Simone Kopeinik

Applying Cognitive Learner Models for Recommender Systems in Sparse Data Learning Environments

Doctoral Thesis

for the attainment of the degree of

Doctor of Engineering Sciences (Dr. techn.)

submitted to

Graz University of Technology

Institute of Interactive Systems and Data Science
Head: Univ.-Prof. Dipl-Inf. Dr. Stefanie Lindstaedt

Supervisor: Univ.-Prof. Dipl-Inf. Dr. Stefanie Lindstaedt
Co-Supervisor: Ass.-Prof. Dipl-Ing. Dr. Elisabeth Lex

Graz, September 2017

Statutory Declaration
Eidesstattliche Erklärung[1]

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, ______________________ ______________________

Date/Datum Signature/Unterschrift

[1]Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008; Genehmigung des Senates am 1.12.2008

Acknowledgement

I wish to express my gratitude to Stefanie Lindstaedt and Dietrich Albert for providing me with an inspiring working environment with interesting research projects and the freedom to personally and professionally evolve. I am also very grateful to Tobias Ley for the interesting collaboration and his support. Further, special thanks are dedicated to Elisabeth Lex for the constructive and always motivating discussions. Her diligent supervision significantly influenced my work. I owe sincere gratitude to Paul Seitlinger, who greatly supported me as a colleague with valuable input and advice, and as a friend who encouraged me in spiritless times. Special mention must be made of my colleagues Dominik Kowald, Lisa Winter, Alexander Nussbaumer, Michael Bedek, Ilire Hassani and Aurora Dimache, whose comments and collaboration within the research work were invaluable.

Thanks are also due to my family who always reminded me, I could still get a proper job. Thanks to all my dear friends who accompanied me during these last rocky years. Heribert, Birgit, Sandra, Jürgen, Chris and Lucia, thank you for listening. Marina thank you for always being there. Finally, I would like to thank my partner Stefan who bravely endured five years of mood swings while gaining high proficiency in the proofreading of research articles.

Regarding the offline datasets, thanks are dedicated to Katja Niemann who provided the datasets MACE and TravelWell. For the KDD15 data, I would like to gratefully acknowledge the organizers of KDD Cup 2015, as well as XuetangX, for making the datasets available. I am also very grateful to Verein für Bildung und Erziehung der Grazer Schulschwestern and particularly Jürgen Mack for the great support in implementing a study in his school lessons.

Abstract

In recent years, various recommendation algorithms have been proposed to support learners in technology-enhanced learning environments. Such algorithms have proven to be quite effective in big-data learning settings (massive open online courses), yet successful applications in other informal and formal learning settings are rare. Common challenges include data sparsity, the lack of sufficiently flexible learner and domain models, and the difficulty of including pedagogical goals into recommendation strategies. Computational models of human cognition and learning are, in principle, well positioned to help meet these challenges, yet the effectiveness of cognitive models in educational recommender systems remains poorly understood to this date. This thesis contributes to this strand of research by investigating i) two cognitive learner models (CbKST and SUSTAIN) for resource recommendations that qualify for sparse user data by following theory-driven top down approaches, and ii) two tag recommendation strategies based on models of human cognition (BLL and MINERVA2) that support the creation of learning content meta-data. The results of four online and offline experiments in different learning contexts indicate that a recommendation approach based on the CbKST, a well-founded structural model of knowledge representation, can improve the users' perceived learning experience in formal learning settings. In informal settings, SUSTAIN, a human category learning model, is shown to succeed in representing dynamic, interest based learning interactions and to improve Collaborative Filtering for resource recommendations. The investigation of the two proposed tag recommender strategies underlined their ability to generate accurate suggestions (BLL) and in collaborative settings, their potential to promote the development of shared vocabulary (MINERVA2). This thesis shows that the application of computational models of human cognition holds promise for the design of recommender mechanisms and, at the same time, for gaining a deeper understanding of interaction dynamics in virtual learning systems.

Zusammenfassung

In den vergangenen Jahren wurde eine Vielzahl an Empfehlungsalgorithmen vorgeschlagen, um Lernenden in technologiegestützten Lernumgebungen zu assistieren. Solche Algorithmen konnten auf großen Datenmengen, wie zum Beispiel in offenen Massen-Online-Kursen, bereits erfolgreich eingesetzt werden. In anderen virtuellen Lernumgebungen blieb die wirkungsvolle Anwendung von Empfehlungssystemen bis jetzt aber eher selten. Dies kann auf typische Probleme des Feldes zurückgeführt werden. Beispiele hierfür sind dünnbesetzte (sparse) Daten, eine mangelnde Flexibilität von User- und Domain-Modellen und die Schwierigkeit, pädagogische Modelle in Empfehlungsalgorithmen einzubeziehen. Kognitive Modelle zur Abbildung von menschlichen Gedächtnisfunktionen und Lernprozessen stellen einen vielversprechenden Ansatz dar, diese Probleme in Angriff zu nehmen. Die Effektivität solcher Modelle zur Anwendung in edukativen Empfehlungssystemen ist jedoch derzeit kaum erforscht. Mit der Untersuchung von i) zwei kognitiven Lerner-Modellen (CbKST und SUSTAIN), die durch ihren theoriegestützten Ansatz auch in Umgebungen mit kleinen Benutzergruppen Erfolg bei der Empfehlung von Lernressourcen versprechen, und ii) zwei kognitiv inspirierten Tag-Empfehlungsstrategien (BLL und MINERVA2), welche die Erstellung von Meta-Daten unterstützen sollen, trägt diese Arbeit zum Forschungsfeld bei. Die Ergebnisse von vier Online und Offline Experimenten weisen darauf hin, dass Empfehlungsstrategien die auf der CbKST, einem fundierten strukturellen Modell zur Wissensrepräsentation, basieren eine positivere Lernerfahrung in formalen Lernumgebungen herbeiführen können. In informellen Settings konnte gezeigt werden, dass SUSTAIN, ein vielseitiges Modell zum Lernen in Kategorien, dazu in der Lage ist, dynamisches interessenbasiertes Lernverhalten abzubilden und weiters *Collaborative Filtering* bei der Empfehlung von Lernressourcen zu verbessern. Bei der Empfehlung von Tags konnten die genannten Ansätze basierend auf individuellen Benutzerdaten (BLL) besonders akkurate

Ergebnisse erreichen, und in Gruppenumgebungen (MINERVA2) zur Bildung eines geteilten Vokabulars beitragen. Diese Arbeit zeigt, dass die Anwendung von Computermodellen zur Abbildung menschlicher Kognition in der Entwicklung von Empfehlungsmechanismen Vorteile verspricht und gleichzeitig zu einem tieferen Verständnis der Interaktionsdynamik in virtuellen Lernsystemen beitragen kann.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Background and Motivation

In everyday life, we are flooded by options of what to eat or buy, which movies to watch or which articles to read. In all these matters, we occasionally struggle to make a decision because we lack knowledge of the topic, or we do not have a sufficient overview of the options available. Thus, we seek advice from experienced peers or sources and trust them to provide us with adequate suggestions. Recommender Systems (RS) are software components that cater for this type of social behaviour (Resnick and Varian, 1997).

In recent years, the application of RS has become very popular in e-commerce platforms, where prominent examples include Amazon (Linden, B. Smith, and York, 2003), YouTube (Davidson et al., 2010) or Netflix (Gomez-Uribe and Hunt, 2016). Integrated in websites, dedicated software services aim to ease the information overload users are typically confronted with, or act as sales assistants, supporting a user's decision-making processes (Konstan, 2004). In order to suggest items of interest or of relevance to a user, the recommendation mechanism's challenge is to properly and continuously predict a user's preferences and needs (Resnick and Varian, 1997).

In Technology-Enhanced Learning (TEL) settings, where learners typically struggle with the organisation, finding and even awareness of relevant learning resources, the recommendation task is more complex (Drachsler, H. Hummel, and Koper, 2007; Anjorin et al., 2012). While e-commerce systems aim to support and influence their users' consumerist behaviour, recommendations in TEL applications should assist learning processes and knowledge acquisition (Drachsler, H. G. K. Hummel, and Koper, 2009). Extracting and drawing on data from learning traces (Duval, 2011), RS assist learners by i) recommending relevant learning activities, learning resources or learning sequences, ii) suggesting like-minded or opposing learning peers (Drachsler, Verbert, et al., 2015), iii) recommending tags to organize self-created or collected content (Klašnja-Milićević, Ivanović, and Nanopoulos, 2015), and iv) predicting learning performance (Drachsler, Verbert, et al., 2015).

To address these tasks, user preferences are just one of many factors to be considered in the design of TEL recommendation strategies. Learner models are required to be more dynamic, since learning goals, interests and knowledge evolve during the learning process. Furthermore, design and development of most recommender systems are strongly context dependent (Drachsler, H. G. K. Hummel, and Koper, 2009). Information like age, language skills and domain expertise are essential when selecting appropriate learning content. For example, a researcher and a primary school child searching for similar topics will not be able to understand the same documents (Keeney and Raiffa, 1993). Besides, dissimilarities between formal and informal learning settings yield different requirements with respect to data models and learning support. In formal learning settings, learning typically happens in closed groups, with structured learning domains and pre-defined learning content and learning goals, whereas informal learning is characterized as an interest driven process without a clear definition of learning goals and thus, learning content (Colardyn and Bjornavold, 2004). While in formal TEL environments, teachers or educational designers typically structure a domain and setup a course, this has proven difficult in informal learning settings, as the set of useful learning resources is typically not known in advance (Drachsler, H. G. K. Hummel, and Koper, 2009). Also, in TEL settings, often only small communities of people actually generate data, and even fewer participate in explicit data contributions (e.g., via generating ratings, tags). This leads to sparse learning data (Buder and Schwind, 2012), which hampers the success of traditional statistical recommendation methods as used in e-commerce systems (Drachsler, H. G. K. Hummel, and Koper, 2009).

While RS have grown into one of the most popular research fields in personalized TEL, as of yet there are no generally suggested or commonly applied recommender system implementations for TEL environments (Drachsler, Verbert, et al., 2015). In fact, the majority of holistic educational recommender systems remain within research labs (Khribi, Jemni, and Nasraoui, 2015). This may be partly attributed to the demanding requirements of the domain, which require learner models and recommendation strategies to be intelligent and very dynamic, incorporating not only pedagogical challenges but also addressing sparse data and cold start problems on a frequent basis.

1.2. Research Questions and Contributions

In this thesis, I explore the conjecture that RS in TEL settings may be more successful if they are based on a thorough understanding of how humans process information. In particular, I focus on the exploration of cognitive user models to recommending learning resources and tags towards the specific requirements of different TEL environments. Scientific work I present was conducted over the course of three research projects, INNOVRET, weSPOT and Merits (see Section 1.3), which determined the learning settings, and accordingly, the requirements of the recommendation strategies. The key challenge was to find theoretically plausible models that cover a great amount of relevant aspects while still being computable on restricted computational resource as often found in educational contexts (Pierce and Cleary, 2016).

1.2.1. Research Questions

In formal learning settings, learning takes place in closed groups with well-structured learning domains and defined learning goals (Colardyn and Bjornavold, 2004). In such learning settings, it is reasonable to follow ontology-based top down approaches for learning recommendations (Drachsler, H. G. K. Hummel, and Koper, 2009). With educational designers, learning domains and activities can be described and tailored to the demands of learning goals and target groups. A structural model of knowledge representation that has proven successful in a variety of TEL settings such as game-based learning (M. D. Kickmeier-Rust et al., 2007), self-regulated learning (C. M. Steiner, Nussbaumer, and Albert, 2009) and work place learning (Ley, Kump, and Gerdenitsch, 2010) is the Competence-based Knowledge Space Theory (CbKST). It provides a theoretically sound framework to model the competences, problems and learning content associated with a knowledge domain, and furthermore the methods to assess and update these structures (Heller et al., 2006; Augustin et al., 2013). Given its well-elaborated theoretical framework, I hypothesise that it is also suitable as a basis for TEL recommendation approaches. This leads to my first research question:

RQ1: **Can a learning resource recommender based on a structural learner model (like the CbKST) improve the learning experience in a formal learning environment?**
To address this research question, a personalisation approach is described in Section 6.1 that combines self-regulated learning with the CbKST as an underlying learner modelling and recommendation strategy. The introduced concept was implemented in form of plug-ins, which were integrated in a Moodle[1] course. A study evaluating the users' perceived learning experience shows encouraging scoring higher than the control group in most aspects of the questionnaire. Most importantly, the usefulness of the provided guidance support scored considerably higher in the experimental group (see Section 6.1.3).

However, in informal and/or social learning settings, where the amount of learning activity is constantly growing and learning interests dynamically change, it has proven infeasible for educational designers to continuously update the structure representing a given knowledge domain (Drachsler, H. Hummel, and Koper, 2007). To address the issue of unstructured learning data, collaborative filtering approaches or hybrid combinations thereof have been suggested (Drachsler, H. G. K. Hummel, and Koper, 2009; Verbert, Drachsler, et al., 2011). However, sparse data problems (Verbert, Drachsler, et al., 2011) and the lack of learning process specific requirements hinder the success of purely statistical methods such as Collaborative Filtering (CF) (Drachsler, H. G. K. Hummel, and Koper, 2009; Verbert, Drachsler, et al., 2011). Therefore, in my second research question, I propose a hybrid recommendation strategy that combines user-based collaborative filtering (CF_U) (Schafer, Frankowski, et al., 2007) and Supervised and Unsupervised STratified Adaptive Incremental Network (SUSTAIN) (Love, Medin, and Gureckis, 2004), a particularly flexible cognitive model of human category learning that captures a learner's dynamic changes in learning interests.

[1] https://moodle.org/

RQ2: **Can a process oriented learner model based on SUSTAIN be applied to improve an existing resource recommendation strategy such as collaborative filtering?**
To address this research question, I introduce a model that slightly adapts the SUSTAIN approach according to the requirements of resource recommendations (see Section 6.2.2). It captures non-linear user-resource dynamics in the form of an unsupervised clustering approach to anticipate learner-specific preferences and decisions on resource engagement. The resource recommendation strategy draws on SUSTAIN to model a user's traces (e.g., items a user has collected in the past) and is further combined with a user-based (CF_U) recommendation strategy to create the hybrid approach SUSTAIN+CF_U. To evaluate the recommendation accuracy of the approach against state-of-the-art recommendation algorithms, an empirical study was conducted on three social bookmarking datasets from BibSonomy, CiteULike and Delicious. The experiments were carried out on these social tagging system datasets, because tagging data is often utilized to describe learning content or to derive semantic topics for resources (Griffiths, Steyvers, Tenenbaum, et al., 2007) by means of LDA (see Section 5.1.2). Plus, these datasets are freely-available for scientific purposes. Evaluation results demonstrate the potential of the approach to personalize and improve user-based CF predictions. Furthermore, to gain insights into which aspects of the SUSTAIN algorithm contribute most to the improved performance, a parameter study was conducted in which the model's main parameters were simulated and observed. The results show that the effect of recency can be inferred from the optimal learning rate and the impact of the dynamic learning approach, i.e., how many semantic clusters work best for a specific dataset.

Encouraged by the promising results of Research Question 2 (RQ2), the performance of the proposed resource recommendation strategies based on SUSTAIN that mimics human behaviour, was tested on datasets from different TEL environments. The performance of SUSTAIN+CF_U was evaluated on social tagging datasets as described earlier. However, in other TEL environments, tagging data is often sparse, which limits the accessibility of learning content in search and recommendation processes (Niemann, 2015). Furthermore, research (Kuhn et al., 2012; Ley and Seitlinger, 2015) indicates that students seek assistance in the tagging process, regarding (i) the take up of the process and therefore, the finding of initial vocabulary and (ii) the achievement of a semantically stable vocabulary amongst their learning peers. The implementation of tag recommendations is one approach to support learners in the tagging process. Thus, in the last RQ, I additionally investigate the suitability of two tag recommendation strategies, drawing on formal process models of human episodic memory that help predict tag choices based on the information of frequency and recency (Base Level Learning Equation (BLL) (John R. Anderson, Bothell, et al., 2004)) and frequency and semantic context (MINERVA2 (Hintzman, 1984)), for TEL settings.

RQ3: **Can resource and tag recommendations that are based on cognitive learner models form a competitive alternative to common statistically based approaches in TEL settings?**

To answer this research question, I completed two experiments that examined whether selected cognitive models can compete with base line recommendation strategies:

An offline data experiment (see Section 7.1) to study the recommender performance of six recommendation algorithms and variations thereof on implicit usage data from six TEL datasets. In this experiment, two scenarios were explored: Firstly, resource recommendation in TEL and secondly, tag recommendation in TEL. The results show that the performance of resource recommendation algorithms strongly depend on dataset characteristics such as the number of users or the extent of item descriptions. For tag recommendations, a hybrid combination of the cognitive-inspired BLL and a most popular approach performs best.

An online experiment (see Section 7.2) to investigate the suitability of three computationally simple tag recommendation strategies in a real-life Inquiry-based Learning (IBL) setting with students (see Section 7.2). Within an IBL environment, two cognitive-inspired algorithms, namely BLL and MINERVA2 (see Section 4.2) were implemented and compared to a most popular tags approach as a baseline. The evaluation was structured according to A/B testing, where one group of students received tag recommendations based on their personal history, while another group of students received recommendations derived from the group's collective tagging history. This experiment shows that tag recommendations based on BLL and individual user's data are most accurate, whereas tag recommendations based on MINERVA2 and collective tagging traces foster the group's agreement on a shared vocabulary (e.g., semantic stabilization).

1.2.2. Contributions

This section provides an overview of the thesis' main contributions to the research field, which are presented in Part II. The research described here was completed as part of my scientific activity at the *Institute of Data Science and Interactive Systems* and results from my work in three different research projects: INNOVRET, weSPOT and Merits (see Section 1.3). Most important collaborators to the presented work are Dietrich Albert, Aurora Dimache, Dominik Kowald, Elisabeth Lex, Tobias Ley, Alexander Nussbaumer, Paul Seitlinger and Lisa Winter. Substantial parts of the research presented in this thesis have been already published in the following 11 publications:

The first paper is a doctoral consortium contribution that outlines preliminary concepts of this thesis. It argues for the necessity of exploring recommendation approaches that go beyond statistical methods as applied in e-commerce systems, describes the proposed research methodology and suggests the application of cognitive models to better address the dynamic nature of learning settings.

1) **Kopeinik, S.** (2015). "Applying Cognitive Learner Models for Recommender Systems in Small-Scale Learning Environments", Presented at *Doctoral Consortium Workshop of EC-TEL 2015.*

The next four publications result from work conducted in the INNOVRET project and relate to Section 6.1. Please note that Aurora Dimache and Attracta Brennan were mainly responsible for the setup of the main course platform (Moodle), the creation and structuring of course content, stakeholder communication and the implementation of evaluation studies. The CbKST services were provided by Alexander Nussbaumer, who kindly extended it to the needs of my personalization tools. Lisa Winter contributed to the elaboration of the underlying pedagogical approach. My contributions include i) the conceptualization and the design of the personalization approach, tailored to the requirements of the available IT infrastructure and the target group's (e.g., heat pump installer) specific needs, ii) the technical implementation of the approach within the Moodle platform, iii) assistance in the structuring of the CbKST-based learning domain and the maintenance of the learning environment, and iv) significant contributions to the design of the evaluation studies.

2) Winter, L. C., **Kopeinik, S.**, Albert, D., Dimache, A., Brennan, A., & Roche, T. (2013, August). "Applying Pedagogical Approaches to Enhance Learning: Linking Self-Regulated and Skills-Based Learning with Support from Moodle Extensions". In *Advanced Applied Informatics (IIAIAAI), 2013 IIAI International Conference*, pp. 203-206. IEEE.
3) Dimache, A., **Kopeinik, S.**, Brennan, A., Roche, T., Winter, L. C., & Albert, D. (2014). "Innovative Online Vocational Training of Renewable Energy Technologies (INNOVRET)". *International Journal of Information & Education Technology*, 4(1), pp. 127-131.
4) **Kopeinik, S.**, Nussbaumer, A., Winter, L. C., Albert, D., Dimache, A., & Roche, T. (2014, July). "Combining Self-Regulation and Competence-Based Guidance to Personalise the Learning Experience in Moodle". In *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference*, pp. 62-64. IEEE.
5) Dimache, A., Roche, T., **Kopeinik, S.**, Winter, L. C., Nussbaumer, A., & Albert, D. (2015). "Suitability of Adaptive Self-Regulated e-Learning to Vocational Training: A Pilot Study in Heat Pump System Installation". *International Journal of Online Pedagogy and Course Design (IJOPCD)*, 5(3), pp. 31-46.

The next two publications present the investigation of the SUSTAIN approach that is described in Section 6.2. The theoretical underpinning and conceptual development of this work was carried out by Paul Seitlinger. Elisabeth Lex and Tobias Ley contributed to the conceptual development and the discussion of results. Dominik Kowald completed the calculations of comparative algorithms within the TagRec framework. The analysis and interpretation of recommendation accuracy metrics results from a collaborative effort of the authors. My contributions include i) the implementation of the SUSTAIN approach within the TagRec framework, ii) the adaptation and optimization of the approach towards the requirements of the application in web-resource recommendations, iii) the performance of a parameter study and its interpretation, to better understand the underlying dynamics of the model.

6) Seitlinger, P., Kowald, D., **Kopeinik, S.**, Hasani-Mavriqi, I., Ley, T., & Lex, E. (2015). "Attention Please! A Hybrid Resource Recommender Mimicking Attention-Interpretation Dynamics". In *Proceedings of the 24th International Conference on World Wide Web Companion*, pp. 339-345. International World Wide Web Conferences Steering Committee.
7) **Kopeinik, S.**, Kowald, D., Hasani-Mavriqi I., & Lex, E. (2017). "Improving Collaborative Filtering Using a Cognitive Model of Human Category Learning", *The Journal of Web Science* 2(4), pp. 45-61.

In Section 7.1 the performance of a number of cognitive-inspired tag and resource recommendation strategies is investigated on six TEL datasets. Please note that Elisabeth Lex contributed with the conceptual development and analysis of evaluation results of this work, Dominik Kowald completed the calculations of comparative algorithms within the TagRec framework. My contributions include i) the collection and preparation of representative TEL datasets, ii) the implementation of TEL related recommendation algorithms, iii) the analysis and interpretation of evaluation results.

8) **Kopeinik S.**, Kowald, D.,& Lex, E. (2016). "Which Algorithms Suit Which Learning Environments? A Comparative Study of Recommender Systems in TEL". In *European Conference on Technology Enhanced Learning*, pp. 124-138. Springer International Publishing.

Research work presented in Section 7.2 was conducted in the course of the weSPOT project. Please note that Paul Seitlinger contributed significantly to the design and setup of the evaluation studies. Elisabeth Lex and Tobias Ley contributed to the interpretation and presentation of study results. My contributions include i) the design, implementation and integration of the applied tag recommendation approaches, ii) the implementation of study related software services (e.g., logging) iii) the implementation of multiple real-life studies and iv) the analysis and interpretation of evaluation results.

9) **Kopeinik, S.**, Bedek, M., Firssova, O., Mack, J., Albert, D. (2015). "Introducing Technology-Enhanced Inquiry-Based Learning to Support Science Education in Secondary Schools: A Teacher Perspective". In *Proceedings of the 7th International Conference on Education and New Learning Technologies*, pp. 6035-6045.
10) **Kopeinik, S.**, Lex, E., Seitlinger, P., Albert, D., & Ley, T. (2017), "Supporting collaborative learning with tag recommendations: a real-world study in an inquiry-based classroom project", In *Proceedings of the 7th International Learning Analytics & Knowledge Conference*, pp. 409-418. ACM Press.

The last paper provides an overview of the TagRec framework and a survey of scientific studies that have been completed using the framework so far. The framework as such was developed by Dominik Kowald. However, I contributed with the implementation of a number of TEL related recommendation algorithms and the realisation of offline data studies described in this thesis.

11) Kowald, D., **Kopeinik, S.**, & 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, Adapation and Personalization (UMAP'2017)*. ACM.

1.2.3. Other Relevant Publications

Publications listed in this Section do not directly relate to research questions posed in this thesis but constitute important groundwork for presented contributions. It also considerably contributed to my deeper understanding of the research field. A full list of chronologically ordered publications can be found in Attachment A.1.

1. Trattner, C., Kowald, D., Seitlinger, P., **Kopeinik, S.**, & Ley, T. (2016). "Modeling Activation Processes in Human Memory to Predict the Use of Tags in Social Bookmarking Systems". *The Journal of Web Science*, 2(1), pp. 1-16.

2. Kowald, D., **Kopeinik, S.**, Seitlinger, P., Ley, T., Albert, D., & 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*, pp. 55-74. Springer, Cham.
3. Kowald, D., Seitlinger, P., **Kopeinik, S.**, Ley, T., & 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*, pp. 75-95. Springer, Cham.
4. Bedek, M. A., **Kopeinik, S.**, Prünster, B. & Albert, D. (2015). "Applying the Formal Concept Analysis to introduce guidance in an inquiry-based learning environment". In *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference* (pp. 285-289). IEEE.
5. **Kopeinik, S.**, Bedek, M., Öttl, G., & Albert, D. (2013). "Competence Analyser: A portable GUI tool for modelling domain and learner knowledge". In *Proceedings of the 21th International Conference on Computers in Education*, pp. 133-138.
6. **Kopeinik, S.**, Nussbaumer, A., Bedek, M., & Albert, D. (2012). "Using CbKST for Learning Path Recommendation in Game-based Learning". In *Proceedings of the 20th International Conference on Computers in Education*, pp. 26-30.

1.3. Research Environment

In this thesis, I present research and software development work that I conducted throughout a number of research projects. The projects intersect in their research field of adaptation and personalization in the context of learning. A brief description of these projects follows.

INNOVRET (Innovative Online Vocational Training of Renewable Energy Technologies) focused on the development of an online learning environment to support people in their training on heat pump systems, with the intention of addressing the special requirements of vocational education. To this end, the project developed a web-based adaptive e-learning environment that considers individual knowledge levels, learning progress and learning goals by integrating a CbKST-based learning resource recommendation strategy with a self regulated learning approach. Resulting software components were embedded in Moodle[2], the project's Learning Management System (LMS) of choice. The project was supported by the Life-Long-Learning programme of the European Commission with the grant number: LLP/LdV/TOI/2011/IRL-501. Relevant scientific concepts and outcomes of this research are described in Section 6.1 and have partly been published in L.-C. Winter et al. (2013), Dimache, Kopeinik, et al. (2014), and Kopeinik, Nussbaumer, L. C. Winter, et al. (2014) and Dimache, Roche, et al. (2015).

weSPOT (Working Environment with Social and Personal Open Tools for inquiry) aimed at creating a flexible and adaptable Inquiry-Based Learning (IBL) environment to support educators in the application of IBL within their classroom and curricular setting. For this purpose, an open social networking platform (elgg[3]) was developed further and enhanced with IBL specific tools and learning analytics software. My contributions relevant to this work encompass domain modelling and recommendation mechanisms to support students in their browsing and tagging behaviour. The project has been funded by the European Commission under the 7th framework programme with the grant number: ICT/STREP-318499. Relevant research outcomes of this project are presented in Section 7.2 and have partly been published in Kopeinik, Bedek, et al. (2015) and Kopeinik, Lex, et al. (2017).

Merits (MEmory Retrieval In Tagging: A model of Social and Semantic Influence) is supported by the Austrian Science Fund (FWF) under the grant number: P25593-G22. The project investigated tagging processes on the web, taking into account human cognitive processes and corresponding virtual social environments as context variables. It especially focused on the impact of semantic categorization and the effects of recency in memory retrieval. Cognitive models investigated in this project have been implemented as tag and resource recommendation strategies, which we investigated on offline and online data. Research supported by the

[2]https://moodle.org/
[3]https://elgg.org/

Merits project is covered in Sections 6.2 and 7.2 and has partly been published in Kowald, Kopeinik, Seitlinger, et al. (2015) and Kopeinik, Lex, et al. (2017) and Seitlinger, Kowald, et al. (2015) and Kopeinik, Kowald, Hasani-Mavriqi, et al. (2016).

1.4. Structure

The remaining Chapters of this thesis are structured in four parts:

Part I *Related Work*, provides an overview of the related work framing the topic of this thesis. General concepts of RS are introduced in Chapter 2. Chapter 3 elaborates on existing work on RS in TEL environments, pointing out characteristic factors and challenges of the application domain.

Part II *Cognitive Modelling for Learning Recommendations*, consists of the scientific core of this work. In Chapter 4, four selected cognitive models are introduced divided into models for learning recommendations (Section 4.1) and models for tag recommendations (Section 4.2). Chapter 5 provides details on the evaluation setting of conducted offline (Section 5.1.1) and online experiments (Section 5.2). A description of baseline algorithms and applied metrics follows (Section 5.3). Chapter 6 gives examples on how models that capture a learners knowledge acquisition can be exploited to assist the learning processes in TEL environments with recommendations. Section 6.1 relates to Research Question 1 (RQ1), and investigates whether a personalization approach that combines the principles of Self-regulated Learning (SRL) with CbKST-based learning resource recommendations can improve the perceived learning experience in a formal learning environment. An offline data study that inspects the potential of the category learning model SUSTAIN to improve CF (Section 6.2) tackles RQ2. The experimental part of this thesis concludes with Chapter 7, which presents the design and results of two evaluation studies addressing Research Question 3 (RQ3): Section 7.1 describes a comparative offline study contrasting cognitive-inspired recommendation strategies with state-of-the-art algorithms. In Section 7.2 a real-life study investigating cognitive-inspired tag recommendation algorithms, is reported.

Part III *Conclusion and Future Work*, provides in Chapter 8, a summary of conducted work, in Chapter 9, a reflection on problems and limitations of the conducted research and an outline of future work is presented.

Part IV *Appendix*, provides a full list of publications and the questionnaire that was used in the online study of RQ1.

Part I.

Related Work

2. Recommender Systems

Recommender systems are software components that predict and suggest items that are potentially interesting or relevant to a user. These suggestions are based on tailored personalization services that aim to ease the information overload in supporting a user's decision-making processes (Konstan, 2004). This is in line with social behaviour where inexperienced people seek advice from experienced peers (Resnick and Varian, 1997).

The first generation of recommender systems focused on collaborative filtering approaches that generate suggestions based on similar users' taste (Jannach et al., 2010), to help individual users identify relevant e-mails (Goldberg et al., 1992), news articles (Resnick, Iacovou, et al., 1994) or music (Shardanand and Maes, 1995). Former used keyword-based filtering approaches were lacking means to i) rank retrieved documents and to ii) filter non-text documents (Konstan, 2004). Recommender systems were applied to address these challenges with the consideration of user preferences in the form of content ratings. Succeeding research explored a greater variation of user interactions to train recommendation models within observed environments (Schafer, Konstan, and Riedl, 1999). Nowadays, recommendation algorithms typically consider data about users, items and user-item interactions (Jesús Bobadilla et al., 2013). Also, the field of application has expanded greatly and encompasses domains like e-government, e-business, e-commerce/e-shopping, e-library, e-learning, e-tourism, e-resource services and e-group activities (Lu et al., 2015).

This Chapter provides a brief introduction into general concepts of recommender systems. It starts with examples of recommendation goals, explains the main characterisations of recommender systems, elaborates on typical problems of the research area and describes how the quality of recommendations can be measured.

2.1. Recommendation Goals

Recommender systems are most popular in e-commerce systems where the three main recommendation goals are defined as (Schafer, Konstan, and Riedl, 1999):

Converting browsers into buyers. By recommending items a user might want or like, a recommender system makes these items explicit to the user. This is a facilitation of the user's search process which can support the user, who would typically just browse, in finding and buying an item.

Improving cross sell. The recommender aims to increase the amount of items that a user already consumes, for instance, by suggesting items relating to items of a user's purchasing list.

Gaining loyalty. With the provided personalization, the recommender system aims to increase the perceived usefulness of the site. This can bring an added value to customers and encourage their revisit (Schafer, Konstan, and Riedl, 1999).

In other domains, these goals may be different. Examples include the improvement of daily routines for chronic patients in e-health recommender (Hidalgo et al., 2014), or the support of learning and teaching processes in the TEL domain (Erdt, Fernandez, and Rensing, 2015).

2.2. Paradigms of Recommender Systems

Traditionally, recommender systems have been segmented into basic concepts, which are, according to Jannach et al. (2010), hereinafter briefly described:

Collaborative recommendation. The approach grounds on the calculation of similarities either to other users or items in a user's history. The basis is a user-item matrix that depict users' interests in items either through ratings or interaction data. For collaborative recommendations, user-based and item-based collaborative filtering recommendations are the most fundamental ones. The user-based approach recommends items that users with similar taste liked in the past, whereas, the item-based approach draws on the similarity of items to those a user liked in the past. No information about the item itself is needed.

Content-based recommendation. Recommendations are based on content information about the recommended items. Content information is prioritized and matched with user preferences. Information about the items can be provided either manually through meta-data or it can be extracted computationally. Content-based recommendation approaches do not rely on other users data and thus, cater well in application domains with small user communities.

Knowledge-based recommendation. In this case knowledge about the user and the item is required. This is often collected explicitly through input forms or in dialogues, where users manually specify their requirements and the relative importance of these requirements. According to this information, items can be filtered.

Hybrid recommendation. To cover a greater amount of aspects in a recommendation setting and to compensate for drawbacks of single recommendation strategies, it often makes sense to combine multiple approaches. This leads to hybrid recommendation systems.

Collaborative Filtering is considered the most popularly implemented approach and thus has progressed the farthest (Burke, 2007). While traditional techniques like collaborative filtering, content-based and knowledge-based methods have been researched extensively, recent inquiries elaborate and focus on social network-based recommender systems, fuzzy-set-based recommender systems, context awareness-based recommender systems and group recommender systems (Lu et al., 2015).

Another common classification forms the distinction between memory-based and model-based recommendation algorithms. Memory-based algorithms generate their suggestions by applying algorithms on the aggregated data structure. Model-based algorithms on the other hand, first perform calculations to generate a user model. Then, recommendations are generated based on this user model (Breese, Heckerman, and Kadie, 1998).

