et al., 2009] to test the semantic imitation model of social tagging, and in [Seitlinger and Ley, 2012] to find evidence that both semantic and lexical memory mechanisms play a role when users choose tags. The latter has also been investigated in [Seitlinger et al., 2013].

However, these studies also have limitations as they have to rely on controlled settings and specific study designs. Thus, to generalize the findings of these studies and to relate them to naturally occurring tag distributions, the models need to be tested in real-world folksonomy settings [Seitlinger et al., 2013]. Therefore, in this thesis, a tag recommendation algorithm, which implements basic mechanisms of human memory theory, is proposed and evaluated using large datasets gathered from real-world social tagging systems (see Chapter 3).

2.2 Tag-Based Recommender Systems

This section gives an overview over state-of-the-art tag-based recommender systems. Since this thesis focuses on tag recommendations in Research Questions 1 to 3, the

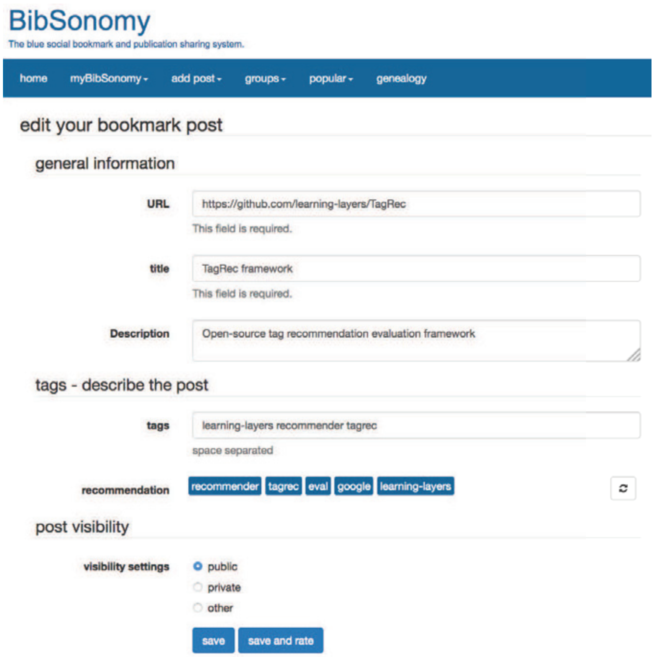


Figure 2.3: Example of a simple tag recommendation mechanism in BibSonomy (retrieved 07-July-2017). In the "recommendation" field, the tagging interface shows popular tags associated with the currently bookmarked resource and the current user to assist the user during the bookmarking process.

focus of this section is on tag recommendation algorithms. Apart from that, hashtag recommendation approaches are discussed as a prerequisite for Research Question 4. This section is based on P1 [Kowald, 2015], P5 [Trattner et al., 2016] (Section 2.2.1) and P10 [Kowald et al., 2017b].

2.2.1 Tag Recommendations

Tag recommendations aim at supporting users in providing meaningful tags for their bookmarked resources. It was shown that tag recommendations reduce the cognitive effort from generation to recognition [Gemmell et al., 2009]. Figure 2.3 illustrates this process, in which a user tags a Web link in BibSonomy and is supported via a list of tag recommendations.

To date, the following two approaches have been established: (i) content-based, and (ii) folksonomy-based tag recommendation algorithms. In relation to the topic of this thesis, a third strand of tag recommendation algorithms, namely cognitive-inspired approaches, is investigated as well [Trattner et al., 2016].

Content-Based Tag Recommendations

Content-based tag recommendation systems aim at analyzing the content of the target resource to identify tags that could be used to describe this resource. Among these systems, one of the most recognizable works is a study conducted by [Heymann et al., 2008]. The paper illustrates that page text is a significantly better predictor for the user's social tags than anchor texts or surrounding hosts of Web links. This was explored for tags gathered from the bookmarking system Delicious (see Section 3.2.1). The same effect has been validated in the work of [Lipczak et al., 2009, Lipczak and Milios, 2010, Lipczak and Milios, 2011, Lin et al., 2015]. Apart from that, tag recommendations based on visual content has been studied by [Lindstaedt et al., 2008, Lindstaedt et al., 2009].

Another relevant and recent research in this area has been contributed by [Lorince and Todd, 2013, Floeck et al., 2010, Moltedo et al., 2012], who show on a theoretical and empirical level that existing tags (e.g., tag clouds) influence the way people generate their own tags for a target resource.

However, while content-based tag recommendation algorithms provide powerful mechanisms for recommending tags, they rely on content data of the resources, which is often not included in publicly available social tagging datasets (see Section 3.2). Furthermore, [Rendle et al., 2009] proved that personalized folksonomy-based approaches outperform the theoretically best unpersonalized method, to which content-based algorithms typically belong. Therefore, this thesis focuses on folksonomy-based rather than content-based approaches [Kowald, 2015, Trattner et al., 2016].

Folksonomy-Based Tag Recommendations

In contrast to content-based tag recommendation approaches, folksonomy-based ones do not analyze the content of the bookmarked resource but rather use past tag assignments collected in the system. The probably most notable research in this context was presented by [Hotho et al., 2006] who introduced an algorithm called FolkRank (FR), which uses the structures of folksonomies for searching and ranking entities. These rankings can also be used to recommend tags.

Subsequent studies of [Marinho and Schmidt-Thieme, 2008] and [Hamouda and Wanas, 2011] show how the classic Collaborative Filtering (CF) approach can be adopted for the recommendation of tags. Significant studies of [Rendle and Schmidt-Thieme, 2010, Wetzker et al., 2010], [Krestel et al., 2009, Krestel and Fankhause, 2010] and [Rawashdeh et al., 2013, Pujari and Kanawati, 2012] introduce a factorization model, a Latent Dirichlet Allocation (LDA) model and a link prediction model, respectively, to recommend tags to users.

Although these approaches perform reasonably well, they are computationally expensive compared to simple "most popular tags" approaches [Jäschke et al., 2007, Jäschke et al., 2008]. Furthermore, they ignore recent observations with regard to social tagging systems, such as the variation of the individual tagging behavior over time [Yin et al., 2011b].

To that end, related research has made the first promising steps towards more accurate folksonomy-based algorithms that also account for the variable of time. For example, [Yin et al., 2011a, Zhang et al., 2012] have shown that these time-dependent approaches outperform other tag recommendation algorithms such as FolkRank (see Section 3.3). BLL_{AC}+MP_r introduced in this thesis also incorporates the variable of time using a power function.

Cognitive-Inspired Tag Recommendations

Cognitive-inspired tag recommendation systems utilize models of human cognition in order to model and predict the next tag assignments a user is going to apply. To date, there is only a small body of research available in the area of cognitive-inspired tag recommendations. One illustrative example is presented in [Stanley and Byrne, 2013]. In this work, the authors use the activation equation of the cognitive architecture ACT-R [Anderson et al., 2004] to predict the reuse of tags for posts in the Question & Answering site StackOverflow[1]. In contrast to BLL_{AC}+MP_r, which also relies on the activation equation of ACT-R to predict the reuse of tags, the

[1]http://stackoverflow.com/

approach of [Stanley and Byrne, 2013] works in an unpersonalized manner. Thus, tag suggestions are solely contextualized to the content of the current StackOverflow post. Furthermore, the temporal information is ignored in this approach.

In their follow-up work [Stanley and Byrne, 2016], the authors worked on these limitations and presented a personalized tag prediction model using the activation equation of ACT-R. The authors tested their approach not only on a dataset gathered from StackOverflow but also on a dataset gathered from the microblogging service Twitter[2].

However, in contrast to this thesis, the scope of [Stanley and Byrne, 2016] is a more psychological one rather than a technical one. Thus, it is the goal of their work to compare the predictions of an ACT-R-based model with a random permutation vector-based model. On the contrary, this thesis aims to contribute to the research area of tag-based recommender systems by improving the current state-of-the-art. Furthermore, BLL_{AC}+MP_r also incorporates tag imitation processes, which is not the case in [Stanley and Byrne, 2016].

Apart from this strand of research, the author of this thesis has also contributed to the design of other cognitive-inspired tag recommendation methods presented in [Seitlinger et al., 2013] and [Kowald et al., 2015b]. These methods are based on a computational model of human categorization called MINERVA2 [Hintzman, 1984] in order to process a network constituted by a input, hidden and output layer. In [Seitlinger et al., 2013], the 3Layers (3L) algorithm was presented, which uses categories assigned to the current resource in order to recommend tags of semantically similar resources. This approach is extended in [Kowald et al., 2015b] to create 3LT, which enriches 3L in order to also incorporate temporal processes of tag usage (see Section 3.3).

However, while 3L and 3LT provide good results in terms of recommender accuracy and ranking, they rely on category information of the resources, which are not available in most public social tagging datasets. Therefore, in [Seitlinger et al., 2013, Kowald et al., 2015b] it was shown that LDA topics derived from the tags in the datasets can also be used as resource categories. Although the presented evaluation results proved this claim, the LDA topic creation process is computationally expensive. This could lead to problems when integrating this approach into a live tag recommendation setting, in which runtime and memory consumption are crucial

[2] `https://twitter.com/`

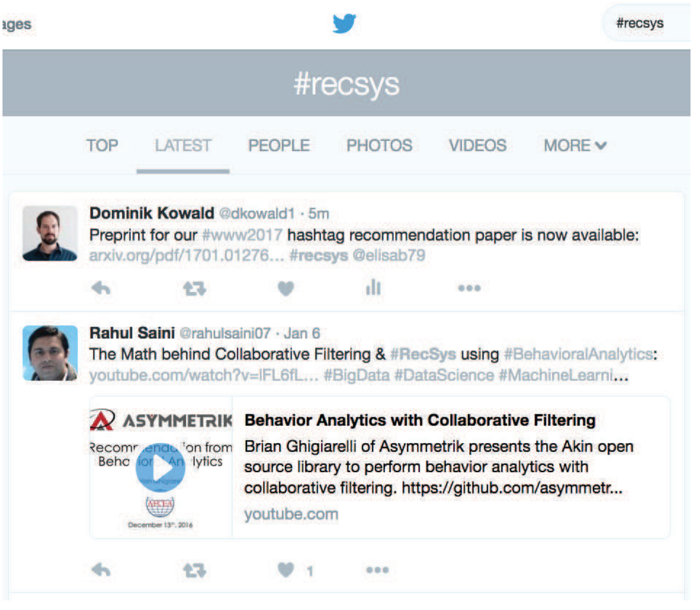


Figure 2.4: Example of hashtags in Twitter (retrieved 07-July-2017). The interface allows for searching tweets based on a given hashtag, in this example "#recsys".

factors [Kowald and Lex, 2015].

2.2.2 Hashtag Recommendations

Hashtags are freely-chosen keywords starting with the hash character "#" to annotate, categorize and contextualize Twitter posts, which are also known as tweets [Kowald et al., 2017b]. In contrast to social tags, which are mainly used to index resources for later retrieval, hashtags have a more conversational nature and are used to filter and direct content to certain streams of information [Huang et al., 2010, Romero et al., 2011]. This process is illustrated in Figure 2.4, in which tweets are searched based on the hashtag "#recsys" to receive tweets related to the research area of recommender systems.

There is already a large body of research available that focuses on the recommendation of hashtags in Twitter. One illustrative example is the work presented in [Godin et al., 2013], in which hashtag recommendations are provided by categorizing tweets into general topics using LDA-based topic modeling. The approach then recommends the hashtags that best fit the topics of a new tweet. The authors evaluate their approach using a qualitative study, in which they ask persons if the recommended hashtags describe the topics of a tweet and could be used to semantically enrich it. In 80% of the cases, they provide a suitable hashtag from a selection of five possibilities. Other similar approaches using topic models are presented in [She and Chen, 2014, Wang et al., 2014, Xu et al., 2015, Efron, 2010]. Moreover, a related algorithm based on a hashtag classification scheme is proposed in [Jeon et al., 2014].

The most notable work in the context of hashtag recommendations is probably the content-based SimRank approach presented in [Zangerle et al., 2011] and [Zangerle et al., 2013]. The authors use a content-based similarity statistic to calculate similarities between tweets and identify suitable hashtags based on these similarity scores. They show that SimRank improves prediction accuracy by around 35% compared to a popularity-based approach. In [Kywe et al., 2012], a personalized extension of SR is presented, in which the authors combine it with user-based Collaborative Filtering. Apart from that, a content-based hashtag recommendation algorithm for hyper-linked tweets is proposed in [Sedhai and Sun, 2014].

Related research has studied temporal effects on hashtag usage, for instance in the context of popular hashtags in Twitter [Lin and Mishne, 2012, Lehmann et al., 2012, Tsur and Rappoport, 2012, Ma et al., 2012]. For example, in [Ma et al., 2012], the authors aim to predict if a specific hashtag will be popular on the next day. By formulating this task as a classification problem, they find that both content features (e.g., the topic of the hashtag) and context features (e.g., the users who used the hashtags) are effective features for popularity prediction. A similar approach is presented in [Yang and Leskovec, 2011], in which the authors uncover the temporal dynamics of online content (e.g., tweets) by formulating a time series clustering problem. One of the very few examples of a time-aware hashtag recommendation approach is the recently proposed algorithm described in [Harvey and Crestani, 2015].

The authors extend the content-based SimRank approach [Zangerle et al., 2011] with a personalization technique by means of Collaborative Filtering and further consider the temporal relevance of hashtags. To account for this temporal relevance, they divide the hashtags into two categories: "organizational" ones, which are used over a long period of time and "conversational" ones, which are used only during a short time span (e.g., for a specific event).

In Chapter 7 of this thesis, a cognitive-inspired hashtag recommendation algorithm is presented, which builds upon BLL_{AC}+MP_r and thus, incorporates the activation equation of the cognitive architecture ACT-R.

2.3 Summary

In this chapter, the research related to this thesis was presented. This included research on social tagging and tag recommendations. With respect to social tagging in general, the motivation and the underlying cognitive processes behind social tagging were discussed. Here, two types of tagging models can be distinguished: (i) the personal tagging model, and (ii) the collaborative tagging model. Apart from that, research has shown that the way users choose tags for annotating resources strongly corresponds to processes in human memory and its cognitive structures, which underpins the relevance of this thesis.

With respect to tag recommendations, three types of current state-of-the-art algorithms exist: (i) content-based approaches, (ii) folksonomy-based approaches, and (iii) cognitive-inspired approaches. The algorithm proposed in this thesis contributes to the small body of research conducted in the area of cognitive-inspired approaches. Moreover, there is currently a lot of interest in hashtag recommendation algorithms, which is investigated in detail in Chapter 7 of this thesis.

In the next chapter, insights from related work will be used to propose a methodology, which is used to tackle the four research questions of this thesis.

Chapter 3

Methodology

"Effective and meaningful evaluation of recommender systems is challenging." [Herlocker et al., 2004]

The aim of this chapter is to present the methodology that is used to answer the research questions of this thesis. Therefore, common practice in research on recommender systems is followed to build on an offline evaluation study design. This means that publicly available data collections (see Section 3.2) are used in order to study the relation between activation processes in human memory and tagging behavior of users on a large scale (Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*). Additionally, these data collections are used to conduct an offline tag recommender study (see Section 3.4) to address Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"*, and Research Question 3, *"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"*. For addressing Research Question 4, *"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"*, data collections from Twitter are used as described in Chapter 7.

Parts of this chapter have been published in *P2* [Kowald and Lex, 2016] (Section 3.2), *P5* [Trattner et al., 2016] (Sections 3.3, 3.2 and 3.4.2), *P8* [Kowald and Lex, 2015] (Sections 3.2, 3.4.1 and 3.4.2) and *P6* [Kowald et al., 2014a] (Section 3.5).

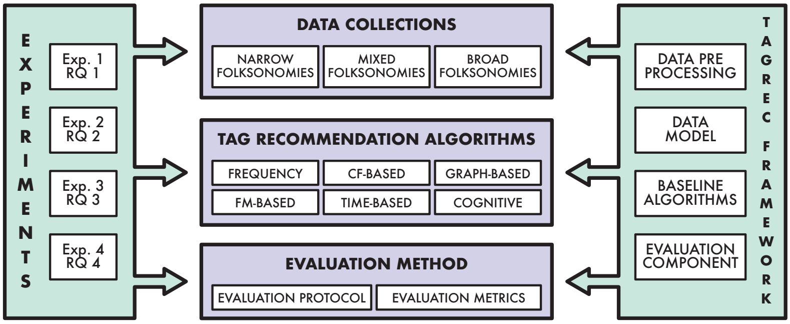


Figure 3.1: Schematic illustration of the methodology used in this thesis to conduct the four proposed experiments. Therefore, the *TagRec* framework supports various data collections, tag recommendation algorithms and a consistent evaluation method.

3.1 Experiments

In order to address the four research questions of this thesis, four experiments were conducted:

- *Experiment 1:* Analyzing the influence of activation processes in human memory on tag reuse (see Chapter 4).
- *Experiment 2:* Designing and evaluating a cognitive-inspired algorithm for tag reuse prediction (see Chapter 5).
- *Experiment 3:* Implementing and evaluating a hybrid approach for tag recommendations in real-world folksonomies (see Chapter 6).
- *Experiment 4:* Utilizing and evaluating the approach for recommending hashtags in Twitter (see Chapter 7).

These experiments are based on a common methodology, which is visualized in Figure 3.1. As shown, this methodology consists of three main parts: (i) data collections (see Section 3.2), (ii) tag recommendation algorithms (see Section 3.3), and (iii) a consistent evaluation method (see Section 3.4). All three parts are supported by the *TagRec* tag recommendation evaluation framework (see Section 3.5). Since *TagRec* is open-source software, this facilitates the reproducibility of the experiments.

Please note that methodological aspects, which specifically address *Experiment 4*, are described in Chapter 7.

3.2 Data Collections

In this section, the data collections analyzed in this thesis are described. This includes descriptions of the social tagging systems from which the datasets have been gathered, the applied preprocessing steps and the final dataset statistics.

This section is based on *P2* [Kowald and Lex, 2016], *P5* [Trattner et al., 2016] and *P8* [Kowald and Lex, 2015]. Please note that the Twitter data collections are described in Section 7.2.

3.2.1 Social Tagging Systems

For the purpose of this thesis and to foster reproducibility, investigations focused on the six well-known social tagging systems Flickr, CiteULike, BibSonomy, Delicious, LastFM and MovieLens. The publicly available datasets gathered from these systems have also been used in many of the related works in tag-based recommender systems and can be seen as the state-of-the-art benchmarking datasets [Jäschke et al., 2007, Doerfel and Jäschke, 2013].

Furthermore, these systems do not only differ in terms of their domain type (i.e., images, URLs, citations, music and movies) and size but also in terms of their folksonomy type: (i) narrow, (ii) mixed, and (iii) broad (see [Kowald and Lex, 2016]). In a broad folksonomy, typically, many users annotate a particular resource, whereas in a narrow folksonomy only the user who has uploaded the resource is permitted to apply tags to it [Helic et al., 2012]. A mixed folksonomy is a folksonomy that cannot be strictly assigned to the narrow or broad case (i.e., typically only a few users annotate a particular resource).

With respect to tag recommendation methods already utilized in these systems, most of them only use very simple methods based on tag popularity (i.e., recommend the most frequently used tags of the target user and / or target resource). The only system, which uses a more sophisticated method is BibSonomy, in which the FolkRank algorithm [Hotho et al., 2006] is integrated, which ranks tags based on the well-known PageRank statistic adapted to folksonomies. Thus, an unbiased evaluation of this thesis' approaches is ensured since neither time-based nor cognitive-inspired tag recommendation algorithms are already implemented in these social tagging systems.

Narrow Folksonomies

In a narrow folksonomy, only the user who has uploaded the resource is permitted to apply tags to it. In this thesis, the narrow folksonomy Flickr is analyzed.

Flickr. Flickr[1] is an image hosting and sharing platform, which also offers online community elements. The Flickr dataset used in this thesis was crawled and provided by the University of Koblenz[2] within the Tagora EU project[3]. The dump from 2010-January-07 contains 28,153,045 bookmarks, 319,686 users, 28,153,045 resources, 1,607,879 tags, and 112,900,000 tag assignments.

Mixed Folksonomies

Mixed folksonomies are networks that cannot be strictly assigned to the narrow or broad case. In this thesis, datasets from CiteULike, BibSonomy and Delicious are used for evaluation.

CiteULike. CiteULike[4] is a scientific reference management system, which gives free access to their data to researchers for non-commercial uses[5]. The whole data dump from 2015-February-03 consists of 4,658,570 bookmarks, 92,374 users, 3,589,546 resources, 899,056 tags and 19,062,426 tag assignments.

BibSonomy. The dataset of the social bookmark and publication sharing system BibSonomy[6] is freely available and can be downloaded for scientific purposes[7]. From the 2015-January-01 dump, we utilized all tags assigned to bookmarks and references (i.e., bibtex), which resulted in 772,112 bookmarks, 10,180 users, 683,482 resources, 199,594 tags and 2,981,038 tag assignments.

[1] `http://www.flickr.com/`
[2] `https://www.uni-koblenz.de/FB4/Institutes/IFI/AGStaab/Research/DataSets/PINTSExperimentsDataSets/`
[3] `http://www.tagora-project.eu/`
[4] `http://www.citeulike.org/`
[5] `http://www.citeulike.org/faq/data.adp`
[6] `http://www.bibsonomy.org/`
[7] `http://www.kde.cs.uni-kassel.de/bibsonomy/dumps`

Delicious. The dataset of the social bookmarking Web service Delicious[8] is freely available for scientific purposes via the University of Koblenz (see Flickr dataset). The dump from 2010-January-07 contains 47,208,747 bookmarks, 532,924 users, 17,262,480 resources, 2,481,698 tags and 140,126,586 tag assignments.

Broad Folksonomies

In a broad folksonomy, typically many users annotate a particular resource. Two examples of such folksonomies are LastFM and MovieLens.

LastFM. LastFM [9] is a social Web portal for browsing, annotating and discovering music. The social tagging dataset of LastFM[10] was published by GroupLens Research[11] in the course of 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011)[12]. The dump from 2011-May-12 consists of 71,062 bookmarks, 1,892 users, 12,522 resources, 9,748 tags and 186,474 tag assignments. It has to be noted that this is not a complete dataset of LastFM but a crawl of 1,892 users and its metadata in the system.

MovieLens. MovieLens is a movie recommender system developed by GroupLens Research (see LastFM). The 10m movie-rating dataset of MovieLens has become one of the most utilized datasets for research on recommender systems. For this thesis, a subset of this dataset was used, in which only users with tagging data are included[13]. The tagging data dump from 2009-January-05 contains 55,484 bookmarks, 4,009 users, 7,601 resources, 15,237 tags and 95,580 tag assignments.

3.2.2 Dataset Preprocessing

In the course of this thesis, three methods for dataset preprocessing were taken into account: (i) sampling, (ii) *p*-core pruning, and (iii) tag cleaning.

[8] `http://www.delicious.com/`
[9] `http://www.last.fm/`
[10] `http://files.grouplens.org/datasets/hetrec2011/hetrec2011-lastfm-2k.zip`
[11] `http://grouplens.org/`
[12] `http://ir.ii.uam.es/hetrec2011/`
[13] `http://files.grouplens.org/datasets/movielens/ml-10m.zip`

Sampling

To reduce computational effort, a dataset sampling technique proposed by [Gemmell et al., 2009] has been applied to the very big Flickr, CiteULike and Delicious datasets. Thus, for Flickr and Delicious, 3% of the user profiles (i.e., all the bookmarks of these users) and for CiteULike, 15% of the user profiles were randomly selected. According to [Gemmell et al., 2009], when following this pruning method, experiments on larger dataset samples (or even the whole datasets) provide nearly identical trends in the algorithmic results. A similar behavior was shown in [Trattner et al., 2016], where multiple samples were drawn from the datasets that all provided almost the same recommender evaluation results.

p-core Pruning

In order to avoid a biased evaluation and to simulate a real-world folksonomy setting, no p-core pruning methods have been applied to the datasets as suggested by [Doerfel and Jäschke, 2013, Doerfel et al., 2016]. This p-core pruning is an iterative process, where in each iteration all resources, tags and users are deleted that occur less than p times in a dataset. This algorithm terminates when no more tag assignments can be deleted, which ensures that all resources, tags and users can be found at least p times in the remaining core [Batagelj and Zaveršnik, 2002]. Based on this definition, it becomes clear that even a small p-core of 2 would delete a lot of bookmarks and thus, substantially distort the underlying distribution of the datasets. Simulating a real-world folksonomy setting is especially important for the development of live recommender services [Kowald and Lex, 2015].

Tag Cleaning

Since automatically generated tags affect the performance of the tag recommender algorithms, all of these tags were removed from the datasets (e.g., *no-tag*, *import*, etc.). Apart from that, all tags have been decapitalized. To follow common practice in tag recommender research, more sophisticated tag cleaning methods, such as stemming, have not been applied (see [Krestel and Fankhause, 2010]).

Folksonomy	Dataset	$\|U\|$	$\|R\|$	$\|T\|$	$\|Y\|$	$\|B\|$	$\|B\|/\|U\|$	$\|B\|/\|R\|$
Narrow	Flickr	9,590	856,755	125,119	3,328,590	856,755	89.338	1.000
Mixed	CiteULike	18,474	811,175	273,883	3,446,650	900,794	48.760	1.110
	BibSonomy	10,179	683,478	201,254	2,986,396	772,108	75.853	1.129
	Delicious	15,980	963,741	184,012	4,266,206	1,447,267	90.567	1.501
Broad	LastFM	1,892	12,522	9,748	186,474	71,062	37.559	5.674
	MovieLens	4,009	7,601	15,238	95,580	55,484	13.839	7.299

Table 3.1: Summary of the real-world folksonomy datasets used in this thesis. Here, $|U|$ is the number of users, $|R|$ is the number of resources, $|T|$ is the number of tags, $|Y|$ is the number of tag assignments and $|B|$ is the number of bookmarks/posts in the datasets. The datasets are sorted from the narrowest one to the broadest one, where the degree of narrowness is given by the average number of bookmarks assigned to a resource (i.e., $|B|$ / $|R|$).

3.2.3 Dataset Statistics

The final dataset statistics after all preprocessing steps are shown in Table 3.1. Based on the number of bookmarks, the biggest dataset is Delicious, whereas the smallest one is MovieLens. The same is true for the average number of bookmarks per user, where the highest number (around 91) can be found in Delicious and the smallest one (around 14) in MovieLens.

Interestingly, the highest narrowness degree (i.e., the average number of bookmarks per resource, $|B|/|R|$) is available in MovieLens. Based on this narrowness degree, the datasets can also be categorized into the three folksonomy types (i.e., narrow, mixed and broad) discussed in this thesis.

3.3 Tag Recommendation Algorithms

A rich set of 20 folksonomy-based tag recommendation algorithms were chosen, implemented and evaluated in the course of this thesis. The algorithms were selected based on their popularity in the community, performance and novelty (see also [Marinho et al., 2011, Balby Marinho et al., 2012]). A summary of the algorithms and their relation to the factors of interest of this thesis (i.e., frequency, recency, semantic context and social influences, see Chapter 4) is shown in Table 3.2. Please note that AC$_u$, BLL, BLL$_{AC}$ and BLL$_{AC}$+MP$_r$, which have been developed in the course of this thesis, are not described in this section but in Chapters 5 and 6.

Algorithm	Frequency	Recency	Semantic Context	Social Influence
MP				x
MP_u	x			
MP_r				x
$MP_{u,r}$	x			x
CF			x	x
APR	x		x	x
FR	x		x	x
LDA			x	x
FM	x		x	x
PITF	x		x	x
MR_u		x		
GIRP	x	x		
GIRPTM	x	x		x
3L	x		x	
3LT	x	x	x	
3LT+MP_r	x	x	x	x
$\mathbf{AC}_u$			x	
BLL	x	x		
$\mathbf{BLL}_{AC}$	x	x	x	
$\mathbf{BLL}_{AC}$+$\mathbf{MP}_r$	x	x	x	x

Table 3.2: Overview of the 20 algorithms evaluated in this thesis. The algorithms are assigned to the factors of tag usage frequency, recency, semantic context and social influences. Please note that novel algorithms that have been developed in the course of this thesis are visualized in bold.

Additionally, Table 3.3 at the end of this section presents an overview of the hyperparameters of the algorithms as used in the experiments. This section is based on *P5* [Trattner et al., 2016]. Please also note, that the algorithms for hashtag recommendations are described in Section 7.5.1.

3.3.1 Frequency-Based Algorithms

Frequency-based tag recommender algorithms are solely based on tag popularity. These approaches are highly computationally efficient but neglect other factors such as temporal effects.

MostPopular (MP)

This approach recommends for any user $u \in U$ and any resource $r \in R$ the same set of tags $\widetilde{T}(u, r)$. This set of tags is weighted by the frequency in all tag assignments Y [Jäschke et al., 2008]:

$$\widetilde{T}_k(u, r) = \arg\max_{t \in T}^{k}(|Y_t|) \tag{3.1}$$

MP is a completely unpersonalized method but has the advantage that it is able to provide tag suggestions for cold-start users and resources (i.e., users or resources without any tagging data available).

MostPopular$_u$ (MP$_u$)

The *most popular tags by user* approach suggests the most frequent tags in the tag assignments of the user Y_u [Jäschke et al., 2008]:

$$\widetilde{T}_k(u, r) = \arg\max_{t \in T_u}^{k}(|Y_{t,u}|) \tag{3.2}$$

MP$_u$ typically provides good results in narrow folksonomy settings, where the tagging process is solely influenced by the individual behavior.