2.3. Tag Recommendations

Tagging is a simple mechanism that allows users to individually and socially annotate resources. It is an important feature of the Social Web, and has demonstrated to improve search considerably Dellschaft and Staab, 2012 by providing users with a simple tool to collaboratively organize content Körner et al., 2010. This makes it a very convenient vehicle to share, discover and recover resources in the web (Xu et al., 2006; Heymann, Ramage, and Garcia-Molina, 2008). Web platforms that allow users to upload and tag resources to share them with other users are called Social Tagging Systems (STS). Examples of prominent STS that support the sharing of web-bookmarks, articles, pictures and music, respectively, are Delicious [1], BibSonomy [2], Flickr [3] and Last.fm [4] (Balby Marinho et al., 2012). Contrary to indexing mechanisms with controlled vocabularies, tagging allows for unrestricted extension of verbalism. This means that STS are not bound to the use of predefined language or terminology, but its classification vocabulary grows with its users' interactions. This entails advantages, such as the support of an adaptive level of granularity but also challenges such as a lack of quality assurance (Mathes, 2004). Also, users often do not tend to provide tags thoroughly or regularly. Therefore, tag recommendations can be provided, suggesting selected words during the tagging process and with this, assists users in choosing appropriate tags for their resources (Jäschke et al., 2008).

2.3.1. Folksonomy

The annotation structure emerging from STS is called folksonomy. The folksonomy describes how users U, resources R and tags T relate to each other, which is defined as a function $F := (U, R, T, Y)$. Y depicts the relation between user, resource and tags segmented in posts or bookmarks. One post characterizes the set of tags $T_{u,r}$ a user u assigns to a resource r (Balby Marinho et al., 2012).

Furthermore, a folksonomy can be generated in different ways. Commonly, a distinction between narrow and broad folksonomies is applied. Narrow folk-

[1] https://del.icio.us/
[2] https://www.bibsonomy.org/
[3] https://www.flickr.com/
[4] https://www.last.fm/

sonomies emerge if resources are only tagged by a single user. For instance in systems where the user uploads self-created content. Broad folksonomies on the other hand, emerge from system where one resource is tagged by a multitude of users, as for instance in social bookmarking systems. (Helic et al., 2012)

2.3.2. Semantic Stability

When user annotate resources without drawing on a controlled vocabulary it is not assured that they will reach a common understanding of terminology to describe resources or resource attributes. Such a common understanding, is an essential criterion for the useful application of tagging systems as means to organize content (Macgregor and McCulloch, 2006). The implicit agreement of users on a vocabulary that is stable over time is called semantic stability (Wagner et al., 2014). As summarized in (Wagner et al., 2014), there is a multitude of metrics to evaluate semantic stability in different contexts. However, only few methods are yet suited for narrow folksonomies, where items are tagged only by the uploading user. Lin et al. (2012) presents the Macro Tag Growth Method (MaTGM) that measures social vocabulary growth at a systemic level, looking at the social tagging system as a whole.

2.4. Typical Problems

This Section introduces typical problems that have been identified in the application of traditional recommender systems (Jannach et al., 2010).

2.4.1. The Sparse Data Problem

A multitude of recommender system applications rely on similarity measures of either users, items or user-item interactions such as ratings or bookmarking. When available data solely covers a restricted number of user-item pairs, this is called sparse data problem. It appears, when a system governs a large number of items and/or users and single users only interact with a small set of items (Papagelis, Plexousakis, and Kutsuras, 2005). The problem becomes even more evident, if there is a large number of items and a small number of users, as often the case in learning settings (Verbert, Manouselis, Ochoa, et al., 2012).

2.4.2. The Cold Start Problem

The cold start problem describes the challenge of recommending items when a system is initially started and lacks usage data, a new user enters the system or a new item is added (Abbas, Zhang, and Khan, 2015). This is also referred to as out-of-matrix prediction, which is in contrast to in-matrix prediction, where data of items and users already exist (Wang and Blei, 2011). Figure 2.1 illustrates data matrices for the two cases. In Figure 2.1b user u_n is a cold start user, items i_{n-1} and i_n are cold start items.

	i_1	i_2	i_3	i_{n-1}	i_n
u_1	1	?	0	?	0
u_2	0	?	0	1	1
u_3	?	1	1	0	1
u_4	1	0	1	1	?
u_{n-1}	0	1	0	0	0
u_n	?	1	?	0	1

(a) in-matrix prediction

	i_1	i_2	i_3	i_{n-1}	i_n
u_1	1	0	0	?	?
u_2	0	1	0	?	?
u_3	1	1	1	?	?
u_4	1	0	1	?	?
u_{n-1}	0	1	0	?	?
u_n	?	?	?	?	?

(b) out-of-matrix prediction

Figure 2.1.: The two tables, adapted from Wang and Blei (2011), show data of users U interacting with items I, where 0 indicates that a user did not like the item, 1 the opposite and ? that no interaction between user and item has taken place so far.

New user problem:

The new user problem identifies the challenge to overcome the period of time until a sufficient amount of user data is collected that allows a meaningful understanding of a user's preferences within a domain. This is a prerequisite for generating recommendations accurately (Adomavicius and Tuzhilin, 2005). Strategies to address this issue include according to Adomavicius and Tuzhilin (2005) and Rashid et al. (2002)

- suggesting popular or demographically relevant items.
- initializing a user profile through the explicit demand of user data for new users (i.e. a user has to complete a number of ratings or tasks when first entering the system).

- the use of trust-based social network data, where the system recommends what people in a user's social network liked.
- withholding recommendations until a reasonable data basis is established.

New item problem:

An item, which is newly added to a system does not have any usage data available. This is particularly problematic for collaborative filtering applications, where the recommendation depends on how many people liked or used an item in the past (Adomavicius and Tuzhilin, 2005). Strategies to address this issue include according to Adomavicius and Tuzhilin (2005)

- using content-based approaches to introduce new items that are similar to items users liked.
- recommending new items to random users to collect ratings.

2.4.3. Over-specialization

Recommendation algorithms learn a user's preferences and accordingly suggest items the user might be interested in. When suggested items become too similar to items the learner has already engaged with in the past, this is called over-specialization. An example would be a news recommender that suggests an article about an incident the user is already informed about (Jannach et al., 2010), or a travel recommender that only suggests destinations of a user's travel history. Item-based CF approaches are particularly prone to over-specialization (McNee, Riedl, and Konstan, 2006). Algorithms such as *topic diversification* (Ziegler et al., 2005), can be applied to generate more diverse recommendation lists and lead to higher user satisfaction.

2.5. Evaluation of Recommender Systems

The development and application of recommender systems is motivated by generating suggestions with high quality (Jannach et al., 2010). This Section surveys different metrics to measure this quality. Accuracy metrics are often applied to optimize recommendation algorithms. However, arguments questioning the sufficiency but also effectiveness of accuracy metrics are discussed. Present research tends to incorporate factors that consider the user satisfaction more elaborately, as for example diversity and novelty of items. (Ziegler et al., 2005)

Accuracy

Accuracy is segmented into i) accuracy of prediction, ii) accuracy of classification and iii) accuracy of ranks (Jannach et al., 2010).

i) **Accuracy of prediction** is used to determine the correctness of a predicted user preference towards one specific item. The corresponding metric, *mean absolute error* (MAE) calculates the average deviation of users' predicted ratings and users' actual ratings.
ii) **Accuracy of classification** determines the quality of a recommended list of items. In this case, most prominently used metrics are *recall* and *precision*.
iii) **Accuracy of ranks** additionally considers the position of correctly recommended items in a list. It is based on the assumption that users might pay less attention to lower ranked items than to those located on the top of the list. Popularly used metrics are the *rankscore* (Breese, Heckerman, and Kadie, 1998) and the *lift index* (Ling and C. Li, 1998).

Coverage

This metric is usually seen in relation to the domain in which the recommended items are modelled. According to Ge, Delgado-Battenfeld, and Jannach (2010) coverage is subdivided into

i) **prediction coverage**, which is defined as the percentage of items that the algorithm is able to recommend. This depends on the algorithm's rules and mechanisms.
ii) **catalog coverage**, which reveals the percentage of items that are effectively recommended to a user. Both metrics can be refined, and weighted by including the factor of usefulness into the formula. Further explanations and definitions hereto are presented in Ge, Delgado-Battenfeld, and Jannach, 2010.

Diversity

When a list of items is recommended to a user, these items might be most accurate if they show the highest similarity to already purchased items. For instance in an online book store, this can lead to a recommendation list of books including a single author the user has purchased before. In practice, this may cause low user satisfaction because the user algorithm neglects other interests of the user. To measure this phenomenon, Ziegler et al. (2005) introduces the *intra-list similarity* metric, which is inverse to diversity.

Novelty and Serendipity
In some situations people might be comfortable with being suggested a familiar item to purchase. In other cases, recommendations might have no value to users because they suggest items they already know or consume and thus, constitute obvious choices. In the later case, novelty can be introduced as a measurement for the *non-obviousness* of recommended items (Herlocker et al., 2004). A related measure is serendipity. Serendipitous recommendations are characterized as being surprising and unexpected to users while still remaining in their field of interest, being useful and relevant. (Ge, Delgado-Battenfeld, and Jannach, 2010)

McNee, Riedl, and Konstan, 2006 argue that the quality of recommendation, considering the aspect of satisfaction, should be enhanced by investigating a more user centred than strictly mathematical approach. Accordingly, user models and profiles need to take a more essential role. For instance by distinguishing between novel and expert users that very likely have different expectations and needs, recommendations can be better tailored to the individual (McNee, Riedl, and Konstan, 2006).

2.6. Conclusion

This Section provided a brief introduction on basic concepts of recommender systems as relevant to this thesis. It intends to facilitate the understanding of concepts and argumentation presented in remaining Chapters. In Part II a variety of recommendation strategies is applied in offline and online studies. These recommendation strategies relate to the four basic paradigms of recommender systems (i.e., collaborative, content-based, knowledge-based, hybrid) that are briefly explained in Section 2.2. A subsequent introduction to basic principles of tagging and tag recommendations and their potential shortcomings is relevant to RQ3, where cognitive-inspired tag recommendations were tested in TEL settings. Problems of recommender systems that are particularly relevant to this work are the sparsity of user data and the cold start of new items. These problems are particularly popular in TEL environments, which often operate with small user communities and a growing extent of learning content (Verbert, Manouselis, Ochoa, et al., 2012).

The next Chapter will delve into the topic focusing on developments of recommender systems in the context of TEL.

3. Recommender Systems in TEL

In TEL, recommender systems aim to improve the learning process and support the achievement of individual learning goals via personalized suggestions (Manouselis, Drachsler, Vuorikari, et al., 2011). To this end, learning recommendations encompass the suggestion of learning sequences, learning goals and entire classes, and on a more granular level the recommendation of learning tasks, activities, resources and peers (Erdt, Fernandez, and Rensing, 2015). When the number of available learning options grows to a point where learners are overwhelmed in finding relevant options and taking adequate choices, this kind of learning support becomes essential for the learner (Manouselis, Drachsler, Vuorikari, et al., 2011) and furthermore has shown to reduce the tutoring workload for educational staff during course time (Santos and Boticario, 2015). In contrast to e-commerce systems, their objective is not the selling of products but the meaningful support of learning endeavours.

This Section first explains characteristics of the TEL domain and how these characteristics influence requirements of recommendation strategies. Then, differences between informal and formal learning settings are outlined. Subsequently, underlying learner models are explained and related work in TEL recommender research is surveyed. The Section proceeds with a discussion of typical problems in the field and concludes with a brief explanation how this Chapter relates specifically to topics of this thesis.

3.1. Characteristics of Learning Recommendations

Recommender systems in learning differ from e-commerce applications in many ways (Manouselis, Drachsler, Vuorikari, et al., 2011). The following paragraphs provide a glimpse at prominent aspects.

i) **Environmental Conditions**

While algorithms that are used by popular search engines show high recall in recommendations, their precision in TEL recommendations remains relatively low (Drachsler, H. G. Hummel, and Koper, 2008). This can be attributed to their poor consideration of context which is essential in learning. In TEL, **recommendations are highly context dependent** (Manouselis, Drachsler, Vuorikari, et al., 2011). For example, when an expert user and a primary school kid are searching for information on the same topic, context information such as age and expertise are crucial in selecting appropriate learning resources. (Keeney and Raiffa, 1993) TEL recommenders typically aim to meet the requirements of specific learning settings, where context comprises computing infrastructure, the geographical user location, time, physical conditions of the learner's environment, information about learning activities and resources, a learner model, and social relations of the learner (Verbert, Manouselis, Ochoa, et al., 2012). In workplace learning context can also constitute the task a learning is engaged with at the time (Lindstaedt, Scheir, et al., 2008). Moreover, recommendations are typically calculated on the basis of historic user interaction. In e-commerce systems this corresponds to a user's purchases in the past, whereas in learning settings this might be prior knowledge (Drachsler, H. G. K. Hummel, and Koper, 2009). To overcome the cold start problem in e-commerce systems such as for instance NETFLIX [1], new users are asked to complete an initial rating of product items (movies, tv-shows, ...) to indicate their preferences. In some learning settings (e.g., informal learning), this is not applicable, since the body of valuable learning activities is mostly unknown to the user. Thus, **the cold start problem cannot be addressed with initial ratings**. (Drachsler, H. Hummel, and Koper, 2007)

ii) **User model**

E-commerce systems typically predict items towards certain user preferences. In TEL, **user preferences are not the most prominent factor** and may not necessarily be in line with learning goals and other stakeholders' interests. Reasons could be found in the difficulty of certain learning ac-

[1] https://www.netflix.com

tivities that learners tend to avoid or the contradicting nature of learning resources towards a learners' beliefs. However, it is important that learning recommendations also suggest items outside learners' comfort zones and thus, their preference profiles. An example constitutes information that contradicts a person's existing beliefs. While learners often try to avoid such resources, they are perceived as challenging and might inspire critical thinking (Buder and Schwind, 2012).

iii) **Recommendation Goal**
In e-commerce, a recommendation system accomplishes its task when the user purchases an item. Learning recommendations aim to support learning processes that evolve over time. There is no final state but rather a next competence level to achieve. Also, learning is a very dynamic process and learner models need to be flexible (Drachsler, H. G. K. Hummel, and Koper, 2009). Learner characteristics like proficiency level, learning interests and goals change throughout the learning process (with context and time). Furthermore, in order to support the learning process, pedagogical strategies need to be considered in the design of the recommendation approach. For instance, the implementation of hybrid recommendation strategies, combining pedagogical or cognitive models with top down approaches like collaborative filtering allows the system to exploit the benefits of both methods. Another option forms the rule based selection of recommendation strategies to accommodate the learning context or the pedagogically instructed neighbourhood definition (e.g., prior knowledge, learning style, demographics) with a top-down approach (Drachsler, H. Hummel, and Koper, 2007). Thus, **recommendation goals are complex and require well-conceived recommendation strategies**.

Due to its special requirements (highlighted in bold) it is strongly recommended to avoid the direct adoption of commercial recommendation strategies. Buder and Schwind (2012) discuss TEL recommender systems as a twofold process. On the one hand they need to be tailored towards the learning system that provides specific educational context, and on the other hand address social constructs which lead to biases in information search.

Given these requirements, a TEL recommendation strategy needs to be adapted to the specifics of a domain. This includes characteristics of the learning environment, the user models and the recommendation goals, which according to Drachsler, H. G. K. Hummel, and Koper (2009) demands consideration of the following questions:

- Which data is available in my environment?
- What do I know about available items?
- What do I know about the user?
- Which process do I want to support with my recommendation strategy?

Furthermore, learning domains and learning environments are very diverse among different learning contexts (e.g., informal and formal learning) and therefore, in respect to the posed questions, warrant different kinds of models and recommendation strategies (Verbert, Manouselis, Ochoa, et al., 2012). This is discussed in the next Section.

3.1.1. Formal and Informal Learning Settings

Colardyn and Bjornavold (2004) define formal, non-formal and informal learning according to European standards. In line with Drachsler, H. G. K. Hummel, and Koper (2009), in the context of this work, the categorization is simplified by only distinguishing between formal and informal learning. Accordingly, informal learning is considered as all instances of learning that do not fit the scope of formal learning:

Formal Learning is embedded in a pre-defined and structured context typically in form of closed groups. It delivers well-chosen learning content often based on curricula.

Informal Learning is not necessarily intentional and happens outside designated learning settings. It is not bound to place or time. The process does not entail predefined learning goals or structured learning content.

In accordance with their heterogeneousness, the requirements of recommender systems for formal and informal learning settings differ (Drachsler, H. G. K. Hummel, and Koper, 2009):

i) **Environmental Conditions**
Recommendations are typically calculated on the basis of historic user interaction. In formal learning, prior knowledge and competences can be modelled and assessed by different evaluation procedures or inferred from previously completed courses. Based on fine grained learner models and meta-data enriched learning items, recommendations are generated according to the learners expected knowledge state. However, educational design to maintain such structures is usually done by domain experts and thus, it is very cost intensive. Therefore, it is not applicable to informal learning settings where the number of learning items constantly grows and learning interests dynamically change.

ii) **User model**
According to Verbert, Manouselis, Drachsler, et al. (2012) learner models ideally include information about a user's knowledge level, learning and cognitive styles, interests, goals and background. Such information is often provided in formal learning environments through curricula based teaching and learning, prior exams or passed classes, learning ontologies and so on. In informal learning, learner information needs to be created dynamically. Thus, bottom up approaches collect all kinds of learner data to aggregate and feed them into dynamic process models.

iii) **Recommendation Goal**
In both settings, the recommendation goal is to support the learning process with best fitting items that are provided to learners in a structured way. However, recommendation strategies need to differ significantly to reach these goals. As mentioned in i) informal learning is based on unstructured data mainly collected or created by learners' themselves. For instance, in formal learning settings, ontologies that model the learning domain can support the monitoring of proficiency development (Denaux, Dimitrova, and Aroyo, 2004). However, in informal learning settings, top down approaches are often not applicable since learning activities are not finite. This means, learner models are required to be more flexible and learning items need to be enriched with meta-data and educational meaning to follow learning theories or strategies. (Drachsler, H. G. K. Hummel, and Koper, 2009)

3.2. Learner Models

Personalized learning recommendations ground on information about individual learners and their behaviour (Hämäläinen and Vinni, 2010). This is depicted in learner models that represent what the learning system knows about single users. Learner models are used to track the progress of the learner and predict their next action, knowledge state or difficulty. What is modelled and to which extent, is defined by the adaptation task (Brusilovsky and Millán, 2007). Learner models in TEL recommendation engines usually consist of static and dynamic (derived from learning activities) learner information (Drachsler, H. G. K. Hummel, and Koper, 2009).

Brusilovsky and Millán (2007) identify the most informative and thus popular features of individual learner models as follows:

- **Knowledge.** Depending on adaptation goals, either structural (domain related) or procedural (problem solving) knowledge is modelled in different granularities. Models should be able to accommodate learner's knowledge increase (learning) as well as decrease (forgetting).
- **Interest.** While not very meaningful in formal learning settings, the modelling of learner interests became extremely popular with the gain of available information and the advent of educative internet platforms such as bookmarking sites or wikis, which support informal learning.
- **Goals.** Goals are particularly complex to model. They are constantly changing with the task and the learning process. Moreover, the spectrum of goals ranges from the immediate need for information to long term learning goals.
- **Background.** Describes learner knowledge or experience that is not depicted in the domain model. Examples are computer literacy, language abilities or educational background. For instance a colleague graduate might need other learning items than a manual worker training on the job. Background typically encompasses stable features that are stated explicitly by learners themselves or supervisors.
- **Traits.** A learner's trait model encompasses stable user features such as personality traits, cognitive styles, cognitive factors or learning styles. Such features can be part of the learner profile and initialized due to standardized

assessment procedures, self- and peer assessment.

More recent research (Goodwin, 2017) structures components of learner models according to four main categories that distinguish on the x-axis between state and trait variables, and on the y-axis between content independent and content dependent variables. In line with Bloom's distinction of learning domains (Bloom et al., 1956), the structure is further subdivided into the three categories cognitive, psychomotor and affective. This is illustrated in Table 3.1.

	Measurement Category	**Trait**	**State**
Content Dependent	Cognitive	Prior cognitive experience, knowledge or training	Comprehension of presented concepts
	Psychomotor	Relevant prior psychomotor experience or training	Measures of skill improvement
	Affective	Relevant fears, likes, goals, attitudes	Arousal and emotions caused by the learning session
Content Independent	Cognitive	Intellect/aptitude, meta-cognitive skills, memory	Attention, cognitive workload
	Psychomotor	Physical strength, stamina,sensory acuity	Endurance and fatigue
	Affective	Personality traits, general test anxiety	Arousal and emotions independent of the learning session

Table 3.1.: Learner model components adapted from Goodwin (2017)

Content dependent variables describe attributes that are particularly relevant to the content of the learning domain, whereas content independent variables refer to domain independent states and traits of a learner. In line with the definition of Brusilovsky and Millán (2007), traits refer to stable user attributes that usually don't develop during the learning process. In this model, background knowledge and experience is also characterized as a user trait. User states, on the other hand, are assumed to change and progress during a learning episode. (Goodwin, 2017) Relevant to the context of this thesis are cognitive traits and states, in particular learners' knowledge and knowledge evolution, memory and attention.

3.2.1. Cognitive Learner Models

As previously described, a learner model's cognitive attributes can be divided in content dependent and content independent user states and traits (see Table 3.1). Content dependent cognitive learner models, thus, represent a learner's knowledge and experiences with regard to the learning domain. (Goodwin, 2017) They are applied to a number of adaptive educational systems that aim to model knowledge or competences (Herder, Sosnovsky, and Dimitrova, 2017) of a learner. A review on learner and skill modelling in intelligent learning environments (Desmarais and Baker, 2012) identifies Bayesian Networks (Ghahramani, 2001), Item Response Theory (Drasgow and Hulin, 1990), Bayesian Knowledge Tracing (Corbett and John R. Anderson, 1994) and Knowledge Spaces (Falmagne and Doignon, 1999) among the most prominent modelling approaches. In the field of intelligent tutoring systems, also theory-driven approaches like ACT-R (John R Anderson et al., 1990) are commonly applied to model the process of acquiring and applying knowledge of different learning domains like programming or mathematics (Herder, Sosnovsky, and Dimitrova, 2017).

Other approaches go beyond the modelling of knowledge and consider domain independent variables such as meta-cognitive skills and cognitive workload. Such research is presented in Aleven et al. (2006), who used the ACT-R architecture to model meta-cognitive skills, or Biswas et al. (2010) that infers users' self-regulated learning strategies from activity patterns. Others investigate users' learning styles and navigational patterns to approximate their Working Memory Capacity (WMC) (Miller, 1994) and, based on knowledge about students' WMC and principles of the Cognitive Load Theory (Paas and Sweller, 2014), suggest different instructional strategies (Chang et al., 2015).

Models used in this research are further described in Chapter 4 and encompass two theory-driven approaches to reflect a learner's knowledge acquisition for learning resource recommendations (see Section 4.1), and two models for tag recommendations (see Section 4.2) that simulate how humans access information from memory.

3.2.2. Learner Modelling and Assessment

To depict the features of a learner model in the learning system, different user modelling approaches can be used. Most commonly used are overlay models (Brusilovsky and Millán, 2007). Overlay models are an extended and likewise tailored version of the domain model. Extended, because an overlay model may encompass more attributes than the domain model (e.g., not only a competence structure but also proficiency levels of the competences). Tailored, because it is often restricted to the subset of domain items in the learning goal of a user. (Herder, Sosnovsky, and Dimitrova, 2017) For the modelling of user interests, approaches such as key word vectors or concept-level models are suggested (Brusilovsky and Millán, 2007). Other approaches are scalar models that provide a mean value for the user's expertise level, bug models that extend overlay models (Carr and Goldstein, 1977) with the representation of misconceptions and genetic models (Goldstein, 1979) that capture a learner's knowledge evolution.

The cornerstone of meaningful personalization is the system's knowledge of relevant learner traits and states. Accordingly, learner models need to be continuously evaluated and updated to reflect the learners' progress. To this end, f can be requested explicitly (e.g., through test-based assessment, ratings) or implicitly through user monitoring (Oard and Kim, 1998; Brusilovsky and Millán, 2007), which in TEL research is also referred to as evidence based assessment (Valerie J Shute and Zapata-Rivera, 2008; Reimann, M. Kickmeier-Rust, and Albert, 2013).

Explicit Feedback is characterized through explicitly requested input from the user, typically by dedicated user interfaces. The data can be used to feed different kind of models. User profiles for instance can help overcome cold start in memory based recommendation approaches. However, this only works if user profiles are thoroughly designed and completed. Also, this requires additional effort and self-awareness from the learner. (Drachsler, H. Hummel, and Koper, 2007) Other applications suggest to apply stereotype models to initialize structured learner models (Embarak, 2011). Stereotype models aim to create simplistic representations of users that allow to assign them to a number of pre-defined groups. According to the group membership of a learner, an initial state of the learner model is estimated. (Embarak, 2011) Test-based assessment is very narrow in its evaluation and interrupts the concentration of a user. It is therefore not applicable in unstructured, informal learning situations such as work place learning. (Lindstaedt, Günter Beham, et al., 2009) Denaux, Dimitrova, and Aroyo (2004) suggest the application of diagnostic dialogues which are supposed to be less intrusive than pre-tests of knowledge.

Implicit Feedback is collected as user-system interaction data. In learning systems the approach is also referred to as evidence-based assessment. The data collection is typically carried out on multiple sources and continuous observations. Evidence-based assessment is defined in general rules and is therefore more flexible than test-based designs where test-items are developed to evaluate specific knowledge areas, competences or skills (V. J. Shute, 2009). As learner models form the basis for a multitude of recommendation strategies, problems that are typically faced in evidence based assessment are also present in TEL recommenders and have been previously discussed in Section 3.1.

Denaux, Dimitrova, and Aroyo (2004) list the following four challenges for evidence based assessment:

- **the cold-start of users**, as there is no data available when a learner first enters the system.
- **the consideration of prior knowledge**, which might influence the level of understanding when a learner engages with a learning resource.
- **the model's accuracy**, because of the interpretation of interaction is prone to error and semantic misconceptions.
- **the dynamic nature of learning**, as learners' goals, preferences and knowledge can change during the learning process and this change may not be traced.

One popular example of evidence based assessment is the Adaptive Hypermedia Architecture (AHA) (De Bra et al., 2003) that suggests an overlay model where competences are inferred from user actions. According to the approach each engagement with learning items of a user indicates a competence gain.

The relationship between learning item and competence is determined in the learning items' meta-data. Lindstaedt, Günter Beham, et al. (2009) suggest the implementation of knowledge indicating events (KIE) to assess a user's expertise in informal workplace learning settings. KIE are specified as all tracked user interactions that occur during working tasks, which expands the data base for the user assessment. An underlying model, which is based on the CbKST (Korossy, 1997), is used to relate KIE with user knowledge or skills. A detailed description of how to develop a supporting learning context model is given in Ley, Ulbrich, et al. (2008).

A typical characteristic of social personalization techniques that aim to overcome restrictions of traditional adaptation and recommendation systems, is the complementary use of collaborative learner data such as shared items and tags, user public profiles, social connections and logs of users' social activities (Brusilovsky and Tasso, 2004).

In the course of this thesis, explicit feedback is used to feed an overlay learner model in a formal learning setting (test-based assessment) and to support learning object annotation (tags). Implicit feedback, on the other hand is exploited to reflect knowledge evolution in informal settings.

3.3. Review on Recommender Systems in Technology-Enhanced Learning

In the last two decades, recommender systems have been investigated and applied in a variety of TEL settings (Drachsler, Verbert, et al., 2015). There have been multiple excellent reviews on existing recommendation approaches, describing the role of recommender systems in the TEL context, giving an overview of the state of the art and aiming to categorize implementations of recommender systems according to status and evaluation. The first prominent review in a series has discussed TEL recommender problems by studying 20 recommender system applications (Manouselis, Drachsler, Vuorikari, et al., 2011). This initial paper was enhanced by Manouselis et al. (2012) who covers 42 recommenders and additionally introduces a classification frame for recommender system approaches in TEL. Drachsler, Verbert, et al. (2015) picks up previous work, extending the state of the art with a total of 82 investigated recommendation approaches. This also leads to an update of the classification framework proposed in Manouselis et al. (2012). However, none of these reviews includes tag recommender systems in the application of TEL. Surveys like Klašnja-Milićević, Ivanović, and Nanopoulos (2015) additionally discuss the potential of collaborative tagging environments and tag recommender systems for TEL.

Most relevant to this work are ontology-based and collaborative filtering strategies for learning resource recommendations, and approaches to support the annotation of learning resources (tagging).

3.3.1. Ontology-based Approaches

So far, research in TEL recommender systems that recommend resources on the basis of learners' knowledge states focuses on the application of static domain models that require an explicit definition of underlying ontologies or concepts (Drachsler, Verbert, et al., 2015). Accordingly, described in this Section is research that applies ontologies for learning recommendations. In this field, a variety of approaches has been reported. Schmidt (2004) calculates knowledge gaps based on competence requirements derived from an ontology and available competences (learner context models). Learning objects are described via metadata that include related competencies and pre-requisites thereof. Learning programs consisting of ranked learning items are suggested to the user. Similar work has been conducted by Specht (2000) that adapts to the learner's knowledge level, user interest and media preferences, and L.-p. Shen and R.-m. Shen (2004), who presents a system that recommends resources according to a competence gap analysis and a rule-based recommendation strategy. An agent-based framework to recommend suitable learning resources on the basis of learner models and competence-resource relations is presented in (Marino and Paquette, 2010). However, all these approaches do not focus on the theoretical underpinning for the development and maintenance of underlying knowledge structures.

A well-elaborated mathematically sound framework for structuring knowledge in such systems are the Knowledge Space Theory (KST) (Falmagne and Doignon, 1999) and its competence based extension, the CbKST (Korossy, 1997). An early prototype of personalized content presentation based on the KST is introduced in (Hockemeyer, Held, and Albert, 1997). Learners are presented with learning items that can be addressed with their expected prior knowledge. While this research describes a conceptual approach for knowledge representation it still lacks an initial assessment of competences, details on user model updates and evaluation studies. One commercial implementation, resulting from this line of research is *ALEKS*[2]. The adaptive learning system focuses on teaching mathematical and algebraic concepts and offers a multitude of courses tailored to different educational levels. Interesting work is presented in Lindstaedt, Scheir, et al. (2008), where a CbKST based domain model is used to feed knowledge services, recommending expert users and learning resources. Further research using the CbKST to recommend learning items in a game-based learning context is outlined in Kopeinik, Nussbaumer, Bedek, et al. (2012). However, none of the approaches explore the combination of self-regulated learning (Zimmerman, 2002) and competence based guidance in formal learning settings.

3.3.2. Collaborative Filtering and Hybrid Extensions

In order to cope with infinite and unstructured learning content, recommender systems based on e.g. CF and hybrid extension of CF have been suggested for informal learning settings (Drachsler, H. G. K. Hummel, and Koper, 2009).

Verbert, Drachsler, et al. (2011) investigate the impact of different similarity measures and neighbourhood sizes on user- and item-based collaborative filtering. The research is conducted on datasets that originate from the data TEL challenge [3] at ECTEL 2010. It shows that Tanimoto-Jaccard reaches the best results and, furthermore, investigates the potential of implicit user data to recommend resources with user-based collaborative filtering. When tested on four different dataset (three within the learning context and one from the commercial domain), results point out that the performance of the algorithms highly depends on the dataset properties. In particular, the sparsity of data is identified as a crucial factor. Another important contribution to this field is presented in Fazeli et al. (2014) showed on two TEL datasets MACE and OpenScout, how an integration of social interaction can improve collaborative filtering approaches. Jesus Bobadilla,

[2]https://www.aleks.com
[3]http://adenu.ia.uned.es/workshops/recsystel2010/datatel.htm

Serradilla, Hernando, et al. (2009) proposes an extension of CF with proficiency scores. These scores should be used to weight the influence of a recommendation according to the user's estimated knowledge. However, the approach has not been tested on representative datasets. A combination of recommendation algorithms weighted by users themselves is described in Zapata et al., 2013. The approach is tailored to the needs of learning object repositories and has only been tested on one dataset.

Other research investigates the combination of collaborative filtering and learning styles (Bourkoukou and El Bachari, 2016), to predict learner preferences in a structured learning domain, the combination of competence based structuring of learning material with user-based collaborative filtering (Cazella, Reategui, and Behar, 2010) or the computationally expensive combination of collaborative filtering based on different data sources and data mining techniques (Salehi, 2013). However, none of these approaches take into account the requirements of informal learning but rather are designed for formal settings where the potential set of learning resources is known and well described.

3.3.3. Tag Recommendations

Tagging, a mechanism to socially annotate resources, is a very convenient vehicle to share, discover and recover resources in social learning environments (Heymann, Ramage, and Garcia-Molina, 2008). Furthermore, tagging has demonstrated its potential to enrich awareness and reflection of students. An empirical study conducted by Kuhn et al. (2012) indicates that tagging supports learning in IBL by helping students organize, retrieve and reflect upon the content of learning resources they found on the Web (e.g., learning videos) or generated themselves (e.g., blog entries). Ley and Seitlinger (2015) further show that individual learning (e.g., amount and strengths of associations around a concept) intertwines with a process on the collective level, i.e., the development of a shared and semantically stable terminology. Put differently, the more the students succeed in choosing similar tags for particular domain concepts, the higher the individual learning outcome will be. Therefore, it can be concluded that learning should benefit from processes helping students converge in the naming of learning resources and develop a shared tagging vocabulary.

Taking a more technical stance, Bateman et al. (2007) suggests tagging as a means to semantically describe learning resources, as it forms a suitable alternative to the poorly available learning-object metadata that is typically created by expert users. This is also a way of tackling the often criticized lack of support for the self-organization of learning content in TEL environments (Bateman et al., 2007).

One strategy to encourage tagging is to apply tag recommendation mechanisms that conventionalize tagging habits by displaying tags already used in the past (e.g., Font, Serrà, and Serra, 2015). However, despite the tremendous amount of recommendation approaches that have been suggested (Khribi, Jemni, and Nasraoui, 2015), TEL research on tag recommendation mechanisms and its potential for learning is still widely unattended (Klašnja-Milićević, Ivanović, and Nanopoulos, 2015).

Tag Recommendation Approaches

Considerable experiments exploring learning resource annotation through tags are presented in Lohmann et al. (2007), in which generally the suitability of tagging within the learning context was investigated. Results claim guidance to be an important factor for the success of tagging. Diaz-Aviles et al. (2011) investigated automated tagging of learning objects utilizing a computationally expensive variant of Latent Dirichlet Allocation (Blei, Ng, and Jordan, 2003) and they evaluated the tagging predictions in a user study. In Niemann (2015), an approach to automatically tag learning objects based on their usage context was introduced, which builds on Niemann and Wolpers (2013). It shows promising results towards the retrospective enhancement of learning object meta-data. However, their approach cannot be used in online settings as it is based on the context of resources in user sessions. For this thesis, only tag recommendation algorithms are relevant that can be utilized also in online settings.

3.4. Evaluation

Unlike in e-commerce applications, TEL recommenders do not have standardized evaluation procedures or datasets (Drachsler, H. G. K. Hummel, and Koper, 2009). Depending on the requirements and context of a specific recommendation strategy, there are differences with respect to evaluation goals and methodologies. Erdt, Fernandez, and Rensing (2015) distinguish between three rough categories of evaluation measures:

1. **Recommender System Performance**
   This category does not differ from the evaluation of commercial recommender systems, as it focuses on technical evaluation criteria. Drachsler, H. G. K. Hummel, and Koper (2009) suggests to investigate recommendation accuracy (recall and precision), coverage and performance (computation effort).
2. **User-Centric Effects**
   This takes into account metrics from a user's experience with the recommender system. It does not necessarily correlate with technical evaluation criteria but focuses on subjective user perceptions about the system. Pu, Chen, and R. Hu (2011) presents a framework for user-centric evaluation where usability and user satisfaction play central roles.
3. **Effects on Learning**
   This differs from technical measures and takes a rather pedagogical stance. It includes the evaluation of learning performance (e.g., prior vs. posterior knowledge, course completion), learning effectiveness and learning efficiency, which investigate the quantity of learning activities that are completed within a certain time and the time needed to achieve a certain learning goal, respectively, and the influence of recommendations on the learning motivation (drop-out rate) (Manouselis, Drachsler, Vuorikari, et al., 2011).

In TEL, these measurements are typically tested in three types of evaluation settings: i) offline studies, where experiments are conducted on existing datasets that are typically extracted from a deployed system, ii) user or laboratory studies, which take place in controlled environments and iii) online studies or real life testing, where users of the target group interact with the system in realistic application scenarios. (Erdt, Fernandez, and Rensing, 2015)

Santos and Boticario (2015) suggests a layered evaluation approach that is highly entangled with the design process of a recommendation engine. According to different stages of the design process, the proposed evaluation methodology extends from focus groups and interviews (user-centric effects) to data analysis (recommender system performance) and the evaluation of knowledge gains (effects on learning). It provides a holistic framework on how to design and evaluate a recommender system based on educational rules. Erdt and Rensing (2014) proposes crowdsourcing to evaluate user-centric effects like novelty, diversity and perceived relevance of recommended resources. Crowdsourcing (Howe, 2006) describes the outsourcing of a specified work task to an open crowd. This facilitates the access to a large amount of people or users. However, one constraint in TEL evaluation tasks is that their motivation is not learning but completing a paid task, which might influence the reliability of their responses (Erdt and Rensing, 2014).

Also other factors in online and in offline evaluations involve uncertainty. For instance, in online settings, users do not rate items they are not recommended, thus, the system lacks information about items the it missed to suggest. In offline settings on the other hand, items users might have liked but did not find themselves are categorized as false recommendations. (Jannach et al., 2010)

3.5. Typical Challenges

While a tremendous amount of recommender system approaches have been presented and investigated over the last fifteen years since the emergence of the research field, there are no generally suggested or commonly applied recommender system implementations for TEL environments (Drachsler, Verbert, et al., 2015). In fact, the majority of holistic educational recommender systems remain within research labs.

This may be partly attributed to the fact that personalization requires complex and domain dependent user and resource models and demands recommendation goals that are i) in line with pedagogical or teaching goals and ii) designed for specific virtual learning environments and their functionalities. Thus, it is not easily possible to transfer a recommender system from one learning environment to another (Drachsler, H. Hummel, and Koper, 2007; Santos and Boticario, 2015). Also, due to the complex nature of TEL recommendations, many approaches require runtime-intensive computations or unavailable, expensive information about learning domains, resources and learner preferences. Particularly, in informal learning settings, information like ontologies, learning object meta-data and even user ratings are very limited. (Manouselis, Drachsler, Vuorikari, et al., 2011) Learning data is characterized by sparsity, due to small user communities, and an extending amount of learning content, which entails new item cold start problems (Verbert, Manouselis, Ochoa, et al., 2012).

Furthermore, the development of new recommendation engines is hindered by present challenges of conducting meaningful evaluations. One issue is the cost intensity of online experiments which hampers meaningful research studies and outcomes. As a consequence, results cannot be easily confirmed by repeating a study. Also, the variation of conditions within a study group is troublesome, because it might decrease user satisfaction and negatively influence students' learning motivation. (Drachsler, H. G. K. Hummel, and Koper, 2009)

Data-set driven research may constitute a proper alternative, but it can only evaluate the accuracy of predictions, whereas items that learners do not find themselves might constitute more useful recommendations (McNee, Riedl, and Konstan, 2006). Datasets are generally hard to find and often incomplete (Manouselis, Drachsler, Vuorikari, et al., 2011). Thus, the lack of sufficiently large TEL datasets is still considered a formidable challenge in evaluation and further development of recommendation algorithms (Verbert, Drachsler, et al., 2011).

Overall, the challenge to design a general, transferable recommender solution still remains (Khribi, Jemni, and Nasraoui, 2015; Herder, Sosnovsky, and Dimitrova, 2017).

3.6. Conclusion

This Section provided an overview on research in the field of TEL recommender systems that is most relevant to this thesis. Initially, the differences of the research field and its requirements compared to those of e-commerce recommender systems are discussed. The challenging nature of the task to recommend learning items is outlined, in respect to environmental conditions, user modelling and recommendation goals. This line of reasoning emphasizes the need for more flexible and well-conceived recommendation strategies that cannot be simply transferred from e.g., the e-commerce domain. One major issue is context dependency, which is also manifested in structural differences between formal and informal learning settings. The specific characteristics of the learning settings and their typical problems are highly relevant to research conducted in this thesis, because requirements (e.g., creating flexible user models that capture a learner's evolving knowledge state), design and development of here investigated recommendation strategies aim to address particular needs of formal and informal learning settings.

With regard to Goodwin (2017), learner model components can be categorized according to states and traits of learners that are related to domain specific categories (cognitive, psychomotor and affective). Particularly relevant to this research are cognitive learner models that represent user states and traits related to knowledge, memory and attention. While a variety of cognitive modelling approaches have been used in adaptive educational hypermedia, TEL recommender systems still focus on static ontology based methods to model learners' knowledge states. This has proven a suitable method for formal learning settings, where students learn complex concepts of a well-known knowledge domain, such as maths. In such environments, they benefit from being diagnosed in terms of their knowledge and competence states as well as from being guided to learn along these states. However, in respect to informal learning settings, the literature review shows a demand for more flexible learner models that, due to the uncertainty of learning content and goals, adapt to the evolving learning trajectory of a user. To this end, Drachsler, H. G. K. Hummel, and Koper (2009) has proposed the application of recommender systems based on Collaborative Filtering (CF) and hybrid extensions of CF. Current implementations of such CF extensions exploit a variety of data sources and modelling strategies, but most of them do not consider specific requirements of informal learning settings. In particular they focus on well-described learning content which is hardly available in informal settings. In summary, current research lacks the extension of data-driven CF approaches with psychologically meaningful, dynamic learner models.

Another key challenge of informal learning setting, where the amount of learning resources grows during the learning process, is the sparsity of learning object meta-data. This hinders the finding and recommendation of relevant learning resources. In principle, tagging support bears great potential to promote the creation of learning object meta-data in form of user created verbal annotations (tags). However, research in TEL recommender systems barely investigates the support of tagging.

The following Part II details the conceptual and experimental work of this thesis. The upcoming Chapters describe four theoretically plausible models that cover a substantial amount of relevant aspects to address the requirements posed for recommender systems in TEL, as well as implementations and evaluations of learning resource recommendation and tag recommendation approaches based on these models.

Part II.

Cognitive Modelling for Recommendations in TEL

4. Cognitive Models

Cognitive modelling describes the specification of runnable computer models that approximate cognitive processes, mechanisms and representations (Sun, 2008). According to Bechtel, Graham, and Balota (1998) cognitive models are divided into three main categories: i) computational, ii) mathematical and iii) verbal-conceptual models. In principal, the three types of models differ with respect to their detail of formalization: Computational models are algorithmic descriptions capturing processes of human performance. Mathematical models are considered a subset of computational models. They comprise mathematical equations, which formalize relationships between entities. Despite their typical lack of process details, they can mostly be implemented as computer models. Verbal-conceptual models, on the other hand, describe such processes, entities and their relationships in rather natural language. (Sun, 2008)

Within this thesis, the focus is on the application of computational models in the domain of learning, which are further also referred to as (cognitive) learner models. Altogether, four models are studied in this thesis. For the recommendation of learning resources, two models are applied that capture the dynamic development of a learner's knowledge state: i) a structural model that is build upon the representations of knowledge spaces i.e. CbKST (Heller et al., 2006), and ii) a process-oriented model that simulates how humans' learn according to categories i.e. SUSTAIN (Love and Medin, 1998). For the recommendation of tags, two cognitive theories are explored that model humans' retrieval of words from memory (i.e. MINERVA2 (Hintzman, 1984), BLL (John R. Anderson, Bothell, et al., 2004)). Both approaches have been suggested for the application in tag recommender systems (Seitlinger, Ley, and Albert, 2013; Kowald, Seitlinger, Trattner, et al., 2014).

4.1. Models for Learning Recommendations

In this Section, first the structural approach CbKST is described in Section 4.1.1, and second, the process-oriented SUSTAIN model is introduced in Section 4.1.2, with respect to their suitability to cater resource recommendation strategies.

4.1.1. Competence based Knowledge Space Theory

The Competence based Knowledge Space Theory (CbKST) (Korossy, 1997) represents a student's knowledge state as an estimation of a set of competences which she or he has acquired up to a given point in time. This set-theoretical approach maps observable behaviour on latent learning states and thus, allows to infer a student's skills from her or his problem solving behaviour. The formalization of such a mapping benefits both, the development of adaptive assessment procedures and the recommendation of learning objects.

The CbKST is a competence-based extension of the Knowledge Space Theory (KST) (Doignon and Falmagne, 1985; Falmagne and Doignon, 1999). The KST depicts a set-theoretical framework with an underpinned mathematical theory. It aims to model students' response behaviour in knowledge assessment procedures, trying to imitate the adaptive assessment procedure applied by teachers when orally examining a student's state of knowledge (knowledge state). For instance, when a teacher assesses a student's knowledge, the teacher would ask a question, and then, as a reaction to the student's answer, select the subsequent question, and so on and so forth. This adaptive process enables the teacher to assess a student's knowledge in a particular domain, while asking only a subset of available questions. This bears two major advantages: First, a reduced amount of time needed for completing an assessment, and second, a lower number of assessment items that is required to complete it. The application of such an adaptive assessment, where students are only confronted with a subset of questions, also hampers the success of learning approaches that target the memorization of assessment tasks (Falmagne, Albert, et al., 2013). To this end, the KST specifies a model to organize a learning domain (e.g., school subject) as a finite set of problems or tasks (also referred to as assessment items). Each assessment item can be either solved or not. The resulting problem space serves as a basis to assess students' knowledge. Accordingly, student abilities within a subject can be formalized as the number of assessment items a student is able to master. In this way, a student's knowledge state at a given point in time is described (Falmagne and Doignon, 1999).

Formally, a Knowledge Structure K consist of a finite set of problems Q and surmise relations between them that indicate solution dependencies. A surmise relation is a transitive and reflexive relation on problem types that indicate knowledge or skill prerequisites. For instance, one needs to know how to do additions to solve a multiplication.

Further on, K is defined as a collection of subsets of Q, which are conditioned by Q's surmise relations. These subsets are called Knowledge States. Figure 4.1 presents an example of a Knowledge Structure with the problem set $Q = a, b, c, d, e$ and the assumed possible Knowledge States

$$K = \{\emptyset, \{b\}, \{c\}, \{b,d\}, \{b,c\}, \{a,c\}, \{b,c,d\}, \{a,b,c\}, \{a,c,d\}, \\ \{a,b,c,d\}, \{a,b,c,e\}, Q\} \tag{4.1}$$

.

Links between knowledge states indicate possible learning paths to navigate from the naive knowledge state $\emptyset$ to a full mastery of knowledge Q. (Doignon, 1994; Heller et al., 2006; Falmagne, Albert, et al., 2013). Building upon knowledge structures and knowledge spaces, adaptive assessment procedures and learning paths can be derived. For this, the knowledge structure needs to be well-graded. This means that in a successive step from one knowledge state to another, the knowledge state may only expand by one question type.

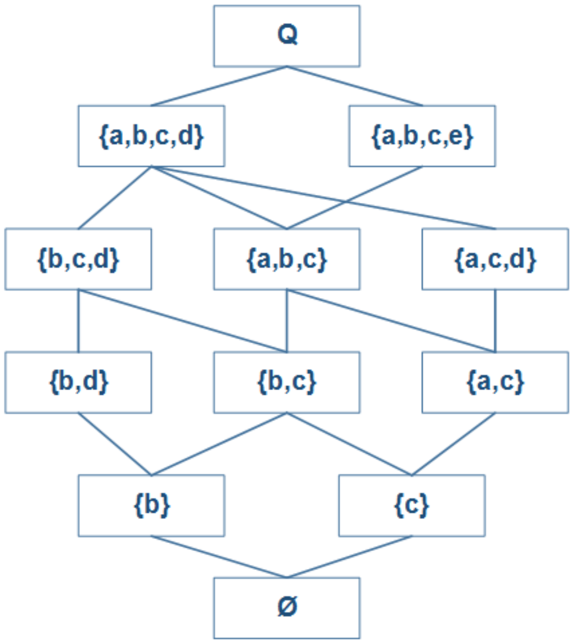


Figure 4.1.: Example of a Knowledge Structure adapted from Doignon (1994).

The probabilistic assessment procedure determines students' knowledge states, taking into account surmise relations between assessment items, in order to infer on the students' abilities of linked items. It selects the first assessment item with an expected medium difficulty, i.e. the item is part of approximately half of the knowledge states. After that, questions are selected iteratively: Every time a learner completes an assessment item, the learner's knowledge state is updated with either a positive or a negative indication of knowledge. The next question is selected upon the resulting knowledge state, which is calculated according to a probability distribution on the domain's knowledge states. In this way, it aims to shorten the assessment procedure by asking as few questions as possible to attain a learner's probabilistic knowledge assessment. The knowledge assessment continues until a peak in the likelihood function positively indicates the existence of a specific knowledge state (Doignon, 1994; Heller et al., 2006).

Initially, the theory focused on the adaptive assessment process. Later it broadened the approach with methods to adaptively present students with learning material that addresses their knowledge states and thus, also caters teaching applications in adaptive learning systems (Hockemeyer, Conlan, et al., 2003; Heller et al., 2006).

Specifically, the CbKST (Korossy, 1997) is a competence-based extension of the behaviourist and descriptive KST. The theory concentrates on the identification of these latent skills and competences and hierarchical relations between them (i.e. prerequisite relations) (Heller et al., 2006). A prerequisite relation can be explained such as: If competence C_1 is a prerequisite of competence C_2, a person who shows competence C_2 is presumed to show competence C_1 as well. This also implies: A person who wants to develop competence C_2, is required to develop competence C_1 beforehand.

The set of competences C of a field of knowledge and their prerequisite relations represent a competence structure. As illustrated in Figure 4.2, a competence structure can be illustrated as a transitive and non-cyclic graph, where nodes are represented as competences and links as prerequisite relations.

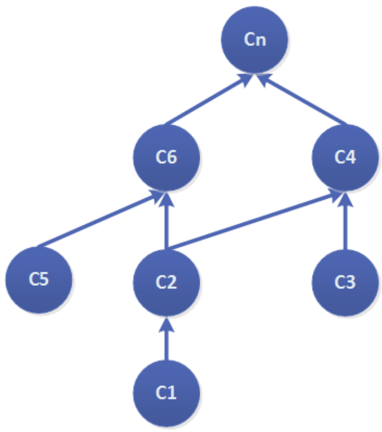


Figure 4.2.: Competence structure represented as directed graph. Nodes are competences and links between nodes are prerequisite relations.

To allow the support of learning actions, the model additionally includes a set of learning resources which are organized in a learning structure. Thus, the CbKST domain model consists of three parts which are aligned to each other (Heller et al., 2006):

i) a **competence structure** that consists of an interrelated set C of competences
ii) a **learning structure** that consists of an interrelated set L of learning resources
iii) a **knowledge structure** that consists of an interrelated set Q of assessment items

A detailed concept on how to align competences, learning resources and assessment items is presented in Heller et al. (2006).

To identify a learner's competence state (i.e. the subset of C a learner shows at an instance), a probabilistic assessment of the knowledge state based on the knowledge structure can be applied and matched to according competences, as described in Doignon (1994). However, a more efficient method has been introduced by Augustin et al. (2013) as the Simplified Updating Rule (SUR). Instead of calculating student's probabilistic knowledge states, the approach calculates the probability of student's to demonstrate unique competences, while still considering their dependencies. Thus, a CbKST learner model holds a probabilistic value, determining the likelihood of each competence in the domain model to be shown by the learner. The updating algorithm maintains the hierarchical competence structure, defining that a competence C_{i-1} which is a prerequisite of a competence C_i shows a greater probabilistic value than the latter one (Korossy, 1999; Augustin et al., 2013).

4.1.2. SUSTAIN

SUSTAIN (*Supervised and Unsupervised STratified Adaptive Incremental Network*) is a flexible network model of human category learning that is introduced and thoroughly discussed in Love, Medin, and Gureckis (2004) and Love and Medin (1998). By means of a clustering approach, it represents the way humans build up and extend their category representations when learning on the basis of examples. It assumes that learning goals and learning tasks are defined through the nature of training examples that further influence the development of categories.

The key elements of the model are flexibility and simplicity, which are supported by the number of hidden units (i.e., clusters) that is not chosen in advance, but discovered incrementally over the learning period.

Initially, the model starts very simple with one cluster that represents the first example. It then grows with the complexity of the problem space. In other words, the model recruits a new cluster whenever a new learning example cannot be accommodated in one of the already existing clusters.

SUSTAIN is described as a three layer network model, which consists of

i) an **input layer** that encodes the input stimulus I,

ii) an **intermediate layer** that is a set of clusters H representing a learner's conceptual understanding in form of categories,

iii) an **output layer** that predicts the category H_j an input stimulus I belongs to.

Depending on the requirements of the learning task, the model provides support for either unsupervised or supervised learning approaches. The two approaches mainly differ through their means of cluster recruitment:

Supervised learning requires an external feedback mechanism that verifies whether a new example was categorized correctly, or not. A false categorization is interpreted as an error and leads to a new cluster recruitment.

Unsupervised learning builds upon the similarity of the input stimulus to clusters depicted in the cluster set as learning criterion. Therefore, it does not require an explicit feedback mechanism to develop a categorical representation of the learning trajectory. Instead, it works as follows: If a given input stimulus' similarity to the existing clusters is below a threshold value τ, it is assumed that the input cannot be sufficiently represented in the existing cluster set. This leads to a new cluster representing the input stimulus.

In order to categorize an input stimulus, a learner's existing clusters compete amongst each other. To this end, an activation value H_j^m is calculated for each cluster. H_j^{act} reflects the similarity of a cluster H_j to an input stimulus I. The classification decision of the model is based on the highest activated cluster, which accordingly, predicts the input stimulus' category.

The approach offers additional parameters to adjust to the peculiarities of different datasets. Relevant to this work are:

The attentional focus r is a constant that represents a person's capability to focus on information aspects or features relevant to a given task, while suppressing minor features of that particular task. To capture a user's specific preferences for certain aspects, the attentional focus r is enhanced by attentional tunings (i.e., tunings of the attentional focus on input features that evolve with encounters with new exemplars).

The learning rate η determines the influence of an input stimuli on its accommodating cluster and consequently defines how fast the algorithm learns new patterns.

The learning threshold τ influences the classification procedure. It is assumed that a cluster H_j sufficiently explains an input stimulus only if its activation $H_j^{act} > \tau$.

4.2. Models for Tag Recommendations

In this sections, two cognitive models are presented that mimic the way humans access information from memory. Both models simulate process-oriented approaches. The activation equation in Section 4.2.1 focuses on the availability of information in a person's declarative memory, whereas MINERVA2 (see Section 4.2.2) regards memorization processes in a person's long term memory.

4.2.1. Activation Equation

The activation equation has been introduced as part of the Adaptive Control of Thought-Rational (ACT-R) theory, which is presented in John R. Anderson, Bothell, et al. (2004).

ACT-R is a cognitive architecture grounding on the assumption that all components in a human mind act in concert to generate behaviour. To that end, the ACT-R theory hypothesizes about how different parts of humans' minds work together and depicts this in a proposed architecture.

The ACT-R architecture consists of a number of collaborating modules, as illustrated in Figure 4.3. In the centre of the model is the production system that coordinates the information flow between different modules. To lighten the working load on the production system, each module is endued with a buffer that holds its most relevant information. This is the data the production system is aware of and reacts to. While authors explicitly state the uncertainty of the number of existing modules, the ACT-R architecture considers and describes four instances in detail:

The Intentional Module monitors a person's current goals and intention.

The Declarative Module is responsible for a person's information retrieval from memory.

The Visual Module interfaces with the external world and identifies and tracks objects in a person's visual field.

The Manual Module interfaces with the external world to coordinate a person's hand movement.

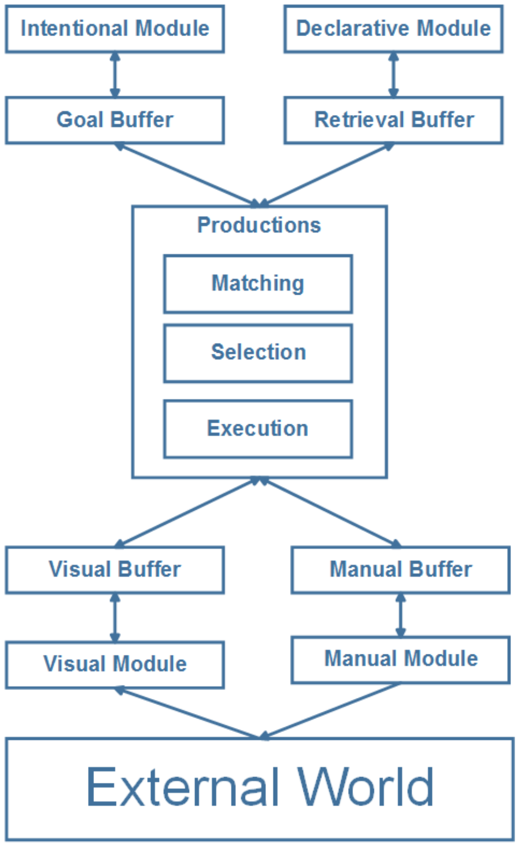


Figure 4.3.: Illustrates the ACT-R architecture adapted from John R. Anderson, Bothell, et al. (2004). Central to this work is the declarative model which handles the retrieval of information from declarative memory.

This work focuses on the declarative module of the ACT-R architecture. In particular, for this thesis, the activation equation 4.2 that formulates the availability of elements in a person's declarative memory is exploited. This has been commonly used to model memory recall tasks (Mozer and Lindsey, 2016) and has been proposed in the context of tag recommendations (Kowald, Seitlinger, Trattner, et al., 2014). A thorough theoretical survey and derivation of the activation equation is presented in John R. Anderson and Schooler (1991).

The basic assumption of the model is that the availability or usage frequency of information in memory is mainly related to three factors (John R. Anderson and Schooler, 1991):

i) **Frequency:** How often has a person been exposed to a memory trace in the past?
ii) **Recency:** How recently did these exposures take place?
iii) **Pattern of prior exposure:** In which context did a memory trace appear?

Additionally, it suggests that each incident of a memory causes an individual activation and contributes to its base level activation, which decays according to a power law function over time. The total activation of a memory trace at a certain instance, is the sum of all its individual activations as illustrated in Figure 4.4.

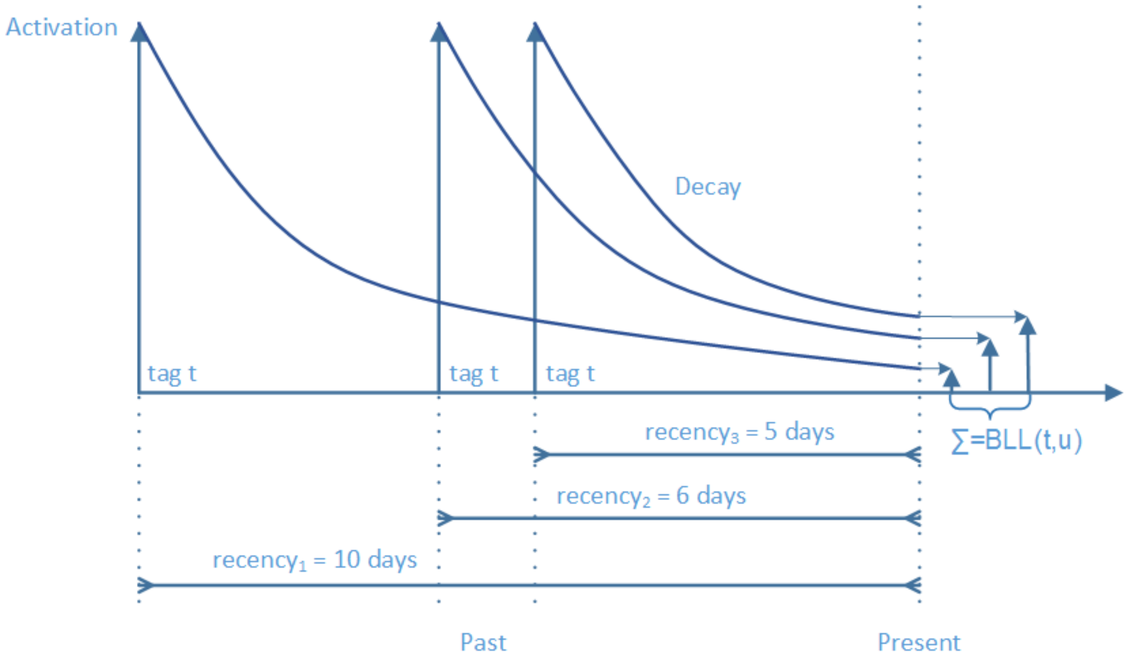


Figure 4.4.: The activation of a memory trace is depicted as an accumulation of the activation of individual instances that decline according to a power law function. The figure is adapted from Trattner et al. (2016a).

Application in Tagging. Kowald, Seitlinger, Trattner, et al. (2014) firstly applied and evaluated the activation equation in a tagging setting. Further research has been conducted for instance in Kowald, Kopeinik, Seitlinger, et al. (2015) and Trattner et al. (2016a).

The activation equation 4.2 comprises first, the base-level activation BLL and second, an associative component that represents the pattern of prior exposure, further referred to as semantic context. To model the semantic context of a resource in a tagging environment, tags other users have assigned to the current resource are considered. In this approach, W_j represents the frequency of appearance of a tag_j. S_{ji} represents the normalized co-occurrence of tag_i and tag_j as an estimate of the tags' strength of association

$$A_i = BLL + \sum_j W_j S_{ji} \tag{4.2}$$

Equation 4.3 determines a tag's usefulness in an individual person's past: with n providing the frequency of the tag being used in the past, and t_j indicating the recency of each tag use. More precisely, t_j is specified as the time since a tag has been used for the j^{th} time. The parameter d models the power law function of forgetting and is set to 0.5 which has turned out to be a reliable estimate of the rate of forgetting across a range or environmental settings and application scenarios (John R. Anderson and Schooler, 1991) .

$$BLL = ln(\sum_{j=1}^{n} t_j^{-d}) \tag{4.3}$$

This results in a weighted list (i.e. according to A_i) of all tags a person has either used in the past or the current resource is associated with.

4.2.2. MINERVA2

MINVERVA2 is a simulation model of episodic memory that has been investigated and suggested in Hintzman (1984). It mainly concentrates on memorization processes of the Long Term Memory (LTM) and claims to accommodate both, individual experiences and generic categories. To this end, it supposes that the LTM is a huge aggregation of episodic memory traces, i.e. representations of events or experiences a person was confronted with in the past. These memory traces are generated through the activation of a set of features in the Short Term Memory (STM). Further, the model assumes a principle of redundancy, where each event leaves behind its unique memory trace, even if an equivalent event is already stored. The role of the STM is constraint to the communication with the LTM which is described as i) sending retrieval cues or "probes" and, ii) retrieving "echo" responses.

When the STM sends a probe to the LTM all memory traces are activated in parallel. However, the degree of activation is given by their similarity to the stimulating probe. Consequently, the responding echo consists of an aggregation of all activated items weighted by their similarity values. Figure 4.5 presents a schematical illustration of this process.

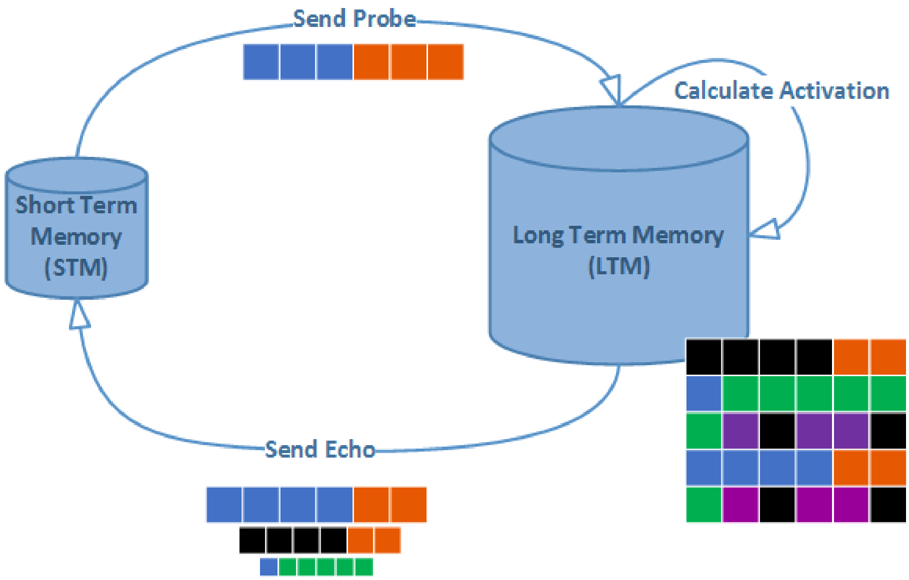


Figure 4.5.: Schematic illustration of information retrieval from long term memory.

The model has been investigated in multiple application areas. Despite its relatively simple approach, it has shown promising results for frequency judgement, schema abstraction, and associative learning (Hintzman, 1984).

Application in Tagging. Seitlinger, Ley, and Albert (2013) have further looked into MINERVA2's schema abstraction approach in order to mimic how humans retrieve words from memory, when confronted with external, contextual stimuli (e.g., topics). More precisely, a computation model for tagging in social learning systems is introduced. Within this application case, a memory trace is represented as two concatenated feature vectors. The first one depicts an input stimulus (i.e., color pattern) the second one its category membership.

As illustrated in Figure 4.6, the suggested approach is a fairly simple model that consists of an input, a hidden and an output layer. The input layer is a vector P (i.e. probe) of n features that describe a resource that has to be tagged. Aligned to it is a vector that represent m tags to describe the resource. The elements of this vector are unknown.

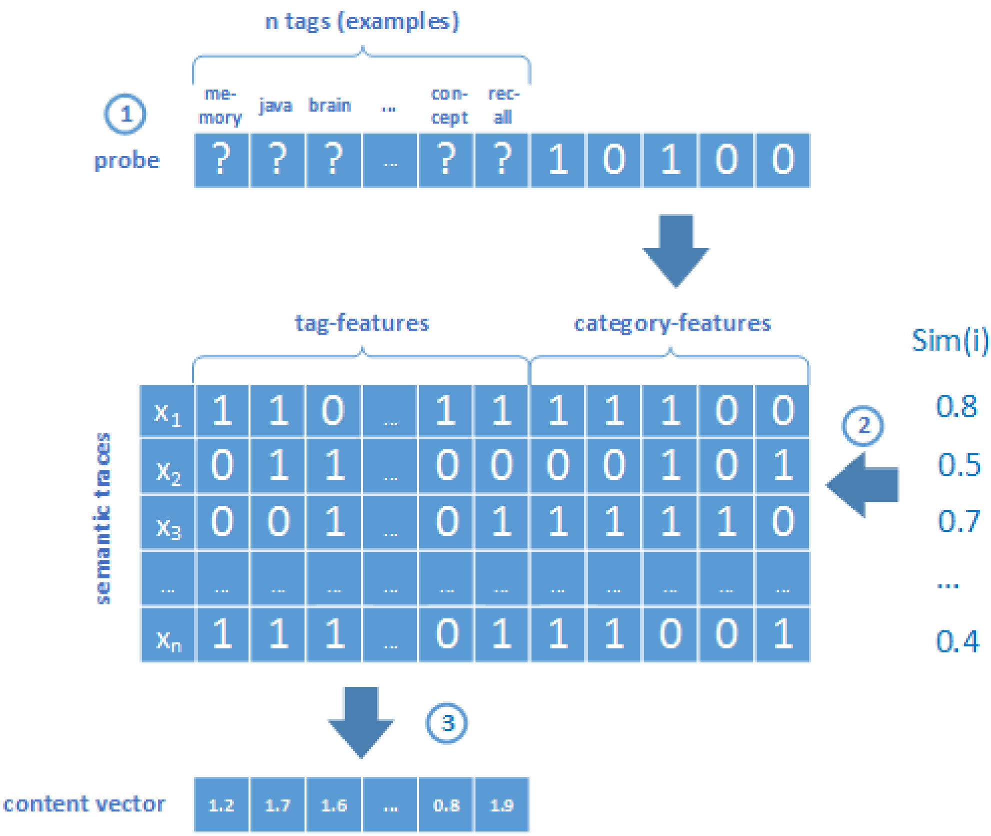


Figure 4.6.: Schematic illustration of the MINERVA2 mechanism applied in a tagging approach, adapted from Seitlinger, Ley, and Albert (2013).