MostPopular$_r$ (MP$_r$)

The *most popular tags by resource* algorithm is the resource-based equivalent to MP$_u$ and weights the tags based on their frequency in the tag assignments of the resource Y_r [Jäschke et al., 2008]:

$$\widetilde{T}_k(u, r) = \arg\max_{t \in T_r}^{k}(|Y_{t,r}|) \tag{3.3}$$

Since MP$_r$ incorporates the tags already assigned to the target resource r, it provides good results in broad folksonomy settings, where a lot of bookmarks are available for each resource.

MostPopular$_{u,r}$ (MP$_{u,r}$)

This algorithm is a mixture of the most popular tags by user and resource approaches:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u \cup T_r}{\arg\overset{k}{\max}}(\beta|Y_{t,u}| + (1-\beta)|Y_{t,r}|) \tag{3.4}$$

The β parameter can be used to balance the influence of the user and the resource components [Jäschke et al., 2008]. However, for the experiments of this thesis, it was set to .5 to give equal importance to both components.

3.3.2 Collaborative Filtering

Collaborative Filtering (CF) is one of the most frequently used algorithms in the area of recommender systems [Schafer et al., 2007]. There are two variants of it, user-based CF, which is based on user similarities, and resource-based CF, which is based on resource similarities. In the course of this thesis, both variants were implemented and evaluated but user-based CF provided better results in almost all settings. Thus, only results for user-based CF are reported in this thesis.

User-Based Collaborative Filtering (CF)

User-based CF algorithms aim to find similar users for the target user u and recommend items of these so-called neighbors. [Marinho and Schmidt-Thieme, 2008] described how the classic Collaborative Filtering (CF) approach can be used for tag recommendations. Since folksonomies have ternary relations (i.e., users, resources and tags), the classic CF approach cannot be applied directly. Thus, the neighborhood N_u of a user u is formed based on the tag assignments in the user profile Y_u. Furthermore, in CF-based tag recommendations only the subset V_r of users that have tagged the active resource r are taken into account when calculating the user neighborhood. The set of k recommended tags can then be determined based on the tags used by the users in this neighborhood [Marinho and Schmidt-Thieme, 2008, Jäschke et al., 2007]:

$$\widetilde{T}_k(u,r) = \underset{t \in T_{N_u}}{\arg\overset{k}{\max}}(\sum_{v \in N_u} sim(Y_u, Y_v) \cdot \delta(v,r,t)) \tag{3.5}$$

where $\delta(v, r, t) = 1$ if $(v, r, t) \in Y$ and 0 otherwise. The only variable parameter here is the number of users in the neighborhood, which has to be set in advance. For the experiments, a neighborhood size $|N_u|$ of 20 was used as suggested in related work [Gemmell et al., 2009]. Other values for $|N_u|$ were tested but CF has not generated significant higher values of accuracy when setting $|N_u| > 20$.

There are different ways to calculate the similarity $sim(Y_u, Y_v)$ between two users u and v. For the experiments, the simple Jaccard similarity coefficient was used. The Okapi BM25 similarity measure [Parra and Brusilovsky, 2009, Parra-Santander and Brusilovsky, 2010, Xu et al., 2008] was tested as well but this one reached almost the same results as Jaccard with a significantly higher computational effort.

Resource-Based Collaborative Filtering (CF$_r$)

CF$_r$ is very similar to user-based CF but, in contrast to it, is based on resource similarities. The set of most similar resources S_r for resource r is calculated based on the resources that have been tagged by user u. More formally, this is given by:

$$\widetilde{T}_k(u, r) = \underset{t \in T_{S_r}}{\arg\overset{k}{\max}}\left(\sum_{s \in S_r} sim(Y_r, Y_s) \cdot \delta(u, s, t)\right) \tag{3.6}$$

where $\delta(u, s, t) = 1$ if $(u, s, t) \in Y$ and 0 otherwise. As in the case of user-based CF, the neighborhood size $|S_r|$ is set to 20 and the similarity between resources is calculated using the Jaccard coefficient.

3.3.3 Graph-Based Algorithms

Graph-based tag recommender approaches incorporate the graph structure of the folksonomy to calculate tag recommendations. In the course of this thesis, two graph-based methods have been developed and evaluated, (i) Adapted PageRank and (ii) FolkRank. Since FolkRank provided better results than Adapted PageRank in the settings evaluated, only the results of FolkRank are reported.

Adapted PageRank (APR)

[Hotho et al., 2006] adapted the well-known PageRank algorithm [Page et al., 1999] in order to rank the nodes within the graph structure of a folksonomy. This is based on the idea that a resource is important if it is tagged with important tags by important users. Therefore, the folksonomy has to be converted into an undirected graph where the set of nodes s is the disjoint union of all users U, resources R and tags T: $s = U \cup R \cup T$. The co-occurrences of users and resources, users and tags, and resources and tags are treated as weighted edges in this graph and can also be represented as an adjacency matrix A. The update of the weightings is done using the following formula where $\vec{p}$ is a preference vector and d is a variable to set its impact [Hotho et al., 2006]:

$$\vec{w} \leftarrow dA\vec{w} + (1 - d)\vec{p} \tag{3.7}$$

For recommending tags, the preference vector $\vec{p}$ is used to give higher weights to the target user and resource of the recommendation task. While all other users and resources get a weight of 1, they get a weight of $1 + |U|$ and $1 + |R|$ [Jäschke et al., 2007].

FolkRank (FR)

The FolkRank algorithm is an extension of the APR approach that was also proposed by [Hotho et al., 2006]. This extension gives a higher importance to the preference vector $\vec{p}$ using a differential approach, where $\vec{w}^{(0)}$ is the weighting vector calculated using the APR algorithm with $\vec{p} = 1$ and $\vec{w}^{(1)}$ is the result with a $\vec{p}$-setting as described above. Taken togehter, $\vec{w}$ is given by:

$$\vec{w} = \vec{w}^{(1)} - \vec{w}^{(0)} \tag{3.8}$$

The FR version used in this thesis is based on an open-source Java implementation provided by the University of Kassel[14]. In this implementation, the preference vector weight d is set to .7 and the maximum number of iterations l is set to 10 [Jäschke et al., 2007].

3.3.4 Factorization Models

Two algorithms based on factorization models were implemented and evaluated in the course of this thesis: First, Latent Dirichlet Allocation (LDA) and second,

[14] `http://www.kde.cs.uni-kassel.de/code`

Pairwise Interaction Tensor Factorization (PITF). Since PITF has established itself as the leading method for tag recommendations, it is also one of the most important baselines for BLL_{AC}+MP_r. Factorization Machines (FM) have also been evaluated in the course of this thesis but PITF always provided better results. Thus, only results for LDA and PITF are reported.

Latent Dirichlet Allocation (LDA)

LDA is a probability model that helps to find latent topics for documents where each topic is described by words in these documents [Krestel et al., 2009]. This can be formalized as follows:

$$P(t_i|d) = \sum_{j=1}^{Z} \left(P(t_i|z_i = j) \cdot P(z_i = j|d) \right) \tag{3.9}$$

Here $P(t_i|d)$ is the probability of the ith word for a document d and $P(t_i|z_i = j)$ is the probability of t_i within the topic z_i. $P(z_i = j|d)$ is the probability of using a word from topic z_i in the document. The number of latent topics Z is determined in advance and defines the level of granularity. The LDA results for this thesis were calculated for $Z = 1000$ topics.

When using LDA for tag recommendations, documents are either users or resources described by tags. This means that based on the tag vectors of the users and resources, these entities can also be represented with the topics identified by LDA. The LDA-based tag recommender suggests then the top-k tags associated with these topics. LDA was implemented using the Java framework Mallet[15] with Gibbs sampling and 2000 iterations as suggested in the framework documentation and by related work (e.g., [Krestel et al., 2009]). Moreover, only topics with a minimum probability value of .001 were considered in order to reduce noise.

Pairwise Interaction Tensor Factorization (PITF)

This approach proposed by [Rendle and Schmidt-Thieme, 2010] is an extension of Factorization Machines (FM) [Rendle, 2010] that explicitly models the pairwise interactions between users, resources and tags. PITF determines a prediction score

[15] `http://mallet.cs.umass.edu/topics.php`

$s(u, r, t)$ based on factorizations of these relationships:

$$s(u, r, t) = \sum_{f}^{k_U} \hat{u}_{u,f}^{T} \cdot \hat{t}_{t,f}^{U} + \sum_{f}^{k_R} \hat{r}_{r,f}^{T} \cdot \hat{t}_{t,f}^{R} + \sum_{f}^{k_T} \hat{u}_{u,f}^{R} \cdot \hat{r}_{r,f}^{U} \tag{3.10}$$

where k_U, k_R and k_T are the dimensions of factorization. The PITF results presented in this thesis were calculated using the open-source C++ tag recommender framework provided by the University of Konstanz[16]. The dimensions of factorization k_U, k_R and k_T were set to 256, the learning rate α was set to .01, the regularization constant λ was set to .0 and the number of iterations l was set to 50 as suggested by the framework documentation.

Additional experiments with factors of 64, 128 and 512, and with more and less than 50 iterations were conducted but across all datasets, the setting of 256 factors and 50 iterations showed almost always the best results. More specifically, a number of factors less than 256 decreased the results significantly whereas a number of factors higher than 256 did not result in any higher estimates while varying the number of iterations. The same is true for α and λ.

3.3.5 Time-Based Algorithms

With respect to time-based algorithms, the GIRPTM approach proposed by [Zhang et al., 2012] is the most prominent one in the field of tag recommendations. Furthermore, this one is especially of interest for this thesis since in contrast to BLL_{AC}+MP_r, GIRPTM models the time component using an exponential function rather than a power function. Thus, besides PITF, GIRPTM is the most important baseline of this thesis. Additionally, a simple "most recent tags" approach was utilized.

MostRecent$_u$ (MR$_u$)

MR$_u$ recommends the k most recently used tags of user u. It is formally given by [Campos et al., 2014]:

$$\widetilde{T}_k(u, r) = \underset{t \in T_u}{\arg\overset{k}{\max}}(time(Y_{t,u})) \tag{3.11}$$

[16] `http://www.informatik.uni-konstanz.de/rendle/software/tag-recommender/`

where $time(Y_{t,u})$ is the timestamp of the most recent assignment of tag t by user u.

Temporal Tag Usage Patterns (GIRP)

This time-dependent tag-recommender algorithm was presented by [Zhang et al., 2012] and is based on the frequency and the temporal usage of a user's tag assignments. In contrast to the BLL and BLL$_{AC}$ approaches presented in this thesis, GIRP models the temporal tag usage with an exponential function rather than a power function, which can be formalized as follows:

$$s(u,r,t) = d^f_{u,t,p} \cdot (d^l_{u,t,p})^{-d^f_{u,t,p}} \cdot \frac{Y_{t,u}}{\sum\limits_{t' \in Y_u} Y_{t',u}} \tag{3.12}$$

where $d^f_{u,t,p}$ denotes the distance from the current bookmark index p to the first bookmark index f (i.e., 1) and $d^f_{u,t,p}$ denotes the distance from p to the last bookmark index l (i.e., p - 1) for user u and tag t. This term is multiplied with the relative tag frequency of t used by u to account for the factor of tag frequency (i.e., MP$_u$).

GIRP with Tag Relevance to Resource (GIRPTM)

This is an extension of the GIRP algorithm [Zhang et al., 2012] using the relative tag frequency of t for the target resource r (i.e., MP$_r$). This is achieved using a linear combination in the same way as also done in the BLL$_{AC}$+MP$_r$ approach of this thesis:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u \cup T_r}{\arg\overset{k}{\max}}(\beta \cdot s(u,r,t) + (1-\beta)\frac{Y_{t,r}}{\sum\limits_{t' \in Y_r} Y_{t',r}}) \tag{3.13}$$

where β was set to .5 to give equal weights to both components.

3.3.6 Cognitive-Inspired Algorithms

Three cognitive-inspired algorithms have been utilized for the evaluation of this thesis. In contrast to BLL$_{AC}$+MP$_r$, these methods incorporate another type of data, namely category information by means of LDA topics. In this thesis, only the results of the best performing approach (i.e., 3LT+MP$_r$) are reported.

The main disadvantage of these three algorithms compared to BLL_{AC}+MP_r is that they rely on the computationally expensive calculation of LDA topics since social tagging datasets typically do not provide category information.

3Layers (3L)

The 3L tag recommender algorithm is based on a mechanism from MINERVA2, a computational theory of human categorization [Hintzman, 1984], to process a network constituted by a input, hidden and output layer. In contrast to other approaches, 3L incorporates not only tagging data but also category information in form of latent topics (i.e., calculated by means of LDA with $Z = 1000$ topics). The prediction score $s(u, r, t)$ for 3L is given by [Seitlinger et al., 2013]:

$$s(u, r, t) = \sum_{i=1}^{l} (L_{i,t} \cdot A_i) \tag{3.14}$$

where $L_{i,t}$ is an entry in the lexical tag matrix, which indicates if tag t has been used by user u in bookmark i. A_i is an activation value, which reflects the Cosine-similarity between the semantic topic vector of resource r and the resource tagged in bookmark i.

Time-Based 3Layers (3LT)

3LT is a time-dependent extension of 3L. Similar to BLL_{AC}, the time component is calculated using the BLL equation of [Anderson and Schooler, 1991]. Formally, 3LT is given by [Kowald et al., 2015b]:

$$s(u, r, t) = \sum_{i=1}^{l} (L_{i,t} \cdot BLL(t) \cdot A_i) \tag{3.15}$$

where $BLL(t)$ is the BLL value of tag t for user u.

Algorithms	Parameters	Value
CF	$\|N_u\|$	20
FR	d	.7
FR	l	10
LDA, 3LT+MP$_r$	Z	1000
PITF	k_U, k_R, k_T	256
PITF	l	50
PITF	α	.01
PITF	λ	.0
MP$_{u,r}$, GIRPTM, 3LT+MP$_r$, **BLL$_{AC}$+MP$_r$**	β	.5
3LT+MP$_r$, **BLL, BLL$_{AC}$, BLL$_{AC}$+MP$_r$**	d	.5

Table 3.3: Hyperparameter settings of the algorithms as used in the experiments. These parameters were chosen based on values available in the literature, own experimentation and framework descriptions and were used across all six datasets to ensure generalizable results, which should also hold across a number of other datasets.

Time-Based 3Layers with MP$_r$ (3LT+MP$_r$)

3LT+MP$_r$ combines 3LT with MP$_r$ using the relative tag frequency of t for r [Kowald et al., 2015b]:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u \cup T_r}{\arg\max^k}(\beta s(u,r,t) + (1-\beta)\frac{Y_{t,r}}{\sum\limits_{t' \in Y_r} Y_{t',r})} \tag{3.16}$$

where β was set to .5 to give equal weights to both components.

3.3.7 Hyperparameters

The hyperparameters of the algorithms, as used in the experiments, are summarized in Table 3.3. The hyperparameters were chosen based on values available in the literature and the experiments conducted in the course of this thesis. The parameters were not optimized towards the given datasets although the author understands that this is a necessary precondition for arriving at an optimal performance of each algorithm.

However, the aim of this thesis is not so much in optimizing the algorithmic performance but to arrive at generalizable, stable and traceable conclusions that hold across a number of datasets. The author would be concerned that by optimizing parameters for concrete dataset characteristics, generalizability, stability and traceability would be compromised.

3.4 Evaluation Method

In this section, the evaluation method for conducting the tag recommender study of this thesis is described. This includes the evaluation protocol (Section 3.4.1) and metrics (Section 3.4.2), which are used to conduct the four experiments of this thesis. This section is based on *P5* [Trattner et al., 2016] and *P8* [Kowald and Lex, 2015].

3.4.1 Evaluation Protocol

In order to evaluate the tag recommender approaches, a leave-one-out method as proposed by related work (e.g., [Jäschke et al., 2007]) is used. Thus, to split the datasets into training (i.e., B_{train}) and test sets (i.e., B_{test}), each user's most recent bookmark in time is determined and removed from the original dataset. The now reduced version of the original dataset is used for training, and the newly created one for testing the algorithms. Each bookmark in the test set consists of a collection of one or more tags applied by a target user to a target resource, which are further referred as relevant tags. In order to make sure that at least one bookmark per user is available for training, only users with more than one bookmark are considered for the test set. Thus, users with less than two bookmarks (i.e., non-cold-start users) are only represented in the training set.

This protocol is a plausible simulation of a real-world environment since it uses the most recent bookmark of each user for testing instead of a random one. Thus, the protocol retains the chronological order of a user's bookmarks, which makes it a suggested offline evaluation procedure for time-based recommender systems [Campos et al., 2014, Song et al., 2008]. Additionally, the leave-one-out strategy ensures that each user with more than one bookmark is represented in the test set, which would not be the case for other methods such as 80/20 splits (i.e., using randomly 80% of the data for training and the rest for testing).

3.4.2 Evaluation Metrics

In order to validate the approaches in a real-world folksonomy setting, a rich set of metrics have been applied to measure recommendation accuracy, diversity, novelty and computational costs. These metrics have been chosen based on related literature [Gunawardana and Shani, 2009, Konstan, 2004, Baeza-Yates et al., 1999].

Recommendation Accuracy

Recommendation accuracy is typically measured using the three metrics Recall, Precision and F1-score.

Recall (R). Recall is calculated as the number of correctly recommended tags divided by the number of relevant tags [Van Rijsbergen, 1974]:

$$R@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \frac{|\widetilde{T}_k(u,r) \cap T(u,r)|}{|T(u,r)|} \tag{3.17}$$

where $\widetilde{T}_k(u,r)$ denotes the k recommended tags and $T(u,r)$ the list of relevant tags of a user u for resource r that is determined by the bookmark in the test set B_{test}. Thus, Recall is a measure for the completeness of the recommendations and rises with the number of recommended tags k.

Precision (P). Precision is calculated as the number of correctly recommended tags divided by the number of recommended tags k [Van Rijsbergen, 1974]:

$$P@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \frac{|\widetilde{T}_k(u,r) \cap T(u,r)|}{k} \tag{3.18}$$

Based on this definition, Precision is a measure for the usefulness of the recommendations and falls with the number of recommended tags k. Recall and Precision are typically presented in the form of Recall / Precision plots for $k = 1$ - 10.

F1-score (F1). F1-score combines Recall and Precision into one score [Van Rijsbergen, 1974]:

$$F1@k = 2 \cdot \frac{P@k \cdot R@k}{P@k + R@k} \tag{3.19}$$

F1-score can be defined as the harmonic mean of Recall and Precision and thus, reaches its highest value typically at $k = 5$ (i.e., F1@5)[17].

Recommendation Ranking

Since Recall, Precision and F1-score only take the proportion of correctly identified recommendations into account, metrics that also measure the ranking of these recommendations have been defined.

Mean reciprocal rank (MRR). MRR is the sum of the reciprocal ranks of all relevant tags in the list of recommended tags [Rawashdeh et al., 2013]:

$$MRR@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \frac{1}{|T(u,r)|} \sum_{t \in T(u,r)} \frac{1}{rank(t)} \tag{3.20}$$

This means that a high MRR is achieved if relevant tags occur at the beginning of the recommended tag list.

Mean average precision (MAP). MAP is an extension of the Precision metric that additionally looks at the ranking of recommended tags. MAP is described in the subsequent formula, where B_i is 1 if the recommended tag at position i is among the relevant tags and 0 otherwise [Rawashdeh et al., 2013]:

$$MAP@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \frac{1}{|T(u,r)|} \sum_{i=1}^{k} B_i \cdot P_{u,r}@i \tag{3.21}$$

where $P_{u,r}@i$ depicts Precision@i calculated for user u and resource r.

Normalized Discounted Cumulative Gain (nDCG). nDCG is another ranking-dependent metric that not only measures how many tags can be correctly predicted but also takes the position of the tags in the list of recommended tags with length k into account. The nDCG metric is based on the *Discounted Cumulative Gain (DCG)*, which is given by [Järvelin and Kekäläinen, 2000]:

$$DCG@k = \sum_{i=1}^{k} \left(\frac{2^{B_i} - 1}{log_2(1+i)}\right) \tag{3.22}$$

[17] http://www.kde.cs.uni-kassel.de/ws/dc09/evaluation

where B_i is 1 if the i^{th} recommended tag is a relevant one and 0 otherwise. nDCG@k is calculated as DCG@k divided by the ideal DCG value iDCG@k, which is the highest possible DCG value that can be achieved if all the relevant tags would be recommended in the correct order. It is given by [Järvelin and Kekäläinen, 2000]:

$$nDCG@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \left(\frac{DCG@k}{iDCG@k}\right) \tag{3.23}$$

All three metrics MRR, MAP and nDCG rise with the number of recommended tags k and thus, are reported for $k = 10$ in this thesis.

Recommendation Diversity and Novelty

Recommender evaluation strategies have mainly focused on recommendation accuracy and ranking, and neglect other important factors that a recommender system should be aware of [Kowald and Lex, 2015]. Therefore, recent research on tag recommender evaluation has promoted the importance of diversity and novelty in tag recommendations [Belém et al., 2016].

Average IntraList Distance (AILD). Tag recommender diversity is measured by means of the AILD metric as defined in [Vargas and Castells, 2011]. According to this metric, the dissimilarity of two recommended tags t_i and t_j is given by the relative difference by means of the Jaccard coefficient between the sets of resources to which the tags were applied (i.e., $dist(t_i, t_j)$). More formally, this is given by:

$$AILD@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \frac{2}{k^2 - k} \sum_{i=1}^{k} \sum_{j=i+1}^{k} dist(t_i, t_j) \tag{3.24}$$

This means that a set of tags is diverse if the tags were used for different sets of resources [Belém et al., 2013].

Average Inverse Popularity (AIP). The novelty of the recommended tag list is calculated using the AIP metric. As defined by [Belém et al., 2013], a recommended tag t_i is novel if it was not previously used to annotate the target resource r.

$$AIP@k = \frac{1}{|B_{test}|} \sum_{u,r \in B_{test}} \frac{1}{K} \sum_{i=1}^{k} \frac{1}{log(1+i)} \cdot IFF_i \tag{3.25}$$

where IFF_i is the Inverse Feature Frequency defined as $IFF_i = log(\frac{Y_r+1}{Y_{t,r}+1})$ [Belém et al., 2011] and K is a normalization factor based on the maximum IFF score IFF_{max}. Thus, the lower the popularity of a tag for a resource, the higher its novelty. Both AILD and AIP are reported for $k = 10$ recommended tags.

Computational Costs

Since recommendations should not only be accurate, diverse and novel but also be provided in (near) real-time, the computational costs in terms of runtime and memory consumption need to be determined.

Runtime. The runtime of the algorithms was measured in two ways. First, the runtime complexity was determined by means of $\mathcal{O}$-notations and second, the runtime was measured in milliseconds [ms] for the complete workflow of the algorithms (i.e., including training and testing).

Memory. The memory consumption of the algorithms was measured in Megabytes [MB]. Since the memory consumption typically varies over the workflow of an algorithm, the maximum value of memory consumption over time is reported in this thesis. Runtime and memory were measured on an IBM System x3550 M4 server with one Intel(R) Xeon(R) CPU E5-2640 v2 @ 2.00GHz and 256GB RAM using Ubuntu 12.04.2 and Java 1.8.

3.5 The TagRec Framework

In this section, a tag recommendation evaluation framework termed *TagRec* is presented, which is one of the core contributions of this thesis. This section is based on *P6* [Kowald et al., 2014a] and *P7* [Kowald et al., 2017a].

In order to ensure reproducibility, all four experiments mentioned in Section 3.1 have been conducted by using the *TagRec* evaluation framework, which was developed in course of this thesis. Thus, *TagRec* implements all dataset processing methods, algorithms, evaluation protocols and metrics described in this chapter.

In general, the purpose of *TagRec* is to provide the research community with a standardized framework that supports all steps of the development and evaluation process of tag-based recommendation algorithms in a reproducible way. This includes methods for data processing, data modeling and recommender evaluation.

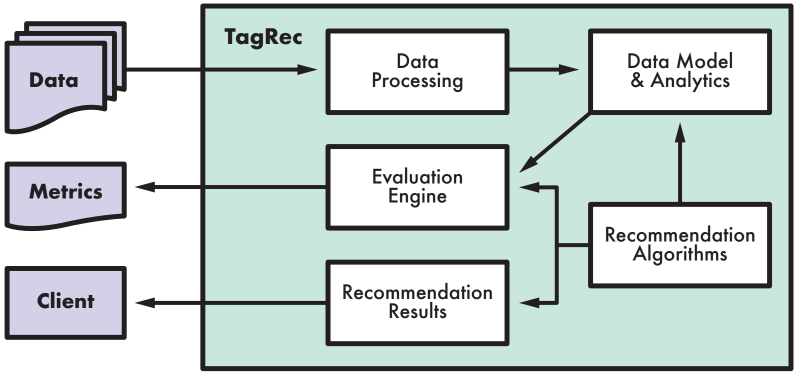


Figure 3.2: System architecture of the *TagRec* framework. The framework consists of five main components: (i) Data Processing, (ii) Data Model & Analytics, (iii) Recommendation Algorithms, (iv) Evaluation Engine, and (v) Recommendation Results [Kowald et al., 2017a].

Thus, *TagRec* aims (i) to increase the transparency in tag-based recommender research (see also [Said and Bellogín, 2014]), and (ii) to decrease the workload associated with developing novel tag-based recommender algorithms by providing researchers with an easy-to-use and easy-to-extend framework. To date, *TagRec* has supported the recommender development and evaluation processes in two large-scale European research projects, which have been published in 17 research papers [Kowald et al., 2017a].

Fully implemented in the Java programming language, *TagRec* is open-source software that can be freely downloaded from GitHub[18] for scientific purposes. Figure 3.2 shows the system architecture of *TagRec*, which consists of five main components briefly described in the remainder of this section.

3.5.1 Data Processing

TagRec offers various methods for data preprocessing: (i) parsing and processing of social tagging datasets (see Section 3.2) into the frameworks's internal data format, (ii) p-core pruning (see Section 3.2.2), (iii) training / test set splitting (see Section

[18] `https://github.com/learning-layers/TagRec/`

3.4.1), and (iv) creating Latent Dirichlet Allocation (LDA) [Krestel and Fankhause, 2010] topics for category-based algorithms such as 3Layers [Seitlinger et al., 2013, Kowald et al., 2015b].

3.5.2 Data Model and Analytics

The data model of *TagRec* is generated from a folksonomy that represents the bookmarks (i.e., the combination of user-id, resource-id, timestamp and assigned tags) in a dataset. Furthermore, the data model is fully object-oriented, and provides distinct classes and powerful methods for modeling and analyzing the relationship and interactions between users, resources and tags. This includes for example the number of times a specific tag has been assigned to a target resource and the time since the last usage of a specific tag in the tag assignments of a target user.

Apart from that, the data model of *TagRec* is connected to Apache Solr[19] and thus, enables fast access to content-based data of entities.

3.5.3 Recommendation Algorithms

Along with the state-of-the-art approaches for tag-based recommendations (see Section 3.3), *TagRec* contains a set of algorithms based on models derived from human cognition to predict tags in folksonomies. All algorithms implement a common interface, which makes it easy to develop and integrate new approaches. Moreover, connectors to external libraries based on factorization machines[20] are offered to use frameworks in other programming languages such as C++.

3.5.4 Evaluation Engine

The evaluation engine quantifies the quality of implemented recommendation strategies based on training / test set splits of a dataset with respect to standard metrics known from the area of recommender systems (see Section 3.4.2). Moreover, the evaluation engine offers data post-processing functionality that can, for example, limit the evaluation to users with a given minimum or maximum number of bookmarks or to users with certain tagging behavior (e.g, categorizers versus describers

[19] `http://lucene.apache.org/solr/`
[20] `https://cms.uni-konstanz.de/informatik/rendle/software/tag-recommender/`

[Körner et al., 2010b]).

For evaluating an algorithm, *TagRec* needs to be provided with three parameters, where the first one specifies the algorithm, the second one specifies the dataset directory and the third one specifies the file name of the dataset sample. For example, *java jar tagrec.jar cf bib bib_sample* runs Collaborative Filtering on a sample of the BibSonomy dataset. The calculated evaluation metrics are then either written to a "metrics" file or printed to the console.

3.5.5 Recommendation Results

As indicated in Figure 3.2, the recommendation results generated by the different algorithms can be forwarded either to the evaluation engine or directly to a client application (e.g., a graphical user interface).

3.6 Summary

In this chapter, the methodology, which is used to conduct the four experiments for answering the research questions of this thesis, was presented. Since this thesis builds on an offline study design, this included descriptions of the evaluated social tagging datasets, evaluation protocol, evaluation metrics and tag recommendation algorithms.

With respect to the data collections, the narrow folksonomy Flickr, the mixed folksonomies CiteULike, BibSonomy and Delicious, and the broad folksonomies LastFM and MovieLens are analyzed. In order to split these datasets into training and test sets, a time-based leave-one-out evaluation protocol is used.

Based on these splits, a rich set of evaluation metrics is utilized in order to measure the prediction accuracy, ranking, diversity, novelty and computational costs of tag recommendations. These metrics are then used for comparisons of 20 tag recommendation algorithms evaluated in course of this thesis. This includes simple frequency-based methods, classic Collaborative Filtering, graph-based algorithms, factorization models as well as modern time-based and cognitive-inspired approaches.