The hidden layer constitutes the LTM and stores a user's semantic traces as feature vectors, where each trace is depicted as one vector X_i in a matrix. A vector constitutes of two parts: a tag-feature vector A_i and a topic-feature vector S_i. S_{ik} is the activation of feature k in trace i.

If a tag j was present in a memory trace depicted as X_i, the tag activation value a_{ij} is set to 1, and 0 otherwise. With the input vector acting as stimuli, the activation of single tags can be calculated. To this end, the cosine similarity for the input feature vector P and each topic-feature vector S_i in the matrix is calculated, following equation 4.4:

$$Sim_i = \frac{\sum_{k=1}^{n}(P_k \cdot S_{ik})}{\sqrt{\sum_{k=1}^{n} P_k^2} \cdot \sqrt{\sum_{k=1}^{n} S_{ik}^2}} \tag{4.4}$$

where P_k and S_{ik} are components of vector P and S_i respectively. Secondly, the the activation value t_j^{out} of tag j as the weighted sum of tag activation values over all traces in the dataset is calculated.

$$t_j^{out} = \sum_i Sim_i \cdot a_{ij} \tag{4.5}$$

The output layer is a weighted list of tags.

4.3. Conclusion

This Section provided a survey on cognitive modelling concepts as relevant to this thesis. It starts with a notion of the basic principals of the research field. Then, four cognitive models are discussed that are investigated as user models for learning recommendations. The selected theories are either applied for the recommendation or learning resources or tags.

The personalization of learning with selected learning content, presumes inferences of a learner's current and future abilities. Thus, for learning resource recommendations, in this thesis methods are considered that are able capture the dynamics of a learner's knowledge state. The Competence-based Knowledge Space Theory (CbKST) is an ontology-based model to organize learning domains in form of competence structures with assigned learning content and assessment items. These structures are typically created by domain experts, who shape specific learning domains and contexts. While it promises precise descriptions of the teaching domain, the process is cost intensive and restricted to pre-defined knowledge areas. In learning settings, where the learning process is unstructured such models are difficult to apply. This is why the second model (i.e. SUSTAIN) constitutes a more dynamic approach. It mimics human category learning by means of unsupervised clustering. The model is particularly flexible, because the number of clusters evolve according to a user's learning trajectory.

For the implementation of tag recommendations, this work draws on two process models that have been previously suggested for their application in tag recommender systems: i) BLL that considers frequency, recency and semantic context of memory traces to calculate the availability of according information in a person's declarative memory and ii) MINERVA2, which mimics information retrieval from long term memory through the recognition of semantic patterns.

The following Chapters elaborate on experiments that implement these four cognitive models in recommendation approaches within the context of learning. Evaluations take place either in online learning situations or based on offline datasets. In the next Chapter, the evaluation studies, baseline algorithms and evaluation metrics are introduced.

5. Evaluation Studies and Preliminaries

Within the course of this thesis different evaluation strategies have been implemented to study the effectiveness of a number of cognitive user models to serve as a basis for recommendation approaches within learning settings. This Section describes applied evaluation procedures, datasets and baseline algorithms. Content of this section has been partly published in Kopeinik, Kowald, Hasani-Mavriqi, et al. (2016) and Kopeinik, Kowald, and Lex (2016).

5.1. Offline Studies

The term offline studies is used to describe experiments conducted on collected log datasets.

5.1.1. Evaluation Protocol

To evaluate the algorithms on offline datasets, a common evaluation protocol in recommender system research is pursued (Kowald and Lex, 2015; Seitlinger, Kowald, et al., 2015).

Each user's activities are sorted in chronological order according to their timestamps. The timestamp indicates the point in time when an activity is traced in the system. Then, each user's activities are either assigned to a training- or a test-set. For each user, the most recent posts are used for testing and the rest for training. This simulates a real-world environment, where future interactions are predicted based on interactions in the past (Campos, Diez, and Ivan Cantador, 2013).

At this point, the protocol slightly differs with the type of recommendation being evaluated. In case of tag recommendations, the latest post of a user is put into the test-set, while the remaining posts are put into the training set (Kowald and Lex, 2015). In this context, a post describes all tags assigned by one user to one resource. To ensure that there is enough training data available per user, only users with at least two available posts are considered for the test-set.

When evaluating resource recommendations, 20% of the most recent activities of a user are selected for testing. The remaining 80% are used for training (see Seitlinger, Kowald, et al. (2015)). In the case of resource recommendations, only users with at least five available activities are considered. This procedure avoids a biased evaluation as no data is deleted from the original datasets.

To evaluate the performance of an approach in comparison to baseline methods, the top-n recommended resources for an algorithm and a user are compared with the relevant resources in the test-set. To this end, a variety of evaluation metrics are applied that are well-known in recommender systems research (Parra and Sahebi, 2013; Herlocker et al., 2004). A description can be found in Section 5.3.2.

5.1.2. Deriving semantic topics for resources.

Categories or semantic topics describing Web resources may be required as input for recommendation approaches. If datasets do not explicitly contain such properties for resources, Latent Dirichlet Allocation (LDA) (Blei, Ng, and Jordan, 2003) can be applied to generate an external categorization. LDA is a probability model that helps find latent semantic topics for documents (i.e., resources). In case of social tagging data, the model takes all tags of a dataset as input and provides an identified topic distribution per resource as output. To perform LDA, the Java framework Mallet[1] was applied. As suggested in related work (e.g.,Krestel, Fankhauser, and Nejdl (2009)) parameters were set to Gibbs sampling and l = 2000 iterations. In order to reduce noise and to meaningfully limit the number of assigned topics, the number of latent topics Z was set to 500 (see alsoKintsch and Mangalath (2011)). Furthermore, only topics that show a minimum probability value of .01 are considered to describe a resource. LDA can be formalized as

[1] `http://mallet.cs.umass.edu/`

follows:

$$P(t_i|d) = \sum_{j=1}^{Z} P(t_i|z_i = j)P(z_i = j|d) \tag{5.1}$$

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.

5.1.3. Identifying candidate resources.

Within the scope of this work, the term *candidate resources* describes a set of resources that serves as an item pool when calculating most suitable items for a user. In this context, *User-based Collaborative Filtering (CF_U)* (Schafer, Frankowski, et al., 2007) was used to identify $n = 100$ candidate resources for each user.

CF_U is implemented in two steps: First, by means of the cosine-similarity measure (see Zheng and Q. Li (2011)) most similar users (k nearest neighbours) of a target user are identified. Second, resources of the identified neighbours, unknown to the target user are selected. This pre-selection assumes that two users that had similar taste in the past, will also share this taste in the future and consequently, will engage with similar resources (Schafer, Frankowski, et al., 2007). As suggested in social tagging system literature (Gemmell et al., 2009), CF_U is calculated on the dataset's binary user-resource matrix with a neighbourhood size $k = 20$.

More formally, the prediction value $CF_U(u, r)$ for a target user u and a resource r is given by equation (5.2):

$$CF_U(u, r) = \sum_{v \in V_{u,r}} sim(u, v) \tag{5.2}$$

where $V_{u,r}$ is the set of most similar users of u that have bookmarked r. $sim(u, v)$ is the cosine similarity value between u and v.

5.1.4. Datasets

Table 5.1 summarizes the dataset properties such as posts, users, resources, tags, topics and their relations, as descriptive statistics. For the purpose of this study, *sparsity* is used to designate the percentage of resources that are not described by topics or tags. A more elaborate description of the datasets follows.

Delicious.

Delicious [2] is a free social web-service platform to collect, share and organize resources on the web. In this thesis, a publicly available dataset [3] from 2011 is used. This dataset encompasses information about social relations, bookmarked websites and user's bookmarking and tagging behaviour. The dataset does not include topics.

BibSonomy.

The university of Kassel provides SQL dumps[4] of the open social bookmarking and resource sharing system *BibSonomy* for the research community, in which users can share and tag bookmarks and bibliographic references. Four log data files are available, that report users' tag assignments, bookmark data, bibliographic entries and tag to tag relations. This dataset consists of tag assignment data that was retrieved in 2015 (*Benchmark Folksonomy Data from BibSonomy* 2013/2015). It does not include topics.

CiteULike.

CiteULike is a social bookmarking system for managing and discovering scholarly articles. Since 2007, CiteULike datasets[5] are published on a regular basis. The dataset for this study was retrieved in 2013 (resource recommendation dataset) and 2015 (tag recommendation dataset). Three log data files report on users' posting of articles, bibliographic references, and group membership of users. Activation data of user posts, including tags, have been used for this study since topics are not available in the dataset.

[2] https://del.icio.us/

[3] `http://files.grouplens.org/datasets/hetrec2011/hetrec2011-delicious-2k.zip`

[4] `http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/`

[5] `http://www.citeulike.org/faq/data.adp`

KDD15.

This dataset originates from the KDD Cup 2015[6], where the challenge was to predict dropouts in Massive Open Online Courses (MOOCs). The MOOC learning platform *XuetangX* was founded in 2013 by Tsinghua University and hosts more than 360 Chinese and international courses. Data encompasses course dates and structures (courses are segmented into modules and categories), student enrolments and dropouts and student events. For the purpose of this thesis, the event types *problem*, *video* and *access* that indicate a student's learning resource engagement were used. There are no tags in this dataset but categories that can be utilized as topics.

MACE.

In the MACE project [7], an informal learning platform was created that links different repositories from all over Europe to provide access to meta-data-enriched learning resources from architecture. 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). Unfortunately, up to now, it has not been possible to gain access to competence and topic data. However, data regarding users' access of learning resources and tagging behaviour has been used in this thesis.

TravelWell.

Originating from the Learning Resource Exchange platform [8], the 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 which both include additional information about the learning resource (Vuorikari and Massart, 2010). Data relevant to this work encompasses user names, resource names, timestamps, tags and categories.

[6] http://kddcup2015.com/information.html
[7] `https://www.fit.fraunhofer.de/en/fb/cscw/projects/mace.html`
[8] `http://lreforschools.eun.org`

Apsodle.

Aposdle is an adaptive work integrated learning system that stems from the Aposdle EU project. The target user group are workers from the innovation and knowledge management domain. The data used in this thesis was collected in a workplace evaluation that also included a context-aware resource recommender. Three files with user activities, learning resource descriptions with topics but no tags and a domain ontology were published (Guenter Beham, Stern, and Lindstaedt, 2010). The dataset contains data from six users. Within this thesis, the user actions *VIEW_RESOURCE* and *EDIT_ANNOTATION* are considered as indications for learning resource engagements.

Table 5.1.: Properties of the datasets that were used in offline evaluation studies. $|P|$ depicts the number of posts, $|U|$ the number of users, $|R|$ the number of resources, $|T|$ the number of tags, $|Tp|$ the number of topics, $|AT_r|$ the average number of tags a user assigned to one resource, $|ATp_r|$ the average number of topics describing one resource, $|AR_u|$ the average number of resources a user interacted with, $|AU_r|$ the average number of users that interacted with a specific resource. The last two parameters SP_t and SP_{tp}, describe the sparsity of tags and topics, respectively.

	$\|P\|$	$\|U\|$	$\|R\|$	$\|T\|$	$\|Tp\|$	$\|AT_r\|$	$\|ATp_r\|$	$\|AR_u\|$	$\|AU_r\|$	SP_t	SP_{tp}
Delicious	59651	1819	24075	23984	0	4.2	0	32.8	2.4	0	100
BibSonomy	82539	2437	28000	30889	0	4.1	0	33.8	3	0	100
CiteULike	105333	7182	42320	46060	0	3.5	0	14.7	2.5	0	100
KDD15	262330	15236	5315	0	3160	0	1.8	17.2	49.4	100	1.1
TravelWell	2572	97	1890	4156	153	3.5	1.7	26.5	1.4	3.2	28.7
MACE	23017	627	12360	15249	0[9]	2.4	0	36.7	1.9	31.2	100
Aposdle	449	6	430	0	98	0	1.1	74.8	1	100	0

[9]Generally the dataset contains topics but unfortunately, at this point, we do not have them available.

5.2. Online Studies

Within this context, the term online studies describes controlled experiments conducted with representative end-users, either implemented as laboratory or real-world settings. In contrast to recommender evaluations using offline data, online experiments allow the evaluation of support tasks rather than prediction tasks. In other words, the algorithm is applied in real-time, supporting the learning while engaging in the learning process. The learner decides whether to accept the systems suggestions or not, thus judging its helpfulness. (McNee, Riedl, and Konstan, 2006)

5.2.1. Evaluation Protocol

While the detailed experimental setups vary with the specifics of the two online studies, the overall protocol is based on the principles of A/B testing (e.g.Ronand Kohavi and Longbotham (2016)). A/B testing is an approach commonly used in controlled online experiments (Ron Kohavi et al., 2014). In line with the protocol, study participants are at first randomly split in two groups: group A and group B. Then, both groups are instructed to complete similar tasks, while being provided with different variations of the online environment. However, the study setup differs only with respect to the effects that are investigated. Due to this, change appearing with new properties can be measured by for instance comparing log data, questionnaires or task outcomes between the two groups (Ronand Kohavi and Longbotham, 2016).

Ethics

As suggested by the American Psychological Association [10], online studies presented in this thesis have been conduced in line with common ethical standards and requirements. This includes an initial presentation of the research project and complementary information on the research process and activities, informed consent of each participant and their legal representatives (if applicable), a clear understanding of voluntary participation, the collection of a minimum of bi-

[10] http://www.apa.org

ographic data and the anonymous analysis, interpretation and publication of data.

5.3. Baseline Algorithms and Metrics

This Section outlines a number of selected algorithms and performance metrics that have been applied in comparative evaluation studies of this thesis.

5.3.1. Baseline Algorithms

The here described algorithms MP, CF, CB_T, and WRMF constitute standard baseline algorithms. UCBSim was selected because it has been proposed in the context of TEL before.

Most Popular (MP).

MP is a simple approach to rank items according to their frequency of occurrence (Jäschke et al., 2007; Parra and Sahebi, 2013). The algorithm can be implemented on user-based, resource-based or group-based occurrences and is labelled respectively, as MP_U, MP_R and MP. $MP_{U,R}$ describes a linear combination of MP_U and MP_R. Unless implemented on user-based data, it represents a non-personalized recommendation approach that suggests the same set of items to any user.

Collaborative Filtering (CF).

CF is a memory-based recommendation approach that filters and ranks items according to other users' preferences. The CF algorithm can be implemented in different variations. User-based Collaborative Filtering (CFU) calculates the neighbourhood of users U to find items that are new to a user by considering items that similar users engaged with in the past (Schafer, Frankowski, et al., 2007). The neighbourhood is defined by the k most similar users calculated by the cosine-similarity measure on the binary user-resource matrix. Resource-based Collaborative Filtering (CF_R), which is also known as Item-based CF, identifies potentially interesting resources for a user by computing similarities between resources. Hence, this approach processes the resources a user has engaged with in the past in order to find similar resources to recommend (Sarwar et al., 2001).

Tag recommendations require the triple: (user, resource, tag). In this thesis, for tag recommendations the adaptation of CF_U as suggested by Marinho and Schmidt-Thieme (2008) was used. Accordingly, the neighbourhood of a user is determined through a user's tag assignments instead of resource engagements. As suggested in the literature (Zheng and Q. Li, 2011; Gemmell et al., 2009), k is set to 20 for all CF implementations.

Content-based Filtering (CB).

CB recommendation algorithms rate the usefulness of items by determining the similarity between an item's content and the target user profile (Basilico and Hofmann, 2004). For the experiments of this thesis, the approach (CB_T) is implemented either by using topics, if available, or otherwise by using tags to describe the item content (see Section 5.1.4 for dataset properties). The similarity between the item vector and the user vector is calculated by the cosine-similarity measure.

Weighted Regularized Matrix Factorization (WRMF).

WRMF is a model-based recommender method for implicit data (e.g., posts) based on the state-of-the-art Matrix Factorization (MF) technique. MF factorizes the binary user-resource matrix into latent user- and resource-factors, which represent these entities, in a common space. This representation is used to map resources and users and thus, to find resources to be recommended for a specific user. WRMF defines this task as a regularized least-squares problem based on a weighting matrix, which differentiates between observed and unobserved activities in the data (Y. Hu, Koren, and Volinsky, 2008). Results for WRFM presented in this work have been calculated using the MyMediaLite 3.10 framework[11] (2013-09-23) with k = 500 latent factors, l = 100 iterations and a regularization value λ = .001.

[11] http://www.mymedialite.net/

Usage Context-based Similarity (UCbSim).

The algorithm was introduced by Friedrich et al. (2007) and further discussed in the TEL context by Niemann and Wolpers (2013) and Niemann and Wolpers (2015). The approach is inspired by paradigmatic relations known in lexicology, where the usage context of a word is defined by the sequence of words occurring before or after it in the context of a sentence. The equivalent to a sentence in online activities is defined as a user session, which describes the usage context. In line with Niemann and Wolpers (2013), the significant co-occurrence of two items i and j is calculated by the mutual information (MI):

$$MI_{i,j} = log_2 \frac{O}{E} \tag{5.3}$$

where O is the number of observed co-occurrences and E the number of expected co-occurrences. The similarity ($sim_{i,j}$) between two objects is given by their cosine-similarity, where each object is described as a vector of its 25 highest ranked co-occurrences. For this study, the algorithm recommends resources that are most similar to the resources users engaged with in their last session. Further, a session is assumed to be completed if no user interaction is observed for 180 minutes.

5.3.2. Metrics

In this research, the metrics recall, precision and f-measure are applied to evaluate the accuracy of recommendation approaches. They have been selected, due to their popularity in recommender system research (Marinho, Hotho, et al., 2012; Verbert, Drachsler, et al., 2011). nDCG was selected as the most suitable measure to evaluate the quality of item rankings (Sakai, 2007).

Recall.

Recall (R) indicates the fraction of the k recommended items that are relevant to a user (i.e., correctly recommended items), to all items relevant to a user. I_u represents items a user u engaged with and $\hat{I}_u$ items that were recommended to the user.

$$R@k = \frac{|I_u \cap \hat{I}_u|}{|I_u|} \tag{5.4}$$

Precision.

The precision (P) metric indicates the fraction of the k recommended items that are relevant to the user.

$$P@k = \frac{|I_u \cap \hat{I}_u|}{|\hat{I}_u|} \tag{5.5}$$

F-measure.

The F-measure (F) calculates the harmonic mean of recall and precision. This is relevant as recall and precision normally do not develop symmetrically.

$$F@k = 2 \cdot \frac{(P@k \cdot R@k)}{(P@k + R@k)} \tag{5.6}$$

nDCG.

Discounted Cumulative Gain (DCG) is a ranking quality metric that calculates usefulness scores (gains) of items based on their relevance and position in a list of k recommended items and is calculated by

$$DCG@k = \sum_{i=1}^{k} (\frac{2^{B(i)} - 1}{log_2(1+i)}) \tag{5.7}$$

where $B(i)$ is 1 if the i^{th} recommended item is relevant and zero if not. To allow comparability of recommended lists with different item counts, the metric is normalized. nDCG is calculated as DCG divided by the ideal DCG value iDCG, which is the highest possible DCG value that can be achieved if all relevant items are recommended in the correct order, formulated as $nDCG@k = \frac{DCG@k}{iDCG@k}$ (Sakai, 2007).

Mean average precision (MAP)

This metric is described to explain the specific use case of tag recommendations, as in this work, it was not applied to evaluate resource recommendations.
Given a list of tags T relevant to a user u and a resource r, the mean average precision metric (MAP) is calculated, combining the ranking of a list of recommended tags $\hat{T}_{u,r}@k$ with their precision values. The metric is applied as depicted in the subsequent formula, where B_k is set to 1 if the recommended tag at position k

occurs in the list of relevant tags $T_{u,r}$, it is set to zero otherwise (Rawashdeh et al., 2012):

$$MAP@k = \frac{1}{|T_{u,r}|} \sum_{k=1}^{|\hat{T}_{u,r}@k|} B_k \cdot P_{u,r}@k \tag{5.8}$$

6. Learning Resource Recommendations (RQ1 & RQ2)

This Chapter deepens the discussion on the two cognitive learner models which have been introduced in Section 4.1. It suggests exemplary application approaches of resource recommendation strategies to personalize TEL environments and is split into two parts: First, Section 6.1 presents an approach of using the CbKST to recommend learning resources in Moodle, a popular learning management system. The approach's suitability to enhance the perceived learning experience of the target user group is evaluated in a laboratory study. In Section 6.2, SUSTAIN is applied to enhance a collaborative filtering resource recommendation approach by considering a person's learning dynamics. Furthermore, an offline study is presented on three social bookmarking datasets that reports recommender accuracy of the resulting SUSTAIN+CF_U model and which is used to investigate underlying dynamics of the model. The two Sections address RQ1 and RQ2, respectively.

6.1. Structural Learner Models: CbKST for Ontology-based Resource Recommendation

Recommender systems in TEL are highly dependent on environmental variables such as the target group, the learning environment, learning goals and the domain they are applied in (Drachsler, H. G. K. Hummel, and Koper, 2009). Thus, in many cases, it is not possible to transfer a recommender system from one context to another without the adaptation of underlying learning goals and models, as elaborated earlier. Domain and learner models are therefore considered a crucial factor in the adaptation of learning environments. For this thesis, the CbKST (see Section 4.1.1) is proposed as a framework that can be applied to create domain and learner models incorporating the assignment of learning resources and learning goals. This section introduces and discusses an approach to design a resource recommendation strategy based on the CbKST.

The content of this Section has been partly published in L.-C. Winter et al. (2013), Dimache, Kopeinik, et al. (2014), and Kopeinik, Nussbaumer, L. C. Winter, et al. (2014) and Dimache, Roche, et al. (2015).

6.1.1. Approach

For this approach, the CbKST is used to create a domain model that describes the learning domain as competences and interrelations between those competences (see Figure 4.2). This model is generated by domain experts, i.e., teachers. In a first step, all competences are identified that are relevant to the learning domain. Then, these competences are aligned according to a prerequisite structure. These prerequisite relations between competences indicate a hierarchy grounded on part of relationships of learning content. For instance, if students want to solve multiplications, they need know how to perform additions before.

This hierarchy structures the learning process through the introduction of suggested sequences. Furthermore, a mapping of Learning Resources (LR) and Assessment Items (AI) to competences (see Figure 6.1) specifies the relation between competences and learning content. As suggested in the literature (e.g., Brusilovsky and Millán (2007)) the domain model is further used to instantiate learner models.

CbKST-based Learning Resource Recommendations

CbKST-based learning resource recommendations rest upon a predefined learner model. Such a learner model comprises of an instance of a domain model, consisting of interrelated competences with assigned LR and AI, probabilistic competence values, a learning goal and a memory of learning resource interactions.

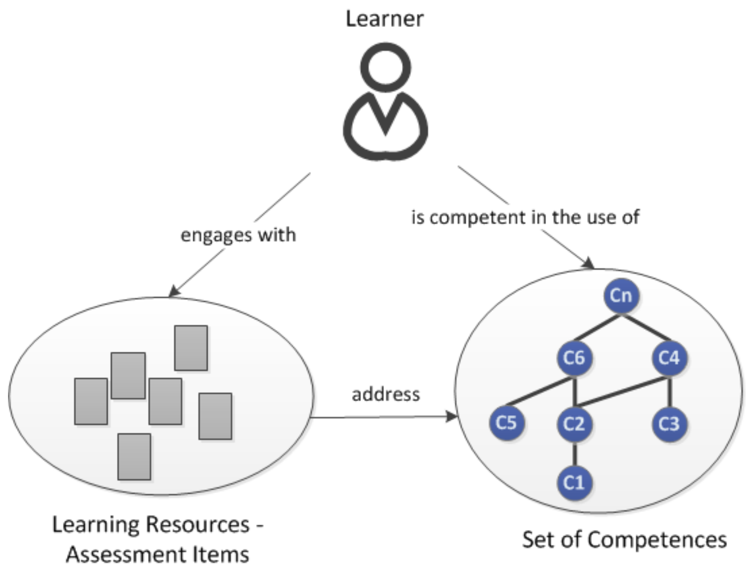


Figure 6.1.: Relation of elements in a CbKST-based adaptation model, adapted from Kopeinik, Nussbaumer, Bedek, et al. (2012). It shows that the learner is modelled by means of a competence structure. Learning resources and assessment items are aligned with competences. The learner engages with learning resources to become competent in targeted knowledge areas.

The learner model is initialized with an estimation of a learner's competence state. This can either be ground on peer- or self- assessment or the completion of automatic assessment procedures, such as supported by the CbKST framework. In a CbKST-based adaptive assessment procedure, learners complete assessment items to evaluate their competence state i.e., a certain set of competences a learner demonstrates. In this process, a posed question is selected relative to a learner's assumed competence state. That means that the correctness of the answer to the first posed question (which has a moderate difficulty) is determining the selection of the next question and so on. Thus, questions are presented in an adaptive way, which is appropriate for a learner's knowledge at the time. (Falmagne, Albert, et al., 2013)

The adaptive recommendation of learning resources builds upon the adaptive assessment. This means, the presented content is selected in accordance with the learner's estimated current competence state, as determined in the most recent assessment procedure. Being aware of a learners competence state, learning resources with an expected medium level of difficulty are suggested. To mathematically infer a student's competence probabilities the Simplified Updating Rule (SUR) (Augustin et al., 2013) was applied in this research. According to the SUR, for each competence in a learner model, a probabilistic value is calculated that indicates whether a student shows a competence or not. Competences with a probability value close to 0.5 are assumed medium difficult to a learner.

Self-Regulated Learning

Self-Regulated Learning (SRL) plays an increasing role in the modern world of education. It emphasises the learners' ability to control and regulate their own learning process. Most researchers define SRL as the way individuals control their own feelings, thoughts and behaviours which are oriented towards goal achievement (Zimmerman, 2000; Zimmerman, 2002; Zimmerman and Schunk, 2012). The SRL process is accomplished in a proactive way, in which a learner's self-regulation of cognitive, metacognitive and motivational processes (within an educational context) is emphasised (Zimmerman, 2002; Hetzner et al., 2011). Effectively, this means that the learners 'direct' their own way of learning based on their own decisions. Meta-cognition, which is required for self-reflection, plays a crucial role in this model. Zimmerman (2002) described meta-cognition as the "awareness of and knowledge about one's own thinking". Meta-cognitive strategies are therefore defined as the attempt to be aware of one's own lack of knowledge. The process of SRL can be divided into three phases (Zimmerman, 2002):

1. **Forethought Phase:** This phase takes place before the actual learning. The learner initiates the learning endeavour by actively thinking about it. This includes activities like analysing and structuring tasks, setting learning goals or being mindful of intrinsic motivations. In this work, this phase is further referred to as *Planning Phase*.
2. **Performance Phase:** Within this phase the actual learning takes place while the user holds on to planned learning strategies and means to observe the learning process. In this work, this phase is further referred to as *Learning Phase*.
3. **Self-Reflection Phase:** In addition to acquiring domain knowledge, the learner applies meta-cognitive activities when taking control over- and

reflecting on learning. In this work, this phase is further referred to as *Reflection Phase*.

6.1.2. Combining the CbKST and Self-Regulated Learning for Personalisation in Moodle

Within the scope of the INNOVRET project (Innovative Online Vocational Training of Renewable Energy Technologies) an online training solution was developed tailored to the requirements of heat pump installers. The approach combines the benefits of SRL and Competence-based learning on the basis of the CbKST.

The combination of SRL and CbKST enhances the benefits for the learner by enriching the learning experience with a combination of self-management, reflection and guidance. Inspired by ideas described in Nussbaumer, Gütl, and Hockemeyer (2007), an adaptive learning approach was designed, implemented and integrated in Modular Object-Oriented Dynamic Learning Environment (Moodle). A schematic design of the model is presented in Figure 6.2.

Each SRL phase is supported by a Moodle plug-in, which rests upon principles of the CbKST. In other words, INNOVRET's (SRL) tools are based on a structured competence model, complemented by corresponding learning resources and assessment items (AIs), which form the CbKST Domain Model. SRL takes place when learners select a learning target in the Planning Tool (Planning Phase), browse through the list of learning resources that are suggested by the Recommendation Tool (Learning Phase), and use the Learning Progress Tool to reflect upon the learning progress they have achieved in previous sessions (Reflection Phase).

Figure 6.3 shows the design of the learning process. It depicts the link between the CbKST services and the three SRL phases.

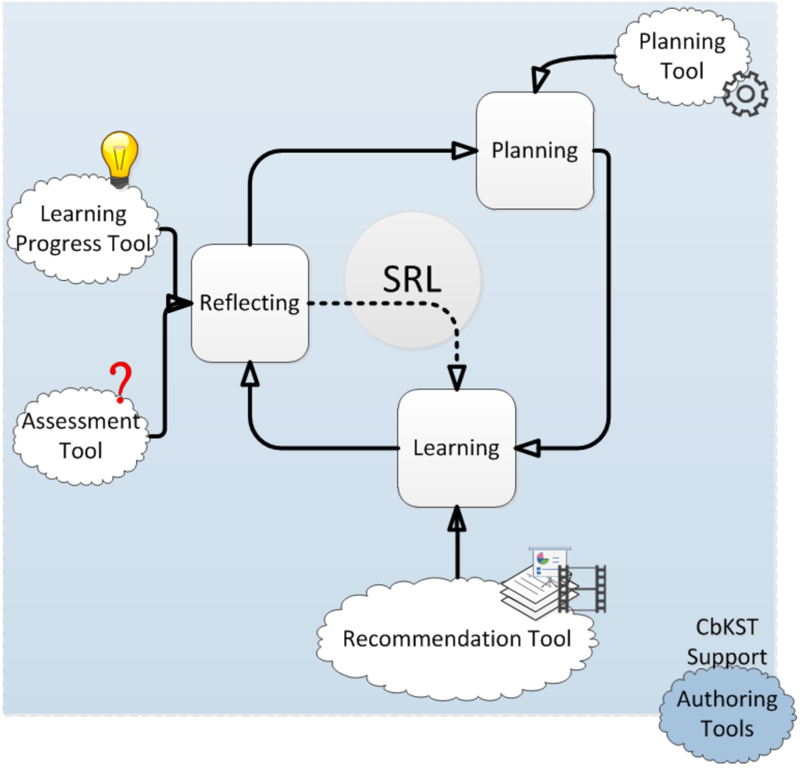


Figure 6.2.: Schematic overview of the combination of SRL and CbKST. Each SRL phase (Planning, Learning, Reflecting) is supported by a Moodle plug-in which rests upon principles of the CbKST.

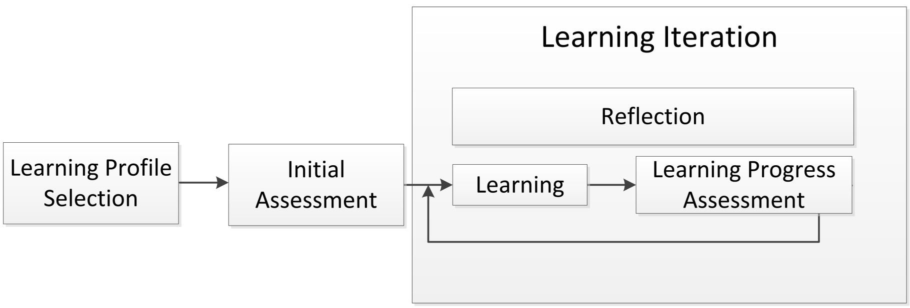


Figure 6.3.: Shows a complete learning process.

A more detailed description of this coupling is given in the following paragraphs.

1. **Planning**
   In the planning phase, the CbKST learner model is initialized. This constitutes of two steps:

Learning Profile Selection: Learners select a learning profile. The learning profile consists of a set of competences the learner aims to show after completing a learning cycle. This set of competences is a subset of competences of the underlying domain model.

Initial Assessment: The learner completes a CbKST-based adaptive competence assessment that determines probability estimates that represents the learner's current competence state (i.e. the set of competences in the learning profile a learner already demonstrates). This data is used to initialise the learner model of the Recommendation Tool.

2. **Learning**
   After completing the initial assessment, the learner model has been initialized with probability values allocated to single competences. The specific learner model is thereby instantiated from the domain model. It consists of competences, interrelations between competences and probability values that indicate the likelihood to which a learner may demonstrate a competence. In the learning phase, the Recommendation Tool suggests learning resources based on probability levels of the learner model that are derived from the CbKST algorithm. In this way, learning resources are selected that have a medium level of difficulty for the learner. Competences with a probability value of 0.5 are assumed to be medium difficult to engage with.
3. **Reflecting**
   Visualisations of the learner's competence state derived from completed assessments, and the learner's learning history support reflection and awareness.

   Learning Progress Assessment: Competence probabilities in the learner model are updated with results of every assessment. This is done by applying the SUR in a learner's learning profile with positive values for correctly answered questions and negative values for incorrect answers. The newly updated learner model serves as a basis for the next learning iteration. Log data and assessment results are visualized to support users reflection processes.

The learning process is further divided into learning iterations (see Figure 6.4). A learning iteration is defined as the period of time between two complete consecutive assessments. Thus, it consists of a learning phase and the assessment of the achieved learning progress during this phase. A reflection plug-in is offered to reflect on learning process and progress.

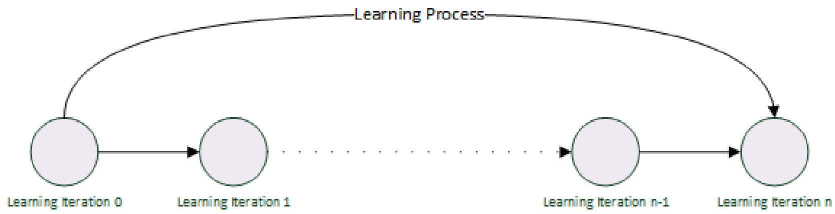


Figure 6.4.: Users engage in learning iterations until reaching a defined learning goal.