One of the core contributions of this thesis was the development of the *TagRec* evaluation framework, which was used to conduct the four experiments of this thesis.

Thus, *TagRec* implements all dataset processing methods, algorithms, evaluation protocols and metrics described in this thesis.

In the next chapter of this thesis, parts of the presented methodology are used to address Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*.

Chapter 4

Research Question 1: The Influence of Activation Processes in Human Memory on Tag Reuse

"What memory is inferring is something we call the need probability, which is the probability that we will need a particular memory trace now." [Anderson and Schooler, 1991]

In this chapter, the influence of activation processes in human memory on the reuse of tags is analyzed. Thus, it is the aim of this chapter to shed light on Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*, which is a prerequisite for the design of the cognitive-inspired tag recommendation algorithm BLL_{AC}+MP_r. To do so, at first the importance of activation processes in human memory for social tagging is discussed (see Section 4.1). This discussion leads to three factors of interest that influence the information access in human memory: (i) usage frequency, (ii) usage recency, and (iii) semantic context. Using the six datasets presented in Chapter 3, the influence of these three factors on tag reuse is evaluated (see Section 4.2). Within this evaluation, it is also validated if the factor of recency can be better modeled via a power or an exponential function.

Parts of this chapter have been published in *P2* [Kowald and Lex, 2016] (Section 4.2.1), *P3* [Kowald et al., 2014b] (Section 4.1), *P4* [Kowald et al., 2015a] (Section 4.1) and *P5* [Trattner et al., 2016] (Sections 4.1 and 4.2.2).

4.1 Activation Processes in Human Memory

Human memory is very adaptive to make appropriate memory units quickly available and thus, tunes the activation of its units to statistical regularities of the environment (e.g., [Anderson and Schooler, 1991, Cress et al., 2013]). This means that the more useful a memory unit has been and the stronger it is related to the current context (i.e., environmental cues), the higher is its activation level and hence, probability of being retrieved. This probability of being retrieved is also referred as "need probability" in human memory theory [Anderson and Schooler, 1991].

The aim of Research Question 1 is to analyze if these activation processes also determine a user's tagging behavior and if the need probability of a tag can be derived from estimates of its activation in the user's memory. According to [Anderson et al., 2004], the activation of a memory unit (e.g., a tag) should depend on at least two variables: (i) the general usefulness of this memory unit (see Section 4.1.1), and (ii) its associations to the current semantic context (e.g., to elements of the resource to be tagged - see Section 4.1.2). This means that a memory unit is more likely to be brought into consciousness, if it was important in the past and if it fits the current topic the user is dealing with.

This section is based on *P3* [Kowald et al., 2014b], *P4* [Kowald et al., 2015a] and *P5* [Trattner et al., 2016].

4.1.1 General Usefulness of a Memory Unit / Tag

The issue of how human memory ensures a fast and automatic information retrieval from its huge long-term memory has been extensively examined by memory psychology (e.g., [Anderson et al., 2004]). Essentially, human memory is tuned to the statistical structure of an individual's environment and keeps available those memory units that have been generally useful in the past. It has been shown that this general usefulness depends on how *frequently* and *recently* (i.e., the time since the last usage) the memory unit has been needed in the past (i.e., the need probability) [Anderson and Schooler, 1991].

Social tagging provides an illustrative example of the strong interplay between external, environmental and internal memory structures (e.g., [Held et al., 2012]). For instance, the development of generative models of social tagging demonstrated that the probability of a tag i being applied can be modeled through the preferential attachment principle (e.g., [Dellschaft and Staab, 2008]): the higher the frequency of i's past occurrence in the tagging environment is, the more likely it will be reused by an individual.

Additionally, the same probability is also a function of i's recency, which is the time elapsed since i last occurred in the environment [Cattuto et al., 2007]. Based on this, it can be assumed that these two factors (i.e., usage frequency and recency) also influence the reuse of tags [Trattner et al., 2016].

4.1.2 Usefulness of a Memory Unit / Tag in the Current Semantic Context

The current semantic context is given by any contextual element that is important in the current situation (e.g., related tags). Through learned associations based on a spreading activation mechanism in human memory [Anderson, 1983], the contextual elements are connected with a memory unit i (e.g., a tag) and can increase i's activation depending on the strength of association between i and these contextual elements.

To simplify matters, in the case of social tagging, the current semantic context is defined as the tags associated with a given resource r due to previous tag assignments of other users. Thus, the strength of association can be derived from the number of co-occurrences between i and these resource tags [Kowald et al., 2015a].

Thus, in the next section, the assumption is evaluated if a strong relation between activation processes in human memory and the reuse of social tags exists. Specifically, in Section 4.2.1, it is evaluated if the three factors of interest that have been discussed in this section really influence the reuse of tags in social tagging systems.

4.2 Influencing Factors of Tag Reuse

In this section, the evaluation results of analyzing the influence of activation processes in human memory on tag reuse are presented. The goal of this evaluation is to address Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*. To achieve this goal, the evaluation is split in two parts: firstly, the influence of tag frequency, recency and current semantic context on tag reuse is discussed and secondly, the question is addressed if a power-law or exponential distribution is better suited for modeling the time-dependent decay of tag reuse.

This section is based on *P2* [Kowald and Lex, 2016] and *P5* [Trattner et al., 2016].

4.2.1 The Influence of Tag Frequency, Recency and Semantic Context on Tag Reuse

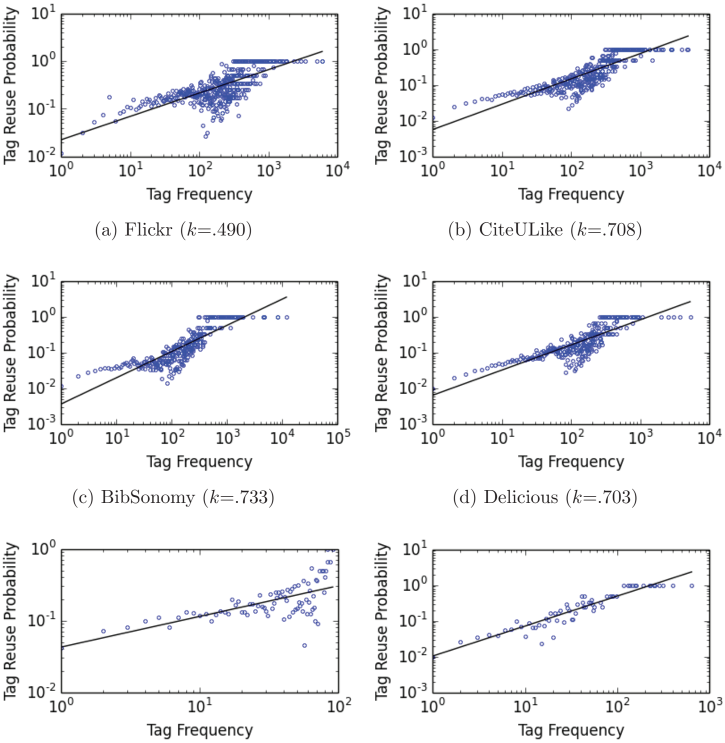
In order to quantify the influence of usage frequency, recency and semantic context on the reuse of tags, the tag assignments of the first $n-1$ bookmarks in the training set (i.e., reflecting the past) of a user u were compared with the tag assignments of u's n^{th} bookmark in the test set (i.e., reflecting the future). For detailed descriptions of the datasets and the dataset splitting method, please refer to Sections 3.2 and 3.4.1.

In the remainder of this section, it is described how this procedure was conducted for each individual factor (i.e., tag frequency, tag recency and semantic context).

Tag Frequency

In the case of tag frequency, for each tag t of user u, the number of times t was used by u in the training set was counted (i.e., the frequency value). Then, it was determined if t was also reused by u in the test set (i.e., the reuse probability of t by u). This procedure was repeated for the tags of each user $u \in U$. Finally, to obtain a statistically reliable value, all tags with the same frequency value were pooled together and the mean reuse probability of these tags was calculated and reported.

In Figure 4.1, the mean tag reuse probability is plotted over the tag frequency on a log-log scale. Additionally, the linear regression model on this data (i.e., $y = k * x + d$) is visualized and the corresponding slope k is provided. Across all six datasets, it can be seen that $k > 0$, which means that the more frequently a tag was used in the past, the higher its reuse probability.



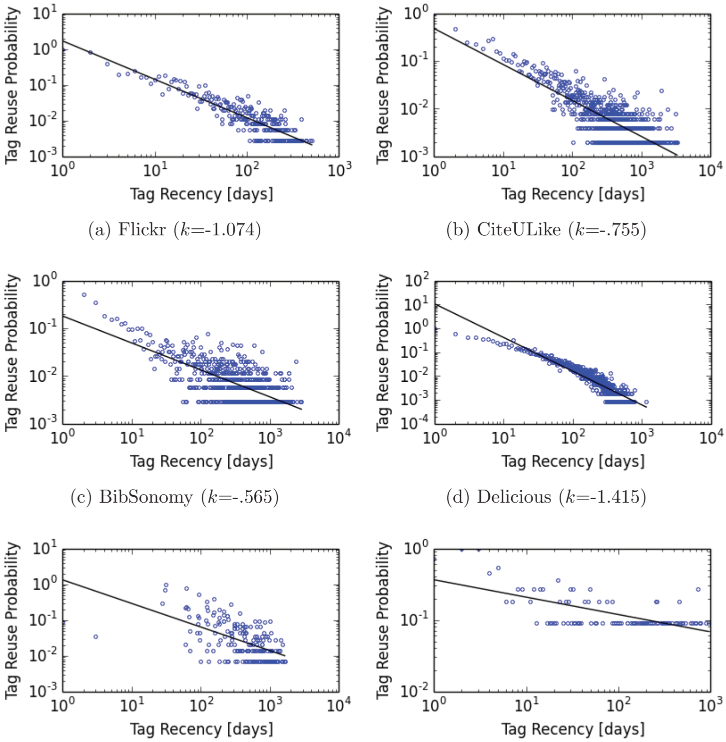
(a) Flickr (k=.490)
(b) CiteULike (k=.708)
(c) BibSonomy (k=.733)
(d) Delicious (k=.703)
(e) LastFM (k=.427)
(f) MovieLens (k=.838)

Figure 4.1: The influence of usage frequency on tag reuse. The *more frequently* a tag was used in the past ($k > 0$), the higher its reuse probability.

Tag Recency

In the case of tag recency, a similar procedure as for the tag frequency was followed. Thus, the days elapsed since the last use of tag t by user u were calculated. Then, as in the case of tag frequency, the tags with the same recency values were pooled together and the mean reuse probability was calculated and reported.

In Figure 4.2, the mean tag reuse probability is plotted on a log-log scale with the corresponding linear regression model, but this time not over the tag frequency

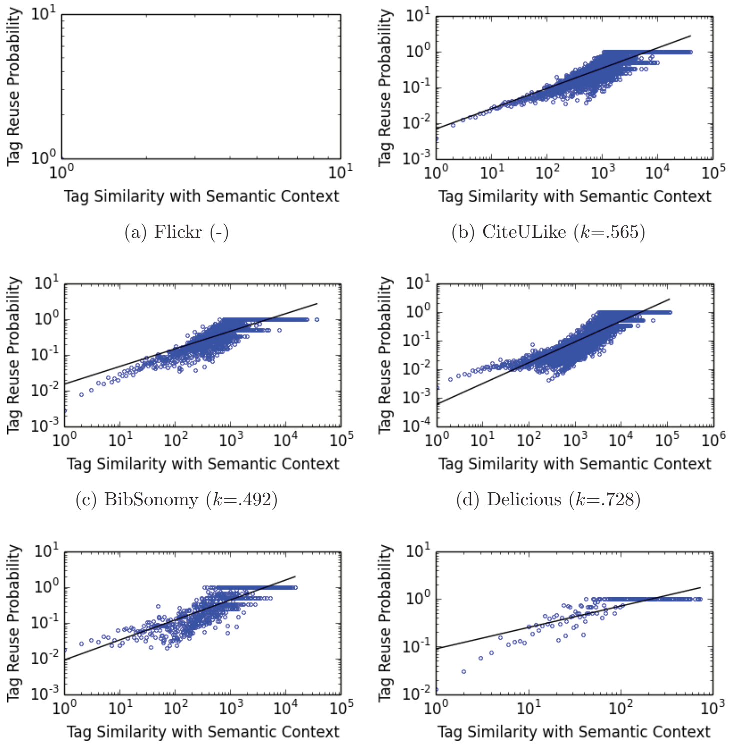


(a) Flickr (k=-1.074)
(b) CiteULike (k=-.755)
(c) BibSonomy (k=-.565)
(d) Delicious (k=-1.415)
(e) LastFM (k=-.659)
(f) MovieLens (k=-.243)

Figure 4.2: The influence of usage recency on tag reuse. The *more recently* a tag was used in the past ($k < 0$), the higher its reuse probability.

but the tag recency in days. Here, a different behavior as for tag frequency can be observed since slope $k < 0$ across all six datasets. This means that *more recently* a tag was used in the past, the higher its reuse probability.

One interesting finding in this respect is that the plot of LastFM differs from the plots of the other datasets in the fact that there are only a few tags with small tag recency values available. This could be a consequence of the crawling strategy of the LastFM dataset sample since, in contrast to the other datasets, this one is



(a) Flickr (-)

(b) CiteULike (k=.565)

(c) BibSonomy (k=.492)

(d) Delicious (k=.728)

(e) LastFM (k=.558)

(f) MovieLens (k=.448)

Figure 4.3: The influence of the current semantic context on tag reuse. The *more similar* a tag is to tags in the *current semantic context* ($k > 0$), the higher its reuse probability. Please note that there is no semantic context in Flickr since in this narrow folksonomy only own images can be tagged.

not a complete dump of all tag assignments in the system but a crawl of the tagging data of specific users (see Section 3.2).

Current Semantic Context

As already mentioned in this chapter, in this thesis, the current semantic context for social tagging is defined as the tags that are already associated with the currently tagged resource r. Thus, to identify the influence of the semantic context on tag reuse, a tag co-occurrence value between t and the tags assigned to r was determined by means of a spreading activation mechanism (see Section 4.1.2). Then, as in the case of tag frequency and recency, the tags with the same tag co-occurrence values were pooled together and the mean reuse probability was calculated and reported.

Figure 4.3 shows the mean tag reuse probability and its linear regression model plotted over the tag similarity with the semantic context on a log-log scale. Similarly as in the case of tag frequency, slope $k > 0$, which means that the *more similar* a tag is to tags in the *current semantic context*, the higher its reuse probability. It has to be noted that there is no semantic context in the Flickr dataset since here only users who have uploaded a resource (i.e., an image) are allowed to tag it (i.e., Flickr is a narrow folksonomy, see Section 3.2).

4.2.2 Power-Law Versus Exponential Recency Decay

This section addresses the question as to whether the effect of recency decays according to a power or an exponential function. The same question has already been investigated in a different context (i.e., re-occurrence of words in New York Times headings) in [Anderson and Schooler, 1991]. The authors found that the power function produces a better fit. Up to now, research on time-aware recommender systems has not applied a power function to model the temporal tagging patterns of users. Surprisingly, state-of-the-art approaches utilized only linear or exponential distributions (e.g., [Huang et al., 2014, Zheng and Li, 2011, Zhang et al., 2012, Yin et al., 2011b, Yin et al., 2011b, Campos et al., 2014]).

It is therefore the aim of this section to investigate as to whether the results obtained by [Anderson and Schooler, 1991] generalize to social tagging environments. This is achieved via a traditional least squares fitting test and a more sophisticated likelihood ratio test.

R^2	Flickr	CiteULike	BibSonomy	Delicious	LastFM	MovieLens
Exponential	.587	.306	.220	.756	.330	.111
Power	**.802**	**.672**	**.495**	**.873**	**.342**	**.437**

Table 4.1: Power-law versus exponential recency decay using a least squares fitting test. Across all six datasets, except of LastFM, the R^2 estimates clearly speak in favor of a power-law distribution for explaining the time-dependent decay of tag reuse.

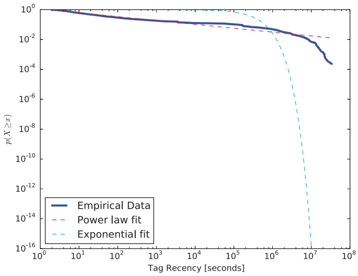
Least Squares Fitting

The traditional least squares fitting test is a rather easy way to check for power-law and exponential distributions in data, and was also used by [Anderson and Schooler, 1991] to check for re-occurrences of words in New York Times headings. This test compares the least squares fitting (i.e., R^2 values) of the linear regression in the log-log transformed data (see Figure 4.2) and the log-linear transformed data. In cases where the R^2 value is larger for the log-log transformed data, a power distribution produces a better fit and in cases where the R^2 value is larger for the log-linear transformed data, an exponential distribution is better suited to explain the data.

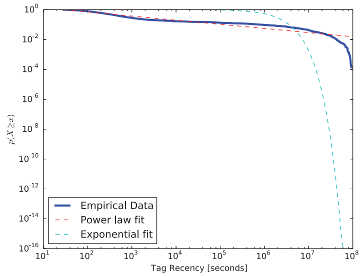
Table 4.1 shows the R^2 estimates for both cases and for all six datasets analyzed in this thesis. It can be seen that the R^2 values are generally higher in the power-law case than in the exponential one, which speaks in favor of the power function for explaining the time-dependent decay of tag reuse. This is especially the case for Flickr, CiteULike, BibSonomy, Delicious and MovieLens. Only in the case of LastFM, the difference between the exponential and the power-law fit is not clear and thus, needs to be investigated in more detail.

Likelihood Ratio Test

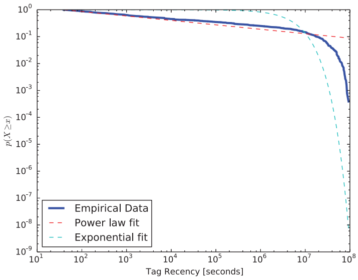
[Clauset et al., 2009] has shown that the least squares-based method can lead to misinterpretations and thus, suggests a likelihood ratio-based test for determining distributions in empirical data. To validate the results of the least squares fitting test, the python package *powerlaw* [Alstott et al., 2014], which implements the method of [Clauset et al., 2009], is used. As shown in Figure 4.4, in all datasets, except of LastFM, the estimated power function provides a better fit for the data than an exponential function. To test for statical significance, the likelihood ratio R between the two observed functions and the empirical data was calculated. Here,



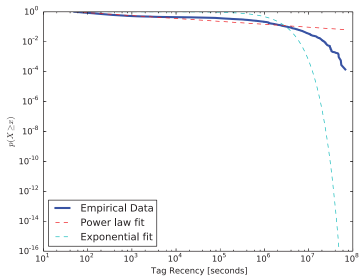
(a) Flickr (R=114, x_{min}=2)



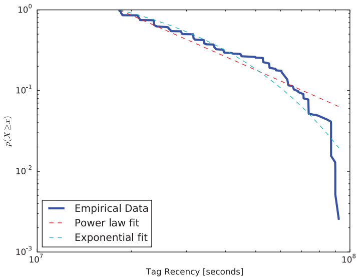
(b) CiteULike (R=122, x_{min}=28)



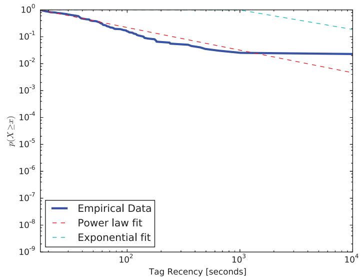
(c) BibSonomy (R=48, x_{min}=39)



(d) Delicious (R=65, x_{min}=50)



(e) LastFM (R=-2, x_{min}=18,313,200)



(f) MovieLens (R=13, x_{min}=17)

Figure 4.4: Power-law versus exponential recency decay using a likelihood ratio test. In all datasets, except of LastFM, a power function provides a better fit than an exponential one ($R > 0$ with $p < .001$).

$R > 0$ means that the data is statistically more likely to follow a power distribution rather than an exponential one. As presented in Figure 4.4 this is the case in all datasets, except of LastFM.

As already mentioned, the crawling strategy of the LastFM dataset sample is different than for the other datasets since LastFM is not a complete dump of all tag assignments in the system (see Section 3.2). Hence, it can be assumed that this is the reason for the large x_{min} value in the case of LastFM, which produces also the negative R value. Furthermore, it should be noted that the decay in Flickr is more pronounced than, for example, in BibSonomy, which might imply that scientific topics in BibSonomy (e.g., research on recommender systems) do not change as fast as topics of photos of different leisure events (e.g., pictures of the last weekend).

From this pattern of results, it can be concluded that the findings revealed by [Anderson and Schooler, 1991] generalize to social tagging environments: the effect of recency on the reuse probability of tags is more likely to follow a power-law distribution than an exponential one.

4.3 Summary

In this chapter, the influence of activation processes in human memory on the reuse of tags was analyzed. Based on human memory theory, the usefulness of a memory chunk (e.g., a tag) depends on at least three factors: (i) past usage frequency, (ii) past usage recency (i.e., the time since the last usage), and (iii) the current semantic context. Thus, it has been the aim of this chapter to evaluate the assumption of this thesis that there exists a strong relation between activation processes in human memory and the reuse of social tags (i.e., Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*).

With respect to this research question, the results presented in this chapter lead to the following four findings:

1. The *more frequently* a tag was used in the past, the higher its probability of being reused.

2. The *more recently* a tag was used in the past, the higher its probability of being reused.

3. The *more similar* a tag is to tags in the *current semantic context*, the higher its probability of being reused.

4. The effect of *recency on the reuse probability of tags* is more likely to follow a *power-law distribution* than an exponential one.

Based on these findings, the strong relation between activation processes in human memory and the use of tags in social tagging systems is verified, which positively answers Research Question 1. Additionally, these findings are a prerequisite for designing a cognitive-inspired approach for tag reuse predictions and tag recommendations. More specifically, the activation equation of the cognitive architecture ACT-R [Anderson, 1996, Anderson et al., 2004] integrates the factors of frequency, recency and semantic context using a power function and a spreading activation-based model to determine the usefulness of chunks in human memory.

Hence, the next chapter of this thesis discusses if this equation can be utilized for tag reuse predictions in order to address Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"*.

Chapter 5

Research Question 2: Designing a Cognitive-Inspired Algorithm for Tag Reuse Prediction

"Activation-Level = Base-Level + Contextualized-Priming."
[Anderson, 1996]

This chapter presents a cognitive-inspired algorithm for tag reuse prediction, which is one of the core contributions of this thesis. Based on the findings highlighted in Chapter 4, which showed that the reuse of tags in social tagging systems highly corresponds to activation processes in human memory, the activation equation of the cognitive model ACT-R (see Section 5.1) is utilized to design a novel tag reuse prediction algorithm termed BLL_{AC} (see Section 5.2). Furthermore, BLL_{AC} is evaluated (see Section 5.2.3) using the methodology presented in Chapter 3. These evaluation results contribute to Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"*, which provides the basis for the implementation of a hybrid approach for tag recommendations in real-world folksonomies (see Chapter 6).

Parts of this chapter have been published in *P2* [Kowald and Lex, 2016] (Section 5.2.3), *P3* [Kowald et al., 2014b] (Section 5.2), *P4* [Kowald et al., 2015a] and *P5* [Trattner et al., 2016] (Sections 5.1 and 5.2).

5.1 Formalizing the Activation of Memory Units

In this section, background information on the formalization of the activation of memory units is presented. This includes a short description of the cognitive architecture ACT-R as well as the activation equation, which is part of ACT-R and also the basis for the tag reuse prediction algorithm presented in this thesis (i.e., BLL_{AC}). This section is based on *P4* [Kowald et al., 2015a] and *P5* [Trattner et al., 2016].

5.1.1 The Cognitive Architecture ACT-R

ACT-R, which is short for "Adaptive Control of Thought – Rational", is a cognitive architecture developed by John Robert Anderson [Anderson, 1996, Anderson et al., 2004, Anderson et al., 1997]. ACT-R deals with defining and formalizing the basic cognitive operations of the human mind (e.g., the access on information in human memory).

Figure 5.1 schematically illustrates the main architecture of ACT-R. In general, ACT-R differs between short-term memory modules, such as the working memory module, and long-term memory modules, such as the declarative and procedural memory modules. Via a sensory register (i.e., the ultra short-term memory), encoded information is passed to the short-term working memory module, which interacts with the long-term memory modules. In case of the declarative memory, (i) the encoded information can be stored, and (ii) already stored information can be retrieved. In case of the procedural memory, the information can be matched against stored rules that can lead to executions [Wheeler, 2014].

Thus, declarative memory holds factual knowledge (e.g., what something is) and procedural memory consists of sequences of actions (e.g., how to do something). This thesis focuses on the declarative part, which contains the well-known activation equation of human memory.

5.1.2 The Activation Equation

Consider a user retrieving a unit from her memory, such as a tag that she has used previously. To derive its usefulness in the current context, the activation level A_i of this memory unit i has to be determined. According to the following activation

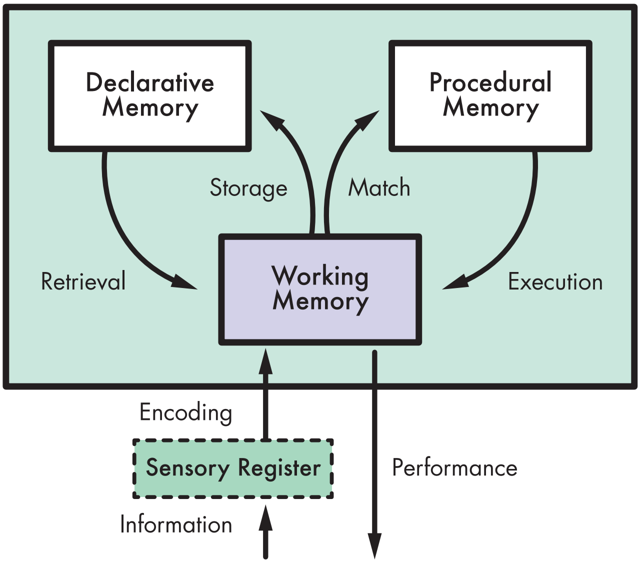


Figure 5.1: Schematic illustration of the cognitive architecture ACT-R. In general, ACT-R differs between short-term memory modules (i.e., working memory) and long-term memory modules (i.e., declarative and procedural memory). This figure was adapted from [Wheeler, 2014].

equation, which is part of the declarative module of the cognitive architecture ACT-R, the usefulness of i is given by:

$$A_i = B_i + \sum_j W_j \cdot S_{j,i} \tag{5.1}$$

The B_i component represents the *base-level* activation and quantifies the general usefulness of a unit i by considering how frequently and recently it has been used in the past. It is given by the base-level learning (BLL) equation:

$$B_i = ln(\sum_{j=1}^{n} t_j^{-d}) \tag{5.2}$$

where n is the frequency of the unit's occurrences and t_j is the recency (i.e., the time in seconds since the j^{th} occurrence of i). For example, if a user has applied the two tags "recognition" and "recommender" with equal frequency but "recommender" has dominated the user's recent bookmarks, the equation predicts a higher activation level for "recommender". The exponent d accounts for the power-law of forgetting, which means that each unit's activation level caused by the j^{th} occurrence decreases in time according to a power function [Anderson et al., 2004].

The second component of Equation 5.1 represents the associative component (i.e., the *contextualized priming*) that tunes the base-level activation of the unit i to the current semantic context. The context is given by any contextual element j important in the current situation (e.g., the tags "memory" and "recollection"). Through learned associations, the contextual elements are connected with tag i and can increase i's activation depending on the weight W_j and the strength of association $S_{j,i}$.

To simplify matters, the tags associated with a given resource r (due to previous tag assignments of other users) are used as the contextual elements. The weight W_j is derived from the number of times tag j has been assigned to resource r, and $S_{j,i}$ is derived from the number of co-occurrences between the tags i and j. The next section contains a more detailed and formal description of all calculation steps related to the task of predicting tag reuse.

5.2 A Tag Reuse Prediction Algorithm Based on Activation in Human Memory

The analysis presented in Chapter 4 revealed that the three factors (i) tag frequency, (ii) recency, and (iii) semantic context greatly influence the reuse probability of tags in social tagging systems (see Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*). Furthermore, it was shown that the factor of recency can be best modeled by using a power-law distribution.

It was therefore decided to implement the activation equation [Anderson et al., 2004] for predicting the reuse of tags and for addressing Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for ac-*

tivation processes in human memory, be exploited to develop a model for predicting the reuse of tags?". This is achieved via three algorithms.

The first algorithm is termed BLL as it implements the BLL equation in the form of a tag reuse prediction approach using the two factors of frequency and recency. This approach should not only be able to outperform a solely frequency-based method such as MP_u [Jäschke et al., 2007] but also an alternative time-dependent algorithm termed GIRP [Zhang et al., 2012], which models the temporal decay of tag reuse using an exponential function rather than a power function.

Algorithms two (i.e., AC_u) and three (i.e., BLL_{AC}) use the associative component of the activation equation to account for the current semantic context of tag usage. While AC_u solely implements this associative component, and thus accounts only for the current semantic context, BLL_{AC} combines it with the BLL equation to integrate all three factors (i.e., frequency, recency and semantic context) via the activation equation (see Equation 5.1).