The next section focuses on the implementation of plug-ins and services within the Learning Management System Moodle.

Environment and Plug-ins

The described personalization strategy and its components were implemented as plug-ins communicating with CbKST WebServices. Plug-ins are embedded in Moodle.

Environment. Moodle[1] (Modular Object-Oriented Dynamic Learning Environment) is an open source Learning Management System (LMS), which is capable of supporting high levels of interaction, web visibility, online social networking, and knowledge exchange between learners (Despotović-Zrakić et al., 2012). Moodle includes many features that improve pedagogical quality (Aydin and Tirkes, 2010), such as communication and collaboration tools or student tracking tools.

Furthermore, the toolset allows teachers and course developers to create and manage online courses modularly. It supports a variety of different LR formats and question types (herein further referred to as assessment items (AIs)). Despite the afore-mentioned features, Moodle is usually course based and does not cater to the individual needs of students (Albert, Nussbaumer, and C. Steiner, 2008). However, its extensibility through the integration of plug-ins and modules, allows for the introduction of personalised learning support, which gives the learner more freedom to control their own learning process. This caters to the needs of

[1]https://moodle.org/

non-homogeneous student groups (Wilson et al., 2007) while taking advantage of, and integrating with the existing infrastructure.

Personalization Plug-ins. Here presented plug-ins were designed and implemented in the course of the INNOVRET project (see Section 1.3). The design of user interfaces and learning resources were highly influenced by the project's target user: A learner working full-time as a trained plumber, typically lacking in computer skills and unaccustomed to learning at a desk. Accordingly, the learning application intends to be as straight forward as possible, providing users with pedagogical and adaptive features while hiding unnecessary complex information such as competence structures.

Functionality offered to learners include: initiating learning processes, determining learning goals with the aid of predefined job profiles, completing assessments and consuming learning resources which are recommended according to their learner model and target learning profile.

Initially, the learner is presented with a Graphical User Interface (GUI) (see Figure 6.5) to select a target learning profile (e.g., gaining basic knowledge in heat pump installation) from a list of options.

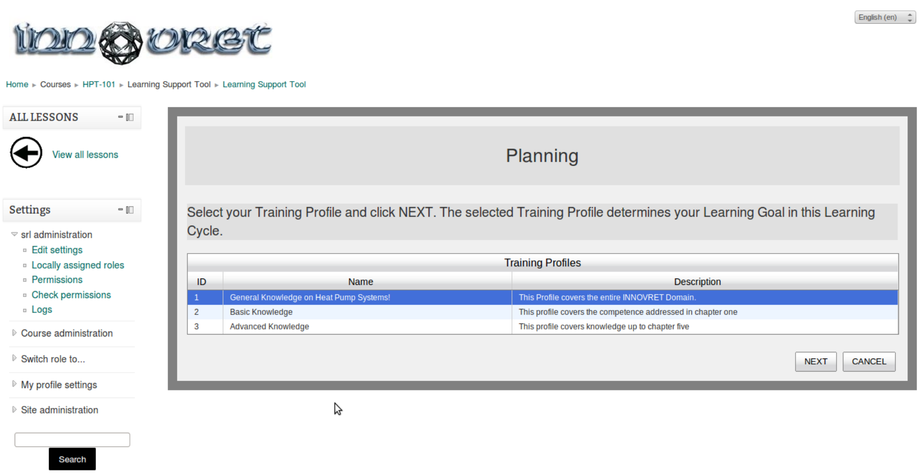


Figure 6.5.: Selecting a learning profile

Training profiles are defined by the course instructor and consist of any number of competences that are part of the competence domain. They are read from xml configuration files and can be altered at any time. After the user selects the training profile, data is sent to the CbKST backend service where the learning profile is constructed by limiting the general competence domain to the selected set of competences. This is already taken into account for the initial competence assessment where the learner is not asked questions about the whole domain but only about the competences in his/her learning profile. The initial competence assessment is adaptive. Depending on a learner's probabilistic competence values and the consistency of answers there will be more or less questions asked until at least 90 % of a user's competences show probabilistic values that indicate either positive or negative tendencies (i.e. $p<40$ or $p>60$). Additionally, a question limit can be set that acts as break rule if the limited number of questions exceeds. The assessment plug-in illustrated in Figure 6.6 builds upon Moodle's question engine and thus, supports Moodle's question formats.

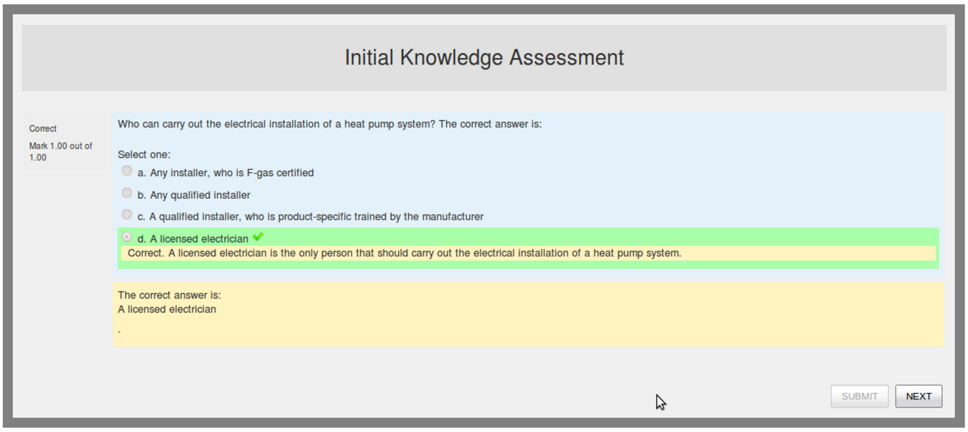


Figure 6.6.: Initial competence assessment supporting Moodle question formats.

After completing the planning phase of a learning cycle, the learner model has been initialized and the learner is presented with the plug-in's Main Menu (see Figure 6.7).

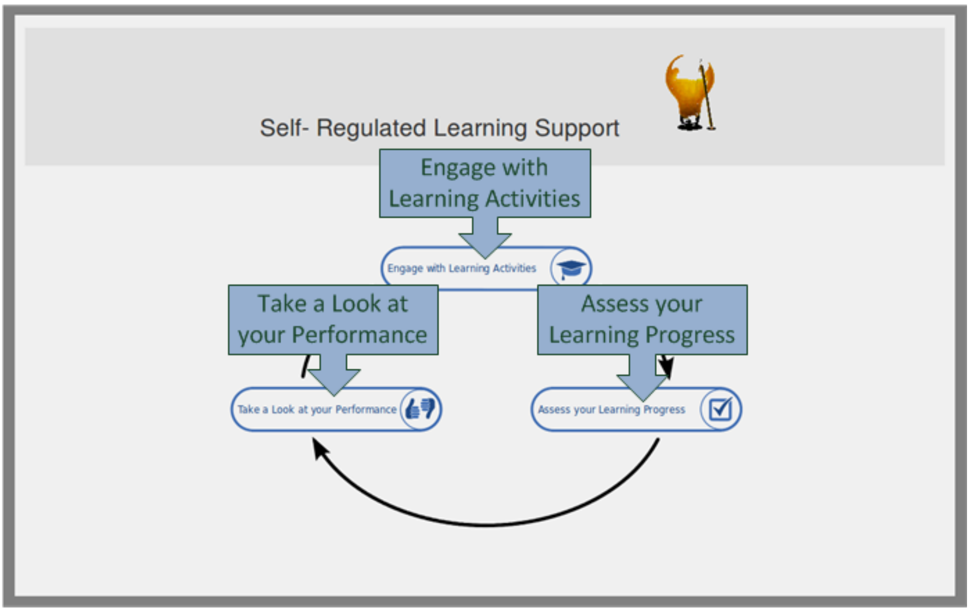


Figure 6.7.: Main Menu of the Plug-in.

The Main Menu offers three options:

I. **Engage with Learning Activities.** In Figure 6.8 the learner is presented with a list of learning resources that the CbKST-based Recommendation Tool selects according to the learner's last determined competence state. To this end, the plug-in integrates with standard Moodle plug-ins to present different formats such as pdf, html or ppt. Learners' interaction with the learning resources is logged and fed into the learner model. An assessment of the learning progress can be started by selecting the *Start Assessment* button in the Learning Recommendation view or via the plug-in's Main Menu.

II. **Assess your Learning Process.** Questions presented in a *Learning Progress Assessment* relate to the competences that have been addressed in a learning iteration. For each learning iteration there is only one *Assessment* to be taken, with this the learning iteration ends, a new competence state is calculated and the learner is presented with newly recommended LRs.

III. **Take a Look at your Performance.** In line with a stakeholder requirement analysis the reflection comprises three tabs, as displayed in Figure 6.9.

Graphical representations are implemented using the YUI charts library [2]. When opening the tab, user data is loaded from the CbKST backend service and presented in the charts. Figure 6.9a presents a user's relative learning performance within a learning iteration in relation to his or her experience level (yellow line). The x-axis shows the number of iterations starting with zero. Iteration zero exposes data from a learner's initial knowledge assessment that takes place within the planning phase. A learning iteration is determined by the time the user completes an assessment until the user completes the subsequent assessment. The green bar shows the percentage of properly answered questions in relation to the total number of questions within the assessment. The blue bar shows the percentage of consumed learning objects in relation to the recommended ones. The *experience level* is a mean value calculated over the sum of competence probabilities of the user's learning profile.
Figure 6.9b illustrates the total number of questions answered correctly and incorrectly per learning iteration. Correctly answered questions are displayed in green, others in red. The x-axis provides information on the learning iteration and the y-axis plots the total number of questions.
Figure 6.9c presents learning resources a learner consumed in green and the number of learning resources that have been recommended to the learner in blue.

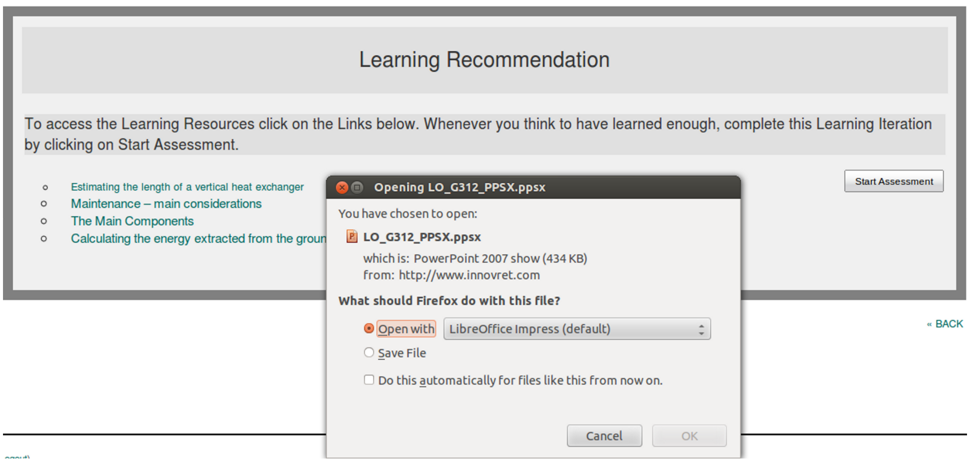
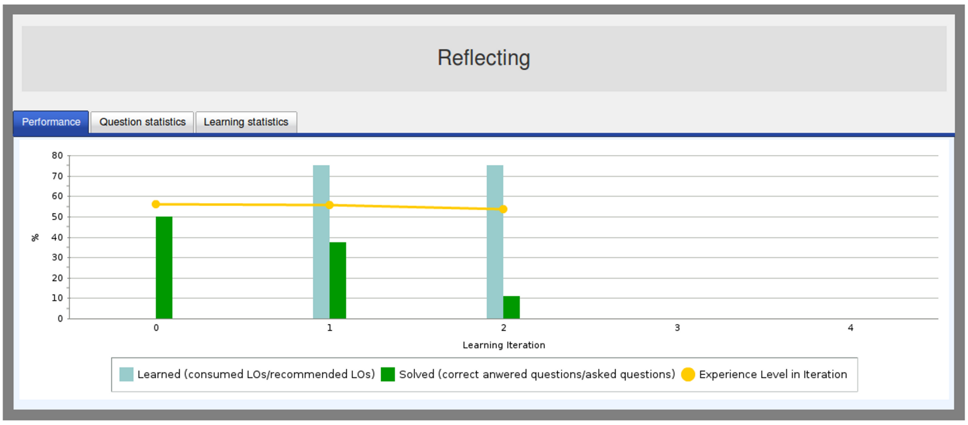


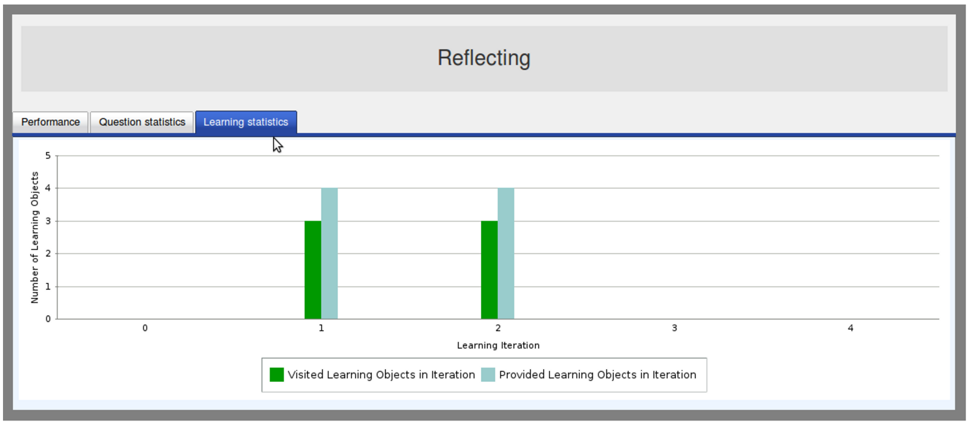
Figure 6.8.: List of Recommended Learning Resources within a Learning Iteration

[2]http://developer.yahoo.com/yui/examples/charts/index.html



(a) Tab one gives an overview of the learning progress.

(b) Tab two shows the total number of questions answered correctly and incorrectly per learning iteration.



(c) Tab three shows the total number of learning objects that were consumed in comparison to the recommended ones.

Figure 6.9.: Reflection of the learning progress. Results are structured according to learning iterations.

Technical Insights

To assure an easy and seamless integration, the personalization strategy was implemented as Moodle plug-in. A clean implementation of the modules is further supported by the application of the Soda plug-in [3]. Soda provides a model-view-controller framework to structure a module's code. YUI [4] was used as java script and CSS library. Whenever possible, the plug-in reuses or integrates with existing Moodle plug-ins, modules and data structure. Examples include presenting the LR or posing questions during the assessment.

Data presented in the plug-in is either retrieved from the CbKST backend service or the Moodle database. Modifications and extensions of data related to the CbKST Logic (i.e. assessment results, target profiles, log data, learner model) are implemented in the CbKST backend, i.e. the Compod Web Service (Nussbaumer, Hillemann, et al., 2015). The communication between the plug-in and the CbKST backend services is realized via REST-based web-services.

6.1.3. Evaluating the User Experience

In the course of this thesis, the approach's suitability to enhance the perceived learning experience of the target user group has been evaluated in a laboratory setting.

Methodology

The proposed learning software was evaluated with fourteen users of the learning system's target group, i.e. male adolescents at the time studying heat pump installation in Ireland. The participants differed in age, proficiency level and computer skills. The study of two hours took place in a computer laboratory, where each participant had access to his own PC. Furthermore, it was following an A/B testing approach and was conducted sensitive to ethical concerns as described in Section 5.2.1.

Accordingly, participants were split randomly in two groups:

[3]http://tech.solin.eu/doku.php?id=moodle:using_soda_to_create_new_moodle_modules
[4]http://yuilibrary.com/

1. The control group A had access to training material via Moodle as a learning management system.
2. The experimental group B used the Moodle version with adaptive learning support tools described in Section 6.1. However, they were accessing the same training material as group A.

Before a two hour studying session, an introduction to the overall approach and the user interface was provided. Then, during the studying session, the participants had to engage in learning activities using their assigned learning environment. The learning content which was specifically developed for the target group, covered technical knowledge about heat pumps, their installation, and maintenance. Following the learning phase the participants were asked to complete a questionnaire on their learning and system experiences (see Appendix B). The questionnaire contained eight to ten questions (eight for the Moodle group) to be answered on a Likert scale from "strongly disagree" to "strongly agree". In addition, a researcher was present for questions and to monitor the participants during the usage of the system. Also, log data of the CbKST group was captured to gain insights into the participant's individual system usage.

6.1.4. Results and Discussion

Table 6.1 presents the post questionnaire and the distribution of participants' answers in percentage terms. The mean values of the results of both groups are illustrated in 6.10, where the answer range is scaled on values from 1 (strongly disagree) to 5 (strongly agree).

Three of the posed questions addressed the overall approach, namely the iterative learning process (Q1), the awareness support (Q2), and the guidance support (Q4). Results of the experimental group reached mean values above average, which suggests that the CbKST-based personalisation of the learning environment leads to a perceived enhancement of the learning experience. Particularly, questions that were articulated in a more precise manner (Q2 and Q4) scored higher. When comparing the two groups in respect to Q4, the experimental group shows better results indicating the added value of the additional guidance.
Q3 and Q5 depicted two negatively posed questions concerning learning problems, namely: if this approach was limiting or stressful. According to the answers,

participants in the experimental group perceived the system as less limiting to their learning than those of the control group. However, there was no significant difference between the groups regarding the perceived stressfulness.
Additional questions targeted the participants' enjoyment in learning (Q6) and their perceived learning success (Q7). In both aspects, the experimental setting scored above average and better than the control group. Q8 and Q10 dealt with usability of the learning environment and the quality of presented learning content, respectively. Mean values of both questions are above average and do not differ from the control group's results. Q9 on the other hand, which was evaluating whether the participants would like to use a system like this in the future, resulted clearly above average with more than 80% of the participants agreement. Also, this score was considerably better in comparison to the control group.

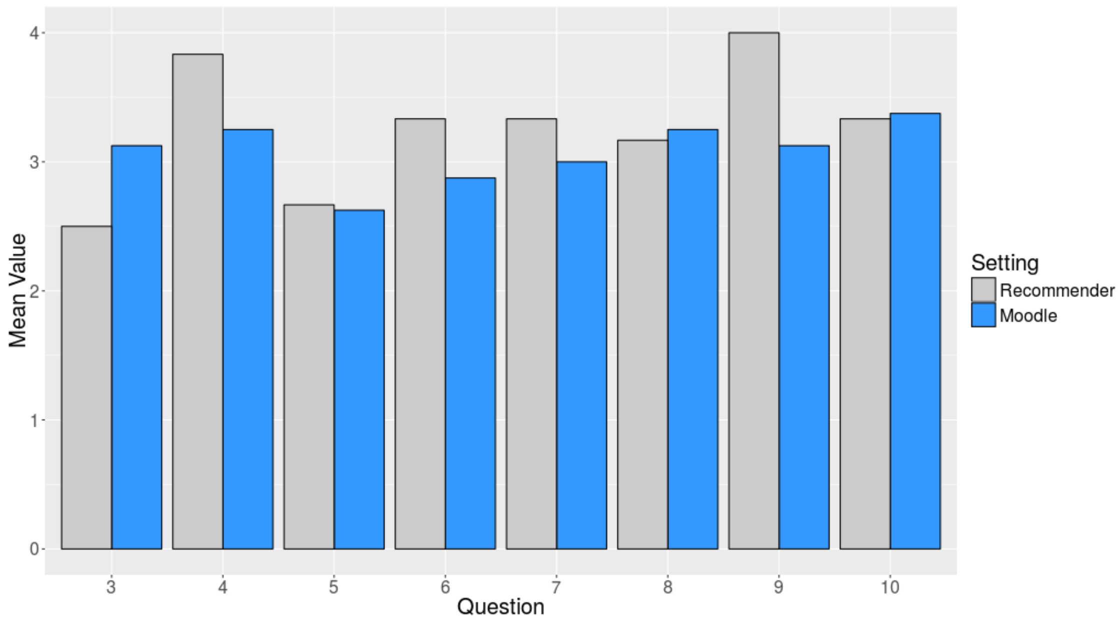


Figure 6.10.: Evaluating the acceptance of the CbKST recommendation setting: mean values of the control group A (N=8) and the experimental group B of the evaluation questionnaire (see Appendix B).

Observations by a researcher revealed that the IT skills of the participants (installers) played an important role in the way they perceived the system and the entire learning experience. Installers, whose IT skills were good or average did not have any problems in navigating and interacting with the system, whereas installers who were not used to using computers found the system itself to be a barrier.

Table 6.1.: Results of the questionnaires completed by learner's using Moodle with the CbKST-based personalisation plug-ins (Experimental Group) and using Moodle without personalisation (Control Group).

Group	Strongly disagree	Disagree	Not sure	Agree	Strongly agree
Q1: The cycle of learning, assessment, and visualisation was good for my learning.					
Experimental	0%	16.7%	33.3%	33.3%	16.7%
Q2: The system supported me to become aware about my learning process?					
Experimental	0%	16.7%	0%	66.7%	16.7%
Q3: The system was limiting my learning.					
Experimental	16.7%	33.3%	33.3%	16.7%	0%
Control	0%	25%	50%	12.5%	12.5%
Q4: The system provided helpful guidance for my learning.					
Experimental	16.7%	0%	0%	50%	33.3%
Control	0%	37.5%	0%	62.5%	0%
Q5: This way of learning was stressful.					
Experimental	16.7%	33.3%	16.7%	33.3%	0%
Control	0%	50%	37.5%	12.5%	0%
Q6: I enjoyed the way of learning with that system.					
Experimental	0%	33.3%	16.7%	33.3%	16.7%
Control	12.5%	12.5%	50%	25%	0%
Q7: I was successful with the learning task.					
Experimental	0%	33.3%	16.7%	33.3%	16.7%
Control	12.5%	25%	12.5%	50%	0%
Q8: The information in the user interface was easy to understand.					
Experimental	16.7%	16.7%	16.7%	33.3%	16.7%
Control	12.5%	12.5%	12.5%	62.5%	0%
Q9: I would like to use a system like this in the future.					
Experimental	0%	0%	16.7%	66.7%	16.7%
Control	12.5%	12.5%	37.5%	25%	12.5%
Q10: I am happy with the quality of the content presentation.					
Experimental	0%	33.3%	16.7%	33.3%	16.7%
Control	0%	25%	25%	37.5%	12.5%

In fact, this is also visible in the rather high standard deviation values, which lie between 0.98 and 1.47 for the ten questions. In interviews, the participants with proper IT skills stated that the CbKST approach would be good and efficient. However, they found that there was room for improvement regarding the simplicity of the user interface. According to the log data of the experimental group, the participants, on average have performed 3.2 learning iterations, visited 9.4 learning resources, followed 82% of the recommended learning resources, and answered 9.2 assessment questions. As mentioned earlier, there was a discrepancy between participants with good and poor IT skills, which became also visible in the usage frequency. However, the fact that users followed 82% of the recommended learning resources indicates that the recommendation strategy was appropriate for the participants.

6.1.5. Conclusion

This Section investigated a recommendation strategy that builds on a CbKST-based learner model to personalize a formal learning environment. To this end, the learning management system Moodle was expanded with plug-ins developed to support and guide learners through a learning domain while promoting self-regulation. The main aim of introducing a CbKST-based recommendation strategy was to support learners within their self-regulated learning process in an efficient and time-saving manner. Based on an assessment of a learner's current competence state, the developed Recommendation Tool presents a set of learning resources that are selected according to an individual's abilities. In this way, the competence-based recommendation algorithm suggests learning resources to the user that are expected to be of medium difficulty, thereby preventing the learner from being overchallenged or underchallenged by the content. The learner may then select the content in a self-regulated manner. The learning progress is accompanied by associated assessments which capture a learner's progress. Based on that, visualizations of the learning progress and system interactions are provided to the learner for reflection.
To investigate the target group's (heat pump installer) perceived learning experience, an experimental study with fourteen users was conducted. An A/B testing approach was applied to compare a standard Moodle course with the extended, personalised Moodle environment. Although the relatively low number of study participants limits the significance of the experiment, the study results are quite encouraging.

These results contributed to RQ1: *Can a learning resource recommender based on a structural learner model (like the CbKST) improve the learning experience in a* *formal learning environment?* and led to the following findings: i) the personalized, CbKST-based learning approach was more appreciated in respect to most of the questioned aspects, and ii) the perceived usefulness of guidance (Q4) that was implemented by the recommendation service, and the learners' readiness to use the system again (Q9) scored notably higher in the CbKST-based learning environment. Based on these findings, RQ1 can be answered positively, as it demonstrates that the CbKST can be successfully applied to implement a resource recommendation strategy in a formal learning setting and moreover, improve the perceived learning experience of target users. However, further research with a larger subject group is needed to corroborate these results.

6.2. Process Oriented Learner Models: Using SUSTAIN to Improve Collaborative Filtering

Collaborative Filtering (CF) is one of the most successful resource recommendation strategies in the web (Bar et al., 2013). It depicts interactions in a user resource matrix, treating the learner as being just another item. This structural simplification can be regarded as abstraction from an individual learner's complexity. It also runs the risk of neglecting nonlinear, dynamic processes going on between different entities, such as a learner's intentional state (e.g., attentional focus, interpretations, decision-making) and resources (e.g., articles) consumed in the past.

SUSTAIN, a learning model built upon theories of human category learning, can differentiate between learners by means of attention and interpretation dynamics that are demonstrated towards observed aspects. This is further referred to as attentional and conceptual processes. Attentional processes describe the cognitive operation that decides, which environmental aspects a person attends to (focuses on) and therefore determines what a person learns, while conceptual processes refer to the development and incremental refinement of a learner's specific model of concepts and its interpretation.

This Section introduces and investigates a hybrid resource recommendation approach termed SUSTAIN+CF_U, to personalize and improve user-based Collaborative Filtering (CF_U). The recommendation strategy is built upon the application of the category learning model SUSTAIN, which is introduced in Section 4.1.2.

The approach has been evaluated by two means: First, it investigates recommender accuracy and determines whether resource recommendations become more accurate if a set of resources identified by CF is processed by SUSTAIN to simulate learner-specific attentional and conceptual processes (see Section 6.2.3).To that end, the unsupervised learning paradigm of SUSTAIN was adapted to fit recommender specific learning tasks. Then it was combined with user-based Collaborative Filtering (CF_U) to create the hybrid approach SUSTAIN+CF_U. The algorithm was compared to SUSTAIN alone, CF_U as well as other state-of-the-art approaches like resource-based CF (CF_R) and an effective Matrix Factorization variant (WRMF) (Y. Hu, Koren, and Volinsky, 2008). Furthermore, to gain insights into which aspects of the SUSTAIN algorithm contributes most to the improved performance, a parameter study in which the model's main parameters are simulated and observed was conducted (see Section 6.2.4).

The content of this Section has been partly published in Seitlinger, Kowald, et al. (2015) and Kopeinik, Kowald, Hasani-Mavriqi, et al. (2016).

6.2.1. Approach

SUSTAIN is a very flexible model that can be applied to a variation of category learning tasks (Love, Medin, and Gureckis, 2004). However, in line with the requirements of the selected learning task (i.e. the recommendation of web resources), the approach focuses on SUSTAIN's unsupervised learning process that implements a clustering mechanism with interconnected input, hidden and output units.

In this thesis, semantic topics that serve as resource description and input features are derived by means of the latent Dirichlet allocation (LDA) model (see Section 5.1.2). Besides, in this work, the algorithms' candidate resources are selected using CF_U.

The approach is split into two phases:

1. **Training:** For each learner, a slightly adapted version of the SUSTAIN model is trained on the learner's resource interaction history (i.e., resources a learner collected in the past).
2. **Test:** To predict resources a learner will engage with at a subsequent point in time, the learner model is applied to rank items of a preselected candidate set. This candidate set includes resources of the learning environment that a learner has not interacted with in the past. It is further confined based on a CF_U ranking.

During training and testing, SUSTAIN maps the input features (e.g., topics identified by Latent Dirichlet Allocation) of a resource to a set of dimensions at the input layer. The activation of each dimension is controlled by the attentional tuning that is learned in the course of the training phase and reflects the importance of the corresponding feature dimension for a specific learner. The hidden layer consists of a set of clusters each representing similar resources encountered in the past. Hence, one cluster corresponds to a learner-specific field of interest. In the test phase, the set and the structure of recruited clusters are treated as fixed measurements that no longer change. The classification decision (i.e., the decision to choose or not choose a given resource) is a function of the activation of the most activated (winning) cluster. Table 6.2 summarizes the notations that are used to describe the hybrid SUSTAIN+CF approach.

Symbol	Description
u	user
v	neighbor in the sense of CF
t	tag
r	resource
c	candidate resource
P	set of posts / bookmarks
U	set of users
$V_{u,r}$	neighbors of user u that bookmarked r
T	set of tags
R	set of resources
R_u	resources of user u
R_v	resources of neighbor v
S_u	similar resources of u based on topics
S_r	similar resources of resource r
C_u	resource candidate set of user u
Z	number of topics (i.e., n dimensions)
k	number of neighbors (CF)
k	number of Matrix Factorization factors
l	number of iterations
I	topic vector of a resource
I_{act}	activated topics of I (i.e., with value 1)
H_j	cluster j in a user's clusters
H_m	most activated (winning) cluster
H_j^{act}	activation value of cluster j
H_m^{act}	activation value of winning cluster m
μ_{ij}	distance to cluster j at dimension i
λ_i	attentional tuning (weight) of dimension i
r	attentional focus parameter
η	learning rate
τ	threshold for the creation of new clusters
$sim(u, v)$	similarity between users u and v
α	weighting parameter of SUSTAIN
$CF_U(u, r)$	Collaborative Filtering value for u and r
$RecRes(u)$	set of recommended resources for user u

Table 6.2.: Overview of notations.

6.2.2. Design: A Hybrid Resource Recommender Based on SUSTAIN

The SUSTAIN approach categorizes web resources by means of topic features. Thus, to describe the web resources' content, 500 LDA topics are derived from tags assigned to resources of the datasets (Griffiths, Steyvers, Tenenbaum, et al., 2007). Section 5.1.2 describes the LDA procedure in more detail. The extracted topics of web resources represent the n input features of the SUSTAIN model. In a second preprocessing step, each user's resources are split into a training set and a test set. This is commonly done in recommender system research, when evaluating recommendation strategies on offline data (see Section 5.1.1).

Then, on the basis of the resources a learner has interacted with in the past (i.e., the training set of a learner), each learner's personal attentional tunings and cluster representations are created in the training phase, and further constitute the learner model. After the training phase, the learner model based prediction algorithm can be applied or evaluated in the testing phase.

To better fit the learning task's specific needs, SUSTAIN's unsupervised clustering approach was slightly adapted. The adaptations of the model impact specifically the training and testing phase. More precisely, due to the comparably high number of input dimensions [5] used in this approach, the model was adjusted by limiting the learning focus to the topics activated by the current learning resource (further referred to as I_{act}). The adaptation of the model led to improved performance results, in comparison to results reported in prior work (Seitlinger, Kowald, et al., 2015).

Training

Following an unsupervised learning procedure, the approach starts simple, with one cluster and expands the number of clusters if necessary. Please note that all SUSTAIN-specific parameter settings are adopted from Love, Medin, and Gureckis (2004) (see Table 6.3).

[5] Although the number of input dimensions is variable, the usage of 500 topics to describe a resource has shown to be most effective.

Function	Symbol	Value
Attentional focus	r	9.998
Learning rate	η	.096
Threshold	τ	.5

Table 6.3.: SUSTAIN's best fitting parameters for unsupervised learning as suggested in Love, Medin, and Gureckis (2004).

For each resource in the training set of a user u, first, the distance μ_{ij} to cluster j at dimension i is calculated as described in equation (6.1):

$$\mu_{ij} = \left| I^{pos_i} - H_j^{pos_i} \right| \tag{6.1}$$

where I is the n-dimensional input vector, which represents the topics of this resource, and vector H_j is cluster j's position in the n-dimensional feature space, which holds a value for each topic and is initially set to $\vec{0}$. In the suggested setup, input and cluster vectors represent 500 topics of which only a few are activated by each resource. Adjusting to this setting, the distance μ_{ij} is set to 1 (maximal distance) for every topic i that is not activated in the input vector ($I^{pos_i} = 0$) and therefore $i \notin I_{act}$ for $I_{act} = \{i \in I \wedge i = 1\}$. In the next step, the approach considers only activated topics $i \in I_{act}$ to calculate the activation value H_j^{act} of the j^{th} cluster by equation (6.2):

$$H_j^{act} = \frac{\sum_{i \in I_{act}} (\lambda_i)^r e^{-\lambda_i \mu_{ij}}}{\sum_{i \in I_{act}} (\lambda_i)^r} \tag{6.2}$$

where λ_i represents the attentional tuning (weight) of dimension i and acts as a multiplier on i in calculating the activation. Initially, vector λ is set to $\vec{1}$ and evolves during the training phase according to equation (6.3) calculated at the end of every training iteration (i.e., after including a resource). r, which is set to 9.998, is an attentional focus parameter that accentuates the effect of λ_i: if $r = 0$, all dimensions are weighted equally.

If the activation value H_m^{act} of the most activated (i.e., winning) cluster is below a given threshold $\tau = .5$, a new cluster is created, representing the topics of the currently processed resource. At the end of an iteration, the tunings of vector λ are updated given by equation (6.3):

$$\Delta\lambda_i = \eta e^{-\lambda_i \mu_{im}} (1 - \lambda_i \mu_{im}) \tag{6.3}$$

where j indexes the winning cluster and the learning rate η is set to .096. In a final step, the position vector of the winning cluster, which holds a value for each of the n topics, is recalculated as described by equation (6.4):

$$\Delta H_m^{pos_i} = \eta(I^{pos_i} - H_m^{pos_i}) \tag{6.4}$$

The training phase is completed when steps (6.1) to (6.4) are subsequently processed for every resource in a user's training set. For each user, this results in a particular vector of attentional tunings λ and a set of j cluster vectors H_j.
In Algorithm 1, the training procedure of the approach is illustrated more formally.