This section is based on *P2* [Kowald and Lex, 2016], *P3* [Kowald et al., 2014b], *P4* [Kowald et al., 2015a] and *P5* [Trattner et al., 2016].

5.2.1 Formalization

In this section, the formalization of the proposed tag reuse prediction algorithm is presented by means of the activation equation. As indicated, this consists of a base-level activation and an associative component.

BLL Equation

For each tag t in a user u's training set B_{train}, the base-level activation $B(t, u)$ of t in u's set of tag assignments (i.e., Y_u) was calculated. Thus, a reference timestamp $timestamp_{u,ref}$, which is the timestamp of the most recent bookmark of user u in seconds, was determined. In this respect, $timestamp_{u,ref}$ corresponds to the most recent timestamp of the user's bookmark that has been selected for the test set B_{test} (see Section 3.4.1).

Next, if $j = 1 \ldots n$ indexes all tag assignments in Y_u, the recency of a tag assignment is given by $timestamp_{u,ref} - timestamp_{t,u,j}$. Hence, the base-level activation

$B(t,u)$ of tag t for user u is given by the BLL equation (see also Equation 5.2):

$$B(t,u) = \ln(\sum_{j=1}^{n} (timestamp_{u,ref} - timestamp_{t,u,j})^{-d}) \tag{5.3}$$

where d is set to .5 based on [Anderson et al., 2004]. Please note that other d values were also tested but this did not lead to better results in terms of recommender accuracy and thus, the value from the literature was kept.

In order to map the values onto a range of 0 to 1 that sum up to 1, a softmax function $\sigma T_u(B(t,u))$ as proposed in related work [McAuley and Leskovec, 2013] was applied:

$$\sigma_{T_u}(B(t,u)) = \frac{\exp(B(t,u))}{\sum_{t' \in T_u} \exp(B(t',u))} \tag{5.4}$$

where t' is a tag in T_u, the set of tags used by user u in the past.

Activation Equation

To investigate not only the factors of tag frequency and recency but also the factor of the current semantic context by means of an associative activation, the activation equation (see Equation 5.1) has been implemented in form of:

$$A(t,u,r) = \sigma_{T_u}(B(t,u)) + \sum_{c \in T_r} (|Y_{c,r}| \cdot S(c,t)) \tag{5.5}$$

where the first part represents the base-level component by means of the BLL equation and the second part represents the associative component evoked by the associative activation.

To calculate the variables of the associative activation (i.e., the contextualized priming) to model a user's current semantic context, the set of tags T_r assigned by other users to the given resource r was incorporated. A user's current semantic context certainly consists of a greater variety of aspects, such as content words in the title or in the page text of the resource. However, since not all social tagging datasets contain title information or page text and other studies have convincingly demonstrated the impact of a resource's prominent tags on a user's tagging behavior (e.g., [Lorince and Todd, 2013]), it was decided to approximate the context by means of other users' tags.

When applying the formula to a recommender system, related literature [Sigurbjörnsson and Van Zwol, 2008, Van Maanen and Marewski, 2009] suggests to use a measure of normalized tag co-occurrence to represent the strength of an association. Accordingly, the co-occurrence between two tags was defined as the number of bookmarks in which both tags are included. To add meaning to the co-occurrence value, the overall frequency of the two tags is also taken into consideration. This is done by normalizing the co-occurrence value according to the Jaccard coefficient following the approach described in [Sigurbjörnsson and Van Zwol, 2008]:

$$S(c,t) = \frac{|B_c \cap B_t|}{|B_c \cup B_t|} \tag{5.6}$$

where $S(c,t)$ is calculated as an association value between a tag t previously given by the target user and a tag c that has been assigned to a resource of interest.

Based on a tag co-occurrence matrix that depicts the tag relations of an entire dataset, information about how many times two tags co-occur ($B_c \cap B_t$) in bookmarks is retrieved and set into relation with the number of bookmarks in which at least one of the two tags appear ($B_c \cup B_t$). The attentional weight W_c of c was set to the number of times c occurred in the tag assignments of the target resource (i.e., $|Y_{c,r}|$).

Hence, the associative component in Equation 5.5 works in a similar way as resource-based Collaborative Filtering in the tag recommender literature [Tso-Sutter et al., 2008]. This means, that tags with a higher similarity to the target resource (measured by tag co-occurrence) get a higher associative activation value than tags with a lower similarity to the target resource (i.e., with a smaller usefulness in the current semantic context).

Tag Reuse Prediction

For tag reuse predictions, the activation levels of a user u's tags can be sorted in order to identify the top-k tags for u bookmarking a resource r. In this respect, three different tag prediction algorithms can be defined. The first one (i.e., BLL) is given by:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u}{\arg\overset{k}{\max}}(\underbrace{\sigma_{T_u}(B(t,u))}_{BLL}) \tag{5.7}$$

and calculates predictions based on the frequency and recency of u's tags in the past. The second one (i.e., AC$_u$) is defined as:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u}{\arg\max^k}(\underbrace{\sum_{c \in T_r}(|Y_{c,r}| \cdot S(c,t))}_{AC_u}) \tag{5.8}$$

where u's tags are ranked according to their usefulness in the given semantic context (i.e., based on co-occurrences with the tags of resource r). Finally, the third one (i.e., BLL$_{AC}$) implements the full activation equation via:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u}{\arg\max^k}(\underbrace{\sigma_{T_u}(A(t,u,r))}_{BLL_{AC}}) \tag{5.9}$$

Hence, BLL$_{AC}$ combines the former two algorithms BLL and AC$_u$ into one approach.

5.2.2 Illustration

In order to further clarify how the equations have been applied to characterize a user's individual tagging history, two simple examples illustrated in Figures 5.2 and 5.3 are provided. This also aims at demonstrating the advantage of BLL$_{AC}$ over conventional "most popular tags" approaches.

BLL Equation

The example in Figure 5.2 shows how the BLL equation provides a more differentiated characterization of a user's tagging pattern than the "most popular tags by user" (i.e., MP$_u$) approach. In this example, a user u applied a tag t three times (i.e., n = 3). It is assumed that she applied the tag ten, eight and seven days ago. The three corresponding recency values are recency$_1$ = 10, recency$_2$ = 8 and recency$_3$ = 7.

The recency of a tag t's use was calculated by subtracting the timestamp of the j^{th} use of t from the timestamp of u's most recent bookmark. Each of the three uses of t activates the corresponding memory unit. In Figure 5.2, the upward directed arrows symbolize this hypothesized activation. Due to the power-law of forgetting, each activation decreases in time (represented by the sloping curves) and each of the three recency values is raised by the power d = -.5 [Anderson et al., 2004].

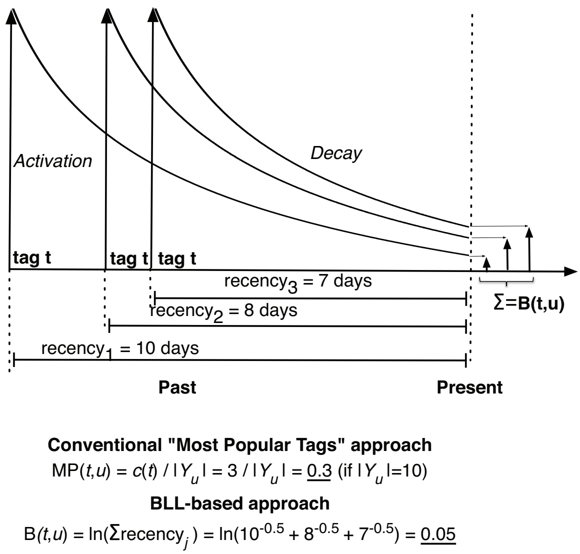


Figure 5.2: Illustration of the BLL equation. Example for applying the BLL equation (i.e., the first component of the activation equation) to estimate the activation value of a tag t and to show the advantage over the conventional "most popular tags by user" (MP_u) approach. This figure was taken from [Trattner et al., 2016] with permission from Paul Seitlinger.

Finally, the base-level activation level of the memory unit for tag t is given by summing the remaining effects of the three tag uses (i.e., $ln(10^{-.5} + 8^{-.5} + 7^{-.5})$), resulting in the base-level activation of .05. To the contrary, a conventional "most popular tags by user" (i.e., MP_u) approach, only takes into account the tag's usage frequency and thus, treats every tag assignment the same, independent of the time elapsed since its use. Given the user's entire set of tag assignments Y_u encompasses 10 assignments, this approach would yield a value of .3 (i.e., 3 / 10).

This should demonstrate that the BLL equation allows for a more differentiated characterization of a user's tagging history than MP_u.

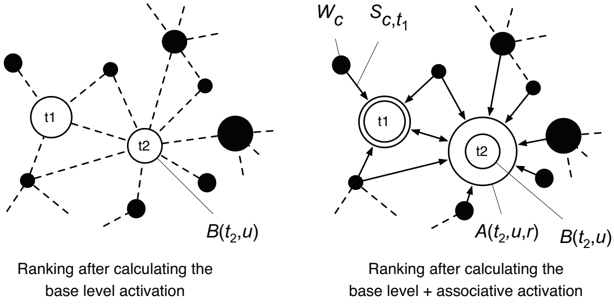


Figure 5.3: Illustration of the activation equation. Example showing the impact of the associative activation (i.e., the second component of the activation equation). Please note that black filled nodes and unfilled nodes represent contextual and target tags, respectively; their sizes represent their attentional weights W_c (in case of contextual tags) and activation (in case of the target tags t_1 and t_2). The edge length represents the co-occurrence-based association strength $S_{c,t}$. *Left panel:* ranking based on base-level activation $B(t,u)$ not taking into account the contextual tags. *Right panel:* refined ranking after considering the associative activation evoked by contextual tags, resulting in the full activation $A(t,u,r)$. This figure was taken from [Trattner et al., 2016] with permission from Paul Seitlinger.

Activation Equation

In the example of Figure 5.3, the additional impact of the associative activation defined by the second component of the activation equation is shown. The associative activation is evoked by the current context (i.e., the tags assigned by preceding users to the target resource – in the following called contextual tags).

The left panel of Figure 5.3 shows two target tags, t_1 and t_2 exhibiting different base-level activation levels (represented by the circle size): t_1 reaches a higher base-level activation and thus, a higher ranking than t_2. This relationship changes when considering the influence of the contextual tags, as schematically visualized in the right panel of Figure 5.3. These contextual tags are represented by the black nodes.

Depending on their weights W_j (represented by the size of the black-filled nodes) and strength of association $S_{j,i}$ (represented by the length of the edges), the contextual tags spread additional associative activation to the target tags t_1 and t_2 (i.e., making them more easily available for retrieval and use). t_2 is stronger associated with the contextual tags and thus, receives stronger associative activation than t_1.

Summed up, it can be seen that t_2 is assigned a higher ranking than t_1 when considering both, the base-level and associative activation by means of the full activation equation.

5.2.3 Evaluation

The aim of this evaluation is to answer Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"*. Thus, BLL_{AC} is compared with algorithms representing its individual components (i.e., MP_u for frequency, MR_u for recency and AC_u for the semantic context), combinations of its components (i.e., GIRP [Zhang et al., 2012] and BLL for frequency and recency) and social influences (i.e., FolkRank (FR) [Hotho et al., 2006]). Additionally, BLL is compared with GIRP in order to verify if a power function is better suited to model the time-dependent decay of tag reuse than an exponential one.

In Table 5.1, the results of this tag reuse prediction evaluation are shown. Across all datasets and metrics, BLL_{AC} provides higher accuracy and ranking estimates than algorithms reflecting its individual components and combinations of its components. In the narrow and broad settings, BLL_{AC} even outperforms FR, which also utilizes social influences by means of recommending other users' tags. In contrast, BLL_{AC} uses other users' tags solely for contextualized priming but just predicts the reuse of the current user's tags.

When further analyzing the results with respect to the folksonomy type of the given social tagging system (i.e., narrow, mixed and broad folksonomies), three different patterns of results can be observed. These observations are also summarized in Table 5.2.

Narrow Folksonomies

In the narrow folksonomy Flickr, the semantic context (i.e, AC_u) has no influence since users are solely tagging their own images. The results show that frequency (i.e., MP_u) and especially recency (i.e., MR_u) can be exploited to efficiently predict a user's tag reuse in this narrow setting. Furthermore, these factors even outperform

Folksonomy	Dataset	Metric	Individual factors MP$_u$	MR$_u$	**AC$_u$**	Combination GIRP	**BLL**	**BLL$_{AC}$**	Social FR
Narrow	Flickr	F_1@5	.371	.464	-	.455	**.470**	.470	.365
		MRR	.392	.506	-	.488	**.512**	.512	.387
		MAP	.509	.671	-	.647	**.680**	.680	.501
		nDCG	.569	.702	-	.686	**.711**	.711	.561
Mixed	CiteULike	F_1@5	.231	.236	.041	.243	.254	**.259**	.250
		MRR	.261	.284	.051	.287	.304	**.312**	.276
		MAP	.307	.333	.059	.335	.358	**.367**	.327
		nDCG	.367	.385	.069	.394	.413	**.422**	.392
	BibSonomy	F_1@5	.253	.252	.063	.262	.269	**.280**	.279
		MRR	.250	.249	.059	.262	.269	**.278**	.269
		MAP	.307	.307	.074	.323	.333	**.346**	.337
		nDCG	.371	.368	.090	.386	.396	**.409**	.408
	Delicious	F_1@5	.173	.179	.108	.190	.203	**.243**	.196
		MRR	.176	.201	.106	.204	.222	**.261**	.184
		MAP	.206	.235	.131	.238	.261	**.312**	.226
		nDCG	.267	.287	.158	.298	.318	**.374**	.292
Broad	LastFM	F_1@5	.193	.189	.202	.198	.202	.251	**.270**
		MRR	.192	.195	.205	.203	.213	**.260**	.257
		MAP	.226	.228	.248	.239	.250	.312	**.313**
		nDCG	.292	.293	.302	.303	.313	.375	**.399**
	MovieLens	F_1@5	.077	.076	.077	.077	.079	.086	**.153**
		MRR	.156	.164	.156	.157	.168	.183	**.243**
		MAP	.159	.168	.160	.160	.172	.188	**.253**
		nDCG	.177	.183	.176	.177	.187	.203	**.319**

Table 5.1: Tag reuse prediction accuracy and ranking results of algorithms that (i) reflect the individual factors of frequency (i.e, MP$_u$), recency (i.e, MR$_u$) and semantic context (i.e., AC$_u$), (ii) combine these factors (i.e., GIRP, BLL and BLL$_{AC}$), and (iii) utilize social influences (i.e, FR).

FR, the algorithm that also utilize social influences by means of recommending popular tags of other users.

When combining frequency and recency, it can be seen that the accuracy of the strong recency factor can only be slightly improved in the case of BLL, which models the time component via a power function, and even decreases in the case of GIRP, which builds on an exponential temporal decay function.

Folksonomy	Frequency	Recency	Semantic Context	Combination	Social
Narrow	+/-	+	-	+/-	-
Mixed	+	+	+/-	+	+/-
Broad	+/-	+/-	+	+/-	+

Table 5.2: Summary of the tag reuse prediction accuracy results showing the performance of the algorithms and their underlying factors with respect to the underlying factors and the given folksonomy type. Please note that "+" indicates a good performance, "+/-" indicates an average performance and "-" indicates a poor performance of a factor / an approach in a specific setting.

Mixed Folksonomies

In the mixed folksonomies CiteULike, BibSonomy and Delicious, a good performance for the factors of frequency and recency (i.e., they nearly reached the accuracy estimates of GIRP and BLL), and an average one for the semantic context is observed. Additionally, the results suggest that a combination of all three factors in the form of BLL_{AC} provides the highest accuracy estimates and outperforms FR.

Again, BLL (and thus, the power function) is apparently better suited to combine frequency and recency than GIRP (and thus, the exponential function).

Broad Folksonomies

Interestingly, the algorithms in the broad folksonomies LastFM and MovieLens show a completely different behavior. In these datasets, since there are a lot of tags assigned by other users to the currently tagged resource, the semantic context has a larger impact on the tag prediction accuracy than in the narrow and mixed settings.

Similarly to the narrow case, the combination of the factors only slightly improves the accuracy of the individual factors. Due to a high number of average posts per resource (5.674 for LastFM and 7.299 for MovieLens – see Table 3.1), FR that recommends popular tags of other users as well, provides the best results in the broad setting.

5.3 Summary

In this chapter, the design of a cognitive-inspired algorithm for tag reuse prediction (i.e., BLL_{AC}) was presented, which is one of the core contributions of this thesis.

This algorithm implements the activation equation of the cognitive architecture ACT-R to model (i) past tag usage frequency, (ii) past tag usage recency, and (iii) the current semantic context using a power-law model. Furthermore, this algorithm was evaluated with respect to reuse prediction accuracy and ranking in order to shed light on Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"*.

With respect to this research question, the results presented in this chapter, which are summarized in Table 5.2, lead to the following four findings:

1. *BLL outperforms GIRP*, which further underlines that the *time-dependent decay of tag reuse is better modeled using a power function than an exponential one.*

2. *BLL_{AC} provides higher accuracy and ranking estimates* than algorithms reflecting its *individual components and combinations of its components.*

3. In the narrow and mixed folksonomy settings, *BLL_{AC} outperforms the well-known FolkRank algorithm.*

4. In the broad folksonomy setting, FolkRank provides the best results, which shows the *importance of incorporating social influences by means of imitating popular tags of other users.*

These findings show that the activation equation of the cognitive architecture ACT-R can be exploited to predict the reuse of tags and thus, Research Question 2 can be positively answered. Furthermore, the strong results of the FolkRank algorithm in broad folksonomy settings indicate the importance of incorporating also social influences into the tag prediction and recommendation process by means of imitating popular tags of other users.

Therefore, the next chapter of this thesis discusses if BLL_{AC} can be extended with social influences by means of tag imitation processes with the aim to contribute to Research Question 3, *"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"*.

Chapter 6

Research Question 3: Implementing a Hybrid Approach for Tag Recommendations in Real-World Folksonomies

"Personalized recommendations will replace the navigation grid on Netflix." [Hunt, 2014]

This chapter describes the implementation process of a hybrid approach for tag recommendations in real-world folksonomies termed BLL_{AC}+MP_r. This hybrid recommendation approach combines the tag reuse prediction algorithm BLL_{AC} presented in Chapter 5, which implements the activation equation of the cognitive architecture ACT-R, with tag imitation processes to not only account for the factors of (i) tag usage frequency, (ii) recency, and (iii) semantic context but also for social influences by means of tag imitation processes (Section 6.1). BLL_{AC}+MP_r is evaluated against various state-of-the-art tag recommendation algorithms (Section 6.2) to contribute to Research Question 3, *"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"*.

Parts of this chapter have been published in *P4* [Kowald et al., 2015a] (Section 6.1,) *P5* [Trattner et al., 2016] (Sections 6.1 and 6.2.2), *P8* [Kowald and Lex, 2015] (Section 6.2.1) and *P9* [Kopeinik et al., 2016b] (Section 6.2.2).

6.1 A Hybrid Tag Recommendation Approach

The findings presented in Chapter 5 showed that the activation equation of the cognitive architecture ACT-R can be exploited to predict the reuse of tags via the BLL_{AC} approach (Research Question 2). Furthermore, the strong results of the FolkRank (FR) algorithm [Hotho et al., 2006] in broad folksonomy settings indicated the importance of incorporating also social influences into the tag prediction and recommendation process by means of imitating popular tags of other users.

Therefore, in this section it is presented how BLL_{AC} can be extended with social influences by means of tag imitation processes in order to realize the hybrid tag recommendation approach $\mathrm{BLL}_{AC}+\mathrm{MP}_r$ (Research Question 3). This section is based on *P4* [Kowald et al., 2015a] and *P5* [Trattner et al., 2016].

6.1.1 Incorporating Tag Imitation Processes

Research on social tagging [Floeck et al., 2010, Seitlinger and Ley, 2012, Seitlinger et al., 2015b, Fu et al., 2010, Fu et al., 2009] has shown that a substantial variance in a user's tag choices can be explained by her tendency to imitate tags previously assigned by other users to a resource. Furthermore, modeling this imitation process allows recommending novel tags, which were not used by the current user in her previous tagging history [Lorince and Todd, 2013, Lipczak, 2012].

In this thesis, tag imitation is realized by taking the most popular tags in the tag assignments of the current resource (MP_r) [Jäschke et al., 2007] into account:

$$\widetilde{T}_k(u,r) = \underset{t \in T_r}{\arg\overset{k}{\max}}(|Y_{t,r}|) \tag{6.1}$$

We have selected MP_r over other methods like Collaborative Filtering (CF), which is also able to suggest novel tags, because as mentioned before, users in social tagging systems are more likely to directly imitate tags that have already been assigned to a target resource. Additionally, this approach was also chosen by other researchers in the field (e.g., [Zhang et al., 2012]).

6.1.2 Combining Activation Processes in Human Memory with Tag Imitation Processes

Finally, the top-k recommended tags for a given user u and resource r (i.e., $\widetilde{T}_k(u, r)$) based on the BLL$_{AC}$+MP$_r$ algorithm are calculated via a linear combination:

$$\widetilde{T}_k(u,r) = \underset{t \in T_u \cup T_r}{\arg\max}^{k}(\underbrace{\beta \underbrace{\sigma_{T_u}(A(t,u,r)}_{BLL_{AC}} + (1-\beta)\sigma_{T_r}(|Y_{t,r}|)}_{BLL_{AC}+MP_r}) \tag{6.2}$$

where β is used to weigh the two components, (i) the activation values $A(t, u, r)$ and (ii) the scores of the most popular tags of the target resource given by MP$_r$. The results presented in this thesis were calculated using $\beta = .5$, thus giving equal weights to both components. Please also note, that both components are normalized by the softmax function σ in order to ensure the same value ranges between 0 and 1 that sum up to 1 (see [McAuley and Leskovec, 2013]).

6.2 Tag Recommendation Evaluation

The aim of this section is to evaluate the BLL$_{AC}$+MP$_r$ approach and thus, to answer Research Question 3, *"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"*. Therefore six datasets are used that are gathered from the real-world folksonomies Flickr, CiteULike, BibSonomy, Delicious, LastFM and MovieLens (see Section 3.2.1). Apart from that, evaluation results in two additional settings (i.e., the ECML PKDD Discovery Challenge 2009 and Technology Enhanced Learning) are presented.

All the evaluation results presented in this section were calculated using the *TagRec* framework (see Section 3.5) and were reported in *P5* [Trattner et al., 2016], *P8* [Kowald and Lex, 2015] and *P9* [Kopeinik et al., 2016b].

6.2.1 Evaluation in Real-World Folksonomies

It was important for the author of this thesis to benchmark the tag recommendation algorithms in the unfiltered datasets without p-core pruning to avoid a biased evaluation and to simulate a real-world folksonomy setting (see Section 3.2.2). This is especially important for the development of live recommender services. The narrowness degrees of the datasets used (see Section 3.2.3) justifies this approach since the average number of bookmarks assigned to a resource is lower than 2 in four of the six datasets (Flickr, CiteULike, BibSonomy and Delicious). This means that even a small p-core of 2 would delete a lot of bookmarks and so, substantially distort the natural structures of these datasets.

Regarding the evaluated algorithms (see Section 3.3), the simplest approaches that are utilized in this experiment are the frequency-based *MostPopular*$_r$ *(MP*$_r$*)* and *MostPopular*$_{u,r}$ *(MP*$_{u,r}$*)* algorithms [Jäschke et al., 2007], and *Collaborative Filtering (CF)* [Marinho and Schmidt-Thieme, 2008] with a neighborhood size of 20. As for algorithms that apply latent factor models, two types of algorithms are chosen: (i) *Latent Dirichlet Allocation (LDA)* [Krestel et al., 2009] with 1000 latent topics, and (ii) *Pairwise Interaction Tensor Factorization (PITF)* [Rendle and Schmidt-Thieme, 2010] with 256 dimensions of factorization.

Another well-known tag recommender approach that was chosen for this experiment is *FolkRank (FR)* [Jäschke et al., 2007]. With regard to time-dependent tag recommenders, four algorithms are included: *Temporal Tag Usage Patterns (GIRPTM)* [Zhang et al., 2012] that works in a more data-driven way and three that are inspired by models of cognitive science. Apart from the two algorithms proposed in this thesis (i.e., BLL$_{AC}$ and BLL$_{AC}$+MP$_r$), the 3LT+MP$_r$ approach is used in this study, which incorporates category information by means of LDA topics into the recommendation process [Seitlinger et al., 2013, Kowald et al., 2015b].

In Table 6.1, the evaluation results of this experiment are presented based on various metrics used to measure the performance of recommender systems (see Section 3.4.2). These results are also summarized in Table 6.5.

Tag Recommendation Accuracy

Tag recommendation accuracy is measured in Table 6.1 via F1 for $k = 5$, and MRR, MAP and nDCG for $k = 10$. In general, these results indicate that the two cognitive inspired algorithms BLL$_{AC}$+MP$_r$ and 3LT+MP$_r$ were the best with regard to recommender accuracy across all six datasets.

With respect to Research Question 3, this means that BLL$_{AC}$+MP$_r$ is able to outperform a set of state-of-the-art baseline algorithms such as CF, LDA, PITF and

	Metric	MP_r	$MP_{u,r}$	CF	LDA	PITF	FR	GIRPTM	3LT+MP_r	**BLL_{AC}**	**BLL_{AC}+MP_r**
Flickr	F1@5	-	.371	.453	.178	.350	.365	.455	**.482**	.470	.470
	MRR	-	.392	.474	.184	.366	.387	.488	**.525**	.512	.512
	MAP	-	.509	.631	.216	.469	.501	.647	**.698**	.680	.680
	nDCG	-	.569	.666	.280	.535	.561	.686	**.727**	.711	.711
	AILD	-	.789	.975	**.980**	**.980**	**.980**	.789	.670	.789	.789
	AIP	-	-	-	-	-	-	-	-	-	-
	Runtime [s]	-	**1**	4,342	1,227	228,868	18,090	2	10,594	5	5
	Memory [MB]	-	4,672	8,488	9,652	**2,502**	9,190	4,974	6,942	6,053	6,053
CiteULike	F1@5	.042	.249	.231	.089	.178	.250	.262	**.277**	.259	.273
	MRR	.043	.277	.263	.086	.207	.276	.303	**.321**	.312	.319
	MAP	.054	.329	.311	.094	.233	.327	.359	**.383**	.367	.380
	nDCG	.063	.392	.359	.138	.294	.392	.420	**.440**	.422	.438
	AILD	.152	.916	.961	**.991**	**.991**	**.991**	.916	.893	.902	.916
	AIP	.142	.952	.960	.983	**.991**	.958	.953	.953	.985	.951
	Runtime [s]	**1**	2	6,315	10,673	343,181	27,305	3	10,796	1,290	1,424
	Memory [MB]	5,725	5,913	9,301	11,943	**3,030**	9,347	6,631	9,474	8,177	8,789
BibSonomy	F1@5	.068	.281	.260	.145	.215	.279	.291	**.307**	.279	.298
	MRR	.054	.268	.248	.143	.218	.269	.282	**.298**	.278	.289
	MAP	.073	.337	.310	.162	.257	.337	.356	**.378**	.346	.365
	nDCG	.091	.407	.369	.219	.327	.408	.425	**.445**	.409	.434
	AILD	.199	.916	.941	.990	**.991**	**.991**	.916	.889	.901	.916
	AIP	.182	.939	.954	.966	.973	.944	.940	.941	**.976**	.937
	Runtime [s]	**1**	2	2,797	9,847	219,573	12,549	2	9,316	502	601
	Memory [MB]	4,811	4,972	9,405	14,012	**2,432**	9,494	5,567	9,137	8,078	8,307
Delicious	F1@5	.135	.238	.243	.182	.199	.196	.261	**.284**	.243	.283
	MRR	.117	.232	.241	.171	.193	.184	.258	**.291**	.261	.290
	MAP	.153	.279	.296	.204	.229	.226	.314	.357	.312	**.358**
	nDCG	.187	.358	.356	.271	.302	.292	.393	.430	.374	**.431**
	AILD	.353	.968	.972	**.999**	**.999**	**.999**	.968	.946	.955	.968
	AIP	.256	.882	.874	.887	.895	.877	.873	.874	**.938**	.863
	Runtime [s]	**1**	3	9,645	15,373	324,737	44,747	4	12,869	395	396
	Memory [MB]	12,198	12,596	35,090	11,894	**3,075**	7,381	13,672	59,425	16,469	17,620
LastFM	F1@5	.199	.258	.226	.258	.276	.270	.263	.279	.251	**.283**
	MRR	.186	.251	.208	.254	.276	.257	.255	.277	.260	**.283**
	MAP	.226	.301	.252	.306	.336	.313	.310	.338	.312	**.344**
	nDCG	.283	.386	.317	.388	.414	.399	.397	.421	.375	**.425**
	AILD	.722	.902	.855	.918	**.919**	**.919**	.902	.900	.840	.902
	AIP	.604	.730	.761	.741	.797	.728	.736	.722	**.866**	.711
	Runtime [s]	**1**	**1**	6	265	8,657	101	**1**	225	**1**	**1**
	Memory [MB]	**80**	92	214	593	87	237	155	3,332	204	301
MovieLens	F1@5	.135	.153	.124	.141	.156	.153	.159	**.162**	.086	.160
	MRR	.211	.260	.198	.233	.264	.243	.251	.263	.183	**.265**
	MAP	.223	.269	.209	.242	.275	.253	.262	.274	.188	**.276**
	nDCG	.271	.328	.254	.296	.324	.319	.326	.336	.203	**.338**
	AILD	.910	.954	.935	**.958**	.957	.957	.954	.954	.726	.954
	AIP	.787	.741	.861	.785	.816	.777	.751	.755	**.976**	.756
	Runtime [s]	**1**	**1**	11	206	6,091	90	**1**	120	**1**	**1**
	Memory [MB]	365	375	1,043	761	**96**	833	434	3,297	500	501

Table 6.1: Accuracy, diversity, novelty, runtime [seconds] and memory consumption [megabytes] estimates of the tag recommender algorithms in the six datasets.