Algorithm 1 Training procedure per user

Initialize a set of cluster $H = \emptyset$
Initialize a vector λ with $\lambda_i = 1$
for every resource topic vector I **do**
    for every cluster $H_j \in H$ **do**
        Calculate μ_j
        Calculate H_j^{act}
    end for
    Identify H_m with max H_m^{act}
    if $H_m^{act} <= \tau$ **then**
        $H_m \leftarrow I$
        $H \leftarrow H \cup \{H_m\}$
    end if
    $\lambda \leftarrow \lambda + \Delta\lambda$
    $H_m \leftarrow H_m + \Delta H_m$
end for
return λ
return H

Testing

In the testing or recommendation phase, learner models are considered static. The learner model is applied to calculate cluster activation values that resources stimulate. The activation value of the highest activated cluster H_m^{act} indicates the relevance of a resource to a learner. For performance reasons potentially relevant resources for a target user u are pre-selected in a candidate set C_u, which consists of the top n resources identified by CF_U. In this work, n is set to 100.

For each candidate c in C_u, H_m^{act} is calculated by equations (6.1) and (6.2) and further combined with relevance scores extracted from CF_u. Further, the values resulting from SUSTAIN and CF_u, are normalized such that $\sum_{c \in C_u} H_m^{act}(c) = 1$ and $\sum_{c \in C_u} CF_U(u, c) = 1$ holds. This leads to the normalized values $\overline{H_m^{act}(c)}$ and $\overline{CF_U(u, c)}$ that are finally put together as shown in equation (6.5) in order to determine the set of k recommended resources $RecRes(u)$ for user u:

$$RecRes(u) = \underset{c \in C_u}{\arg\overset{k}{\max}}(\underbrace{\alpha \underbrace{\overline{H_m^{act}(c)}}_{SUSTAIN} + (1-\alpha)\overline{CF_U(u, c)}}_{SUSTAIN+CF_U}) \tag{6.5}$$

where α can be used to weigh the two components of the hybrid approach. In this work, α is set to .5 in order to equally weight SUSTAIN and CF_U.

Technical Insights

The here described SUSTAIN+CF_U approach has been implemented as part of *TagRec*, an open source, JAVA-based recommender benchmarking framework (Kowald, Kopeinik, and Lex, 2017), which is freely available via GitHub[6]. The framework has been developed and used for the evaluation and development of algorithms in a scientific context. Thus, it is well-suited for offline data studies on tag-based recommendation algorithms. Amongst other algorithms, the implementation of SUSTAIN enables the adjustment of model specific parameters as for instance described in Table 6.3.

6.2.3. Model Validation Based on Recommendation Accuracy

This section describes the methodology that was selected to evaluate and investigate SUSTAIN based on recommender performance metrics. Information regarding datasets, method, metrics and baseline algorithms used in the recommender evaluation study is presented.

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

Datasets

The datasets used for this study, i.e., BibSonomy, CiteULike and Delicious are described in Section 5.1.4. To test the approach in three different settings that vary in their dataset sizes, datasets from the social bookmarking and publication sharing system BibSonomy[7], the citation sharing system CiteULike[8] and the social bookmarking system Delicious[9] were used. For the CiteULike dataset 20% of the user profiles were randomly selected (Gemmell et al., 2009) in order to reduce computational effort. The other datasets were processed in full size. Furthermore, all posts assigned to unique resources, i.e., resources that have only been bookmarked once (see Parra-Santander and Brusilovsky (2010)) were excluded. Statistics of the resulting dataset samples (i.e., after the exclusion of posts assigned to unique resources) as well as training and test sets that are relevant to this experiment are illustrated in Table 6.4 and Figure 6.11.

Dataset	Type	$\|P\|$	$\|U\|$	$\|R\|$	$\|T\|$	$\|P\|/\|U\|$
Bibsonomy	Sample	82,539	2,437	28,000	30,919	34
	Training	66,872	2,437	27,157	27171	27
	Test	15,667	839	11,762	12,034	19
CiteULike	Sample	105,333	7,182	42,320	46,060	15
	Training	86,698	7,182	40,005	41,119	12
	Test	18,635	2,466	14,272	16,332	8
Delicious	Sample	59,651	1,819	24,075	23,984	33
	Training	48,440	1,819	23,411	22,095	27
	Test	11,211	1,561	8,984	10,379	7

Table 6.4.: Properties of the the used dataset samples (including training and test set statistics) for BibSonomy, CiteULike and Delicious. Here, $|P|$ is the number of posts, $|U|$ is the number of users, $|R|$ is the number of resources and $|T|$ is the number of tags.

[7] http://www.kde.cs.uni-kassel.de/bibsonomy/dumps/

[8] http://www.citeulike.org/faq/data.adp

[9] http://files.grouplens.org/datasets/hetrec2011/hetrec2011-delicious-2k.zip

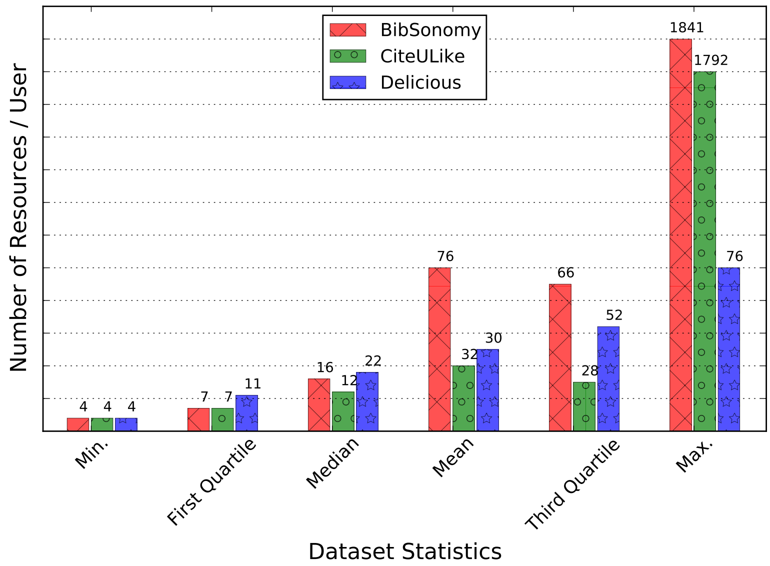


Figure 6.11.: Resource statistics of the training datasets for BibSonomy, CiteULike and Delicious illustrating the number of resources users have engaged with.

Baseline Algorithms

A set of well-known resource recommender baseline algorithms are taken to determine the performance of this novel approach. This encompasses on the one hand algorithms that are similar to the proposed SUSTAIN approach in terms of their processing steps (CF_U and CB_T), but also current state-of-the-art methods for personalized resource recommendations (CF_R and WRMF) along with a simple non-personalised recommendation strategy (MP). Further details on the algorithms and their implementation are provided in Section 5.3.1

Technical Preliminaries

Given the characteristics of the selected datasets, i.e. that resources do not have topics assigned, the approach requires two steps of data preprocessing. First, semantic topics have to be extracted to describe resources. For this purpose LDA was applied with the number of latent topics set to 500 (see also Kintsch and Mangalath (2011)) and a limitation of topics for a resource that show a minimum probability value of .01. The method is described in Section 5.1.2. Second, candidate resources have to be identified. Candidate resources describe the learner-specific set of Web resources that the algorithm considers recommending to a user. This helps reduce the computation time when predicting resources, as the algorithm is only applied on a subset of available items. The identification process that follows a collaborative filtering approach is described in Section 5.1.3.

Methodology

In order to evaluate the approach on the selected offline datasets, each user's chronologically ordered activities are split into a training- and a test-set, where the test-set includes most recent activities (as described in Section 5.1.1).

To compare the performance of the SUSTAIN approach with the selected baseline recommendation strategies (see Section 6.2.3), the top 20 recommended resources that are suggested for each user by each algorithm are contrasted with relevant resources in the test set. To that end, a variety of well-known evaluation metrics Parra and Sahebi (2013) and Herlocker et al. (2004) in recommender systems research, namely, Normalized Discounted Cumulative Gain (nDCG@20), Mean Average Precision (MAP@20), Recall (R@20) and Precision (P@20) are calculated. Moreover, by means of Precision/Recall plots, the performance of the algorithms for different numbers of recommended resources ($k = 1 - 20$) is outlined.

6.2.4. Parameter Investigation to Understand the Dynamics of SUSTAIN

This Section describes the setup and rationale of a parameter investigation that aims to achieve a better understanding of aspects of the SUSTAIN algorithm that contribute to the improved performance. In an initial study that has been reported in Seitlinger, Kowald, et al. (2015) and in the comparative studies that are presented in Section 6.2.5, the best fitting parameters for unsupervised learning as suggested in Love, Medin, and Gureckis (2004) are used. This parameter set results from extensive parameter studies, applying a genetic algorithm to fine tune SUSTAIN for a variety of learning data and learning problems. The paper concluded that SUSTAIN does not show great sensitivity to single parameter values but rather succeeds due to its cognitive plausibility.

However, this learning task differs from the presented studies in multiple aspects, for instance in the amount of training data, in the application domain and most significantly in the format of the input stimuli. In Love, Medin, and Gureckis (2004), the input stimuli are characterized by multiple dimensions of input units. For instance, a dimension (e.g., color) with 3 input units (e.g., green, yellow, blue) could have an input vector of [0,0,1]. When recommending resources, the input are topics describing those web-items. Thus, in this experiment, the input stimuli consist of 500 dimensions (i.e., LDA topics) of binary input units. Another aspect to consider is that data, which is typically available in non-commercial learning environments, and equally, the social bookmarking datasets used in this study, are sparse and premature. With this in mind, a short parameter study was conducted to better understand the underlying dynamics of the adapted approach and to investigate possible inconsistencies. The priority was to look into SUSTAIN's parameters r, η in a first step, but secondly, also to find the best fitting α value to optimally weight the impact of CF_u.

The results in Section 6.2.5 were generated using the default SUSTAIN parameters stated in Love, Medin, and Gureckis (2004), to avoid tuning our approach and thus favouring it over the baseline algorithms. Additionally, the parameter study was performed on separate holdout sets extracted from the training data (using the same method as described in Section 5.1.1) in order to prevent a biased study conducted on the test data.

SUSTAIN. In a first step, plausible ranges for r and η and sequential steps within these ranges are determined. Additionally, the simulation includes the originally suggested values as presented in Table 6.3.
For r, which strengthens the impact of input dimensions by potentiating λ_i (see equation (6.2)), the investigation starts with $r = 1$ as a lower bound. This leads to a simulation with plain λ values. From there, the value is increased linearly with $r = r + 2$ for $r <= 21$. As λ shows rather small values, with a great percentage varying from 1.0 to 1.3, a relatively high value of r seems to be reasonable.
For the learning rate η, the simulation span is set such that $\eta_{min} > \frac{1}{N_{\max}}$ where $N_{\max}$ is the maximal amount of training resources per user. Thus, the learning rate η is set to 7.5 E-4 on the lower bound, while 1 was chosen as an upper bound.

In between those bounds, three learning rates per decimal power were tested. As the median values for resources per user in the experiment's training sets are 12, 16 and 22 (see Figure 6.11), the optimal learning rate is expected to be fairly high.

As described in the original study setup, the parameter study was simplified by treating $\tau = 0.5$ as a fixed value. τ is the threshold responsible for whether a new cluster is formed or not and may range from 0 to 1.

When interpreting the first set of plots, additional questions appeared, such as, to what extent the training datasets and the topic distribution of their users may influence the optimal amount of clusters. This was investigated by inspecting the distribution of clusters and resources per user and dataset that were calculated with the recommended parameter setting that is outlined in Table 6.3. Finally, the performance development of SUSTAIN with different learning rates was investigated while varying τ within its range of 0 and 1, monitoring steps of .1. With respect to insights gained in the first parameter setting, r was set to a fixed value of 9, and the learning rate to a range from .01 to 1.

Weighting CF_U. α is the only parameter that is not part of SUSTAIN, but inversely weights the impact of the SUSTAIN and CF_u components (see equation 6.5). α is set to values between .1 and .9.

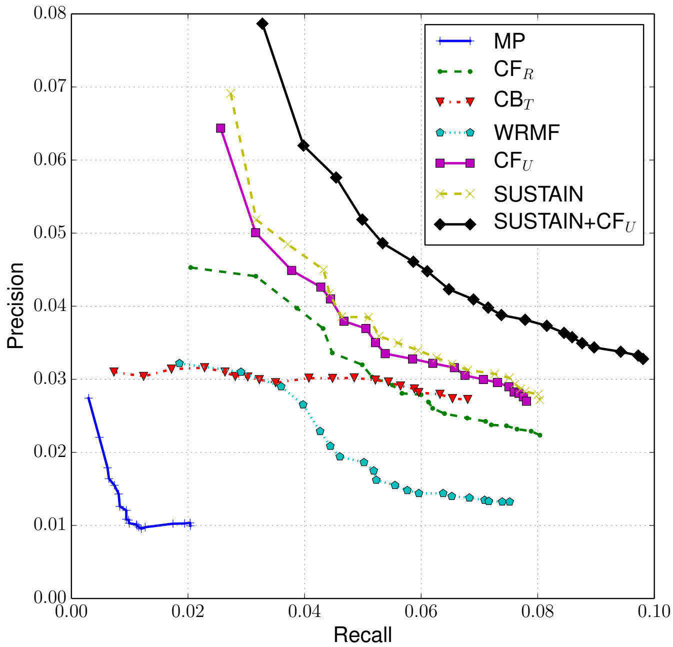
6.2.5. Results and Discussion

This section presents the results of the two conducted experiments.

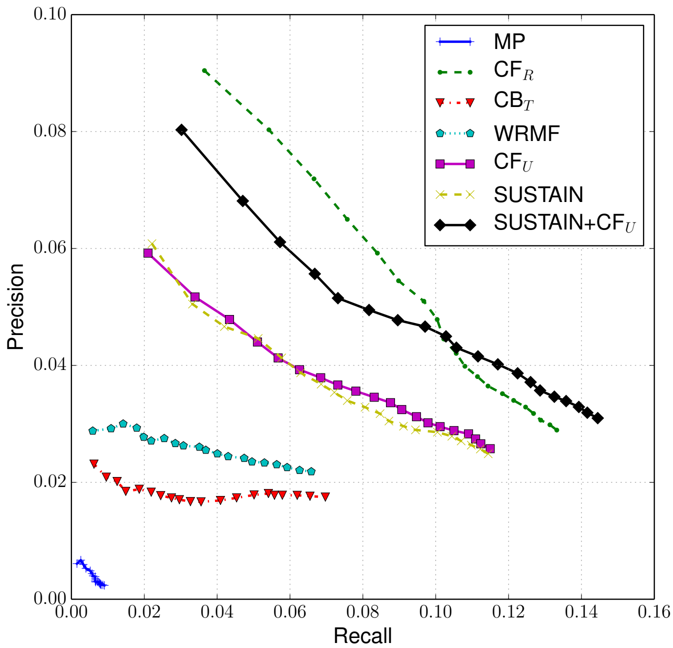
Model Validation Based on Recommendation Accuracy

In order to validate the accuracy of the SUSTAIN+CF_U recommendation approach, a comparison study with a selected set of state-of-the-art resource recommender algorithms was conducted. Results are presented by two means:

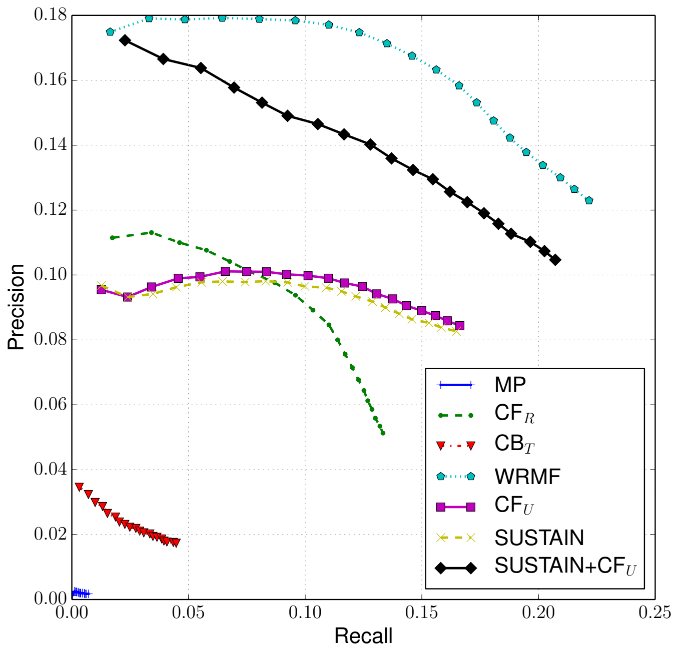
1. Recall/Precision Plots: Figure 6.12 reveals the evolution of accuracy values with a growing number of recommendations (i.e., one to 20). Note that recall (per definition) increases with the number of recommended items.
2. Accuracy Metrics Table: Table 6.5 presents the results of four recommender metrics as would be achieved with a number of 20 recommended items.



(a) BibSonomy



(b) CiteULike



(c) Delicious

Figure 6.12.: Precision/Recall plots for BibSonomy, CiteULike and Delicious showing the recommender accuracy of the proposed approach SUSTAIN+CF$_U$ in comparison to the baseline methods for k = 1 - 20 recommended resources. The results indicate that SUSTAIN+CF$_U$ provides higher Precision and Recall estimates than CF$_U$ (RQ2) and SUSTAIN for each k and in all three datasets. In the case of BibSonomy, SUSTAIN+CF$_U$ even outperforms all baseline methods, including WRMF.

One obvious insight gained from the plots is that the simplest baseline algorithm, i.e., the non-personalized MP approach, achieves very low estimates of accuracy for all datasets. In comparison, the other baseline algorithms reach larger estimates and therefore seem to be successful in explaining a substantial amount of variance in user behaviour.

Interestingly, the performance of the algorithms varies greatly across the three datasets BibSonomy, CiteULike and Delicious. Regarding nDCG@20, a different algorithm performs best in each of the three datasets. For instance, in the case of CiteULike, the best results are achieved with CF$_R$. This can be explained by studying the average topic similarity per user, which is defined as the average pairwise cosine similarity between the topic vectors of all resources a user has bookmarked. This is averaged over all users. In CiteULike the topic similarity (18.9%) is much higher than in BibSonomy (7.7%) and Delicious (4.5%), indicating a more thematically consistent resource search behaviour. We can assume that the higher consistency positively impacts predictions that are based on resources collected in the past, such as CF_R-based predictions.

In regard to Delicious, the users in the dataset are chosen using a mutual-fan crawling strategy (see Iván Cantador, Brusilovsky, and Kuflik (2011)) and thus, are not independent from each other. This is conducive to methods that capture relations between users with common resources by means of high-dimensional arrays, such as WRMF. However, compared to the other algorithms, especially to CF_R and WRMF, SUSTAIN+CF_U demonstrates relatively robust estimates (especially in terms of Precision and Recall) as SUSTAIN+CF_U provides fairly good results in all three datasets. However, particularly good results are achieved on BibSonomy, where it outperforms all baseline algorithms.

Furthermore, the evaluation results indicate that the SUSTAIN+CF_U approach outperforms CF_U and the unpaired SUSTAIN approach in all settings. For instance, in the Precision/Recall plots in Figure 6.12, it is illustrated that there is no overlap between corresponding curves, with SUSTAIN+CF_U always reaching higher values than SUSTAIN and CF_U separately.

Moreover, results of the ranking-dependent metric nDCG@20 in Table 6.5 particularly show a remarkably better value for SUSTAIN+CF_U than CF_U, demonstrating that the proposed approach, through its improved personalization, can be used to successfully re-rank candidate resources identified by CF_U. It can be assumed that this effect occurs as user-based CF does not rank the resources of a neighbour. This possibly leads to a list of recommendations that contains only the resources of a user's nearest neighbour but no ranking within this list. With the hybrid approach, this issue is tackled with SUSTAIN providing activation values for each resource. Consequently, RQ2 that investigates the potential of SUSTAIN to improve collaborative filtering for resource recommendations, can be answered positively.

However, to better understand the algorithms dynamics, an additional parameter study was conducted and is presented in the next Section.

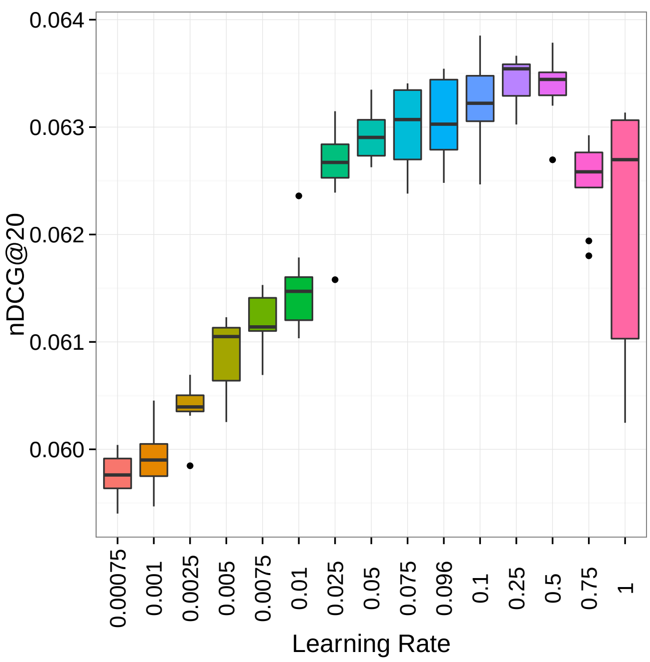
Dataset	Metric	MP	CF_R	CB_T	WRMF	CF_U	SUSTAIN	SUSTAIN+CF_U
BibSonomy	nDCG@20	.0142	.0569	.0401	.0491	.0594	.0628	**.0739**
	MAP@20	.0057	.0425	.0211	.0357	.0429	.0436	**.0543**
	R@20	.0204	.0803	.0679	.0751	.0780	.0902	**.0981**
	P@20	.0099	.0223	.0272	.0132	.0269	.0295	**.0328**
CiteULike	nDCG@20	.0064	**.1006**	.0376	.0411	.0753	.0828	.0977
	MAP@20	.0031	**.0699**	.0170	.0210	.0468	.0503	.0634
	R@20	.0090	.1332	.0697	.0658	.1149	.1344	**.1445**
	P@20	.0023	.0289	.0174	.0218	.0257	.0279	**.0310**
Delicious	nDCG@20	.0038	.1148	.0335	**.1951**	.13	.131	.1799
	MAP@20	.0011	.0907	.0134	**.1576**	.0743	.0936	.1275
	R@20	.0071	.1333	.0447	**.2216**	.1599	.1649	.2072
	P@20	.0017	.0512	.0173	**.1229**	.0785	.0826	.1047

Table 6.5.: nDCG@20, MAP@20, R@20 and P@20 estimates for BibSonomy, CiteULike and Delicious. The results indicate that the proposed SUSTAIN+CF_U approach outperforms CF_U and SUSTAIN in all settings. Furthermore, SUSTAIN+CF_U is able to compete with the computationally more expensive WRMF approach. *Note*: highest accuracy values per dataset over all algorithms are highlighted in bold.

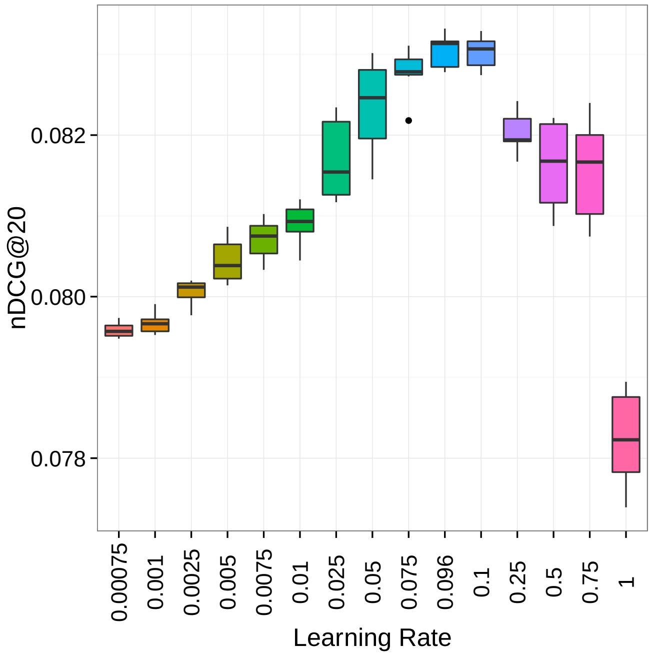
Parameter Investigation to Understand the Dynamics of SUSTAIN

This Section aims to identify the core aspects of the SUSTAIN model that show the greatest effects on the recommendation strategy's performance. It also investigates the impact of user traces and particularities of single datasets as well as the optimal weighting of CF_U and SUSTAIN.

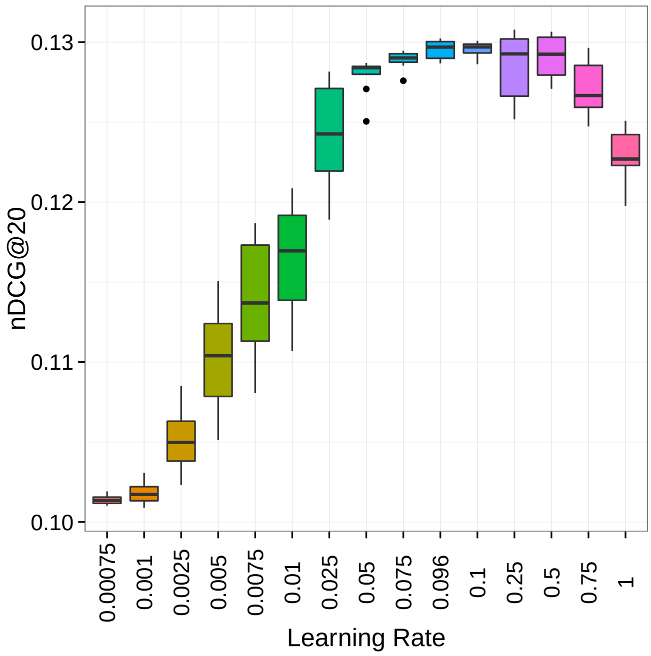
SUSTAIN. In Figure 6.13, results of the first simulation are illustrated. In this setup, $\tau = .5$ is treated as a fixed variable, similar to the original parameter study (see Love, Medin, and Gureckis (2004)), and solely varied learning rate η and attentional focus parameter r within a parameter range, as explained in 6.2.4. The plots show SUSTAIN's performance on the y-axis given as nDCG@20 values and the learning rates on the x-axis. The shape of the box plot indicates the distribution of the performance values caused by a set of different r's, which means, the higher the box plot, the greater the influence of r. Even though some variation can be observed, for the best performing η, the influence of r seems to be marginal in this setting.



(a) BibSonomy



(b) CiteULike



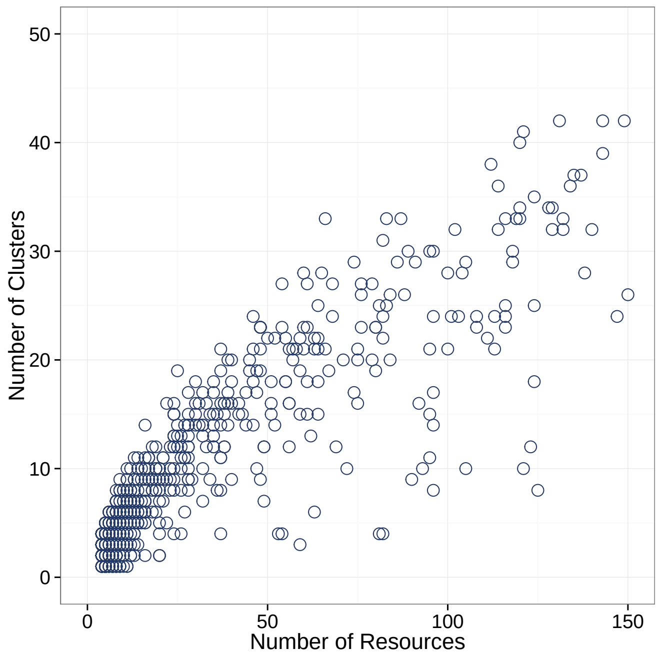
(c) Delicious

Figure 6.13.: Recommendation effectiveness influenced by learning rate and attentional focus parameter.

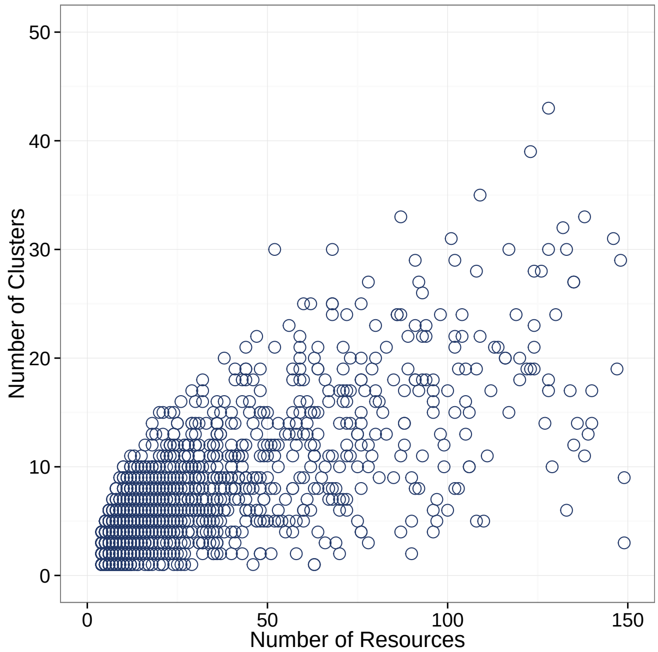
In this use case of SUSTAIN, the learning rate tends to be the most important factor to consider. Two scenarios may occur:

i The learning rate is too small: a user's behaviour cannot be tracked fast enough.
ii The learning rate is too high: the algorithm forgets previous resources too quickly.

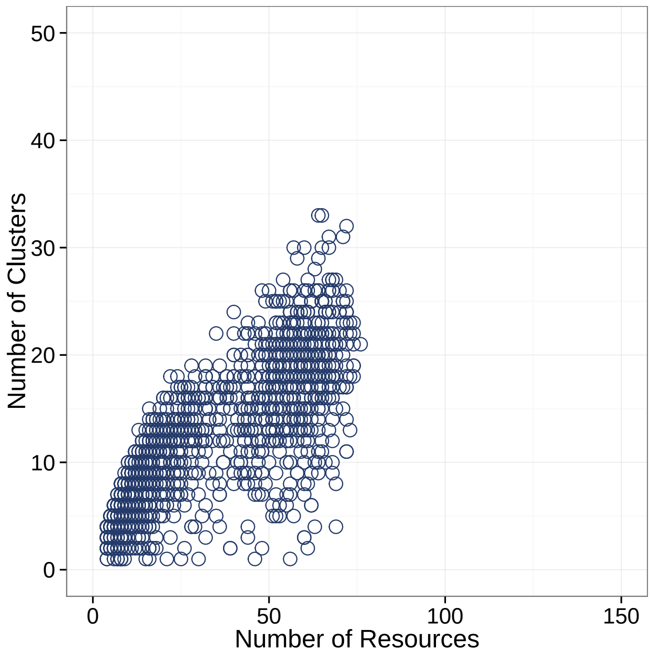
The first scenario is likely to apply to users with few resources, whereas the second scenario is potentially problematic for users with many resources. As illustrated in Figure 6.11, the considered training datasets show a large variation in the distribution of training resources per user. This can be observed within (i.e., Min. vs. Max. Resources/User) and between datasets. However, the common trend shows that about 50 percent of users have less than 25 resources available for training the algorithm. In line with these observations, SUSTAIN's performance peaks at an intermediate value which is around $\eta = .1$. In this use case, when aiming to achieve optimal predictions, it particularly underlines the added value of taking into account the browsing history of a user and not just the most recent item.



(a) BibSonomy



(b) CiteULike



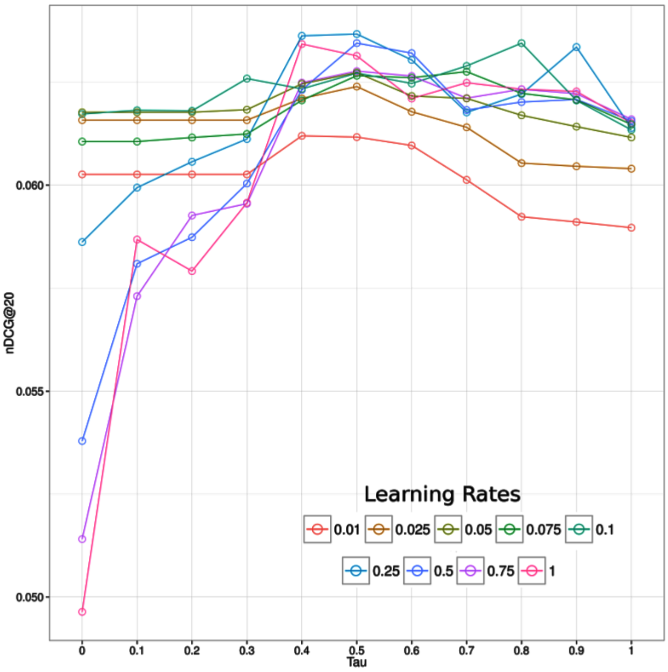
(c) Delicious

Figure 6.14.: Snapshot of the distribution of the clusters and resources appearing with parameters recommended in the literature. Please note that the range of the plots is restricted in order to improve readability. BibSonomy and CiteULike have both about 100 users with more than 150 resources, which are not depicted in this plot.

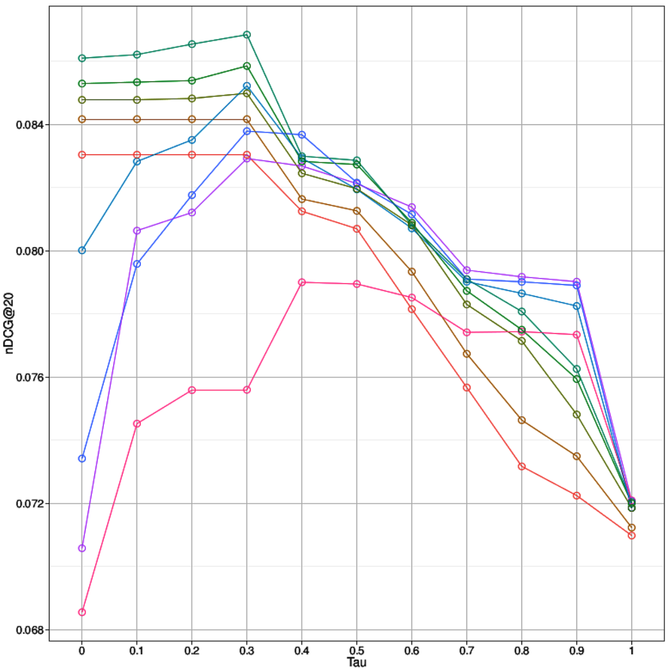
Among the three datasets, the learning rate has the greatest impact on Delicious (note that the ranges of nDCG@20 differ between plots). An explanation of this behaviour can be derived from Figure 6.14, which presents a snapshot of the cluster resource distribution per user and dataset. In the case of Delicious, the overall trend shows that a new cluster is created for each second or third resource. Since in this approach only the cluster with the highest activation learns, the strong influence of the learning rate, or in other words, the need for faster learning per cluster, seems reasonable.

Given that a new cluster is created whenever a new resource is added that cannot be integrated into any of the existing clusters due to a lack of similarities, the cluster distribution also presents the level of topic overlap among the resources of a typical user. For instance, when calculating basic statistics for the resource-to-cluster ratio of Delicious, results show that the average value is 2.8 resources per cluster in comparison to 4.2 resources per cluster for CiteULike. This indicates a large topic overlap between resources of users in CiteULike. In other words, CiteULike users tend to engage in less topics than users of other datasets. Furthermore, a decreasing trend of the resource-to-cluster ratio can be observed, as the number of resources grows. Also, the plot for CiteULike highlights the rather weak relationship between clusters and resources, which signifies a great variety among users.

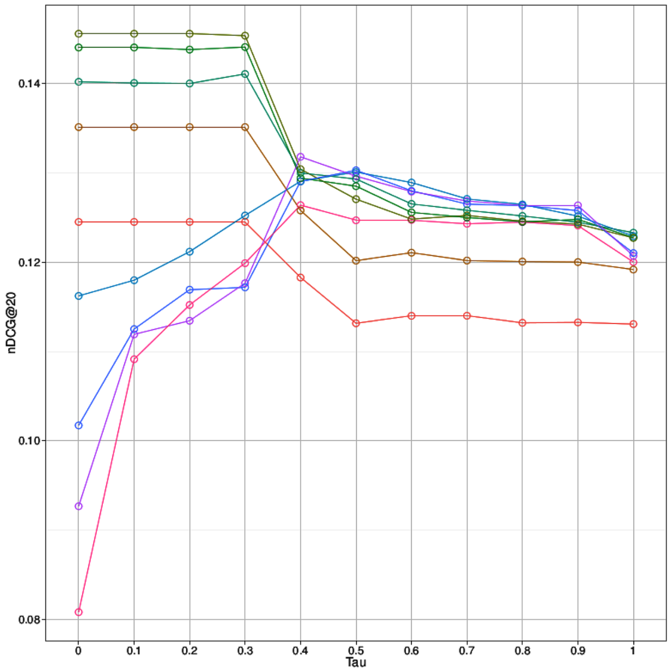
These results lead to the next simulation, which investigates how the number of clusters impacts the performance, and whether a dynamic clustering approach is even necessary for this specific task. In particular, the experiment considers if a different τ can lead to a better performance with the training sets. Thus, the second simulation, considers SUSTAIN's performance development when varying τ and η. This time $r = 9$ was treated as a fixed variable, due to the marginal difference it caused in the first study setup. The findings are illustrated as line charts (see Figure 6.15).



(a) BibSonomy



(b) CiteULike



(c) Delicious

Figure 6.15.: Recommendation effectiveness influenced by learning rate and the number of clusters. The number of clusters is represented on the x-axis. Different colours determine different learning rates.

Regarding the optimal number of clusters, we can see that the three datasets vary greatly in their behaviour. Delicious performs best with only one cluster (i.e., $\tau = 0$), CiteULike and BibSonomy show better results with $\tau = .3$ and $\tau = .5$, respectively.

Delicious is the dataset most sensitive to τ (please note that the plots differ in their ranges of nDCG@20). Again, this might be due to the high variation of topics, which leads to overfitting when too many clusters are formed. BibSonomy exhibits a larger topic overlap per user than Delicious. At the same time, in the case of BibSonomy, there is a much larger amount of training data per user than is the case with Delicious and CiteULike. Figure 6.11 for instance shows that 25 percent of users have between 66 and 1841 resources available for training. CiteULike differs due to its small amount of training data per user. Note the comparably low values for median and third quartile. This results in an optimal number of clusters between one and seven with the mean = 1.05. Thus, results clearly suggest that the optimal number of clusters varies with the properties of the training data. This value relates to the available number of training samples and the topic density.

Weighting CF_U. A simulation varying α from 0 to 1 to find the best fit for the weighting of CF_U to SUSTAIN (see 6.5) was performed. Results identified $\alpha = .65$ as the best fitting value for all datasets. Moreover, all values in the range of .3 to .8 perform close to optimal.

6.2.6. Conclusion

This Section investigated the suitability of a model of human category learning, SUSTAIN (Love, Medin, and Gureckis, 2004), to mimic non-linear user-resource dynamics (i.e., attentional focus and interpretation dynamics) and apply it to recommender web-resources. Offline studies on three social bookmarking datasets (BibSonomy, CiteULike and Delicious) demonstrated the potential of the approach to personalize and improve user-based CF predictions. This improvement can be attributed to the cognitive plausibility of SUSTAIN. The dynamically created user model allows for a more flexible and thorough representation of a user's decision-making on a given set of resources: Reconstructing the user history in the form of an iteratively trained model with history-specific patterns of attentional tunings and clusters does more justice to a user's individuality than a CF-based representation of user-resource relations. Deepening these investigations, it becomes evident that both aspects, i.e., memorization of a user's history as well as clustering, contribute to the algorithm's performance. In more detail, a parameter study revealed that restricting cluster growth by adapting the model's parameter τ can prevent overfitting in sparse data environments. Moreover, study results indicate that the hybrid SUSTAIN+CF_U model is more robust in terms of accuracy estimates than the computationally more expensive Matrix Factorization-based approach WRMF.

These results contributed to RQ2: *Can a process oriented learner model based on SUSTAIN be applied to improve an existing resource recommendation strategy such as collaborative filtering?* and led to the following findings: i) SUSTAIN can be used to capture non-linear user-resource dynamics and in this sense, to depict an individual user's learning process, and ii) the hybrid approach SUSTAIN+CF_U achieves better performance results than SUSTAIN, CF_U and

CF_R and can furthermore compete with the computationally expensive WRMF method. Based on these findings, RQ2 can be answered positively. It shows that a SUSTAIN-based recommendation strategy can improve CF_U in informal learning settings such as provided by social bookmarking systems. However, the approach depends on the availability of learning resource meta-data which, in datasets used here, is given by tags, but may suffer sparsity in other TEL environments (Niemann, 2015).

To better understand the potential of cognitive-inspired recommendation strategies for the TEL domain, the next Chapter presents two studies: First, an offline study comparing six different resource recommendation algorithms on a variety of TEL datasets. It investigates state-of-the-art methods, and further explores the here presented SUSTAIN approach within different learning contexts. One factor that hampers the success of content-based resource recommendation algorithms is the sparsity of learning resource meta-data. A number of cognitive-inspired tag recommendation algorithms are explored to support the creation of user-generated meta-data in TEL settings. Second, a real-life study deepens this investigation, testing the potential of two cognitive-inspired tag recommendation approaches in an online IBL setting.

7. Tag and Resource Recommendations Based on Cognitive Learner Models: An Alternative for TEL Settings (RQ3)

In this Chapter, two studies are described that were designed to investigate the performance of cognitive-inspired tag and resource recommendation strategies in offline and online TEL settings. The comparison with statistically based approaches assesses their suitability for learning recommendations and therewith address RQ3. The offline study is presented in Section 7.1. Here, the recommender accuracy of six recommendation strategies on six TEL datasets originating from different domains is reported. Section 7.2 describes the comparative implementation of tag recommendation mechanisms based on BLL and MINERVA2 in an online TEL setting, i.e., an inquiry-based learning environment.

7.1. A Data-driven Study to Compare Recommender Algorithms in TEL

This section describes an evaluation study that investigates the performance of six recommendation algorithms and variations thereof on implicit usage data from six TEL datasets originating from different application areas. These areas include social bookmarking systems (BibSonomy, CiteULike), Massive Open Online Course (MOOC)s (KDD15), open social learning (MACE, TravelWell) and workplace learning (Aposdle). While existing research investigates the application of implicit usage data-based algorithms (e.g., Verbert, Drachsler, et al. (2011), Fazeli et al. (2014), and Niemann and Wolpers (2013)) on selected datasets, a more extensive comparative study directly opposing state-of-the-art recommendation algorithms had been missing. The study hypothesizes that recommendation algorithms show different performance results depending on learning context and dataset properties, as also suggested in Manouselis, Vuorikari, and Van Assche (2010) and Verbert, Drachsler, et al. (2011). The experiment is divided in two parts distinguished by their application cases:

1. **The recommendation of learning resources:** This study investigates how accurately state-of-the-art resource recommendation algorithms that use only implicit usage data, perform on different TEL datasets. To this end, six datasets from different TEL domains such as social bookmarking, social learning environments, Massive Open Online Courses (MOOCs) and workplace learning could be obtained to evaluate accuracy and ranking of six state-of-the-art recommendation algorithms.
2. **The recommendation of tags:** This evaluation focuses on the three tag recommendation algorithms MP, CF_U and BLL_{AC}, which were implemented in six variations. These variations result from differences in usage data and hybrid combinations of the algorithms. Because not all six datasets have tags (see Table 5.1), the experiment is restricted to BibSonomy, CiteULike, TravelWell and MACE.

The content of this Section has been published in Kopeinik, Kowald, and Lex (2016).

7.1.1. Methodology

The evaluation of the algorithms follows an evaluation protocol splitting the data into training and test sets as commonly done in recommender system research (Kowald and Lex, 2015; Seitlinger, Kowald, et al., 2015). Section 5.1.1 describes the procedure in detail.

Algorithms

For the purpose of this study, six tag and resource recommendation strategies were selected. Three of them incorporate well-established, computationally inexpensive algorithms, namely MP, CF and CB_T. U. UCBSim has been proposed in the context of TEL before. These are considered baseline algorithms and are described in Section 5.3.1. Furthermore, an analysis of the computational complexity of the algorithms is presented in Trattner et al. (2016b).

The remaining two approaches have been proposed and discussed in the context of this thesis (see Chapter 4) and comprise:

- SUSTAIN (Seitlinger, Kowald, et al., 2015; Kopeinik, Kowald, Hasani-Mavriqi, et al., 2016)
- Base Level Learning Equation (BLL_{AC})(Kowald, Kopeinik, Seitlinger, et al., 2015)

SUSTAIN. SUSTAIN has been thoroughly described and investigated in Section 6.2. Here, it is implemented as SUSTAIN and additionally, as hybrid approach SUSTAIN+CF_U, which is a linear normalized combination of SUSTAIN and CF_U.

Base Level Learning Equation. For BLL_{AC} that is presented in Section 4.2.1, the most relevant tags are selected according to the highest activation values. BLL_{AC}+MP_R denotes a linear combination of the BLL_{AC} approach with MP_R.

All algorithms of this study as well as the evaluation methods are implemented in the *TagRec* recommender benchmarking framework (Kowald, Kopeinik, and Lex, 2017).

Datasets

Datasets used in this study are described in Section 5.1.4. The investigated TEL datasets originate from different application areas such as social bookmarking systems (BibSonomy, CiteULike), MOOCs (KDD15), open social learning (MACE, TravelWell) and workplace learning (Aposdle) and thus, vary in their properties. To complete an evaluation as accurately as possible to the real application, pruning on the dataset was avoided, i.e. the datasets were not pre-filtered to exclude users or resources with little interaction data.

Metrics

For the performance evaluation of the selected recommendation algorithms (MP, CF, CB, UCbSim, BLL, SUSTAIN), the metrics recall, precision and f-measure and nDCG were used. All metrics are averaged over the number of considered users in the test set. Section 5.3.2 elaborates on the metrics' implementations.

7.1.2. Results and Discussion

For this Section, six recommendation algorithms with a total of thirteen variations were evaluated, in terms of prediction accuracy (R, P, F) and ranking (nDCG). In this setting, metric @5 (i.e., the 5 highest ranked items are compared to the test set) is considered most relevant, because this seems to be a reasonable number of items to confront a learner with. Additionally, F@10 and nDCG@10 are reported. To best simulate real-life environment, the study was conducted on six unfiltered TEL datasets from different learning settings. The Section is split in two parts. First, the evaluation of algorithms for learning resource recommendations is presented and then, the evaluation of tag recommendations.

Learning Resource Recommendations

Table 7.1 presents the results of the conducted study, illustrating how well different algorithms performed on the selected TEL dataset. In line with Verbert, Drachsler, et al. (2011), who compared the performance of CF on different TEL datasets, it can be observed that the algorithms' performance values strongly depend on the dataset and its characteristics. Solely CF_U shows a stable behaviour over all datasets. As expected, the performance of CF_U is related to the average number of resources a user interacted with.

Table 7.1.: Results of the resource recommender evaluation organized per dataset and algorithm. The datasets BibSonomy, CiteULike and MACE did not include topic information, thus for those three, CB_T and SUSTAIN was calculated using tags instead of topics. The highest accuracy values per dataset are highlighted in bold.

Dataset	Metric	MP	CF_R	CB_T	CF_U	UCbSim	SUSTAIN	SUSTAIN+CF_U
BibSonomy	R@5	.0073	.0447	.0300	.0444	.0404	.0396	**.0530**
	P@5	.0154	.0336	.0197	.0410	.0336	.0336	**.0467**
	F@5	.0099	.0383	.0238	.0426	.0367	.0363	**.0496**
	F@10	.0102	.0380	.0226	.0420	.0351	.0374	**.0497**
	nDCG@5	.0088	.0416	.0270	.0440	.0371	.0392	**.0541**
	nDCG@10	.0103	.0490	.0313	.0509	.0440	.0469	**.0629**
CiteULike	R@5	.0051	**.0839**	.0472	.0567	.0716	.0734	.0786
	P@5	.0048	**.0592**	.0353	.0412	.0558	.0503	.0553
	F@5	.0050	**.0694**	.0404	.0477	.0627	.0597	.0650
	F@10	.0042	.0601	.0362	.0488	.0573	.0530	**.0618**
	nDCG@5	.0048	**.0792**	.0427	.0511	.0686	.0704	.0717
	nDCG@10	.0054	**.0901**	.0504	.0635	.0802	.0815	.0863
KDD15	R@5	.0067	**.4774**	.1885	.4325	.4663	.3992	.4289
	P@5	.0018	.2488	.1409	.2355	**.2570**	.2436	.2377
	F@5	.0029	.3074	.1612	.3050	**.3314**	.3025	.3059
	F@10	.0034	.2581	.1244	.2773	**.3195**	.2756	.2769
	nDCG@5	.0053	**.3897**	.1927	.3618	.3529	.3227	.3608
	nDCG@10	.0081	**.4740**	.2090	.4281	.4465	.3939	.4284
TravelWell	R@5	.0035	.0257	.0174	.0404	.0471	**.0483**	.0139
	P@5	.0127	.0212	.0382	**.0425**	.0297	.0382	.0382
	F@5	.0056	.0232	.0240	.0414	.0365	**.0427**	.0204
	F@10	.0078	.0194	.0304	.0456	.0459	**.0481**	.0429
	nDCG@5	.0072	.0220	.0275	.0305	**.0491**	.0446	.0220
	nDCG@10	.0092	.0239	.0353	.0461	**.0631**	.0544	.0405
MACE	R@5	.0253	.0080	.0016	**.0283**	.0151	.0093	.0222
	P@5	.0167	.0079	.0023	**.0251**	.0213	.0065	.0190
	F@5	.0201	.0079	.0019	**.0266**	.0177	.0076	.0205
	F@10	.0169	.0116	.0031	**.0286**	.0189	.0155	.0241
	nDCG@5	.0248	.0082	.0014	**.0264**	.0165	.0079	.0215
	nDCG@10	.0281	.0136	.0026	**.0357**	.0282	.0157	.0302
Aposdle	R@5	.0	.0	.0	**.0026**	.0	.0	.0
	P@5	.0	.0	.0	**.0333**	.0	.0	.0
	F@5	.0	.0	.0	**.0049**	.0	.0	.0
	F@10	**.0196**	.0	.0151	.0045	.0	.0045	.0045
	nDCG@5	.0	.0	.0	**.0042**	.0	.0	.0
	nDCG@10	**.0152**	.0	.0103	.0042	.0	.0036	.0033

The SUSTAIN algorithm, which re-ranks the 100 best rated CF_U values, uses categories of a user's resources to construct learning clusters. Hence, the extent of the resource's descriptive features (used are either topics, or tags if topics are not available) is crucial to the success of the algorithm. A comparison of numbers presented in Table 7.1 with the dataset statistics of Table 5.1, indicates that an average of at least three features per resource is needed to improve the performance of CF_U.

Similarly, a poor performance of CF_R is reported for MACE, TravelWell and Aposdle, where the average number of users per resource is lower than two. MP as the simplest approach performs widely poor, except for MACE, where it almost competes with the more complex CF_U. This may relate to the number of learning domains covered by a learning environment. MACE is the only learning environment that is restricted to one subject, namely *architecture*.

The results of this study further underline the importance of a dense user resource matrix. In fact, it revealed a strong correlation of .958 ($t = 19.5502$, $df = 34$, p-value $< 2.2e\text{-}16$) between the average number of users per resource ($|AU_r|$) (see Table 5.1) and the performance (F@5) of all considered algorithms but MP. This is especially visible when comparing KDD15 ($|AU_r| = 49.4$) and Aposdle ($|AU_r| = 1$). KDD15 is the only MOOC dataset in the study. It differs predominantly through its density but also through the structural nature of the learning environment, where each course is hierarchically organized in modules, categories and learning resources.

Contradicting Erdt, Fernandez, and Rensing (2015), which suggested to use MOOCs datasets to evaluate TEL recommendations, this study's findings indicate that recommender performance results calculated on MOOCs are not representative for other, typically sparse, TEL environments. This is especially true for small-scale environments such as Aposdle, where the evaluation positively shows that algorithms based on implicit usage data do not satisfy the use case. For Aposdle, which has only six users, none of the considered algorithms showed acceptable results. While approaches based on individual user data (CB_T, SUSTAIN) may work in similar settings, here, the association of topics is very unfortunate. It does not describe the content of a resource but rather the application type (e.g., template). Furthermore, the allocation of topics to resources is poor and on average only 1.16.

Consequently, learning environments that serve only a very small number of users, such as often the case in work place or formal learning settings, should draw on recommendation approaches that build upon a thorough description of learner and learning resources as incorporated in ontology-based recommender systems.

Tag Recommendations

The tag recommender evaluation was limited to the four TEL datasets of our study that feature tags. Contrary to the results of the resource recommender study, a clear winner can be observed, which performs best on all datasets and metrics as depicted in Table 7.2. BLL_{AC}+MP_R combines frequency and recency of a user's tagging history, which is enhanced by context information and consequently, also recommends tags that are new to a user. Because runtime and complexity are considered very important factors in most TEL environments (Manouselis, Vuorikari, and Van Assche, 2010), please note that $\text{MP}_{U,R}$ outperforms the comparably cost-intensive CF_U in three of four settings, and hence forms a good alternative for runtime-sensitive settings. An extensive evaluation of runtime and memory for tag recommendation algorithms can be found in Kowald and Lex (2015).

7.1.3. Conclusion

This Section presented a data-driven study that measures the performance of six established recommendation algorithms and variations thereof on altogether six TEL datasets from different application domains. The datasets' learning settings cover social bookmarking, open social learning, MOOCs and workplace learning.

Table 7.2.: Results of the tag recommender evaluation, in which the cognitive-inspired BLL_{AC}+MP_R clearly outperforms its competitors (*RQ2*). *Note*: the highest accuracy values per dataset are highlighted in bold.

Dataset	Metric	MP_U	MP_R	$MP_{U,R}$	CF_U	BLL_{AC}	BLL_{AC}+MP_R
BibSonomy	R@5	.3486	.0862	.3839	.3530	.3809	**.4071**
	P@5	.1991	.0572	.2221	.2066	.2207	**.2359**
	F@5	.2535	.0688	.2814	.2606	.2795	**.2987**
	F@10	.1879	.0523	.2131	.1875	.2028	**.2237**
	nDCG@5	.3449	.0841	.3741	.3492	.3851	**.4022**
	nDCG@10	.3712	.0918	.4070	.3693	.4095	**.4343**
CiteULike	R@5	.3665	.0631	.3933	.3639	.4114	**.4325**
	P@5	.1687	.0323	.1829	.1698	.1897	**.2003**
	F@5	.2310	.0427	.2497	.2315	.2597	**.2738**
	F@10	.1672	.0294	.1825	.1560	.1797	**.1928**
	nDCG@5	.3414	.0600	.3632	.3457	.4016	**.4140**
	nDCG@10	.3674	.0631	.3926	.3596	.4221	**.4385**
TravelWell	R@5	.2207	.0714	.2442	.1740	.2491	**.2828**
	P@5	.1000	.0366	.1333	.0800	.1300	**.1400**
	F@5	.1376	.0484	.1724	.1096	.1708	**.1872**
	F@10	.1125	.0388	.1356	.0744	.1287	**.1426**
	nDCG@5	.2110	.0717	.2253	.1622	.2525	**.2615**
	nDCG@10	.2411	.0800	.2686	.1730	.2783	**.2900**
MACE	R@5	.1306	.0510	.1463	.1522	.1775	**.1901**
	P@5	.0576	.0173	.0618	.0631	.0812	**.0812**
	F@5	.0799	.0259	.0869	.0893	.1114	**.1138**
	F@10	.0662	.0170	.0692	.0615	.0829	**.0848**
	nDCG@5	.1146	.0463	.1296	.1502	.1670	**.1734**
	nDCG@10	.1333	.0483	.1477	.1568	.1835	**.1902**

In a first experiment, the suitability of three state-of-the-art recommendation algorithms (MP, CF, CB) and two approaches suggested for the educational context (UCbSim, SUSTAIN) were investigated. The algorithms were applied on implicit usage data. According to the study's findings, satisfactory performance values can only be reached for KDD15, the MOOCs dataset. This suggests that standard resource recommendation algorithms, originating from the data-rich commercial domain, are not well-suited to the needs of sparse data learning environments. In the second study, an evaluation of computationally inexpensive tag recommendation algorithms was conducted. To this end, the performance of MP, CF and a cognitive-inspired algorithm, BLL_{AC}, was computed on four datasets. Results show that a hybrid recommendation approach combining BLL_{AC} and MP_R clearly outperforms the remaining methods.

The next Section 7.2 continues with this line of work, deepening the investigation by evaluating tag recommendation strategies in an online learning environment.

7.2. Online Study: Tag Recommendation Algorithms in the weSPOT Project

This Section presents the application and exploration of three computationally simple tag recommendation strategies in an online social IBL environment. Two algorithms, namely BLL and MINERVA2 are based on cognitive models that mimic student's tagging behaviour, taking into account either temporal or semantic context (see Section 4.2). The most popular (MP) tags approach, as the third algorithm, is a computationally simple mechanism that has demonstrated its potential on TEL datasets (Kopeinik, Kowald, and Lex, 2016) and is used as a baseline. A study has been conducted that investigates the effectiveness of the two cognitive-inspired recommendation mechanisms. Furthermore, as the tag vocabulary, on which a student draws to organize and reflect on resources, emerges not only from personal tag choices, but also from those of others, students are expected to benefit from semantic stabilisation (Wagner et al., 2014), a phenomenon that describes the common agreement of tag choices of particular ranges of topics. The more stable the currently evolved tag vocabulary is, the more helpful it should be to share own and exploit others' search results. Thus, this study further elaborates on whether semantic stabilization can be supported by the proposed tag recommendation mechanisms. This inquiry is based upon the hypothesis that in online social learning environments, semantic stability can be fostered by cognitive-inspired tag recommendation approaches. All algorithms are implemented in two variations, i.e. either based on a user's personal, or the group's collective tagging history.

The content of this Section has been partly published in Kopeinik, Bedek, et al. (2015) and Kopeinik, Lex, et al. (2017).

7.2.1. Approach

A very simple, though relatively effective, tag recommendation strategy is the Most Popular (MP) algorithm (Jäschke et al., 2007; Kopeinik, Kowald, and Lex, 2016). However, it can be assumed that a frequency-based, computationally simple recommendation strategy may be even more successful, if it is grounded on a thorough understanding of how humans process information.

In previous work (Kowald, Seitlinger, Kopeinik, et al., 2015; Kopeinik, Kowald, and Lex, 2016), the suitability of two tag recommendation approaches was extensively evaluated via offline studies. Two approaches that aim to imitate cognitive processes of retrieving words from memory have shown to be particularly promising:

BLL implements the Base Level Learning Equation (John R. Anderson and Schooler, 1991), which models the frequency and recency of past tag use.

MINERVA2 (Hintzman, 1984; Seitlinger, Ley, and Albert, 2013), incorporates tag use frequency as well as semantic context.

These approaches have been implemented in an online learning setting, based on data of two origins: A user's personal tagging history (P) and a groups collective tagging history (C). When implementing an algorithm in an online setting, it is very likely that the approach needs to be adjusted to the data available in the particular environment. In this example, the algorithms are adapted to fit the conditions of the inquiry-based learning platform weSPOT, which is described in Section 7.2.2. The adaptation of the algorithms is described subsequently.

BLL

The theoretical model of the activation equation and its proposed application in tagging is described in Section 4.2.1. However, considering the type of data that is provided by the weSPOT learning environment, the calculation of the associative component has to be adapted. The problem arises, as this component is based on tags multiple users have assigned to the very same content or item within the environment. weSPOT, though, is a narrow folksonomy (such as for instance Flickr), where content is generated and tagged only by one user. Therefore, two strategies can be followed:

1. Implementing solely the BLL component.
2. Collecting context information (i.e. tags that are new to a user) by different means, as for instance most frequent tags of the user's inquiry group. This implementation depicts an additional recommendation approach and can be denoted as $BLL_U + MP_G$.

MINERVA2

MINERVA2 is a model that aims to mimic a process of human categorization. Its theoretical foundation was introduced in Section 4.2.2. The tag recommendation mechanism, which is based on this model was firstly described in Seitlinger, Ley, and Albert, 2013. It is represented as a simple network model with an input, a hidden and an output layer. The input layer is a feature vector that describes the resource. Within this environmental setup the input layer consists of attributes the user selects. These attributes were drawn from the inquiry's domain model. For further information on the domain model please see Bedek et al., 2015. The output layer is a list of ranked tags, with a maximum of five suggestions.

7.2.2. Recommending Tags in the weSPOT Environment

The study was implemented in the collaborative online learning environment weSPOT[1]. The weSPOT inquiry space guides students through the inquiry cycle, which models the scientific inquiry process, based on a theoretical framework, in six phases: Question/Hypothesis, Operationalisation, Data Collection, Data Analysis, Interpretation/Discussion and Communication. Each phase further includes dedicated activities as discussed in Protopsaltis et al., 2013.

Figure 7.1 exhibits the weSPOT inquiry space that implements the six IBL phases, providing individual tabs for each phase labelled with (1). Each phase-tab further includes widgets (2) that enable the students to carry out activities relevant to a specific phase in the inquiry-based learning cycle. The platform's side panel (3) provides reflection and support tools that are not related to specific inquiry phases but rather support the entire learning approach. These tools encompass a learning analytics dashboard, an open user model and a learning resource recommendation interface. The weSPOT space supports collaborative learning in defined groups. Thus, each student group is provided with a sub-environment that forms an inquiry space addressing a specified research interest. Teachers take a supportive and administrating role. They setup the inquiry, monitor the students during their learning activities, answer questions and intervene if student's loose track. In the platform, teachers are provided with a configuration interface, to

[1] http://inquiry.wespot.net/

design inquiry spaces by selecting phases (tabs) and activities (widgets) that suit the purpose of their student's inquiry projects. Teachers also add students and initial learning content to the group environment.

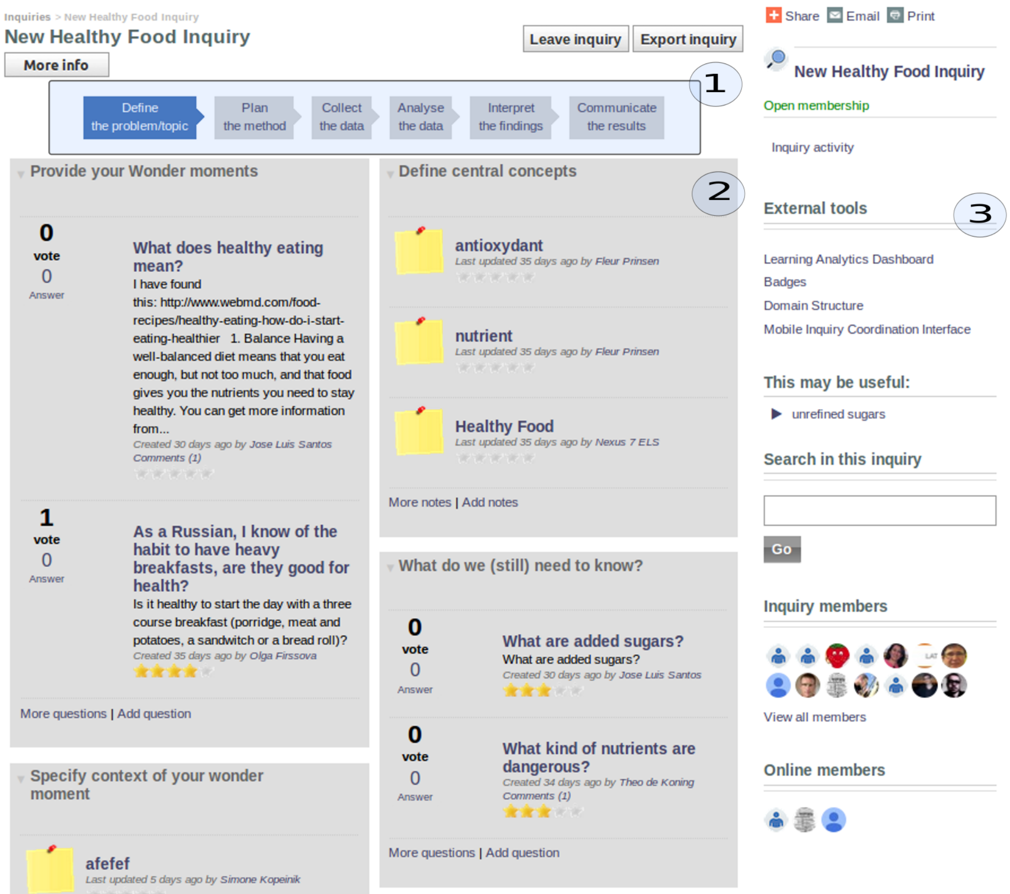


Figure 7.1.: The collaborative online learning platform: weSPOT IBL space. (1) shows the IBL phases that are depicted as one tab each. (2) widgets in one tab. (3) side panel with external (supporting) tools and group information.

While students work on their inquiry projects, they engage in activities that typically create content by, e.g., posting questions, starting or contributing to discussions or by uploading documents and pictures. These and other learning activities are tracked, saved and fed into different user profiles, to be later used in learning analytic diagrams, to issue badges and to provide personalized recommendations of learning resources and tags.

Tagging Interface

Figure 7.2 shows an extended version of the environment's standard input form. The tag recommendation plug-in that extends the form is marked with an orange frame.

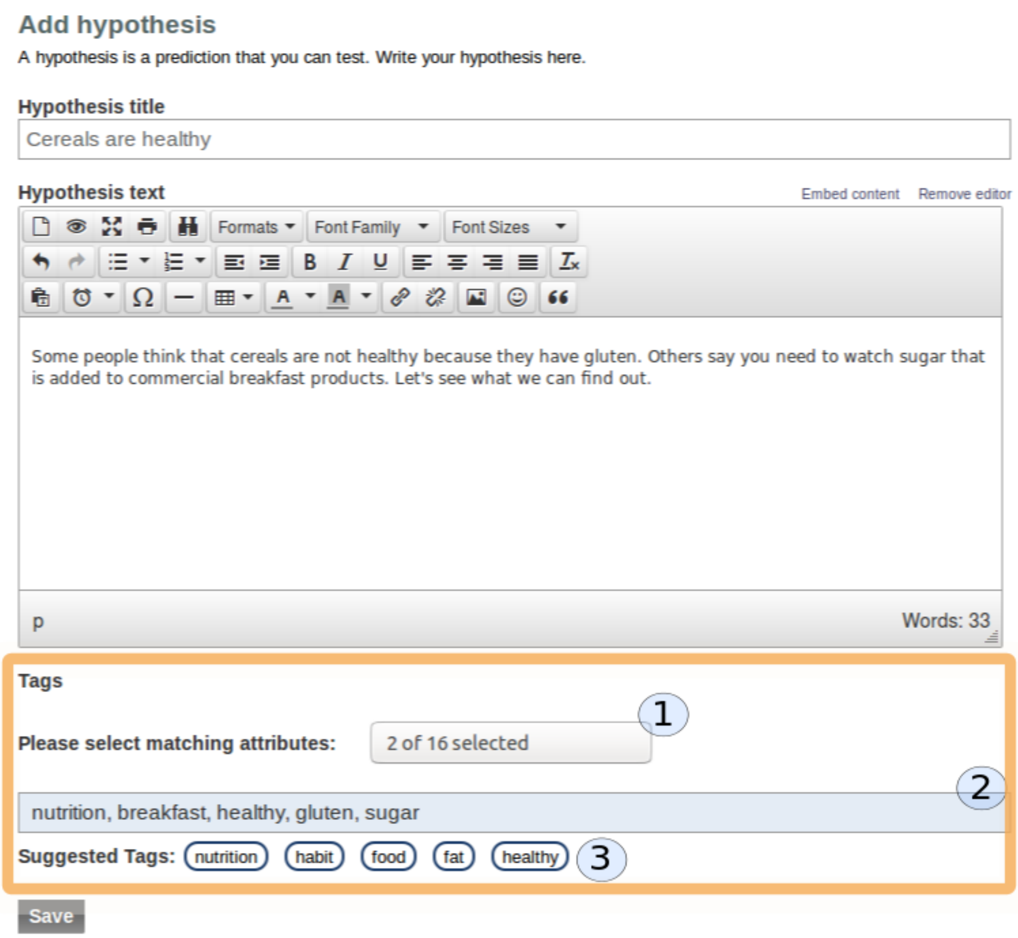


Figure 7.2.: The standard elgg input form extended by our tag recommendation plug-in (marked with the orange frame). After choosing relevant semantic features, students can either select from recommended tags (3) by clicking on the selected item or enter their own tags in the text field (2).