FR. Moreover, BLL_{AC}+MP_r also outperforms GIRPTM, the currently leading time-depended tag recommendation algorithm. Particularly good results are derived with the ranking-dependent metrics MRR, MAP and nDCG. This observation clearly illustrates the advantages of BLL_{AC}+MP_r, which is build upon long-standing models of human memory theory, over the less-theory driven GIRPTM algorithm that also utilizes time information of social tags.

The difference between the narrow, mixed and broad folksonomy settings are especially of interest when comparing these results with previous studies (e.g., [Jäschke et al., 2008, Jäschke et al., 2007, Rendle et al., 2009, Lipczak, 2012]), in which FR and PITF typically had the best recommender accuracy in p-core pruned (i.e., very broad) folksonomies. The results presented in this section also indicate a good performance of FR and PITF in broad folksonomies (i.e., LastFM and MovieLens) but a fairly poor one in the narrow (i.e., Flickr) and mixed (i.e., CiteULike, BibSonomy and Delicious) settings. The opposite is the case for the BLL_{AC} approach presented in this thesis, which strictly recommends individual tags of the current user and thus, performs well in the narrow and mixed settings but is only average in the broad setting.

Tag Recommendation Diversity and Novelty

In Table 6.1, tag recommendation diversity is indicated by the AILD metric and novelty by the AIP metric (see Section 3.4.2).

Tag recommendation diversity. According to the AILD metric, the dissimilarity of two recommended tags is given by the relative difference by means of the Jaccard coefficient between the sets of resources to which the tags were applied [Vargas and Castells, 2011]. In this respect, the most diverse tag recommendations are provided via the classic approaches LDA, PITF and FR.

Tag recommendation novelty. According to the AIP metric, a recommended tag is novel if it was not previously used to annotate the target resource (i.e., if it cannot be found in the resource r's set of tags T_r) [Belém et al., 2013]. With regard to novelty, the strictly individual BLL_{AC} approach outperforms all the other algorithms, which also recommend tags of other users.

Summed up, the cognitive-inspired BLL_{AC}+MP_r approach that has the highest recommender accuracy also provides fair results in terms of these two metrics.

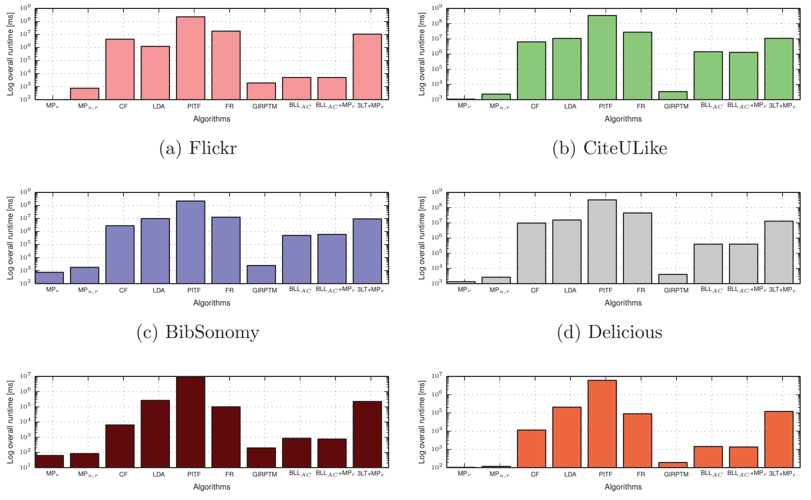
Algorithm	Complexity	Reference
MP	$\mathcal{O}(\lvert Y \rvert)$	[Jäschke et al., 2008]
MP$_u$	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert)$	[Jäschke et al., 2008]
MR$_u$	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert)$	[Campos et al., 2014]
GIRP	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert)$	[Zhang et al., 2012]
BLL	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert)$	[Trattner et al., 2016]
MP$_r$	$\mathcal{O}(\lvert R \rvert \cdot \lvert Y_r \rvert)$	[Jäschke et al., 2008]
MP$_{u,r}$	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert + \lvert R \rvert \cdot \lvert Y_r \rvert)$	[Jäschke et al., 2008]
GIRPTM	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert + \lvert R \rvert \cdot \lvert Y_r \rvert)$	[Zhang et al., 2012]
AC$_u$	$\mathcal{O}(\lvert B \rvert \cdot \lvert Y_b \rvert)$	[Trattner et al., 2016]
BLL$_{AC}$	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert + \lvert B \rvert \cdot \lvert Y_b \rvert)$	[Trattner et al., 2016]
BLL$_{AC}$**+MP**$_r$	$\mathcal{O}(\lvert U \rvert \cdot \lvert Y_u \rvert + \lvert B \rvert \cdot \lvert Y_b \rvert + \lvert R \rvert \cdot \lvert Y_r \rvert)$	[Trattner et al., 2016]
CF	$\mathcal{O}(\lvert U \rvert \cdot \lvert V_r \rvert \cdot \lvert Y_v \rvert)$	[Jäschke et al., 2008]
FR	$\mathcal{O}(\lvert U \rvert \cdot l \cdot (\lvert Y \rvert + \lvert U \rvert + \lvert R \rvert + \lvert T \rvert))$	[Jäschke et al., 2008]
LDA	$\mathcal{O}((\lvert U \rvert + \lvert R \rvert) \cdot \lvert T \rvert \cdot Z)$	[Blei et al., 2003]
3LT	$\mathcal{O}(\lvert R \rvert \cdot \lvert T \rvert \cdot Z + \lvert U \rvert \cdot \lvert Y_u \rvert)$	[Kowald et al., 2015b]
3LT+MP$_r$	$\mathcal{O}(\lvert R \rvert \cdot \lvert T \rvert \cdot Z + \lvert U \rvert \cdot \lvert Y_u \rvert + \lvert R \rvert \cdot \lvert Y_r \rvert)$	[Kowald et al., 2015b]
PITF	$\mathcal{O}(l \cdot \lvert B \rvert \cdot (k_T \cdot \lvert T \rvert^2 + k_U \cdot k_R \cdot k_T))$	[Rendle et al., 2009]

Table 6.2: Computational complexity of BLL$_{AC}$ and BLL$_{AC}$+MP$_r$ compared to state-of-the-art algorithms. Pleae note, that the algorithms are sorted in ascending order according to their complexity.

Computational Costs

In addition to recommender accuracy, diversity and novelty, the computational costs of the tag recommendation approaches were investigated in terms of computational complexity (see Table 6.2), monitored runtime (see Figure 6.1) and memory consumption (see Figure 6.2).

Computational complexity. Table 6.2 shows the complexity of all algorithms in ascending order. It can be seen that the popularity-based algorithms MP, MP$_u$, MP$_r$ and MP$_{u,r}$, which count frequencies by simply iterating over the tag assignments of the user (i.e., Y_u) and / or the resource (i.e., Y_r), provide linear runtime. For the time-based algorithms MR$_u$, GIRP, GIRPTM and BLL, similar behavior can be observed. An additional term is introduced, when calculating AC$_u$, BLL$_{AC}$ and BLL$_{AC}$+MP$_r$. This term describes the initialization of the co-occurrence matrix that holds the semantic context. The matrix is built by iterating over each bookmark b in the set of bookmarks B of a folksonomy and checking the tag assignments of b (i.e.,



(a) Flickr (b) CiteULike

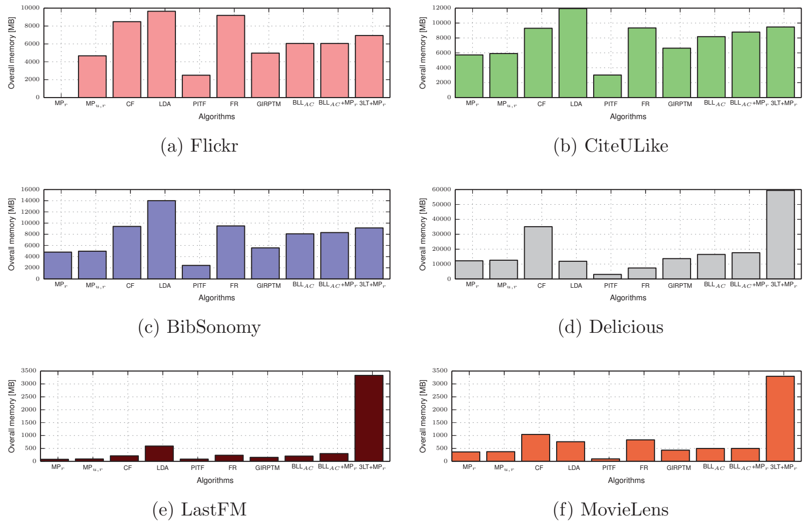
(c) BibSonomy (d) Delicious

(e) LastFM (f) MovieLens

Figure 6.1: Runtime measurements (im milliseconds) of tag recommendation algorithms showing the efficiency of BLL$_{AC}$+MP$_r$.

Y_b) for co-occurences. Even though this calculation step increases the computational complexity of the approach, this step only needs to be performed once, which may be done offline (especially for big datasets) and subsequently, it may not effect the online runtime in a live system.

Moreover, it can be seen that BLL$_{AC}$ and BLL$_{AC}$+MP$_r$ show better performance than state-of-the-art methods such as CF, LDA, FR and PITF. As these cognitive-inspired algorithms rely on relatively little but meaningful operations considering only user tag frequency, recency and semantic context in terms of resource tags, this algorithm outperforms the former. CF on the other hand, processes not only the tag assignments Y_u of the target user, but additionally the tag assignments of each user v in the set of users (i.e., neighbors) that have also tagged the target resource (i.e., V_r). In cases where there are no other users available that have tagged the target resource (i.e., cold-start resources), V_r becomes the set of all users, which then could lead to much higher computational costs as expected. With regard to FR, which depends on the number of nodes $|U|$, $|R|$ and $|T|$, and PITF, which depends on the



(a) Flickr (b) CiteULike

(c) BibSonomy (d) Delicious

(e) LastFM (f) MovieLens

Figure 6.2: Memory consumption measurements (in megabytes) of tag recommendation algorithms showing the efficiency of BLL$_{AC}$+MP$_r$.

dimensions of factorization k_U, k_R and k_T, even multiple iterations l are computed, which leads to higher runtime complexities (see Section 3.3).

When comparing BLL$_{AC}$+MP$_r$ with 3LT+MP$_r$ (i.e., the two best performing approaches in the accuracy experiment), it becomes apparent that 3LT+MP$_r$ has a much higher computational complexity. This is due the fact that 3LT+MP$_r$ relies on a very expensive topic creation process.

Runtime. To furthermore prove the theoretical assumptions made in the complexity analysis, a real runtime experiment was carried out. In particular, an experiment was conducted on an IBM System x3550 M4 server with one Intel(R) Xeon(R) CPU E5-2640 v2 @ 2.00GHz and 256GB RAM using Ubuntu 12.04.2 and Java 1.8 to determine the overall runtime performance of the algorithms. All algorithms were executed as single core, single thread instances to ensure that the measured runtime is not affected by the implementation.

The results of this evaluation (in milliseconds) can be found in Figure 6.1.

As expected, the experiment provides further evidence that the popularity-based approaches, such as MP, MP_u, MP_r and $MP_{u,r}$, the time-dependent approaches MR_u, GIRP and GIRPTM, and also the cognitive-inspired approaches BLL_{AC} and BLL_{AC}+MP_r perform significantly better than the more sophisticated approaches FR, LDA, PITF and 3LT+MP_r.

Memory consumption. The memory consumption measurements (in megabytes) of the algorithms are visualized in Figure 6.2. Here, the lowest memory is required by the PITF approach. One reason for the low memory consumption of PITF is surely the fact that it was developed in C++ (the other approaches were implemented in Java).

With respect to the memory consumption measurements of BLL_{AC}+MP_r, it is much smaller than the ones of CF, LDA and especially 3LT+MP_r. This is again due the fact that 3LT+MP_r relies on the computationally expensive topic modeling process, which has to be calculated for each resource in the dataset.

6.2.2 Evaluation in Related Settings

In order to demonstrate that BLL_{AC}+MP_r does not only provide accurate results in folksonomy settings, this section shows evaluation results in related settings: (i) the ECML PKDD Discovery Challenge 2009 dataset, and (ii) two datasets from Technology Enhanced Learning projects.

Evaluation on the ECML PKDD Discovery Challenge 2009 Dataset

In order to increase the reproducibility of the presented results and to ensure that these results can be compared with the results of other studies, another experiment was conducted on the well-known ECML PKDD discovery challenge 2009 dataset[1].

Experimental setup. This dataset is an rather "old" snapshot (from 2009) of BibSonomy at p-core level 2 consisting of 64,406 bookmarks, 1,185 users, 22,389 resources, 13,276 tags and 253,615 tag assignments, but is used in many related tag recommendation studies. Additionally, the dataset provides already a given train / test set split, which further ensures the comparability of results.

[1] `http://www.kde.cs.uni-kassel.de/ws/dc09/`

Algorithm	$F1@5$	Reference
GIRP	.087	[Trattner et al., 2016]
MP_u	.098	[Rendle and Schmidt-Thieme, 2009]
BLL	.104	[Trattner et al., 2016]
$\mathbf{BLL}_{AC}$	.238	[Trattner et al., 2016]
GIRPTM	.248	[Trattner et al., 2016]
MP_r	.288	[Rendle and Schmidt-Thieme, 2009]
$\mathrm{MP}_{u,r}$	.290	[Rendle and Schmidt-Thieme, 2009]
FR	.290	[Rendle and Schmidt-Thieme, 2009]
CF	.295	[Rendle and Schmidt-Thieme, 2009]
PITF	.302	[Rendle and Schmidt-Thieme, 2009]
$\mathbf{BLL}_{AC}$+$\mathbf{MP}_r$	.308	[Trattner et al., 2016]
Challenge winner	.355	[Rendle and Schmidt-Thieme, 2009]

Table 6.3: $F1@5$ estimates for selected algorithms on the ECML PKDD Discovery Challenge 2009 dataset [Trattner et al., 2016]. It can be seen that BLL_{AC}+MP_r is only outperformed by the winning algorithm (optimized ensemble of Factorization Machines [Rendle and Schmidt-Thieme, 2009]).

The winning algorithm based on the $F1@5$ evaluation metric in this tag recommender challenge was an optimized ensemble of factorization machines algorithms and was proposed by [Rendle and Schmidt-Thieme, 2009]. In Table 6.3, the results presented in [Rendle and Schmidt-Thieme, 2009] together with the results of the novel time-dependent approaches of this thesis are shown.

Results. The $F1@5$ estimates indicate that the dataset and the splitting method is of advantage for resource-based approaches since MP_r clearly outperforms MP_u. Interestingly, GIRP [Zhang et al., 2012], reaches an even lower $F1@5$ score than MP_u, which also indicates that the information of time seems not to be important in this setting. However, BLL reaches a higher $F1@5$ score than MP_u, which again shows the advantage of its power decay function.

Another indication of the importance of the current semantic context in form of resource tags, is given by the very good results of the BLL_{AC} approach, which are similar to the results of GIRPTM. Although, BLL_{AC} still recommends only tags already used by the given user, it adjusts the ranking using already assigned resource tags (i.e., the current semantic context).

Moreover, BLL_{AC}+MP_r reaches a $F1@5$ score of .308 and thus, again outperforms other sophisticated methods such as GIRPTM, CF, FR, FM and PITF. With regard to the final ECML PKDD discovery challenge 2009 ranking, this would result in the 8th position without any optimizations to the dataset or the length of the recommended tag list. Additionally, BLL_{AC}+MP_r is much more efficient in terms of computational complexity than the better performing approaches (especially the ones based on Factorization Machines; see also Table 6.2) and can be executed for this dataset on a single machine in a few seconds.

Summed up, the results of this experiment show that BLL_{AC}+MP_r is capable of providing high estimates of recommender accuracy in different settings without the need of dataset optimization or complex calculation steps.

Evaluation on Technology Enhanced Learning Datasets

In the previous evaluations, the performance of the tag recommendation algorithms has been shown on datasets gathered from large social tagging systems. However, these settings differ from Technology Enhanced Learning (TEL) settings, in which information like ontologies, learning object metadata and even user ratings are very limited [Manouselis et al., 2011]. Specifically, the spectrum of commonly available tracked learner activities varies greatly, but typically includes implicit usage data like learner-ids, some general information on learning resources, timestamps and indications of a user's interest in learning resources (e.g., opening, downloading or bookmarking) [Verbert et al., 2012].

Thus, it is the aim of this section to show the performance of BLL_{AC}+MP_r also in these small and sparse TEL environments.

Experimental setup. Two datasets from the TEL projects MACE and TravelWell are used for this study. Additionally, results for CiteULike and BibSonomy are provided in order to increase comparability.

In the *MACE* project, an informal learning platform was created that links different repositories from all over Europe to provide access to meta-data-enriched learning resources from the architecture domain. The dataset encompasses user activities like the accessing and tagging of learning resources and additional learning resource descriptions such as topics and competences [Stefaner et al., 2007]. The MACE dataset used in this study consisted of 23,017 users, 627 resources and 12,360 tags.

Dataset	Metric	MP_u	MP_r	$MP_{u,r}$	CF	$\mathbf{BLL}_{AC}$	$\mathbf{BLL}_{AC}\mathbf{+MP}_r$
CiteULike	R@5	.3665	.0631	.3933	.3639	.4114	**.4325**
	P@5	.1687	.0323	.1829	.1698	.1897	**.2003**
	F1@5	.231	.042	.249	.231	.259	**.273**
	F1@10	.1672	.0294	.1825	.1560	.1797	**.1928**
	nDCG	.367	.063	.392	.359	.422	**.438**
BibSonomy	R@5	.3486	.0862	.3839	.3530	.3809	**.4071**
	P@5	.1991	.0572	.2221	.2066	.2207	**.2359**
	F1@5	.253	.068	.281	.260	.279	**.298**
	F1@10	.1879	.0523	.2131	.1875	.2028	**.2237**
	nDCG	.371	.091	.407	.369	.409	**.434**
TravelWell	R@5	.2207	.0714	.2442	.1740	.2491	**.2828**
	P@5	.1000	.0366	.1333	.0800	.1300	**.1400**
	F1@5	.137	.048	.172	.109	.170	**.187**
	F1@10	.1125	.0388	.1356	.0744	.1287	**.1426**
	nDCG	.241	.080	.268	.173	.278	**.290**
MACE	R@5	.1306	.0510	.1463	.1522	.1775	**.1901**
	P@5	.0576	.0173	.0618	.0631	.0812	**.0812**
	F1@5	.079	.025	.086	.089	.111	**.113**
	F1@10	.0662	.0170	.0692	.0615	.0829	**.0848**
	nDCG	.133	.048	.147	.156	.183	**.190**

Table 6.4: Results of the tag recommendation evaluation in TEL settings. It can be seen that the cognitive-inspired $BLL_{AC}+MP_r$ clearly outperforms its competitors [Kopeinik et al., 2016b].

Originating from the Learning Resource Exchange platform[2], the *TravelWell* dataset captures teachers' search for and access of open educational resources from a variety of providers all over Europe. Thus, it covers multiple languages and subject domains. Activities in the dataset are supplied in two files with either bookmarks or ratings [Vuorikari and Massart, 2010]. For this study, the bookmarks file was used, which contained 2,572 users, 97 resources and 1,890 tags.

Results. The results of this study are shown in Table 6.4 by means of R@5, P@5, F1@5, F1@10 and nDCG. As expected, the recommendation accuracy for the algorithms is smaller in the two TEL datasets MACE and TravelWell than in CiteULike and BibSonomy. This is due to the fact that these datasets are very small and sparse, which makes it very hard to predict tag usage. However, $BLL_{AC}+MP_r$ can

[2] `http://lreforschools.eun.org`

Algorithm	Accuracy			Diversity	Novelty	Runtime	Memory
	Narrow	Mixed	Broad				
MP_r	-	-		-	-	++	+
$MP_{u,r}$						++	+
CF				+			
LDA	-	-		++		-	-
PITF	-		+	++	+	-	++
FR			+	++			
GIRPTM	+	+	+			++	+
3LT+MP_r	++	++	++			-	-
BLL_{AC}	+	+		-	++	+	
BLL_{AC}+MP_r	++	++	++			+	

Table 6.5: Summary of the evaluation of the tag recommendation algorithms showing the relation between the performance of the algorithms and the given evaluation metric. This table shows tag recommender accuracy in narrow, mixed and broad settings, diversity, novelty, runtime and memory consumption. Please note that "++" indicates best, "+" good, "-" poor and an empty space average performance.

again be identified as the best performing algorithm both in non-TEL and TEL settings.

Because runtime and computational complexity are considered very important factors in most TEL environments [Manouselis et al., 2010], the performance of $MP_{u,r}$, which outperforms the comparably cost-intensive CF approach in three of four datasets, should be emphasized. Hence, it forms a good alternative in runtime-sensitive settings.

6.3 Summary

In this chapter, the implementation process of a hybrid tag recommendation algorithm (i.e., BLL_{AC}+MP_r) for real-world folksonomies was presented. This approach is based on the tag reuse prediction algorithm BLL_{AC} illustrated in Chapter 5 and extends it via incorporating social influences (i.e., imitating and recommending popular tags of other users).

Furthermore, BLL_{AC}+MP_r was validated in several real-work folksonomy datasets and settings by comparing it to state-of-the-art tag recommendation algorithms via various evaluation metrics. The main results regarding this research question are summarized in Table 6.5. These results contributed to Research Question 3, *"Can a* *tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"* and led to the following four findings:

1. BLL_{AC} can be expanded with *tag imitation processes* in order to realize the hybrid tag recommendation algorithm BLL_{AC}+MP_r.

2. BLL_{AC}+MP_r provides the *most robust results over all datasets* with respect to various evaluation metrics and folksonomy settings.

3. BLL_{AC}+MP_r also provides accurate recommendations in related settings such as the *ECML PKDD Discovery Challenge 2009* dataset, and two datasets from *Technology Enhanced Learning* projects.

4. In contrast to related cognitive-inspired tag recommendation approaches such as 3LT+MP_r, *BLL_{AC}+MP_r provides reasonable results with respect to computational costs.*

Based on these findings, Research Question 3 can be answered positively, which shows that BLL_{AC}+MP_r can be applied for various use cases in the area of tag recommendations in social tagging systems. In the next chapter, this approach is generalized for the task of recommending hashtags in Twitter in order to contribute to Research Question 4, *"Given that activation processes in human memory can be modeled to improve tag recommendations, can they also be utilized for hashtag recommendations in Twitter?"*.

Chapter 7

Research Question 4: Utilizing the Approach for Hashtag Recommendations in Twitter

"How do you feel about using # (pound) for groups. As in #barcamp [msg]?" [Messina, 2007]

This chapter describes how activation processes in human memory can be utilized for recommending hashtags in Twitter. Thus, it is the aim of this chapter to demonstrate that the cognitive-inspired tag recommendation algorithm proposed in this thesis (i.e., BLL_{AC}+MP_r) can be generalized for related use cases in the field of tag-based recommender systems, such as hashtag recommendations in Twitter (Section 7.1). Therefore, data collections are crawled from Twitter (Section 7.2) and temporal dynamics are studied in these data collections (Section 7.3) in order to propose two cognitive-inspired hashtag recommendation approach called $\mathrm{BLL}_{I,S}$ and $\mathrm{BLL}_{I,S,C}$ (Section 7.4). These approaches are evaluated against various baseline algorithms (Section 7.5) in order to address Research Question 4, *"Given that activation processes in human memory can be modeled to improve tag recommendations, can they also be utilized for hashtag recommendations in Twitter?"*.

This chapter is based on *P10* [Kowald et al., 2017b].

7.1 Hashtag Recommendations

Over the past years, the microblogging platform Twitter has become one of the most popular social networks on the Web. Users can build a network of follower connections to other Twitter users, which means that they can subscribe to content posted by their *followees* [Myers and Leskovec, 2014, Kwak et al., 2010]. Twitter was also the first social platform that adopted the concept of *hashtags*, as suggested by Chris Messina[1]. Meanwhile, many social platforms, such as Instagram and Facebook, have adopted hashtags as well.

Unsurprisingly, the widespread acceptance of hashtags has sparked a lot of research in the field of *hashtag recommendations* to support users in assigning the most descriptive hashtags to their tweets. As described in Section 2.2.2, existing methods typically utilize collaborative, content and topic features of tweets to recommend hashtags to users. Undoubtedly, these features play an important role in recommending hashtags that best describe a tweet but lack of a way of predicting which hashtags a user will likely apply in a newly created tweet given previous hashtag assignments [Kowald et al., 2017b].

As shown in the previous chapters of this thesis, activation processes in human memory can be modeled by means of the activation equation of ACT-R in order to predict and recommend tags in social tagging systems. Thus, it is the aim of this chapter to show that activation processes in human memory can also be utilized for the recommendation of hashtags in Twitter (Research Question 4). This demonstrates that the cognitive-inspired tag recommendation approach proposed in this thesis (i.e., BLL_{AC}+MP_r) can be generalized for related use cases in the area of tag-based recommender systems.

7.2 Twitter Data Collections

In this section, the data collection procedure and the two datasets used to address Research Question 4 are described. Additionally, individual as well as social hashtag reuse patterns are investigated in these datasets as a prerequisite for the proposed hashtag recommendation approaches. This section is based on P10 [Kowald et al., 2017b].

[1] `https://twitter.com/chrismessina/status/223115412`

7.2.1 Crawling Strategy and Dataset Statistics

Two datasets are crawled using the Search API of Twitter[2]. The final statistics of these datasets are illustrated in Table 7.1.

The first one (i.e., *CompSci* dataset) consists of researchers from the field of computer science and their followees, while the second one (i.e., *Random* dataset) consists of random people and their followees. These two datasets are used to test the hashtag recommendation approaches in two different network settings: (i) a domain-specific one (i.e., the domain of computer scientists), and (ii) a more general one consisting of random Twitter users. The crawling strategy for both datasets comprises of the following four steps:

Crawl Seed Users

Firstly, a list of seed users U_S for each dataset is identified and crawled. In the case of the *CompSci* dataset, the users who were identified as computer scientists in the work of [Hadgu and Jäschke, 2014] are taken. In the case of the *Random* dataset, the Streaming API of Twitter[3] was used in October 2015 to get a stream of tweets. Then, the user-ids are extracted from those tweets to get the list of random seed users. From both user lists, all users with more than 180 followees are removed, which results in $|U_S| = 2{,}551$ seed users for the *CompSci* dataset and $|U_S| = 3{,}466$ seed users for the *Random* dataset. The threshold of using a maximum of 180 followees is chosen because the Twitter Search API only allows 180 requests per 15 minutes, which enables to crawl the tweets of all followees of a seed user within this reasonable time window.

Crawl Followees

Next, these follower relationships are used to crawl the followees F of the seed users in order to create a directed user network for analyzing the social influence on hashtag reuse. Based on the number of seed users, the average number of followees per seed user $|F|/|U_S| = 94$ in the case of the *CompSci* dataset and 72 in the case of the *Random* dataset. Following these notations, the set of followees of user u is denoted as F_u in the remainder of this chapter. Overall, this crawling procedure

[2] `https://dev.twitter.com/rest/public/search`
[3] `https://dev.twitter.com/streaming/overview`

Dataset	$\lvert U_S \rvert$	$\lvert F \rvert$	$\lvert U \rvert$	$\lvert T \rvert$	$\lvert HT \rvert$	$\lvert HTAS \rvert$
CompSci	2,551	241,225	91,776	5,649,359	1,081,403	9,161,842
Random	3,466	252,219	127,112	8,157,702	1,507,773	13,628,750

Table 7.1: Statistics of the *CompSci* and *Random* Twitter datasets. Here, $|U_S|$ is the number of seed users, $|F|$ is the number of followees of these seed users, $|U|$ is the number of distinct users, $|T|$ is the number of tweets, $|HT|$ is the number of distinct hashtags and $|HTAS|$ is the number of hashtag assignments.

results in $|U|$ = 91,776 distinct users for the *CompSci* dataset and $|U|$ = 127,112 distinct users for the *Random* dataset.

Crawl Tweets

In the third step, the 200 most recent tweets of all the users are crawled and the tweets without hashtags are removed. The threshold of a maximum of 200 most recent tweets is set because of another restriction of the Twitter Search API that only allows 200 tweets to be received per a single request. This crawling procedure results in $|T|$ = 5,649,359 tweets for the *CompSci* dataset with an average number of tweets per user $|T|/|U| = 61$, and $|T|$ = 8,157,702 tweets for the *Random* dataset with $|T|/|U| = 64$. The crawled tweets cover a time range from 2007 to 2015 for both datasets.

Extract Hashtag Assignments

In the final step of the crawling procedure, the hashtag assignments are extracted by searching for all words that start with a "#" character. This results in $|HTAS|$ = 9,161,842 hashtag assignments for $|HT|$ = 1,081,403 distinct hashtags in the *CompSci* network and $|HTAS|$ = 13,628,750 for $|HT|$ = 1,507,773 in the *Random* network. Thus, in both datasets, each distinct hashtag is used approximately 9 times on average and each user uses approximately 100 hashtag assignments in her tweets on average. Examples for popular hashtags are *#bigdata*, *#iot* and *#ux* in case of the *CompSci* dataset, and *#shahbag*, *#ff* and *#art* in case of the *Random* dataset.

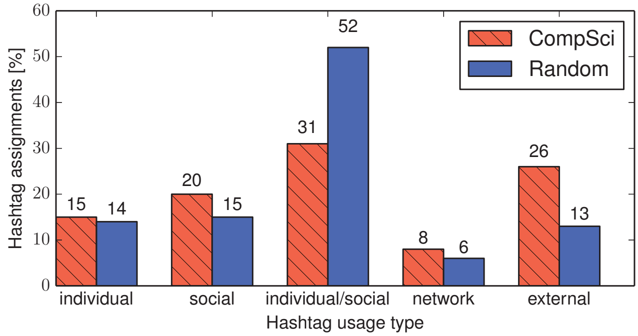


Figure 7.1: Analysis of hashtag usage types in Twitter. For each hashtag assignment, it is studied whether the corresponding hashtag has been used by the same user before in time ("individual"), by some of the users she follows ("social"), by both ("individual/social"), by anyone else in the dataset ("network") or neither of them ("external"). It can be seen that between 66% and 81% of hashtag assignments in both datasets can be explained by individual or social hashtag usage (i.e., the sum of "individual", "social" and "individual/social").

7.2.2 Analysis of Hashtag Usage Types

In these datasets, hashtag reuse practices are studied with the aim of identifying the different types of hashtag usages as a prerequisite for the proposed hashtag recommendation approaches. Specifically, for each hashtag assignment, it is studied whether the corresponding hashtag has either been used by the same user before ("individual"), by some of her followees ("social"), by both ("individual/social"), by anyone else in the dataset ("network") or by neither of them ("external").

The results of this study are shown in Figure 7.1. It can be seen that 66% of hashtag assignments in the *CompSci* dataset and 81% in the *Random* dataset can be explained by individual or social hashtag reuse. This finding further supports the choice to utilize these two types of influences (i.e., individual and social) for recommending hashtags. In contrast to these large numbers, the 6% to 8% of hashtags in the "network" category is relatively small. Interestingly, the amount of "external" hashtags is twice as high in the *CompSci* dataset (i.e., 26%) as in the *Random* one (i.e., 13%). Thus, in these datasets, computer scientists tend to use more hashtags, which have not been previously introduced in the network, than random Twitter users. Because of this, it can be assumed that the recommendation accuracy results would generally be lower in the *CompSci* dataset than in the *Random* one, which will be evaluated in Section 7.5. Summing up, both individual and social hashtags have an impact on users' choice of hashtags for a new tweet.

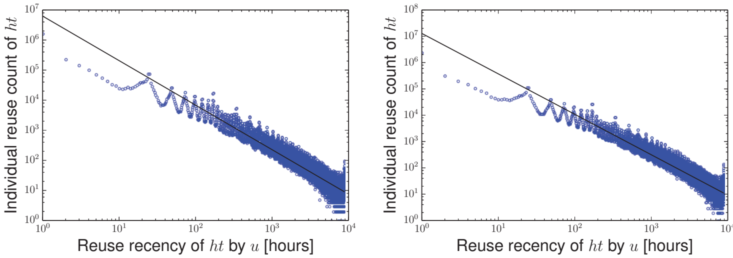
7.3 Temporal Effects on Hashtag Reuse

In this section, it is studied to what extent temporal effects play a role in the reuse of individual and social hashtags in the two datasets (i.e., *CompSci* and *Random*). Specifically, the recency of hashtags assignments (i.e., the time since the last hashtag usage/exposure), as well as whether this effect of time-dependent decay follows a power-law or exponential distribution. In Section 4.2, this was already analyzed in the context of social tagging systems. This section is based on P10 [Kowald et al., 2017b].

7.3.1 Temporal Effects on Individual Hashtag Reuse

The effect of time on individual hashtag reuse is visualized in the plots of Figure 7.2. To put the x-scale of these plots onto a meaningful range, the threshold for the maximum hashtag reuse recency is set to one year (i.e., 8,760 hours). The plots show the individual hashtag reuse count plotted over the reuse recency of a hashtag ht by a user u in hours. Hence, for each hashtag assignment of a hashtag ht by user u, the time since the last usage of ht by u (i.e., the reuse recency) is taken and all hashtag assignments with the same recency value (i.e., the same time difference in hours) are pooled together. The individual reuse count for this recency value is then given by the size of the set of these hashtag assignments.

The two plots show similar results for both datasets and indicate that the more recently a hashtag ht was used by a user u in the past, the higher its individual reuse count is. Interestingly, there is a clear peak after 24 hours in both datasets, which further indicates that users typically use the same set of hashtags in this time span and thus, tend to tweet about similar topics on a daily basis. Furthermore, high R^2 values of nearly .9 can be observed for the linear fits in the log-log scaled plots, which indicates that a large amount of the data can be explained by a power function. This is also suggested by the power-law-based model of the BLL equation



(a) Individual hashtag reuse *CompSci* dataset ($R^2 = .883$)

(b) Individual hashtag reuse *Random* dataset ($R^2 = .894$)

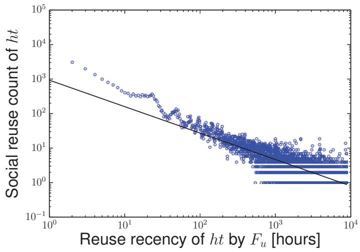
Figure 7.2: The effect of time on individual hashtag reuse for the *CompSci* and *Random* datasets (plots are in log-log scale). The plots show that the more recently a hashtag ht was used by a user u, the higher its individual reuse count (i.e., people tend to reuse hashtags that have been used very recently by their own). Additionally, the R^2 estimates for the linear fits of the data are reported. It can be seen that temporal effects play an important role in individual hashtag reuse in both datasets.

[Anderson et al., 2004]. In contrast, the linear fits in log-linear scaled plots only provide R^2 values of approximately .7, where high values would speak in favor of an exponential function.

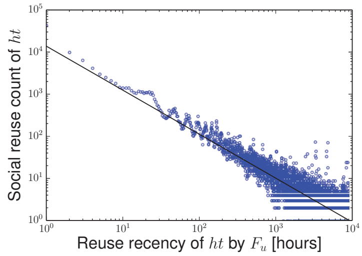
7.3.2 Temporal Effects on Social Hashtag Reuse

The plots of Figure 7.3 show the effect of time on the social hashtag reuse for the *CompSci* and *Random* datasets. These plots are created similarly as the plots of Figure 7.2 but this time the social hashtag reuse count is plotted over the reuse recency of a hashtag ht by the followees F_u of user u. Hence, for each hashtag assignment of ht by u, the most recent usage timestamp of ht by F_u is taken. The difference between this timestamp and the timestamp of the currently analyzed hashtag assignment indicates the time since the last social exposure of ht to u. Again, the threshold for the maximum hashtag reuse recency is set to one year (i.e., 8,760 hours).

In these plots, similar results can be observed since, in both datasets, the more recently a user was exposed to a hashtag, the higher its social reuse count. Furthermore, there is again (i) a clear peak after 24 hours, and (ii) the R^2 values for



(a) Social hashtag reuse
CompSci dataset (R^2 = .689)



(b) Social hashtag reuse
Random dataset (R^2 = .771)

Figure 7.3: The effect of time on social hashtag reuse for the *CompSci* and *Random* datasets (plots are in log-log scale). The plots show that the more recently a user u was exposed to a hashtag ht, which was used by her followees F_u, the higher its social reuse count (i.e., people tend to reuse hashtags that have been used recently in the social network). Additionally, the R^2 estimates for the linear fits of the data are reported. It can be seen that temporal effects play an important role in social hashtag reuse in both datasets.

the linear fits in the log-log scaled plots (i.e., = .7) are larger than in the log-linear scaled plots (i.e., = .4), which speaks in favor of a power function.

7.3.3 Power-Law Versus Exponential Time-Dependent Decay of Hashtag Reuse

The question whether a power or an exponential function is better suited to model the time-dependent decay of hashtag reuse is of interest especially for the design of a hashtag recommendation approach since both types of functions have been used in the area of time-aware recommender systems. While the BLL equation suggests the use of a power function to model the decay of item exposure in human memory [Anderson and Schooler, 1991], related hashtag recommender approaches, such as the one proposed in [Harvey and Crestani, 2015], use an exponential function for this purpose. As already mentioned, the visual inspection of Figures 7.2 and 7.3 and the R^2 values of the linear fits favor a power function. However, [Clauset et al., 2009] has shown that this least squares-based method can lead to misinterpretations and thus, a likelihood ratio-based test is suggested (see also Section 4.2).

Dataset	Parameter	Individual hashtag reuse	Social hashtag reuse
	x_{min}	141	1
CompSci	α	1.699	1.242
	R	**188**	**164**
	x_{min}	141	1
Random	α	1.723	1.269
	R	**235**	**294**

Table 7.2: Power-law versus exponential time-dependent decay in Twitter. It can be seen that a power function provides a better fit than an exponential function ($R > 0$ with $p < .001$) for explaining temporal effects on individual and social hashtag reuse in the two datasets.

Thus, the Python implementation [Alstott et al., 2014] of the method described in [Clauset et al., 2009] is used to validate if a power function produces a better fit than an exponential one. The results of this test are shown in Table 7.2. The main value of interest here is the log-likelihood ratio R between the two functions. It can be seen that $R > 0$ in all four cases with $p < .001$. This means that the power function indeed provides a better fit than the exponential function for explaining temporal effects on individual and social hashtag reuse. The x_{min} and α values of the fits are also provided. In this respect, the α slopes can be used to set the d parameter of the BLL equation (i.e., 1.7 in the individual case and 1.25 in the social case, see Section 7.4). Interestingly, these values are much higher than the suggested value of BLL's d parameter, which is .5 [Anderson et al., 2004]. It can be assumed that this is the case because tweeting is more strongly influenced by temporal interest drifts than other applications studied in the ACT-R community (e.g., [Anderson and Schooler, 1991]).

7.4 Hashtag Recommendation Approach

In the previous section, it was shown that temporal effects are important factors when users reuse individual and social hashtags. In this section, these insights are used as a basis to design the hashtag recommendation approach illustrated in Figure 7.4. Thus, one can distinguish between hashtag recommendations without (*Scenario 1*) and with (*Scenario 2*) incorporating the current tweet t. This section is based on P10 [Kowald et al., 2017b].

Whereas the first variant of this approach solely uses the past hashtags of a user u and/or her followees F_u, the second variant also utilizes the text of the current tweet t. Hence, these two scenarios also differ in their possible use cases since the first one aims to foresee the topics a specific user will tweet about based on the predicted hashtags, whereas the second one aims to support a user in finding the most descriptive hashtags for a new tweet text [Godin et al., 2013].

7.4.1 Scenario 1: Hashtag Recommendations Without the Current Tweet

For the first variant of the approach, the content of the current tweet t is ignored and solely past hashtag usages are utilized. As already stated, the BLL equation coming from the cognitive architecture ACT-R [Anderson et al., 2004] is used for this task. As visualized in *Scenario 1* of Figure 7.4, the BLL equation is adapted for (i) modeling the reuse of individual hashtags (BLL_I), (ii) modeling the reuse of social hashtags (BLL_S), and (iii) combining the former two into a hybrid recommendation approach ($\mathrm{BLL}_{I,S}$).

Modeling Individual Hashtag Reuse

In order to model the reuse of individual hashtags, the individual base-level activation $B_I(ht, u)$ of a hashtag ht for a user u is defined as follows:

$$B_I(ht, u) = \ln(\sum_{j=1}^{n} (TS_{ref} - TS_{ht,u,j})^{-d_I}) \tag{7.1}$$

where n denotes the number of times ht was used by u in the past (i.e., $|HTAS_{ht,u}|$) and the term $TS_{ref} - TS_{ht,u,j}$ states the recency of the jth usage of ht by u. In this respect, TS_{ref} is the reference timestamp (i.e., when recommendations should be calculated) and $TS_{ht,u,j}$ is the timestamp when ht was used by u for the jth time. Based on the results of the analysis presented in Table 7.2, the individual time-dependent decay factor d_I is set to 1.7.

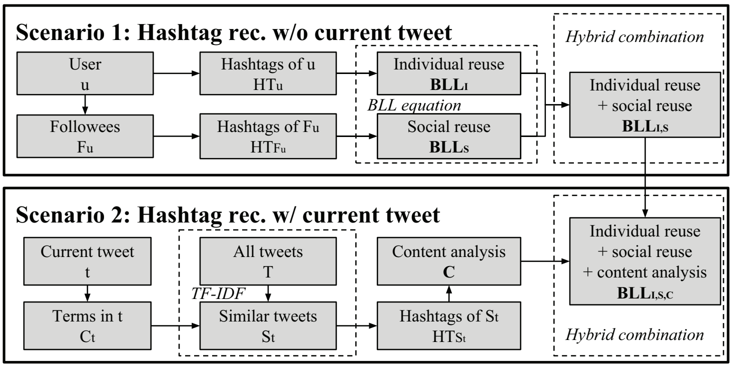


Figure 7.4: Schematic illustration of the cognitive-inspired approach for hashtag recommendations in Twitter. The approach is implemented in two scenarios (i.e., without and with incorporating the content of the current tweet). In *Scenario 1*, the BLL equation is used to realize (i) the individual BLL$_I$ algorithm, (ii) the social BLL$_S$ algorithm, and (iii) the hybrid BLL$_{I,S}$ algorithm, which combines both. In *Scenario 2*, a content analysis is used to identify similar tweets for a currently proposed tweet t and to identify the hashtags of the most similar ones. This content-based tweet analysis is combined with the BLL$_{I,S}$ method to provide personalized and content-aware hashtag recommendations in the form of the hybrid BLL$_{I,S,C}$ approach.

Modeling Social Hashtag Reuse

The reuse of social hashtags is modeled in a similar way but instead of analyzing how frequently and recently a hashtag ht was used by user u, it is analyzed how frequently and recently ht was used by the set of followees F_u of u. Thus, the social base-level activation $B_S(ht, u)$ of ht for u is formulated as follows:

$$B_S(ht, u) = \ln(\sum_{j=1}^{m} (TS_{ref} - TS_{ht,F_u,j})^{-d_S}) \tag{7.2}$$

where m is the number of times ht was used by F_u before the reference timestamp TS_{ref} (i.e., $|HTAS_{ht,F_u}|$). The term $TS_{ref} - TS_{ht,F_u,j}$ states the recency of the j^{th} exposure of ht to u caused by F_u, where $TS_{ht,F_u,j}$ is the timestamp when ht was used by F_u for the jth time. As when modeling the individual hashtag reuse, the social time-dependent decay factor d_S is set based on the results of the analysis presented in Table 7.2 (i.e., $d_S = 1.25$).

Combining Individual and Social Hashtag Reuse

In order to mix these two components in form of a hybrid approach, a linear combination is used in the same way as it was done to realize BLL$_{AC}$+MP$_r$ (see Section 6.1). Hence, in order to be able to add the individual and social base-level activations $B_I(ht, u)$ and $B_S(ht, u)$, these values have to be mapped onto a common range of 0 to 1 that add up to 1. Therefore, the softmax functions $\sigma(B_I(ht, u))$ and $\sigma(B_S(ht, u))$ are defined as proposed by [McAuley and Leskovec, 2013]. This is given by:

$$\sigma(B_I(ht, u)) = \frac{\exp(B_I(ht, u))}{\sum\limits_{ht' \in HT_u} \exp(B_I(ht', u))} \tag{7.3}$$

where HT_u is the set of distinct hashtags used by u. For $B_S(ht, u)$, the softmax function $\sigma(B_S(ht, u))$ can be calculated in the same way but on the basis of HT_{F_u} (i.e., the set of hashtags used by u's followees F_u).

Taken together, the combined base-level activation $B_{I,S}$ for the BLL$_{I,S}$ approach is given by a linear combination:

$$B_{I,S}(ht, u) = \beta \underbrace{\sigma(B_I(ht, u))}_{BLL_I} + (1 - \beta) \underbrace{\sigma(B_S(ht, u))}_{BLL_S} \tag{7.4}$$

where the β parameter can be used to give weights to the two components. Based on experimentation, β is set to .5 in order to equally weigh the individual and social influence. As indicated in Equation 7.4 and Figure 7.4, predictions can also be calculated either solely based on the individual hashtag reuse, referred as BLL$_I$, or the social hashtag reuse, referred as BLL$_S$.

7.4.2 Scenario 2: Hashtag Recommendations With the Current Tweet

As shown in *Scenario 2* of Figure 7.4, the second variant of the recommendation approach aims to provide hashtag suggestions while also incorporating the content of the currently proposed tweet t. Thus, the unpersonalized method proposed by [Zangerle et al., 2011] is used to find hashtags of similar tweets and combined with the $\mathrm{BLL}_{I,S}$ approach to generate personalized and content-aware recommendations.

Content-Based Tweet Analysis

The content of tweets is analyzed in order to find similar tweets for a target tweet t and to extract the hashtags of these similar ones. Therefore, the term frequency-inverse document frequency (TF-IDF) statistic is incorporated, which identifies the importance of a term for a document in a collection of documents. TF-IDF can be further used to calculate the similarity between two documents d and $\overline{d}$ by summing up the TF-IDF statistics of d's terms in $\overline{d}$. When applying this statistic to Twitter, tweets are treated as documents and the similarity between the target tweet t and a candidate tweet $\bar{t}$ is calculated as follows:

$$sim(t, \bar{t}) = \sum_{c \in C_t} n_{c,\bar{t}} \times \log(\frac{|T|}{|\{t' : c \in t'\}|}) \tag{7.5}$$

where C_t are the terms in the text of target tweet t, $n_{c,\bar{t}}$ is the number of times $c \in C_t$ occurs in the candidate tweet $\bar{t}$, $|T|$ is the number of tweets in the dataset and $|\{t' : c \in t'\}|$ is the number of times c occurs in any tweet $t' \in T$. The first factor of this equation reflects the term frequency TF, whereas the second factor of this equation reflects the inverse document frequency IDF [Zangerle et al., 2011].

Based on these similarity values, the most similar tweets S_t for t are identified and the hashtags used in these tweets (i.e., HT_{S_t}) are extracted. For each hashtag $ht \in HT_{S_t}$, a content-based score $CB(ht, t)$ is assigned, which is the highest similarity value within the most similar tweets S_t in which ht occurs. This method is implemented using the Lucene-based full-text search engine Apache Solr 4.7.10[4]. Based on Solr's software documentation and experimentation, the minimum term frequency tf is set to 2 and the minimum document frequency df to 5.

Combining Personalized and Content-Aware Hashtag Recommendations

The personalized $\mathrm{BLL}_{I,S}$ approach is combined with this content-based analysis (C) in order to generate personalized hashtag recommendations (see Figure 7.4). Again, this is achieved via a linear combination of both approaches. Taken together, the

[4]`http://lucene.apache.org/solr/`

top-k recommended hashtags $\widetilde{HT}_{u,t}$ for u and t are given by a linear combination:

$$\widetilde{HT}_{u,t} = \underset{ht \in \overline{HT}_{u,t}}{\arg\max^{k}} (\lambda \underbrace{B_{I,S}(ht, u)}_{BLL_{I,S}} + (1 - \lambda) \underbrace{\sigma(CB(ht, t))}_{C}) \tag{7.6}$$

where $\overline{HT}_{u,t}$ is the set of candidate hashtags for u and t (i.e., $HT_u \cup HT_{F_u} \cup HT_{S_t}$). The λ parameter is used to give weights to the personalized and content-aware components. To that end, λ is set to .3 based on experimentation. Please note that the content-based score $CB(ht, t)$ has to be normalized using the softmax function (see Equation 7.3), whereas $B_{I,S}(ht, u)$ is already normalized (see Equation 7.4). This finally constitutes the personalized hashtag recommendation algorithm termed $\text{BLL}_{I,S,C}$.

7.5 Hashtag Recommendation Evaluation

In this section, the evaluation procedure to address Research Question 4 is presented. This includes the methodology used as well as the results in terms of recommendation accuracy and ranking for the two scenarios (i.e., with and without the current tweet). This section is based on P10 [Kowald et al., 2017b].

7.5.1 Methodology

The methodology is given by the evaluation protocol, evaluation metrics and baseline algorithms used. Although this methodology is based on the the methodology described in Chapter 3, it was necessary to adapt it to the case of hashtag recommendations in Twitter and therefore, the most important aspects are described in this section.

Evaluation Protocol

In order to split the Twitter datasets into training and test sets, the evaluation protocol described in Section 3.4.1 is used. Therefore, for each seed user in the datasets (see Section 7.2) with at least two tweets (i.e., 2,020 users in the *CompSci* dataset and 2,679 users in the *Random* dataset), the most recent tweet of the user is determined and put into the test set. The remaining tweets are then put into the training set. This protocol ensures not only that the hashtags of at least one tweet per user are available for training but also that the chronological order of the data is preserved (i.e., future hashtags are predicted based on usage patterns of past ones). These training and test sets are used in the following two evaluation scenarios:

Hashtag recommendations w/o current tweet. In the first scenario, the content of the currently proposed tweet is ignored (i.e., the one in the test set) and hashtag predictions are provided solely based on the current user-id. Thus, in *Scenario 1*, all test set tweets are evaluated.

Hashtag recommendations w/ current tweet. In the second scenario, the content of the current tweet is also incorporated. In this setting, only the test set entries are evaluated, which do not include retweets (i.e., 954 test set tweets in the *CompSci* dataset and 1,504 test set tweets in the *Random* dataset). The reason for excluding the retweets from the test set in *Scenario 2* is that searching for similar tweets in the training set would result in identical tweets with identical hashtags, which would heavily bias the evaluation (see also [Zangerle et al., 2011]).

Evaluation Metrics

To finally quantify the quality of the algorithms, for each test set entry, the top-10 hashtags an algorithm predicts for the given user u and tweet t (i.e., $\widetilde{HT}_{u,t}$) are compared with the set of relevant hashtags actually used by u in t.

This comparison is done using various evaluation metrics, which have been described in Section 3.4.2. Specifically, Precision (P) and Recall (R) are reported for $k = 1$ to 10 predicted hashtags by means of Precision/Recall plots, and F1-score (F1@5) is reported for $k = 5$ predicted hashtags. Additionally, the three ranking-dependent metrics, namely (i) Mean Reciprocal Rank (MRR), (ii) Mean Average Precision (MAP), and (iii) Normalized Discounted Cumulative Gain (nDCG) are reported for $k = 10$ predicted hashtags.

Baseline Algorithms

The proposed approach is compared to a rich set of state-of-the-art hashtag recommendation algorithms. These algorithms are partly based on the approaches described in Section 3.3. Specifically, results for the following approaches are presented:

$\mathbf{MP}_I$. The *Most Popular Individual Hashtags* algorithm ranks the hashtags based on the frequency in the hashtag assignments of current user u. MP_I is also referred to as Most Popular Tags by User (MP_u) in tag recommendation terms [Jäschke et al., 2008].

$\mathbf{MR}_I$. *Most Recent Individual Hashtags* is a time-dependent variant of MP_I. MR_I suggests the k most recently used hashtags of current user u [Campos et al., 2014]. The BLL_I approach can be seen as an integrated combination of MP_I and MR_I based on human memory theory.

$\mathbf{MP}_S$. The *Most Popular Social Hashtags* algorithm is the social correspondent to the individual MP_I approach [Jäschke et al., 2008]. Thus, MP_S does not rank the hashtags based on the frequency in the hashtag assignments of user u but based on the frequency in the hashtag assignments of user u's set of followees F_u.

$\mathbf{MR}_S$. *Most Recent Social Hashtags* is the time-dependent equivalent to MP_S. MR_S sorts the hashtag assignments of u's followees F_u by time and recommends the k most recent ones. The BLL_S algorithm is a cognitive-inspired integration of MP_S and MR_S.

MP. The unpersonalized *Most Popular Hashtags* approach returns the same set of hashtags for any user. These hashtags are ranked by their overall frequency in the dataset [Jäschke et al., 2008].

FR & CF. The well-known *FolkRank* and *User-based Collaborative Filtering* tag recommendation approaches (see Sections 3.3.3 and 3.3.2) have been adapted for the task of hashtag recommendation by simply treating each tweet as a resource that has been tagged.

SR. *SimilarityRank* is an unpersonalized hashtag recommendation algorithm, which utilizes the content of the currently proposed tweet t [Zangerle et al., 2011]. Similarly to the $\mathrm{BLL}_{I,S,C}$ approach, this is achieved using TF-IDF to determine content-based similarity scores between tweets. These scores are used to recommend the k hashtags that occur in t's most similar tweets.

TCI. *TemporalCombInt* is one of the most recent approaches for personalized hashtag recommendations and also one of the very few approaches that accounts for the effect of time on hashtag usage [Harvey and Crestani, 2015]. TCI builds on a linear combination of SR and CF and incorporates temporal effects by considering

Dataset	Metric	MP_I	MR_I	$\mathbf{BLL}_I$	MP_S	MR_S	$\mathbf{BLL}_S$	MP	FR	CF	$\mathbf{BLL}_{I,S}$
CompSci	F1@5	.086	.098	**.101**	.022	.076	**.118**	.006	.083	.099	**.153*****
	MRR	.136	.188	**.193**	.032	.122	**.187**	.007	.130	.163	**.268*****
	MAP	.143	.195	**.202**	.033	.128	**.205**	.007	.136	.169	**.285*****
	nDCG	.175	.218	**.225**	.046	.154	**.235**	.012	.169	.196	**.324*****
Random	F1@5	.160	.169	**.175**	.072	.103	**.138**	.012	.159	.165	**.208*****
	MRR	.261	.300	**.314**	.109	.159	**.220**	.023	.260	.278	**.361*****
	MAP	.279	.315	**.335**	.116	.171	**.240**	.024	.279	.296	**.389*****
	nDCG	.323	.352	**.370**	.144	.205	**.280**	.035	.324	.333	**.434*****

Table 7.3: Recommender accuracy results for *Scenario 1* of Research Question 4. In this scenario, approaches are compared that ignore the current tweet content. It can be observed that (i) BLL_I outperforms MP_I and MR_I, (ii) BLL_S outperforms MP_S and MR_S, and (iii) $BLL_{I,S}$ outperforms MP, FR and CF. Based on a t-test, the symbols * ($\alpha = .1$), ** ($\alpha = .01$) and *** ($\alpha = .001$) indicate statistically significant differences between $BLL_{I,S}$ and CF.

the time-dependent relevance of a hashtag with respect to the recommendation date. This is done by categorizing the hashtags into "organizational" and "conversational" hashtags, and modeling the decay of temporal relevance using an exponential function. By fitting this model to the collected Twitter data, the two main parameters of the algorithm, η_l and η_h, are set to .1 and .2, respectively.

7.5.2 Results and Discussion

In Section 7.3, it was shown that time is an important factor for hashtag reuse. Because of this, it can be assumed that a time-dependent and cognitive-inspired approach should provide reasonable results compared to other algorithms. The accuracy estimates for the two evaluation scenarios are shown in Tables 7.3 and 7.4, and Figure 7.5.

Scenario 1: Hashtag Recommendations Without the Current Tweet

In our first evaluation scenario, approaches that predict future hashtags without incorporating the content of the currently proposed tweet are evaluated (see Table 7.3). Here, three main results are identified:

$\mathbf{BLL}_I$ > $\mathbf{MP}_I$, $\mathbf{MR}_I$. When predicting individual hashtag reuse, the BLL_I approach is compared to the frequency-based MP_I and the recency-based MR_I algorithms. The results clearly reflect the importance of the time component since MR_I and BLL_I provide higher prediction accuracy and ranking estimates than MP_I for all evaluation metrics across both datasets.

Apart from that, it can be observed that BLL_I outperforms MR_I, which speaks in favor of the cognitive-inspired combination of hashtag frequency and recency by means of the BLL equation.

$\mathbf{BLL}_S$ > $\mathbf{MP}_S$, $\mathbf{MR}_S$. Concerning the prediction of social hashtag reuse, the BLL_S approach is compared to the frequency-based MP_S and the recency-based MR_S methods. Similar to the case of individual hashtag reuse, MR_S and the BLL-based method provide higher accuracy estimates than the solely frequency-based one, but interestingly, this time the differences between these methods is much larger.

This indicates that the time information is especially important in a social setting. This behavior was somehow expected since typically only the most recent tweets of the followees are shown on a user's Twitter timeline. Again, the combination of hashtag frequency and recency by means of the BLL equation provides the best results.

$\mathbf{BLL}_{I,S}$ > MP, FR, CF. Finally, the hybrid $\mathrm{BLL}_{I,S}$ approach is compared to the unpersonalized MP algorithm, the well-known FR method from tag recommender research and classic user-based CF. The first observation that becomes apparent is the poor performance of the unpersonalized MP baseline, which underpins the importance of personalized methods for hashtag recommendation.

Additionally, and more importantly, the hybrid $\mathrm{BLL}_{I,S}$ approach does not only improve its BLL_I and BLL_S components but also provides significantly higher accuracy and ranking estimates than FR and CF. This shows that $\mathrm{BLL}_{I,S}$ is capable of providing reasonable hashtag recommendations solely based on temporal usage patterns of past hashtag assignments.

Scenario 2: Hashtag Recommendations With the Current Tweet

In the second scenario, hashtag recommendation methods that also incorporate the content of the current tweet are evaluated (see Table 7.4). This includes the unpersonalized SR approach, the time-dependent TCI algorithm and the $\mathrm{BLL}_{I,S,C}$ approach. The two main results are:

TCI, $\mathbf{BLL}_{I,S,C}$ > SR. The first main result of the second evaluation scenario is that

Dataset	Metric	SR	TCI	$\mathbf{BLL}_{I,S,C}$
CompSci	F1@5	.139	.182	**.200***
	MRR	.264	.334	**.395*****
	MAP	.283	.354	**.417*****
	nDCG	.299	.385	**.446****
Random	F1@5	.181	.243	**.261***
	MRR	.341	.436	**.489****
	MAP	.374	.472	**.530****
	nDCG	.388	.507	**.562****

Table 7.4: Recommender accuracy results for *Scenario 2* of Research Question 4. In this scenario, approaches that also incorporate the current tweet content are compared. It is shown that $\text{BLL}_{I,S,C}$ outperforms SR and TCI. Based on a t-test, the symbols * ($\alpha = .1$), ** ($\alpha = .01$) and *** ($\alpha = .001$) indicate statistically significant differences between $\text{BLL}_{I,S,C}$ and TCI.

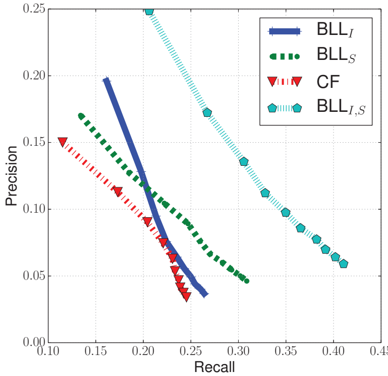
both time-dependent methods TCI and $\text{BLL}_{I,S,C}$ outperform the unpersonalized SR approach. The result was somehow expected since both TCI and $\text{BLL}_{I,S,C}$ extend the TF-IDF-based tweet content analysis of SR with personalization techniques via CF (TCI) or the BLL equation ($\text{BLL}_{I,S,C}$).

$\mathbf{BLL}_{I,S,C}$ > TCI. The second main result of *Scenario 2* is that $\text{BLL}_{I,S,C}$ provides significantly higher accuracy estimates than TCI. This is due to three main differences between these methods: (i) instead of using hashtags of similar users by means of CF for adding personalization, not only individual hashtags of the current user but also social hashtags of the current user's followees are incorporated, (ii) instead of applying the effect of time on a global hashtag level, the time-dependent decay is modeled on an individual and social level, and (iii) instead of modeling this time-dependent decay using an exponential function, a power function is used by means of the BLL equation.

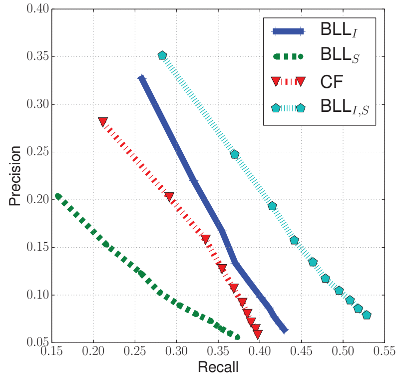
CompSci Dataset Versus Random Dataset

Another interesting finding that can be observed is that all algorithms provide better results for the *Random* dataset than for the *CompSci* dataset. This indicates that the task of predicting hashtags is harder in the domain-specific network of computer scientists than in the network of random users.

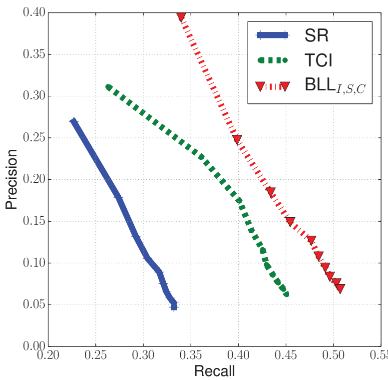
According to Figure 7.1, this makes sense since the amount of "external" hash-



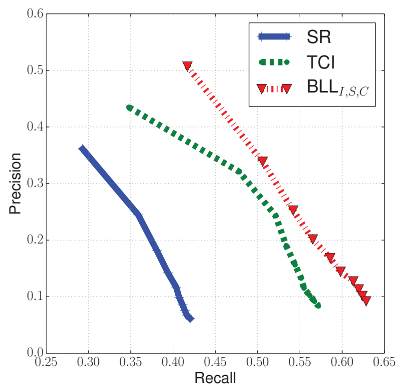
(a) *Scenario 1*: Hashtag recommendation w/o current tweet *CompSci* dataset



(b) *Scenario 1*: Hashtag recommendation w/o current tweet *Random* dataset



(c) *Scenario 2*: Hashtag recommendation w/ current tweet *CompSci* dataset



(d) *Scenario 2*: Hashtag recommendation w/ current tweet *Random* dataset

Figure 7.5: Precision / Recall plots of the two evaluation scenarios showing the accuracy of BLL$_I$, BLL$_S$, CF, BLL$_{I,S}$, SR, TCI and BLL$_{I,S,C}$ for k = 1 - 10 recommended hashtags. It is shown that BLL$_{I,S}$ provides the best results in *Scenario 1* and BLL$_{I,S,C}$ provides the best results in *Scenario 2*.

tags is twice as high in the *CompSci* dataset (i.e., 26%) than in the *Random* one (i.e., 13%). These "external" hashtags could be addressed in the recommendation process by using an additional knowledge source such as the list of currently trending hashtags.

7.6 Summary

In this chapter, the design, implementation and evaluation procedure of a cognitive-inspired approach for hashtag recommendations in Twitter (i.e., $\text{BLL}_{I,S}$) was presented. Based on the findings of the previous chapters of this thesis, this algorithm utilizes activation processes in human memory to account for temporal effects on individual hashtag reuse (i.e., reusing own hashtags) and social hashtag reuse (i.e., reusing hashtags, which has been previously used by a followee). Therefore, an analysis of hashtag usage types in two empirical networks (i.e., *CompSci* and *Random* datasets) crawled from Twitter was conducted, which revealed that between 66% and 81% of hashtag assignments can be explained by past individual and social hashtag usage. By analyzing the timestamps of these hashtag assignments, it was further shown that temporal effects play an important role for both individual and social reuse of hashtags and that a power function provides a better fit to model this time-dependent decay than an exponential function.

Based on this, the BLL equation of ACT-R was utilized to develop $\text{BLL}_{I,S}$ and $\text{BLL}_{I,S,C}$, two algorithms for recommending hashtags. Whereas $\text{BLL}_{I,S}$ aims to recommend hashtags without incorporating the current tweet (i.e., *Scenario 1*), $\text{BLL}_{I,S,C}$ also utilizes the content of the current tweet using the TF-IDF statistic (i.e., *Scenario 2*). Both algorithms were compared to state-of-the-art hashtag recommendation algorithms in order to contribute to Research Question 4 ("Given that activation processes in human memory can be modeled to improve tag recommendations, can they also be utilized for hashtag recommendations in Twitter?").

Summed up, the four main findings of this chapter are as follows:

1. A substantial amount of hashtag assignments in Twitter can be explained by *past individual and social hashtag usage*.

2. *Temporal effects* have an important influence on both *individual as well as social hashtag reuse*.

3. *A power function* is better suited to model this *time-dependent decay* than an exponential one.

4. $\text{BLL}_{I,S}$ outperforms related algorithms in the first evaluation scenario (i.e., *without the current tweet*) and $\text{BLL}_{I,S,C}$ in the second one (i.e., *with the current tweet*).

These findings show that activation processes in human memory can be utilized for the recommendation of hashtags in Twitter. This further demonstrates that the cognitive-inspired tag recommendation approach proposed in this thesis can be generalized for related use cases in the area of tag-based recommender systems, which positively answers Research Question 4. Additionally, this opens up a number of possible research strands for future work, such as the design of cognitive-inspired resource recommender systems, which will (among others) be discussed in the next (and final) chapter of this thesis.

Chapter 8

Conclusion

"One worthwhile task carried to a successful conclusion is worth half-a-hundred half-finished tasks."
(Malcolm S. Forbes, Publisher of Forbes magazine)

This thesis has modeled activation processes in human memory in order to improve tag recommendations. Therefore, the relation between activation processes in human memory and the reuse of tags in social tagging system has been investigated and a novel tag recommendation algorithm termed BLL_{AC}+MP_r has been proposed based on the activation equation of the cognitive architecture ACT-R. This algorithm has been evaluated in six real-work folksonomy datasets to demonstrate that a cognitive-inspired approach is able to outperform state-of-the-art tag recommendation methods that ignore these insights from cognitive science. Finally, the algorithm was adapted and extended for recommending hashtags in Twitter, which shows that activation processes in human memory can be utilized for related use cases in the area of tag-based recommender systems.

This chapter concludes this thesis by summarizing the achieved contributions in Section 8.1 and by discussing the impact of this dissertation research in Section 8.2. Apart from that, remaining open questions as well as possibilities for future work are presented in Section 8.3.

8.1 Contributions

This section describes the main scientific contributions that were achieved in course of this thesis.

8.1.1 Research Question 1: The Influence of Activation Processes in Human Memory on Tag Reuse

With respect to Research Question 1, *"How are activation processes in human memory influencing the tag reuse behavior of users in social tagging systems?"*, the relation between activation process in human memory and the reuse of tags was studied in Chapter 4.

Based on human memory theory [Anderson et al., 2004], the usefulness of a piece of information (e.g., a word or tag) depends on (at least) three factors: (i) past usage frequency, (ii) past usage recency, and (iii) the current semantic context. The results presented in Section 4.2.1 showed that these three factors also influence the reuse of tags in social tagging systems. This means that (i) the more frequently a tag was used in the past, the higher its reuse probability, (ii) the more recently a tag was used in the past, the higher its reuse probability, and (iii) the more similar a tag is to tags in the current semantic context, the higher its reuse probability. Apart from that, it was shown in Section 4.2.2 that the effect of recency on the reuse probability of tags is more likely to follow a power-law distribution than an exponential one, which is in line with the model of time-dependent decay of item exposure in human memory [Anderson and Schooler, 1991].

These findings verified the strong relation between activation processes in human memory and the use of tags in social tagging systems, which positively answered Research Question 1. Additionally, these findings were used as a prerequisite for designing a cognitive-inspired approach for tag reuse predictions and tag recommendations based on the activation equation of the cognitive architecture ACT-R.

8.1.2 Research Question 2: Designing a Cognitive-Inspired Algorithm for Tag Reuse Prediction

Based on the outcomes of Research Question 1, a novel cognitive-inspired approach for predicting the reuse of tags (i.e., BLL_{AC}) was proposed in Chapter 5. This contributed to Research Question 2, *"Can the activation equation of the cognitive architecture ACT-R, which accounts for activation processes in human memory, be exploited to develop a model for predicting the reuse of tags?"*.

Specifically, BLL_{AC} implements the activation equation of the cognitive architecture ACT-R [Anderson et al., 2004] to model (i) past tag usage frequency, (ii) past tag usage recency, and (iii) the current semantic context using a power-law model. This algorithm was evaluated with respect to tag reuse prediction accuracy and ranking using the methodology presented in Chapter 3. The evaluation results presented in Section 5.2.3 showed that BLL_{AC} outperforms not only algorithms reflecting its individual components and combinations of its components but also the well-known FolkRank algorithm [Hotho et al., 2006] in narrow and mixed folksonomy settings. In the broad folksonomy setting, however, FolkRank provided the best results, which showed the importance of incorporating also social influences for tag recommendations by means of imitating popular tags of other users.

Thus, it was proved that the activation equation of the cognitive architecture ACT-R can be exploited to predict the reuse of tags, which positively answered Research Question 2. Furthermore, the strong prediction accuracy results of the FolkRank algorithm in broad folksonomy settings indicated the importance of incorporating also social influences by means of tag imitation processes in order to realize a hybrid tag recommendation approach.

8.1.3 Research Question 3: Implementing a Hybrid Approach for Tag Recommendations in Real-World Folksonomies

The next main scientific contribution of thesis was the implementation and evaluation of a hybrid tag recommendation algorithm for real-world folksonomies termed $\text{BLL}_{AC}+\text{MP}_r$. In this respect, the results presented in Chapter 6 contributed to Research Question 3, *"Can a tag prediction model based on the activation equation be expanded with tag imitation processes in order to improve tag recommendations in real-world folksonomies?"*.

With respect to this research question, the contributions are threefold. Firstly, it was shown that BLL_{AC} can be expanded with tag imitation processes by means of popular tags of other users in order to realize $\text{BLL}_{AC}+\text{MP}_r$. Secondly, according to the evaluation results presented in Section 6.2.1, BLL_{AC}+MP_r provided the most robust results over all datasets with respect to various folksonomy settings and evaluation metrics. Thirdly, these results were validated in related settings such as the ECML PKDD Discovery Challenge 2009 dataset, and two datasets from Technology Enhanced Learning projects (see Section 6.2.2).

Based on these findings, Research Question 3 was answered positively, which showed that an algorithm based on activation processes in human memory can be utilized for various use cases in the area of tag recommendations for social tagging systems.

8.1.4 Research Question 4: Utilizing the Approach for Hashtag Recommendations in Twitter

In order to demonstrate the generalizability of the proposed approach, Research Question 4, *"Given that activation processes in human memory can be modeled to improve tag recommendations, can they also be utilized for hashtag recommendations in Twitter?"*, was investigated in Chapter 7.

Specifically, this research question tackled the design, implementation and evaluation procedure of two cognitive-inspired approaches for hashtag recommendations in Twitter (i.e., $\mathrm{BLL}_{I,S}$ and $\mathrm{BLL}_{I,S,C}$). In this respect, three contributions were achieved. Firstly, it was shown that a substantial amount of hashtag assignments in Twitter can be explained by past individual and social hashtag usage (see Section 7.2). Secondly, the analysis presented in Section 7.3 validated that temporal effects have an important influence on both individual as well as social hashtag reuse. Thirdly, the evaluation conducted in Section 7.5 showed that $\mathrm{BLL}_{I,S}$ outperforms state-of-the-art hashtag algorithms in two evaluation scenarios.

These findings showed that activation processes in human memory can be utilized for the recommendation of hashtags in Twitter, which positively answered Research Question 4. This further demonstrated that the cognitive-inspired tag recommendation approach proposed in this thesis (i.e., BLL_{AC}+MP_r) can be generalized for related use cases in the area of tag-based recommender systems.

8.1.5 TagRec: An Open-Source Tag Recommendation Evaluation Framework

Finally, on the methodological level (see Chapter 3), this thesis contributed with an open-source evaluation framework for tag-based recommender systems called *TagRec*.

The purpose of *TagRec* is to provide the research community with a standardized framework that supports all steps of the development and evaluation process of tag-based recommendation algorithms in a reproducible way. This includes methods for data preprocessing, data modeling and recommender evaluation.

In order to ensure reproducibility, all experiments conducted to address the four research questions of this thesis have been undertaken by using *TagRec*. Apart from that, all described datasets are freely available on the Web and can directly be processed using *TagRec*.

8.2 Impact

The impact of this dissertation research can be found in the areas of (i) recommender systems, (ii) social tagging, (iii) tag recommendation evaluation, and (iv) recommendations in informal learning settings.

8.2.1 Recommender Systems

With respect to the research area of recommender systems, this thesis showed that principles of human cognition (e.g., activation processes in human memory) can be utilized to improve tag recommendations [Kowald, 2015]. By taking into account how humans access information in their memory, cognitive-inspired recommendation strategies were developed as an alternative to data-driven methods such as Collaborative Filtering or Tensor Factorization [Kowald et al., 2014b, Kowald et al., 2015a].

The evaluation results presented in this thesis indicated that cognitive-inspired approaches are able to outperform purely data-driven algorithms not only in terms of recommendation accuracy but also with respect to computational efficiency [Kowald et al., 2015a, Trattner et al., 2016].

These findings had an impact on the research area of recommender systems by motivating the development of related cognitive-inspired recommendation algorithms (e.g., [Seitlinger et al., 2013, Kowald et al., 2015b, Seitlinger et al., 2015a, Kopeinik et al., 2016a]).

8.2.2 Social Tagging

This thesis contributed to the large body of research that analyzes interactions in social tagging systems. Specifically, the relation between activation processes in human memory and the reuse of tags was investigated, which resulted in a set of factors that influence the reuse of tags.

As shown in [Kowald and Lex, 2016], three of these factors are (i) past usage frequency, (ii) past usage recency, and (iii) the current semantic context cues. Especially, the second factor (i.e., recency) has led to further interesting investigations such as the study of temporal effects on hashtag reuse in Twitter [Kowald et al., 2017b].

In this respect, this thesis has also answered the question if a power-law or exponential distribution is better suited to model the time-dependent decay of tag reuse. While related work has used an exponential function for this purpose, the experiments presented in this thesis has shown that a power function should be favored [Kowald and Lex, 2016, Kowald et al., 2017b].

8.2.3 Tag Recommendation Evaluation

This thesis provided the to-date most extensive evaluation of tag recommendation algorithms [Kowald and Lex, 2015]. This included the evaluation of 20 tag recommendation algorithms on 6 datasets by using 10 evaluation metrics. This large-scale study provided the research community with a performance overview of state-of-the-art tag recommendation algorithms in various evaluation settings.

Apart from that, the open-source *TagRec* framework [Kowald et al., 2014a] was developed in course of this thesis. To date, *TagRec* supported the development and evaluation process of tag-based recommender systems in two large-scale European projects described in 17 research papers. This included use cases in the areas of (i) social tag recommendation, (ii) resource recommendations, (iii) recommendation evaluation, and (iv) hashtag recommendations [Kowald et al., 2017a].

8.2.4 Recommendations in Informal Learning Settings

Parts of this dissertation research were conducted in course of the large-scale European project Learning Layers[1] [Ley et al., 2013, Ley et al., 2014], which was about scaling informal learning at the workplace. In this respect, tag recommendations were used to support users in finding descriptive tags to bookmark learning resources. The $\text{BLL}_{AC}+\text{MP}_r$ approach developed in this thesis especially supported this informal learning setting, since $\text{BLL}_{AC}+\text{MP}_r$ not only accounts for the particular expertise of a user (via the BLL_{AC} component) but also for the knowledge of the collective (via the MP_r component)[2] [Santos et al., 2016]. The evaluation results presented in [Kopeinik et al., 2016b] demonstrated that this approach can successfully be applied in both social bookmarking as well as Technology Enhanced Learning settings.

Furthermore, the aforementioned *TagRec* framework was used as a recommendation engine in the Social Semantic Server (SSS)[3] [Dennerlein et al., 2015a, Kowald et al., 2013], which was the main back-end technology in the Learning Layers project. Thus, various informal learning tools such as KnowBrain[4] [Dennerlein et al., 2015b] and Bits & Pieces[5] [Dennerlein et al., 2014] were supported by the recommendation algorithms developed in the course of this thesis.

Finally, it is planned to use *TagRec* in the course of the H2020 European project AFEL[6], which is about analytics for everyday learning. Here, principles implemented in *TagRec* will be used to develop a recommendation algorithm for suggesting learning paths (i.e., ordered lists of connected learning resources for a specific topic) to users.

8.3 Open Questions and Future Work

This thesis opens up a number of promising research directions and possibilities for future work, which are discussed in this section.

[1] `http://learning-layers.eu/`
[2] `http://results.learning-layers.eu/infrastructure/sss/services/tag-rec/`
[3] `https://github.com/learning-layers/SocialSemanticServer`
[4] `https://github.com/learning-layers/KnowBrain`
[5] `https://github.com/learning-layers/BitsAndPieces`
[6] `http://afel-project.eu/`

"Are there other cognitive processes that can be utilized to model the use of tags in social tagging systems?"

This thesis focused on modeling activation processes in human memory by means of the activation equation of the declarative memory module of the cognitive architecture ACT-R. This means, that the tags of a user with the highest activation value in the current context are recommended.

However, the tag choices of a user are also influenced by other cognitive processes, for example, processes defined in the procedural memory module of ACT-R [Anderson et al., 2004]. Here, production rules are used to define actions based on a given information goal (e.g., looking for resources in the research area of "recommender systems"). In this respect, it would be very interesting to extend the current approach with such processes in order to decide if a given tag will be chosen by a given user who is guided by a specific information goal. One way to realize this could be the SNIF-ACT model [Fu and Pirolli, 2007, Pirolli and Fu, 2003, Pirolli et al., 2002].

Apart from that, the current implementation of tag imitation processes in this thesis is solely based on the popularity of other users' tags. Here, the semantic imitation model described in [Fu et al., 2010] could be applied to personalize these tag imitation processes. Similarly, the 3Layers algorithm [Seitlinger et al., 2013] could be used to replace MP_r with tags of semantically similar resources.

"Can content information of the resources be used to model the current semantic context of social tagging?"

One limitation of $\mathrm{BLL}_{AC}+\mathrm{MP}_r$ is that it uses other users' tags as cues to model the current semantic context. Since most publicly available social tagging datasets do not contain content information of resources (e.g., title or description text), this allows $\mathrm{BLL}_{AC}+\mathrm{MP}_r$ to solely utilize tagging data for this purpose. However, this approach only works for non cold-start resources that have already been tagged by other users.

Therefore, one potential strand of future work would be to integrate content data of resources. This idea is supported by the work of [Lipczak and Milios, 2010], in which it is shown that the title of a resource already has a large impact on the choice of tags in social tagging systems.

"Can $BLL_{AC}+MP_r$ be further improved by utilizing the fan effect of the cognitive

architecture ACT-R?"

The associative component of BLL_{AC}+MP_r defines the usefulness of a given tag in the current semantic context via tag co-occurrences. Consequently, general tags have more co-occurrences with other tags than specific ones, which results in a higher spreading activation.

In order to overcome this, [Anderson and Reder, 1999] defines the so-called fan effect, which ensures that the spreading activation of general tags is softened and the one of specific tags is increased. As future work, it is therefore planned to evaluate if the integration of the fan effect into BLL_{AC}+MP_r can further improve its prediction accuracy.

"Can the mixing parameter of the hybrid combination of BLL_{AC}+MP_r be optimized with respect to the given dataset?"

One limitation of BLL_{AC}+MP_r is that the mixing parameter β is set to its default value (i.e., .5), which results in equal weights of both components (i.e., BLL_{AC} and MP_r). However, the evaluation results presented in this thesis showed that the performance of MP_r depends on the given folksonomy type.

Thus, it would be interesting to test if β can be optimized based on the given dataset: for narrow folksonomies, β should be increased to give a higher weight to BLL_{AC}, and for broad folksonomies, β should be decreased to give a higher weight to MP_r.

"Can the offline evaluation results of this thesis be verified in online evaluation settings?"

This thesis followed common practice in the area of recommender systems to build on an offline study design. Although a rich set of baseline algorithms, evaluation metrics and real-world folksonomy datasets were used, only an online study is able to measure the real user acceptance of the generated recommendations [Jäschke et al., 2009].

Thus, it is planned to adapt and use open-source social bookmarking tools such as KnowBrain [Dennerlein et al., 2015b] or SemanticScuttle[7] to conduct a user study, in which BLL_{AC} is compared with other algorithms (e.g., 3Layers [Seitlinger et al., 2013]). Such setting would not only allow to evaluate the user acceptance of the

[7] `http://semanticscuttle.sourceforge.net/`

algorithms but also to test the algorithms in different scenarios (e.g., individual versus collaborative tagging).

"Can the proposed hashtag recommendation approach $BLL_{AC}+MP_r$ be improved if additional user information is taken into account?"

One limitation of the $\text{BLL}_{I,S}$ and $\text{BLL}_{I,S,C}$ hashtag recommendation algorithms is that the reuse of social hashtags is solely modeled by analyzing how frequently and recently a hashtag was used by a user's followees. Thus, these approaches neglect by whom the hashtag was used, which should influence a user's choice to adapt a followee's hashtags.

Thus, for future research, it is planed to extend these approaches by incorporating the social status of the followee. According to [Hasani-Mavriqi et al., 2015, Hasani-Mavriqi et al., 2016], the reputation of the users (e.g., by means of the number of followers) could be used as a proxy to model the social status. In this respect, the social connection strength between the users (e.g., by means of the number of mentions or retweets) could also be integrated to further extend these approaches.

"Can the findings of this thesis be used for the design and implementation of related cognitive-inspired recommender systems?"

This thesis showed that activation processes in human memory can be utilized for both tag recommendations in social tagging systems and hashtag recommendations in Twitter. These findings open up possible research strands for future work, such as the design of cognitive-inspired recommender systems (e.g., for resource recommendations).

A first attempt in this direction was presented in [Lacic et al., 2014b], where the BLL equation of ACT-R is used to improve Collaborative Filtering to recommend resources to users. Another example is the work of [Seitlinger et al., 2015a, Kopeinik et al., 2016a], in which a model of human category learning is used to further personalize Collaborative Filtering-based recommendations.

Thus, one possibility for future work would be to integrate such cognitive-inspired recommender approaches into a generic recommender framework such as ScaR (Scalable Recommendation-as-a-Service)[8] [Lacic et al., 2014a, Lacic et al., 2014c] in order to test their applicability in various application settings (e.g., E-commerce systems).

[8] `http://scar.know-center.tugraz.at/`

Finally, the author of this thesis believes, that general future research of recommender systems should focus on a hybrid combination of (i) cognitive-inspired, and (ii) data-driven (e.g., machine learning-based) approaches. This way, the strengths of both types of algorithms can be combined in order to adapt to the given data (i.e., large-scale versus small-scale settings). This thesis already provided a first step into this research direction.

Bibliography

[Alstott et al., 2014] Alstott, J., Bullmore, E., and Plenz, D. (2014). powerlaw: a python package for analysis of heavy-tailed distributions. *PloS one*, 9(1):e85777.

[Anderson, 1983] Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior*, 22(3):261–295.

[Anderson, 1996] Anderson, J. R. (1996). Act: A simple theory of complex cognition. *American Psychologist*, 51(4):355.

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

[Anderson et al., 1997] Anderson, J. R., Matessa, M., and Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, 12(4):439–462.

[Anderson and Reder, 1999] Anderson, J. R. and Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology General*, 128:186–197.

[Anderson and Schooler, 1991] Anderson, J. R. and Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science*, 2(6):396–408.

[Baeza-Yates et al., 1999] Baeza-Yates, R., Ribeiro-Neto, B., et al. (1999). *Modern information retrieval*, volume 463. ACM press New York.

[Balby Marinho et al., 2012] Balby Marinho, L., Hotho, A., Jäschke, R., Nanopoulos, A., Rendle, S., Schmidt-Thieme, L., Stumme, G., and Symeonidis, P. (2012).

Recommender Systems for Social Tagging Systems. SpringerBriefs in Electrical and Computer Engineering. Springer.

[Batagelj and Zaveršnik, 2002] Batagelj, V. and Zaveršnik, M. (2002). Generalized cores. *arXiv preprint cs/0202039*.

[Belém et al., 2013] Belém, F., Martins, E., Almeida, J., and Gonçalves, M. (2013). Exploiting novelty and diversity in tag recommendation. In *Advances in Information Retrieval*, pages 380–391. Springer.

[Belém et al., 2011] Belém, F., Martins, E., Pontes, T., Almeida, J., and Gonçalves, M. (2011). Associative tag recommendation exploiting multiple textual features. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 1033–1042. ACM.

[Belém et al., 2016] Belém, F. M., Batista, C. S., Santos, R. L. T., Almeida, J. M., and Gonçalves, M. A. (2016). Beyond relevance: Explicitly promoting novelty and diversity in tag recommendation. *ACM Transactions on Intelligent Systems and Technologies*, 7(3):26:1–26:34.

[Blei et al., 2003] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.

[Campos et al., 2014] Campos, P. G., Díez, F., and Cantador, I. (2014). Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2):67–119.

[Cattuto et al., 2007] Cattuto, C., Loreto, V., and Pietronero, L. (2007). Semiotic dynamics and collaborative tagging. *Proceedings of the National Academy of Sciences*, 104(5):1461–1464.

[Chi and Mytkowicz, 2008] Chi, E. H. and Mytkowicz, T. (2008). Understanding the efficiency of social tagging systems using information theory. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 81–88. ACM.

[Clauset et al., 2009] Clauset, A., Shalizi, C. R., and Newman, M. E. (2009). Power-law distributions in empirical data. *SIAM review*, 51(4):661–703.

[Cress et al., 2013] Cress, U., Held, C., and Kimmerle, J. (2013). The collective knowledge of social tags: Direct and indirect influences on navigation, learning, and information processing. *Computers & Education*, 60(1):59–73.

[Dellschaft and Staab, 2008] Dellschaft, K. and Staab, S. (2008). An epistemic dynamic model for tagging systems. In *Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 71–80. ACM.

[Dellschaft and Staab, 2012] Dellschaft, K. and Staab, S. (2012). Measuring the influence of tag recommenders on the indexing quality in tagging systems. In *Proceedings of the 23rd ACM conference on Hypertext and social media*, pages 73–82. ACM.

[Dennerlein et al., 2015a] Dennerlein, S., Kowald, D., Lex, E., Theiler, D., Lacic, E., and Ley, T. (2015a). The social semantic server: a flexible framework to support informal learning at the workplace. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*, page 26. ACM.

[Dennerlein et al., 2014] Dennerlein, S., Rella, M., Tomberg, V., Theiler, D., Treasure-Jones, T., Kerr, M., Ley, T., Al-Smadi, M., and Trattner, C. (2014). Making sense of bits and pieces: A sensemaking tool for informal workplace learning. In *European Conference on Technology Enhanced Learning*, pages 391–397. Springer.

[Dennerlein et al., 2015b] Dennerlein, S., Theiler, D., Marton, P., Rodriguez, P. S., Cook, J., Lindstaedt, S., and Lex, E. (2015b). Knowbrain: An online social knowledge repository for informal workplace learning. In *Design for Teaching and Learning in a Networked World*, pages 509–512. Springer.

[Doerfel and Jäschke, 2013] Doerfel, S. and Jäschke, R. (2013). An analysis of tag-recommender evaluation procedures. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 343–346. ACM.

[Doerfel et al., 2016] Doerfel, S., Jäschke, R., and Stumme, G. (2016). The role of cores in recommender benchmarking for social bookmarking systems. *ACM Trans. Intell. Syst. Technol.*, 7(3):40:1–40:33.

[Efron, 2010] Efron, M. (2010). Hashtag retrieval in a microblogging environment. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 787–788. ACM.

[Floeck et al., 2010] Floeck, F., Putzke, J., Steinfels, S., Fischbach, K., and Schoder, D. (2010). Imitation and quality of tags in social bookmarking systems–collective intelligence leading to folksonomies. In *On collective intelligence*, pages 75–91. Springer.

[Fu, 2008] Fu, W.-T. (2008). The microstructures of social tagging: a rational model. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 229–238. ACM.

[Fu et al., 2010] Fu, W.-T., Kannampallil, T., Kang, R., and He, J. (2010). Semantic imitation in social tagging. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 17(3):12.

[Fu et al., 2009] Fu, W.-T., Kannampallil, T. G., and Kang, R. (2009). A semantic imitation model of social tag choices. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 4, pages 66–73. IEEE.

[Fu and Pirolli, 2007] Fu, W.-T. and Pirolli, P. (2007). Snif-act: A cognitive model of user navigation on the world wide web. *Human–Computer Interaction*, 22(4):355–412.

[Gemmell et al., 2009] Gemmell, J., Schimoler, T., Ramezani, M., Christiansen, L., and Mobasher, B. (2009). Improving folkrank with item-based collaborative filtering. *Recommender Systems & the Social Web*.

[Godin et al., 2013] Godin, F., Slavkovikj, V., De Neve, W., Schrauwen, B., and Van de Walle, R. (2013). Using topic models for twitter hashtag recommendation. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 593–596. ACM.

[Golder and Huberman, 2006] Golder, S. A. and Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of information science*, 32(2):198–208.

[Gunawardana and Shani, 2009] Gunawardana, A. and Shani, G. (2009). A survey of accuracy evaluation metrics of recommendation tasks. *Journal of Machine Learning Research*, 10(Dec):2935–2962.

[Hadgu and Jäschke, 2014] Hadgu, A. T. and Jäschke, R. (2014). Identifying and analyzing researchers on twitter. In *Proceedings of the 2014 ACM conference on Web science*, pages 23–32. ACM.

[Halpin et al., 2007] Halpin, H., Robu, V., and Shepherd, H. (2007). The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web*, pages 211–220. ACM.

[Hamouda and Wanas, 2011] Hamouda, S. and Wanas, N. (2011). Put-tag: personalized user-centric tag recommendation for social bookmarking systems. *Social network analysis and mining*, 1(4):377–385.

[Harvey and Crestani, 2015] Harvey, M. and Crestani, F. (2015). Long time, no tweets! time-aware personalised hashtag suggestion. In *Proceedings of the 2015 European Conference on Information Retrieval*, pages 581–592. Springer.

[Hasani-Mavriqi et al., 2015] Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2015). The influence of social status on consensus building in collaboration networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 162–169. ACM.

[Hasani-Mavriqi et al., 2016] Hasani-Mavriqi, I., Geigl, F., Pujari, S. C., Lex, E., and Helic, D. (2016). The influence of social status and network structure on consensus building in collaboration networks. *Social Network Analysis and Mining*, 6(1):80.

[Held et al., 2012] Held, C., Kimmerle, J., and Cress, U. (2012). Learning by foraging: The impact of individual knowledge and social tags on web navigation processes. *Computers in Human Behavior*, 28(1):34–40.

[Helic et al., 2012] Helic, D., Körner, C., Granitzer, M., Strohmaier, M., and Trattner, C. (2012). Navigational efficiency of broad vs. narrow folksonomies. In

Proceedings of the 23rd ACM conference on Hypertext and social media, pages 63–72. ACM.

[Helic et al., 2010] Helic, D., Trattner, C., Strohmaier, M., and Andrews, K. (2010). On the navigability of social tagging systems. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 161–168. IEEE.

[Helic et al., 2011] Helic, D., Trattner, C., Strohmaier, M., and Andrews, K. (2011). Are tag clouds useful for navigation? a network-theoretic analysis. *International Journal of Social Computing and Cyber-Physical Systems*, 1(1):33–55.

[Herlocker et al., 2004] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53.

[Heymann et al., 2008] Heymann, P., Koutrika, G., and Garcia-Molina, H. (2008). Can social bookmarking improve web search? In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 195–206. ACM.

[Hintzman, 1984] Hintzman, D. L. (1984). Minerva 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, 16(2):96–101.

[Hotho et al., 2006] Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. (2006). Information retrieval in folksonomies: search and ranking. In *Proceedings of the 3rd European conference on The Semantic Web: research and applications*, pages 411–426. Springer-Verlag.

[Huang et al., 2014] Huang, C.-L., Yeh, P.-H., Lin, C.-W., and Wu, D.-C. (2014). Utilizing user tag-based interests in recommender systems for social resource sharing websites. *Knowledge-Based Systems*, 56:86–96.

[Huang et al., 2010] Huang, J., Thornton, K. M., and Efthimiadis, E. N. (2010). Conversational tagging in twitter. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 173–178. ACM.

[Hunt, 2014] Hunt, N. (2014). Techcrunch interview with neil hunt, cpo of netflix. [Online under `https://techcrunch.com/2014/05/19/netflix-neil-hunt-internet-week/`; accessed 07-July-2017].

[Järvelin and Kekäläinen, 2000] Järvelin, K. and Kekäläinen, J. (2000). Ir evaluation methods for retrieving highly relevant documents. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '00, pages 41–48, New York, NY, USA. ACM.

[Jäschke et al., 2009] Jäschke, R., Eisterlehner, F., Hotho, A., and Stumme, G. (2009). Testing and evaluating tag recommenders in a live system. In *Proc. of RecSys'09*, pages 369–372, New York, NY, USA. ACM.

[Jäschke et al., 2007] Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2007). Tag recommendations in folksonomies. In *Knowledge Discovery in Databases: PKDD 2007*, pages 506–514. Springer.

[Jäschke et al., 2008] Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., and Stumme, G. (2008). Tag recommendations in social bookmarking systems. *Ai Communications*, 21(4):231–247.

[Jeon et al., 2014] Jeon, M., Jun, S., and Hwang, E. (2014). Hashtag recommendation based on user tweet and hashtag classification on twitter. In *International Conference on Web-Age Information Management*, pages 325–336. Springer.

[Konstan, 2004] Konstan, J. A. (2004). Introduction to recommender systems: Algorithms and evaluation. *ACM Transactions on Information Systems (TOIS)*, 22(1):1–4.

[Kopeinik et al., 2016a] Kopeinik, S., Kowald, D., Hasani-Mavriqi, I., and Lex, E. (2016a). Improving collaborative filtering using a cognitive model of human category learning. *The Journal of Web Science*, 2(1).

[Kopeinik et al., 2016b] Kopeinik, S., Kowald, D., and Lex, E. (2016b). Which algorithms suit which learning environments? a comparative study of recommender systems in tel. In *Proceedings of the 11th European Conference on Technology Enhanced Learning*, ECTEL '16, pages 124–138. Springer.

[Kopeinik et al., 2017] Kopeinik, S., Lex, E., Seitlinger, P., Albert, D., and Ley, T. (2017). Supporting collaborative learning with tag recommendations: a real-world study in an inquiry-based classroom project. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 409–418. ACM.

[Körner et al., 2010a] Körner, C., Benz, D., Hotho, A., Strohmaier, M., and Stumme, G. (2010a). Stop thinking, start tagging: tag semantics emerge from collaborative verbosity. In *Proceedings of the 19th international conference on World wide web*, pages 521–530. ACM.

[Körner et al., 2010b] Körner, C., Kern, R., Grahsl, H.-P., and Strohmaier, M. (2010b). Of categorizers and describers: An evaluation of quantitative measures for tagging motivation. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 157–166. ACM.

[Kowald, 2015] Kowald, D. (2015). Modeling cognitive processes in social tagging to improve tag recommendations. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 505–509, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

[Kowald et al., 2013] Kowald, D., Dennerlein, S., Theiler, D., Walk, S., and Trattner, C. (2013). The social semantic server - a framework to provide services on social semantic network data. In *I-SEMANTICS (Posters and Demos)*, volume 1026 of *CEUR Workshop Proceedings*, pages 50–54. CEUR-WS.org.

[Kowald et al., 2017a] Kowald, D., Kopeinik, S., and Lex, E. (2017a). The tagrec framework as a toolkit for the development of tag-based recommender systems. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, pages 23–28, New York, NY, USA. ACM.

[Kowald et al., 2015a] Kowald, D., Kopeinik, S., Seitlinger, P., Ley, T., Albert, D., and Trattner, C. (2015a). Refining frequency-based tag reuse predictions by means of time and semantic context. In *Mining, Modeling, and Recommending'Things' in Social Media*, pages 55–74. Springer.

[Kowald et al., 2014a] Kowald, D., Lacic, E., and Trattner, C. (2014a). Tagrec: Towards a standardized tag recommender benchmarking framework. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*, HT '14, pages 305–307, New York, NY, USA. ACM.

[Kowald and Lex, 2015] Kowald, D. and Lex, E. (2015). Evaluating tag recommender algorithms in real-world folksonomies: A comparative study. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pages 265–268, New York, NY, USA. ACM.

[Kowald and Lex, 2016] Kowald, D. and Lex, E. (2016). The influence of frequency, recency and semantic context on the reuse of tags in social tagging systems. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, HT '16, pages 237–242, New York, NY, USA. ACM.

[Kowald et al., 2017b] Kowald, D., Pujari, S. C., and Lex, E. (2017b). Temporal effects on hashtag reuse in twitter: A cognitive-inspired hashtag recommendation approach. In *Proceedings of the 26th International Conference on World Wide Web*, WWW '17, pages 1401–1410, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

[Kowald et al., 2015b] Kowald, D., Seitlinger, P., Kopeinik, S., Ley, T., and Trattner, C. (2015b). Forgetting the words but remembering the meaning: Modeling forgetting in a verbal and semantic tag recommender. In *Mining, Modeling, and Recommending'Things' in Social Media*, pages 75–95. Springer.

[Kowald et al., 2014b] Kowald, D., Seitlinger, P., Trattner, C., and Ley, T. (2014b). Long time no see: The probability of reusing tags as a function of frequency and recency. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW '14 Companion, pages 463–468, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

[Krestel and Fankhause, 2010] Krestel, R. and Fankhause, P. (2010). Language models and topic models for personalizing tag recommendation. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 1, pages 82–89. IEEE.

[Krestel et al., 2009] Krestel, R., Fankhauser, P., and Nejdl, W. (2009). Latent dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on Recommender systems*, pages 61–68. ACM.

[Kwak et al., 2010] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM.

[Kywe et al., 2012] Kywe, S. M., Hoang, T.-A., Lim, E.-P., and Zhu, F. (2012). On recommending hashtags in twitter networks. In *Social Informatics*, pages 337–350. Springer.

[Lacic et al., 2014a] Lacic, E., Kowald, D., Parra, D., Kahr, M., and Trattner, C. (2014a). Towards a scalable social recommender engine for online marketplaces: The case of apache solr. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 817–822. ACM.

[Lacic et al., 2014b] Lacic, E., Kowald, D., Seitlinger, P., Trattner, C., and Parra, D. (2014b). Recommending items in social tagging systems using tag and time information. In *Proceedings of the 1st International Workshop on Social Personalisation co-located with the 25th ACM Conference on Hypertext and Social Media*, HT '14. CEUR-WS.

[Lacic et al., 2014c] Lacic, E., Kowald, D., and Trattner, C. (2014c). Socrecm: A scalable social recommender engine for online marketplaces. In *Proceedings of the 25th ACM conference on Hypertext and social media*, pages 308–310. ACM.

[Lehmann et al., 2012] Lehmann, J., Gonçalves, B., Ramasco, J. J., and Cattuto, C. (2012). Dynamical classes of collective attention in twitter. In *Proceedings of the 21st international conference on World Wide Web*, pages 251–260. ACM.

[Ley et al., 2013] Ley, T., Cook, J., Dennerlein, S., Kravcik, M., Kunzmann, C., Laanpere, M., Pata, K., Purma, J., Sandars, J., Santos, P., et al. (2013). Scaling informal learning: An integrative systems view on scaffolding at the workplace. In *Proceedings of the European Conference on Technology Enhanced Learning*, pages 484–489. Springer.

[Ley et al., 2014] Ley, T., Cook, J., Dennerlein, S., Kravcik, M., Kunzmann, C., Pata, K., Purma, J., Sandars, J., Santos, P., Schmidt, A., et al. (2014). Scaling informal learning at the workplace: A model and four designs from a large-scale design-based research effort. *British Journal of Educational Technology*, 45(6):1036–1048.

[Ley and Seitlinger, 2010] Ley, T. and Seitlinger, P. (2010). A cognitive perspective on emergent semantics in collaborative tagging: The basic level effect. *Social and Semantic web*, page 13.

[Ley and Seitlinger, 2015] Ley, T. and Seitlinger, P. (2015). Dynamics of human categorization in a collaborative tagging system: How social processes of semantic stabilization shape individual sensemaking. *Computers in human behavior*, 51:140–151.

[Lin and Mishne, 2012] Lin, J. and Mishne, G. (2012). A study of "churn" in tweets and real-time search queries. In *Sixth International AAAI Conference on Weblogs and Social Media*.

[Lin et al., 2015] Lin, Y.-l., Trattner, C., Brusilovsky, P., and He, D. (2015). The impact of image descriptions on user tagging behavior: A study of the nature and functionality of crowdsourced tags. *Journal of the Association for Information Science and Technology*, 66(9):1785–1798.

[Lindstaedt et al., 2009] Lindstaedt, S., Mörzinger, R., Sorschag, R., Pammer, V., and Thallinger, G. (2009). Automatic image annotation using visual content and folksonomies. *Multimedia Tools and Applications*, 42(1):97–113.

[Lindstaedt et al., 2008] Lindstaedt, S., Pammer, V., Mörzinger, R., Kern, R., Mülner, H., and Wagner, C. (2008). Recommending tags for pictures based on text, visual content and user context. In *Internet and Web Applications and Services, 2008. ICIW'08. Third International Conference on*, pages 506–511. IEEE.

[Lipczak, 2012] Lipczak, M. (2012). *Hybrid tag recommendation in collaborative tagging systems*. PhD thesis, Dalhousie University Halifax.

[Lipczak et al., 2009] Lipczak, M., Hu, Y., Kollet, Y., and Milios, E. (2009). Tag sources for recommendation in collaborative tagging systems. *ECML PKDD discovery challenge*, 497:157–172.

[Lipczak and Milios, 2010] Lipczak, M. and Milios, E. (2010). The impact of resource title on tags in collaborative tagging systems. In *Proceedings of the 21st ACM conference on Hypertext and hypermedia*, pages 179–188. ACM.

[Lipczak and Milios, 2011] Lipczak, M. and Milios, E. (2011). Efficient tag recommendation for real-life data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(1):2.

[Lorince and Todd, 2013] Lorince, J. and Todd, P. M. (2013). Can simple social copying heuristics explain tag popularity in a collaborative tagging system? In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 215–224. ACM.

[Ma et al., 2012] Ma, Z., Sun, A., and Cong, G. (2012). Will this # hashtag be popular tomorrow? In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 1173–1174. ACM.

[Manouselis et al., 2011] Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., and Koper, R. (2011). Recommender systems in technology enhanced learning. In *Recommender systems handbook*, pages 387–415. Springer.

[Manouselis et al., 2010] Manouselis, N., Vuorikari, R., and Van Assche, F. (2010). Collaborative recommendation of e-learning resources: an experimental investigation. *Journal of Computer Assisted Learning*, 26(4):227–242.

[Marinho et al., 2011] Marinho, L., Nanopoulos, A., Schmidt-Thieme, L., Jäschke, R., Hotho, A., Stumme, G., and Symeonidis, P. (2011). Social tagging recommender systems. In *Recommender Systems Handbook*, pages 615–644. Springer US.

[Marinho and Schmidt-Thieme, 2008] Marinho, L. B. and Schmidt-Thieme, L. (2008). Collaborative tag recommendations. In *Data Analysis, Machine Learning and Applications*, pages 533–540. Springer.

[McAuley and Leskovec, 2013] McAuley, J. and Leskovec, J. (2013). Hidden factors and hidden topics: understanding rating dimensions with review text. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 165–172. ACM.

[Messina, 2007] Messina, C. (2007). Twitter post of chris messina, inventor of the hashtag. [Online under `https://twitter.com/chrismessina/status/223115412/`; accessed 07-July-2017].

[Moltedo et al., 2012] Moltedo, C., Astudillo, H., and Mendoza, M. (2012). Tagging tagged images: on the impact of existing annotations on image tagging. In *Proceedings of the ACM multimedia 2012 workshop on Crowdsourcing for multimedia*, pages 3–8. ACM.

[Myers and Leskovec, 2014] Myers, S. A. and Leskovec, J. (2014). The bursty dynamics of the twitter information network. In *Proceedings of the 23rd international conference on World wide web*, pages 913–924. ACM.

[O'reilly, 2005] O'reilly, T. (2005). What is web 2.0. [Online under `http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html`; accessed 07-July-2017].

[Page et al., 1999] Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: bringing order to the web. *Stanford InfoLab*.

[Parra and Brusilovsky, 2009] Parra, D. and Brusilovsky, P. (2009). Collaborative filtering for social tagging systems: an experiment with citeulike. In *Proceedings of the third ACM conference on Recommender systems*, pages 237–240. ACM.

[Parra-Santander and Brusilovsky, 2010] Parra-Santander, D. and Brusilovsky, P. (2010). Improving collaborative filtering in social tagging systems for the recommendation of scientific articles. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, volume 1, pages 136–142. IEEE.

[Pirolli and Fu, 2003] Pirolli, P. and Fu, W.-T. (2003). Snif-act: A model of information foraging on the world wide web. In *International Conference on User Modeling*, pages 45–54. Springer.

[Pirolli et al., 2002] Pirolli, P., Fu, W.-T., Reeder, R., and Card, S. K. (2002). A user-tracing architecture for modeling interaction with the world wide web. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 75–83. ACM.

[Pujari and Kanawati, 2012] Pujari, M. and Kanawati, R. (2012). Tag recommendation by link prediction based on supervised machine learning. In *Sixth International AAAI Conference on Weblogs and Social Media*.

[Rader and Wash, 2008] Rader, E. and Wash, R. (2008). Influences on tag choices in del. icio. us. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pages 239–248. ACM.

[Rawashdeh et al., 2013] Rawashdeh, M., Kim, H.-N., Alja'am, J. M., and El Saddik, A. (2013). Folksonomy link prediction based on a tripartite graph for tag recommendation. *Journal of Intelligent Information Systems*, 40(2):307–325.

[Rendle, 2010] Rendle, S. (2010). Factorization machines. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*, pages 995–1000. IEEE.

[Rendle et al., 2009] Rendle, S., Balby Marinho, L., Nanopoulos, A., and Schmidt-Thieme, L. (2009). Learning optimal ranking with tensor factorization for tag recommendation. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 727–736. ACM.

[Rendle and Schmidt-Thieme, 2009] Rendle, S. and Schmidt-Thieme, L. (2009). Factor models for tag recommendation in bibsonomy. In *ECML/PKDD 2008 Discovery Challenge Workshop, part of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pages 235–243.

[Rendle and Schmidt-Thieme, 2010] Rendle, S. and Schmidt-Thieme, L. (2010). Pairwise interaction tensor factorization for personalized tag recommendation. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 81–90. ACM.

[Robu et al., 2009] Robu, V., Halpin, H., and Shepherd, H. (2009). Emergence of consensus and shared vocabularies in collaborative tagging systems. *ACM Transactions on the Web (TWEB)*, 3(4):14.

[Romero et al., 2011] Romero, D. M., Meeder, B., and Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 695–704. ACM.

[Said and Bellogín, 2014] Said, A. and Bellogín, A. (2014). Comparative recommender system evaluation: benchmarking recommendation frameworks. In *Pro-*

ceedings of the 8th ACM Conference on Recommender systems, pages 129–136. ACM.

[Santos et al., 2016] Santos, P., Dennerlein, S., Theiler, D., Cook, J., Treasure-Jones, T., Holley, D., Kerr, M., Attwell, G., Kowald, D., and Lex, E. (2016). Going beyond your personal learning network, using recommendations and trust through a multimedia question-answering service for decision-support: A case study in the healthcare. *Journal of Universal Computer Science*, 22(3):340–359.

[Schafer et al., 2007] Schafer, J. B., Frankowski, D., Herlocker, J., and Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer.

[Sedhai and Sun, 2014] Sedhai, S. and Sun, A. (2014). Hashtag recommendation for hyperlinked tweets. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 831–834. ACM.

[Seitlinger et al., 2015a] Seitlinger, P., Kowald, D., Kopeinik, S., Hasani-Mavriqi, I., Lex, E., and Ley, T. (2015a). Attention please! a hybrid resource recommender mimicking attention-interpretation dynamics. In *Proceedings of the 24th International Conference on World Wide Web*, pages 339–345. ACM.

[Seitlinger et al., 2013] Seitlinger, P., Kowald, D., Trattner, C., and Ley, T. (2013). Recommending tags with a model of human categorization. In *Proceedings of the 22nd ACM international conference on Conference on information and knowledge management*, CIKM '13, pages 2381–2386, New York, NY, USA. ACM.

[Seitlinger and Ley, 2012] Seitlinger, P. and Ley, T. (2012). Implicit imitation in social tagging: familiarity and semantic reconstruction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1631–1640. ACM.

[Seitlinger et al., 2015b] Seitlinger, P., Ley, T., and Albert, D. (2015b). Verbatim and semantic imitation in indexing resources on the web: A fuzzy-trace account of social tagging. *Applied Cognitive Psychology*, 29(1):32–48.

[Sen et al., 2006] Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F. M., and Riedl, J. (2006). Tagging, communities, vocabulary, evolution. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 181–190. ACM.

[She and Chen, 2014] She, J. and Chen, L. (2014). Tomoha: Topic model-based hashtag recommendation on twitter. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 371–372. ACM.

[Sigurbjörnsson and Van Zwol, 2008] Sigurbjörnsson, B. and Van Zwol, R. (2008). Flickr tag recommendation based on collective knowledge. In *Proceedings of the 17th international conference on World Wide Web*, pages 327–336. ACM.

[Sinclair and Cardew-Hall, 2008] Sinclair, J. and Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 34(1):15–29.

[Song et al., 2008] Song, Y., Zhang, L., and Giles, C. L. (2008). A sparse gaussian processes classification framework for fast tag suggestions. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, CIKM '08, pages 93–102, New York, NY, USA. ACM.

[Stanley and Byrne, 2013] Stanley, C. and Byrne, M. D. (2013). Predicting tags for stackoverflow posts. In *Proceedings of ICCM*, volume 2013.

[Stanley and Byrne, 2016] Stanley, C. and Byrne, M. D. (2016). Comparing vector-based and bayesian memory models using large-scale datasets: User-generated hashtag and tag prediction on twitter and stack overflow. *Psychological Methods*, 21(4):542.

[Stefaner et al., 2007] Stefaner, M., Dalla Vecchia, E., Condotta, M., Wolpers, M., Specht, M., Apelt, S., and Duval, E. (2007). Mace–enriching architectural learning objects for experience multiplication. In *European Conference on Technology Enhanced Learning*, pages 322–336. Springer.

[Trattner et al., 2016] Trattner, C., Kowald, D., Seitlinger, P., Ley, T., and Kopeinik, S. (2016). 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.

[Trattner et al., 2012] Trattner, C., Lin, Y.-l., Parra, D., Yue, Z., Real, W., and Brusilovsky, P. (2012). Evaluating tag-based information access in image collections. In *Proceedings of the 23rd ACM conference on Hypertext and social media*, pages 113–122. ACM.

[Tso-Sutter et al., 2008] Tso-Sutter, K. H., Marinho, L. B., and Schmidt-Thieme, L. (2008). Tag-aware recommender systems by fusion of collaborative filtering algorithms. In *Proceedings of the 2008 ACM symposium on Applied computing*, pages 1995–1999. ACM.

[Tsur and Rappoport, 2012] Tsur, O. and Rappoport, A. (2012). What's in a hashtag?: Content based prediction of the spread of ideas in microblogging communities. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 643–652, New York, NY, USA. ACM.

[Van Maanen and Marewski, 2009] Van Maanen, L. and Marewski, J. N. (2009). Recommender systems for literature selection: A competition between decision making and memory models. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 2914–2919.

[Van Rijsbergen, 1974] Van Rijsbergen, C. J. (1974). Foundation of evaluation. *Journal of Documentation*, 30(4):365–373.

[Vargas and Castells, 2011] Vargas, S. and Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 109–116. ACM.

[Verbert et al., 2012] Verbert, K., Manouselis, N., Drachsler, H., and Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society*, 15(3):133–148.

[Vuorikari and Massart, 2010] Vuorikari, R. and Massart, D. (2010). datatel challenge: European schoolnet's travel well dataset. In *1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)*.

[Wagner et al., 2014] Wagner, C., Singer, P., Strohmaier, M., and Huberman, B. A. (2014). Semantic stability in social tagging streams. In *Proceedings of the 23rd international conference on World Wide Web*, pages 735–746. ACM.

[Wang et al., 2014] Wang, Y., Qu, J., Liu, J., Chen, J., and Huang, Y. (2014). What to tag your microblog: Hashtag recommendation based on topic analysis and collaborative filtering. In *Web Technologies and Applications*, pages 610–618. Springer.

[Wetzker et al., 2010] Wetzker, R., Zimmermann, C., Bauckhage, C., and Albayrak, S. (2010). I tag, you tag: translating tags for advanced user models. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 71–80. ACM.

[Wheeler, 2014] Wheeler, S. (2014). Learning theories: Adaptive control of thought. [Online under `http://www.teachthought.com/learning/theory-cognitive-architecture/`; accessed 07-July-2017].

[Xu et al., 2015] Xu, J., Zhang, Q., and Huang, X. (2015). Personalized hashtag suggestion for microblogs. In *Chinese National Conference on Social Media Processing*, pages 38–50. Springer.

[Xu et al., 2008] Xu, S., Bao, S., Fei, B., Su, Z., and Yu, Y. (2008). Exploring folksonomy for personalized search. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 155–162. ACM.

[Yang and Leskovec, 2011] Yang, J. and Leskovec, J. (2011). Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 177–186. ACM.

[Yin et al., 2011a] Yin, D., Hong, L., and Davison, B. D. (2011a). Exploiting session-like behaviors in tag prediction. In *Proceedings of the 20th international conference companion on World wide web*, pages 167–168. ACM.

[Yin et al., 2011b] Yin, D., Hong, L., Xue, Z., and Davison, B. D. (2011b). Temporal dynamics of user interests in tagging systems. In *Twenty-Fifth AAAI conference on artificial intelligence*.

[Zangerle et al., 2011] Zangerle, E., Gassler, W., and Specht, G. (2011). Recommending#-tags in twitter. In *Proceedings of the Workshop on Semantic*

Adaptive Social Web (SASWeb 2011). CEUR Workshop Proceedings, volume 730, pages 67–78.

[Zangerle et al., 2013] Zangerle, E., Gassler, W., and Specht, G. (2013). On the impact of text similarity functions on hashtag recommendations in microblogging environments. *Social Network Analysis and Mining*, 3(4):889–898.

[Zhang et al., 2012] Zhang, L., Tang, J., and Zhang, M. (2012). Integrating temporal usage pattern into personalized tag prediction. In *Web Technologies and Applications*, pages 354–365. Springer.

[Zheng and Li, 2011] Zheng, N. and Li, Q. (2011). A recommender system based on tag and time information for social tagging systems. *Expert Systems with Applications*, 38(4):4575–4587.

[Zubiaga, 2009] Zubiaga, A. (2009). What are social tags? [Online under `http://blog.zubiaga.org/2009/02/what-are-social-tags/`; accessed 07-July-2017].

[Zubiaga et al., 2013] Zubiaga, A., Fresno, V., Martinez, R., and Garcia-Plaza, A. P. (2013). Harnessing folksonomies to produce a social classification of resources. *IEEE transactions on knowledge and data engineering*, 25(8):1801–1813.


More Books!
yes
I want morebooks!