Within this implementation, the annotation process consisted of two steps:

1. **The selection of semantic features (attributes)**
   The learner selects semantic features that describe the contribution from a provided dropdown menu (1). Attributes were drawn from the inquiry's

domain model which were provided by the teacher. Further information on the domain model and related tools can be found in Bedek et al., 2015.

2. **The assignment of tags**
   After the student closed the dropdown menu, tag recommendations (3) appeared just below the tags input text field (2). Students could either select from these recommendations or add their own tags manually.

Technical Insights.

The core of the weSPOT environment is an open social online platform that is based on Elgg[2]. Elgg is an open source social networking engine and comes with a framework for the creation of social environments. It is expandable via plug-ins and follows a MVC (Model-View-Controller) pattern which makes it convenient to extend.

When a user enters content (e.g., question, hypothesis, file or discussion entry) to an inquiry space, this happens through an input form which includes a "tag view". The tag recommendation plug-in is an extension of this "tag view" and adaptively suggests tags to users. The tag view (thus also the tag recommendation functionality) is by default included in all plug-ins that allow users to create content, as for instance in discussions, file uploads or blog entries. Figure 7.2 shows an Elgg input form with the recommendation plug-in embodied as marked by the orange frame.

Whenever a form that includes the tag view is loaded, the tag recommendation plug-in is initialized and pulls the data (e.g., suggested tags) via a web-service interface from the backend service. From the pool of recommendation strategies explained in Section 7.2.1 the applied algorithm is selected randomly. The tag recommendations displayed in the plug-in are calculated in a backend web-service component. Within the backend service domain and learner models are created and maintained and based upon these models recommendation algorithms are implemented. The algorithms use previously collected learner interaction data in a Solr-based data store, to assure real-time data . A REST-based web service querying this data store generates personalised recommendations on demand.

[2]https://elgg.org/

7.2.3. Real-life Evaluation Study

Offline data studies have indicated that the modelling of cognitive processes underlying tagging habits leads to an increased accuracy of recommendations (Trattner et al., 2016a). However, offline data studies are limited to evaluating the prediction of user behaviour. The work presented here explores the performance of tag recommendation algorithms in an online, real-world scenario to investigate whether the promising results from offline data generalize to online environments. The experiments took place in form of a field study in which students used an online IBL environment within a realistic school context for a duration of about four weeks.

In order to derive more precise and practical design implications for specific learning settings, two variables underlying the design of these recommenders are systematically varied. The first variable is the type of information the algorithm takes into account to estimate the current probability of a tag being retrieved from the learner's memory. While MP only considers a tag's usage frequency (baseline), the two cognitive-inspired algorithms extend this approach by the information of recency of usage (BLL) and the extent to which a tag matches the current (i.e., the resource's) semantic context (MINERVA2).

With regard to results presented in Font, Serrà, and Serra (2015) it is fair to assume that an increased awareness of peer learners' tag choices will promote the development of a common terminology. Thus, the second variable is the vocabulary, from which the algorithm selects the tags. The experiment distinguishes between:

Personal (P)
: Students receive tag recommendations based on their personal tagging history.

Collective (C)
: Students receive tag recommendations based on the collective tagging history of their learning group.

Methodology

The study investigates the suitability of BLL and MINERVA2 to face the challenges of real-life IBL learning settings. The resulting data sample consists of N=56 students with an age ranging from 15 to 17 years. As summarized in Table 7.3, the independent variables formed a 2 (Vocabulary: Personal vs. Collective; between-subjects) by 3 (Algorithm: MP vs. BLL vs. MINERVA2; within-subjects) design. In addition BLL+MP is considered for the collective vocabulary condition. This recommendation approach is of particular interest, as in offline studies on TEL datasets it clearly outperformed remaining algorithms (Kopeinik, Kowald, and Lex, 2016). The dependent variables were recommender accuracy and semantic stabilization.

Study Setup

To investigate whether in online social learning environments, cognitive-inspired tag recommendation strategies can be applied successfully and furthermore, can foster semantic stability, a real-life evaluation in the context of high school biology lessons, engaging students in IBL projects was conducted. To this end, an online environment for open social inquiry-based learning was used. IBL itself is very well-fitting for the purpose of a collaborative tagging study as, throughout the learning process, students are constantly challenged to find, create, upload and share content. In the course of the study, four secondary school classes with students at the age of 15 to 17 used a dedicated social learning environment to work on their biology projects.

The study design treats semantic stability and recommendation accuracy as dependent variables.

Evaluating Recommender Accuracy

This experiment evaluated the performance of the tag recommendation algorithms MP, BLL and MINERVA2 utilizing the performance metrics recall, precision and f-measure (see Section 5.3.2). When calculating recall and precision, for each post, the relation of tags recommended $\hat{T}_{u,r}$ to a user u for a resource r to the tags that the user assigned to a resource $T_{u,r}$ are considered. All metrics are averaged over the number of considered posts.

Evaluating Semantic Stability

As summarized in Wagner et al. (2014), there is a multitude of metrics to evaluate semantic stability. Only few methods are yet suited for narrow folksonomies, where items are tagged only by the uploading user. Lin et al. (2012) presents the Macro Tag Growth Method (MaTGM) that measures social vocabulary growth at a systemic level, looking at the social tagging system as a whole. In this setting, each IBL group is considered as an isolated social tagging system. MaTGM is applied to compare the tag growth within these systems.

The most representative dataset for this study setup, stems from the school class that comprises the most extensive tag data for both conditions (personal and collective). This dataset is described in Table 5.1 as study *TG*. The posts (tag assignments) of each group were sorted according to their timestamps, ending with the most recent item annotation. Then, the tag growth after each post, is calculated as a value pair $(tg_i, f(tg_i))$, where tg_i is the cumulative number of tags, and $f(tg_i)$ is the cumulative number of unique tags occurring in i posts.

Procedure

Prior to the first lesson, students and their parents were presented with the goals and benefits of using IBL and the online learning environment, as well as with the aims of the study. Afterwards, parents and students were asked for their consent in written and verbal form, respectively. Throughout the process of the study neither participation in the platform nor in tagging was obligatory or contributed to the grading of students.

For the purpose of the study, students of each class were divided in two groups per class, which led to groups of 9 to 18 students, depending on class size. In the first two lessons, students were introduced to the online learning environment (see Section 7.2.2) and the concepts of IBL. Then, each group used the virtual learning environment for at least eight school lessons over a period of four weeks or longer to complete an IBL project. The teacher provided each class' learning groups with similar learning content and learning tasks and acted in a supporting role. The variation between groups is constituted by the nature of tag recommendations.

Tag recommendations of one group are based on individual user's personal tag data, whereas the second group's tag recommender draw on the group's collective tagging traces. Depending on the group, tag recommendation strategies were randomly selected either from the personal or collective pool of recommendation strategies, as illustrated in Table 7.3.

Vocabulary	Algorithm
Personal (P)	MP_U BLL_U $MINERVA2_U$
Collective (C)	MP_G BLL_G $MINERVA2_G$ $BLL_U +MP_G$

Table 7.3.: Students of each class were separated in two groups and consequently received either tag recommendations based on their personal tagging history (P) or based on the collective tagging traces of their inquiry group (C). Condition C was complemented by the mixed approach $BLL_U +MP_G$.

Each student was provided with a tablet computer available during class. Students were encouraged to use the tagging functionality when creating or uploading new content, and were also provided with information in verbal and written form, on how to do that.

Dataset

The datasets used in this study were collected on a dedicated log data server, from which the eight inquiry groups were extracted that participated in the experiment. All groups consisted of students attending a high school in Graz and worked on the projects in the course of biology classes, on altogether four different research topics. As one setting took place in the course of an extra-curricular specialisation, ten students participated twice in the experiment. The data was collected over a period of three school semesters (from spring 2015 to summer 2016).
Although students were provided with initial instructions on the tagging interface and the tagging process itself, quite a few students did not tag at all, or provided tags in unusual ways. Consequently, the datasets were pre-filtered by mainly excluding posts with tags in form of sentences or tags concatenated with special characters. Students with no remaining posts were also excluded from the datasets, which led to the data samples given in Table 7.4.

Research Topic	Vocabulary	$\lvert P\rvert$	$\lvert U\rvert$	$\lvert T\rvert$	$\lvert T_{unq}\rvert$	$\lvert AT_u\rvert$	$\lvert AP_u\rvert$
Soil ecosystems	P	9	6	17	11	2.3	1.5
	C	98	13	177	32	5.4	7.5
Biodiversity in cities	P	8	4	19	9	2.3	2.0
	C	35	14	75	24	3.9	2.5
Renewable resources	P	6	5	29	22	4.6	1.2
	C	12	8	34	19	4.1	1.5
Climate change	P	65	6	232	85	16.8	10.8
	C	83	10	297	86	16.4	8.3

Table 7.4.: Properties of the preprocessed datasets extracted from eight inquiry groups. $|P|$ depicts the number of posts, $|U|$ the number of users, $|T|$ the number of tags, $|T_{unq}|$ the number of unique tags, $|AT_u|$ the average number of tags per user, $|AP_u|$ the average number of posts per user. *Vocabulary* refers to the data the tag recommendations were based on i.e., (U)ser or (G)roup.

To measure the recommender accuracy (RA), all samples under the independent study variable *Vocabulary* were taken together. The resulting dataset properties are presented in 7.5. To evaluate semantic stability according to tag growth (TG), the class was selected, which created the highest number of posts in both inquiry groups. This sample can be considered as the most significant to the investigation.

Aspect	Vocabulary	$\lvert P\rvert$	$\lvert U\rvert$	$\lvert T\rvert$	$\lvert T_{unq}\rvert$	$\lvert AT_u\rvert$	$\lvert AP_u\rvert$
RA	G	228	38	584	153	15.4	6.0
	U	88	18	297	121	16.5	4.9
TG	G	83	10	297	86	16.4	8.3
	U	65	6	232	85	16.8	10.8

Table 7.5.: Properties of the datasets taken into account for the investigation of two aspects: Recommender accuracy (RA) and tag growth (TG). $|P|$ depicts the number of posts, $|U|$ the number of users, $|T|$ the number of tags, $|T_{unq}|$ the number of unique tags, $|AT_u|$ the average number of tags per user, $|AP_u|$ the average number of posts per user. *Vocabulary* refers to the data the tag recommendations were based on i.e., (U)ser or (G)roup.

7.2.4. Results and Discussion

This paragraph presents the results of the two recommender experiments. It comprises an evaluation of the suitability of the here described algorithms for supporting learners' tagging processes. In line with the study design, all algorithms are applied to two modes: Personal (P), where the recommendation strategy draws on a single user's posting history, and collective (C), where the recommendation strategy draws on the prior posts of an entire group.

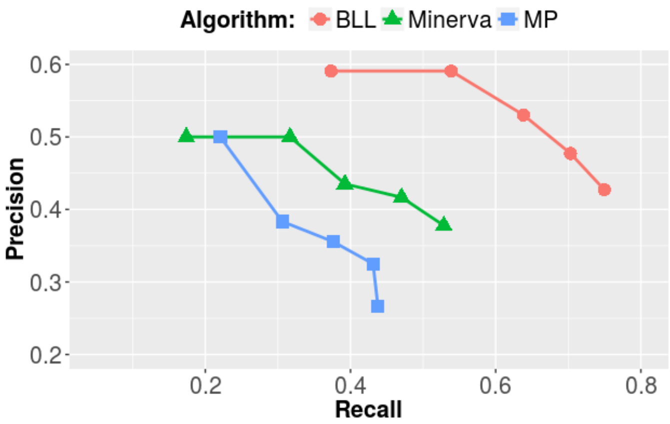
The Accuracy of Cognitive-Inspired Tag Recommendation Strategies in an Online Data Setting

In this paragraph, the results of the conducted evaluation study are presented in respect to recommendation accuracy. Table 7.6 provides the number of observations (see column N_T) and accuracy estimates (R, P and F) for each recommender, as also illustrated in Figure 7.3.

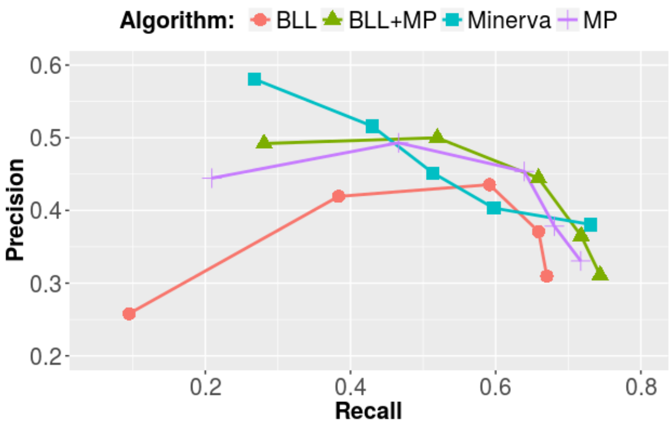
Table 7.6.: Properties of the analysed dataset, structured by the applied algorithm. The algorithms were either calculated on a user's personal word trace P, an inquiry groups collective word traces C or a *Mixed* approach PC considering both type of data, collective and individual. N_T depicts the number of tagged resources, as derived from the online evaluation. The metrics recall, precision and f-measure are mean values and standard deviations of *R@5*, *P@5* and *F@5*, respectively.

	Algorithm	N_T	*P@5*	*R@5*	*F@5*
P	*MP*	30	0.26 (0.25)	0.44 (0.36)	0.31 (0.27)
	MINERVA2	36	0.38 (0.32)	0.53 (0.39)	0.41 (0.31)
	BLL	22	**0.43** (0.28)	**0.75** (0.33)	**0.50** (0.26)
C	*MP*	72	0.33 (0.21)	0.72 (0.36)	0.42 (0.23)
	BLL	62	0.31 (0.23)	0.67 (0.37)	0.39 (0.23)
	MINERVA2	31	0.38 (0.28)	0.73 (0.38)	0.46 (0.30)
PC	$BLL + MP$	63	0.31 (0.21)	0.74 (0.38)	0.41 (0.25)

Resources without tags or with invalid tags (e.g., urls or sentences) were manually excluded from the dataset. Also, due to the cold start problem, initial tag recommendations may have encompassed less than five tags. In order to preserve these early observations and to perform an analysis comprising the whole set of empirical data, all metrics are calculated assuming the presence of five recommended tags, i.e., if a tag recommendation only consists of two tags, the three missing tags are considered as tags the user has not selected (incorrectly recommended tags).



(a) Personal vocabulary condition: tag recommendations are based on a user's personal tagging traces.



(b) Collective vocabulary condition: tag recommendations are based on the learning group's collective tagging traces.

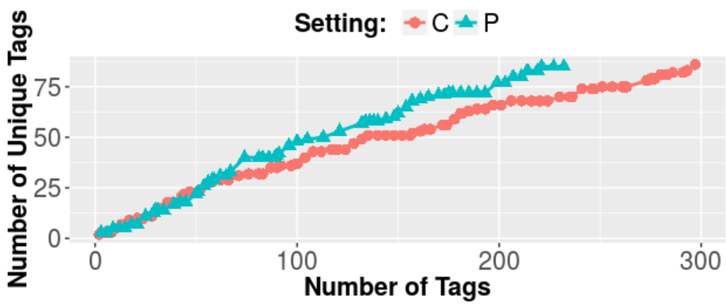
Figure 7.3.: Recall/Precision plots illustrating the accuracy of recommendation algorithms in the personal and the collective vocabulary condition. B_i applied in the personal setting performs best over all considered recommendation approaches. In the collective condition, best results can be achieved for B*i*+MP and MINERVA2.

The table discloses the impact of the two variables algorithm and dataset on performance: BLL appears to reach higher estimates than MINERVA2 (relative to MP) under the personal vocabulary condition, with the opposite being true for the collective condition. In line with this descriptive pattern, a 2 (Personal vs. Collective) $\times$ 3 (MP vs. BLL vs. MINERVA2) ANOVA, a statistical model for the analysis of variance (Hale, 1977), was applied on F. It reveals no significant main effects, either for vocabulary, $F(1,44)$=1.22, *n.s.*, nor algorithm, $F(2,44)$=2.35, *n.s.*, but a significant interaction between these two factors, $F(2,44) = 4.33$, $p < .05$.

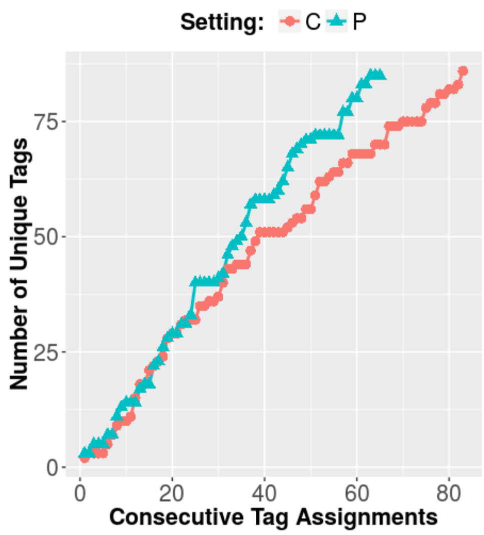
Results indicate that in the personal setting the BLL approach, which mimics the activation of words in a person's memory as a function of frequency and recency, performs best. On the other hand, BLL applied in the collective vocabulary condition performed very poorly. Also, we can see that the recommender MINERVA2 showed better performance in the collective, than in the personal vocabulary condition. While a model that categorizes according to semantic context should be able to depict both, personal and collective data, it is fair to assume that the size of the dataset plays a crucial role. It is fair to assume that the approach will become more accurate with the growing extent of the dataset. Hence, two possible interpretations are: First, MINERVA2 performs better on collective compared to personal tagging traces, as the dataset is likely to be more extensive. Second, the algorithm's performance will enhance with the time of use. This corroborates initial expectations, as it indicates that students' interests within a group differ but are individually relatively stable within the short period of a school project. Students' individual developments within a topic can be further depicted with the introduction of recency, as implemented in BLL_U. A very interesting result constitutes the moderate performance of BLL_U+ MP_G, which is contrary to results from offline TEL data studies such as presented in Kopeinik, Kowald, and Lex (2016), where the approach clearly outperforms remaining recommendation strategies.

The Impact of Individual and Collaborative Tag Recommendations on Semantic Stability

The two plots illustrated in Figure 7.4 present the development of the tag vocabulary on a group level as described in Section 7.2.3. The graphs oppose the tag growth occurring in the collective group vocabulary condition (C) with the tag growth happening in the personal vocabulary condition (P), where students received their tag recommendations either based on collective tag traces or on personal tag traces, respectively. Figure 7.4a depicts the tag growth function according to the Macro Tag Growth Method and shows that while initially the vocabulary growth overlaps in both groups, group C starts to introduce less new vocabulary in relation to tags than group P. In other words, we can observe that students in the collective condition start to pick up the vocabulary of their peers faster.



(a) Tag growth function according to the Macro Tag Growth Method.



(b) Number of unique tags accumulated with consecutive tag assignments.

Figure 7.4.: The plots show the development of tagging vocabulary on a system (inquiry-based learning group) level. The two line graphs depict the between-subject variables of the study, that distinguish between the settings: collective (C) and personal (P).

This result is even stronger when considering that a greater number of users contributed to the tagging data of the collective condition than to the data of the personal vocabulary condition (see Table 5.1). This indicates a positive effect of collective tag recommendations on semantic stabilisation. Figure 7.4a provides additional insights into the timing of the process. We can observe that the two tag growth functions clearly diverge after about 40 added posts.

Datasets

Inspecting Table 5.1, we can observe that the tagging frequency varies greatly among the groups. Students that participated in the study used the environment in the course of biology lessons. However, the IBL project work did not contribute to their grading. Also, they were encouraged to tag but there was no particular monitoring of this process taking place. Thus, some groups showed more motivation and participated more actively in the projects and within the environment than others.

Another aspect is that there is significantly less data available for vocabulary condition P, where users' tag recommendations were based on their individual tagging history. Due to the cold start problem, students in condition P were not provided with initial tag seeds but rather had to come up with their personal tag traces to initiate the tag recommendation process. The resulting lack of tagging support may have played a crucial role when students did not tag their contributions or tagged their contributions in unusual ways (see Section 5.1.4), but this needs further investigations. However, the assumption is in line with previous findings (e.g., Kuhn et al. (2012)) that underline students' need for support in the tagging process.

7.2.5. Conclusion

This Section continued to work on RQ3 by conducting a real-life evaluation investigating the application of tag recommendation approaches from two perspectives:

1. It divided students in two groups that received either tag recommendations based on personal or collective tag traces. In this way, insights into the effect

of collective tag recommendations on the semantic stabilisation process of collective learning groups could be gained.

2. It evaluated the performance of two tag recommendation approaches that imitate human behaviour, in particular the process of human categorization and the retrieval of words from memory. The algorithms, MINERVA2 and base-level learning equation (BLL), as well as MP as a baseline, were applied as within-subject variables, either on the basis of the collective or the personal tagging history.

The experiment's results show that selecting recommendations from the collective vocabulary, i.e., exposing a learner to others' tags, is much more effective to drive stabilization than drawing from the personal vocabulary and thus, displaying only individual tags. Furthermore, the results suggest that searching for relevant tags in the collective's vocabulary benefits strongly from considering usage frequency and semantic context, i.e., from a strategy implemented by MINERVA2. The information of recency, on the other hand, appears to show advantages when aiming to identify relevant tags within the personal vocabulary. One practical design implication therefore is that stabilization within the setting of inquiry-based group learning can be supported well by recommenders that both draw on data of the whole collective and are sensitive to the semantic context of learners' search results in order to estimate tag choice probabilities. In case of an individual learning setting, however, it is promising to apply recommenders that focus on information about time and frequency of past tag choices to predict their current availability in a learner's memory and hence, relevance for the current learning episode.

This Chapter contributes to RQ3: *Can resource and tag recommendations that are based on cognitive learner models form a competitive alternative to common statistically based approaches?* and led to the following results: i) standard resource recommendation algorithms, originating from the data-rich commercial domain are not well suited to the needs of sparse data learning environments, ii) the evaluation of algorithms on MOOCs data sets is not representative for other typically sparse learning environment's, iii) tag recommendations based on BLL and individual user's tagging traces, outperform baseline algorithms in offline and online studies, and iv) MINERVA2, an algorithm considering semantic context and a group's collective tagging data fosters semantic stabilisation in collaborative learning settings.

Based on these finding RQ3 can be partly answered positively, which shows that tag recommender algorithms based on cognitive learner models can compete with common statistically based approaches and thus, are well-suited to perform on sparse learning data. Results of the data-driven learning resource recommendation study expose that none of the algorithms satisfies the use case. An extensive study, comparing different learning settings, proves the insufficiency of standard resource recommenders for sparse data environments, thereby substantiating a common claim in the field. Furthermore it highlights the problem of sparsity in learning content meta-data. The support of user's content annotations via tag recommendations may generate more complete data sets in future learning settings, which increases the suitability of content-based recommendation approaches such as SUSTAIN or CB_T. Moreover, due to limitations of the used TEL datasets, it was not possible to include the CbKST-based recommendation strategy into the here reported data-driven study. Thus, it remains to exploit how the well-founded structural learner model competes with statistically based approaches.

Part III.

Discussion and Future Work

8. Conclusion

In this thesis, the conjecture is explored that RS in TEL settings may be more successful if they are based on a thorough understanding of how humans process information. In particular, the focus is on the exploration of cognitive user models to recommending learning resources and tags towards the specific requirements of different TEL environments. Scientific work, presented in this thesis was conducted over the course of three research projects: INNOVRET, weSPOT and Merits (see Section 1.3) that determined the learning settings and consequently, the requirements of the recommendation strategies. The key challenge was to find theoretically plausible models that cover a large amount of relevant aspects while being still computable on restricted computational resources as often found in educational contexts (Pierce and Cleary, 2016). A short recap of the three research questions and the contributions this thesis has made is given in the following Section.

8.1. Scientific Contributions

This Section is structured according to the research questions posed.

RQ1: **Can a learning resource recommender based on a structural learner model (like the CbKST) improve the learning experience in a formal learning environment?**

This first research question was addressed in Sections 6.1. Research was conducted within the INNOVRET project (see Section 1.3), which specified the demands of the personalization strategy. The challenge was to design a personalization approach, incorporating the specific requirements of formal learning settings for students with diverse backgrounds. To this end, Moodle plug-ins were introduced that combine principles of self-regulated learning with adaptive guidance support. The CbKST was applied as an underlying domain model that allows the generation of overlay learner models and provides means for adaptive assessment. Based on these learner models, the recommendation strategy suggests learning resources tailored to a user's current knowledge state. This can enhance the learning experience by avoiding frustration, caused by over-challenging learning items, and boredom, caused by under-challenging learning items (Csikszentmihalyi, 2014). To capture the effects on the user's learning experience, an experimental study with fourteen users was conducted. An A/B testing approach was applied to compare the perceived learner satisfaction in a standard Moodle course (control group) with the extended, personalised Moodle environment (experimental group). This first evaluation showed encouraging results as the CbKST-based experimental group rated their satisfaction with the system above average and on substantial scales higher than the control group.

While the low number of participants precludes generalization of the results, this work contributes to the research field as a proof of concept and a valid example of how to apply the CbKST as a domain and learner model for learning resource recommendations in formal learning environments. Further studies are planned to corroborate the findings.

RQ2: **Can a process oriented learner model based on SUSTAIN be applied to improve an existing resource recommendation strategy such as collaborative filtering?**

To investigate this research question, a recommendation strategy using SUSTAIN to capture the dynamics of human learning in informal learning settings has been proposed in Section 6.2. The SUSTAIN-based learner model captures attentional foci and interpretation dynamics that evolve along a user's sense-making activities on a self-directed learning path of consecutively discovered learning resources. This is a dynamic approach that fits the requirements of informal learning settings where learners explore unstructured learning content without a pre-defined learning goal (Colardyn and Bjornavold, 2004). The results of the evaluation of the SUSTAIN model and a hybrid version of CF_U and SUSTAIN on three social bookmarking datasets (BibSonomy, CiteULike and Delicious) are presented in Section 6.2.5. The study reveals that SUSTAIN+CF_U outperforms SUSTAIN, CF_U and CF_R on all three datasets. Furthermore, WRMF only reaches higher accuracy estimates in one of the datasets, which indicates that the approach can also compete with this much more computationally expensive method (see Kopeinik, Kowald, Hasani-Mavriqi, et al. (2016)). These results support the hypothesis of Drachsler, H. G. K. Hummel, and Koper (2009) who argue for the application of hybrid recommendation strategies combining CF with pedagogically plausible learning models. To better understand the underlying dynamics of the modelling approach, additionally, a study systematically varying the model's main parameters was conducted. This investigation shows that the memorization of a user's history, as well as the extent of clustering, contribute to the algorithm's performance. More precisely, it indicates that the restriction of cluster growth can prevent overfitting in sparse data environments.

In conclusion, by characterising users' observable search behaviour in terms of underlying learning processes, such as the formation and adjustment of mental categories, the performance of existing recommender mechanisms like collaborative filtering can be improved and at the same time by studying these models, a deeper understanding of underlying interaction dynamics can be achieved. This answers the research question positively.

RQ3: **Can resource and tag recommendations that are based on cognitive learner models form a competitive alternative to common statistically based approaches in TEL settings?**

This research question has been addressed by two means, an offline data study and a real-life study.

Offline data study

The study is described in Section 7.1. It investigates the suitability of six state-of-the-art resource and tag recommendation algorithms and variations thereof on implicit usage data from six TEL datasets. To best simulate conditions of a real-life setting, the experiment evaluates the prediction accuracy of i) learning resource recommenders and ii) tag recommenders on the unfiltered datasets.

Generally, the learning resource recommender experiment shows that the performance of applied recommendation algorithms strongly depends on learning context and dataset properties. This is in line with findings of prior research

(Manouselis, Vuorikari, and Van Assche, 2010). Also, a strong correlation between the average number of users per resource and the algorithms' performance values highlight the importance of a dense user-resource matrix. Among all algorithms, satisfactory values can only be reported for KDD15 which is the only MOOC dataset in the study. It differs predominantly not only through its density but also through the structural nature of the learning environment, where each course is hierarchically organized in modules, categories and learning resources. Thus, this study's findings indicate that recommender performance results calculated on MOOCs are not representative for other, typically sparse, TEL environments. This is especially true for small-scale environments such as Aposdle, where the evaluation positively shows that algorithms based on implicit usage data do not satisfy the use case. For Aposdle, which has only six users, none of the considered algorithms showed acceptable results. While approaches based on individual user data (CB_T , SUSTAIN) may work in similar settings, this may be hindered due to the unfortunate association of topics, which do not describe the content of a resource but rather the application type (e.g., template) and a poor allocation of topics to resources. As a consequence of these finding, one may conjecture that learning environments that serve only a very small number of users should either draw on a thorough description of learner and learning content, such as incorporated in ontology-based approaches, or strongly support the annotation of relevant learning content, for instance with tag recommendation algorithms.

The offline tag recommender study of three algorithms, implemented as six variations based on usage data and hybrid combinations, identifies a cognitive-inspired recommendation algorithm combined with a popularity-based approach as most successful. BLL_{AC}+MP_R combines frequency and recency of a user's tagging behaviour with context information and thus, also recommends tags that are new to a user. In runtime sensitive-settings $MP_{U,R}$ may constitute a good approach, as it outperforms the comparably cost-intensive CF_U in three of four datasets.

While existing research investigates the application of implicit usage data-based algorithms (e.g., Verbert, Drachsler, et al. (2011), Niemann and Wolpers (2013), and Fazeli et al. (2014)) on selected datasets, a more extensive comparative study directly opposing state-of-the-art recommendation algorithms had been missing. Overall, this experiment revealed that standard resource recommendation algorithms are not well-suited to the needs of sparse data learning environments. The SUSTAIN algorithm that enhances CF_U showed promising results in the social bookmarking dataset of RQ2. Due to the lack of descriptive learning resource features (tags or topics) the approach performed poorly in most of the other TEL datasets. Tag recommendation algorithms on the other hand performed reasonably well and have been further explored in a real-life study.

Real-life Study

The aim of the real-life study presented in Section 7.2 was i) to investigate the effectiveness of cognitive-inspired tag recommendation algorithms that have proven successful in offline experiments, in an online environment and ii) to explore the potential of these tag recommendation mechanisms to support semantic stabilisation. To this end, a tagging recommendation mechanism was implemented within a real-life IBL setting. Learners are supported in the annotation of uploaded or self-created content in form of a tag recommendation plug-in that provides personalized tag recommendations retrieved from a web-services component. The study investigated the potential of two cognitive-inspired tag recommendation approaches, namely the base-level learning equation (BLL) and MINERVA2 (see Section 4.2 and 7.2.1) and MP_U in two settings. The three algorithms are implemented either using collective or individual tag traces. Results indicate that the application of recommenders using collective tagging traces fosters semantic stabilisation in collaborative learning settings. This leads to more adequate annotation and deeper learning (Ley and Seitlinger, 2015). The consideration of frequency and semantic context as applied in MINERVA2 further contributes to the adequacy of tagging recommendations. In respect of individual learning, it reveals that the frequency and recency based approach (BLL) performs best.

Due to the sparseness of data in online settings, the experiment was restricted to the three algorithm setup. Nonetheless, results show that the cognitive-inspired algorithms can compete with MP that outperformed CF_U in three of four offline datasets.

8.2. Impact

With the increasing integration of technologies in education and the growing extent of educational content on the Web, information retrieval and filtering has become a crucial task in TEL systems. To support learners in finding relevant learning content, recommender systems have developed into one of the most prominent research strands in TEL (Drachsler, Verbert, et al., 2015). However, the modelling of learners and their contexts is a complex endeavour which still warrants extensive research. Thus, research questions addressed in this work are highly relevant, as they contribute to a deeper understanding of challenges concerning learner models and recommendation strategies in different learning and recommendation contexts.

More precisely, the impact of this dissertation research can be found in the areas of

i) **learning resource recommendation in TEL:** On the basis of differences between formal and informal learning settings, it presents recommendation approaches that address the specifics of the learning environment. In formal learning settings, where learning content and learning goals are typically known in advance, a CbKST-based approach is suggested and implemented as a prototypical implementation of Moodle plug-ins. The CbKST as a recommendation strategy is a continuation of work conducted by Albert, Nussbaumer, and C. Steiner (2008) and Nussbaumer, Hillemann, et al. (2015) who investigated the method to guide users by providing open learner models, consisting of competence structures and associated learning and assessment items. The conceptual approach of this recommendation strategy specifically addresses users that may be overwhelmed by the confrontation with concepts of competences and prerequisite relations between those competences. Moreover, the plug-ins created are open source and can be further used in adaptive Moodle courses.
For the setting of informal learning environments, the thesis explores SUSTAIN, a particularly flexible model of knowledge representation. The detailed offline analysis of the SUSTAIN model contributes to the understanding of interaction mechanisms between learners and learning content in virtual learning environments. Furthermore, SUSTAIN+CF_U gives a con-

crete example of how to combine theory-driven top down methods with data-driven bottom up strategies and thus, exploit knowledge gained from the individual and collective user behaviour.

ii) **data-driven resource and tag recommender evaluation in TEL:** First, an extensive study comparing different learning settings proves the insufficiency of standard resource recommenders for sparse data environments, thereby substantiating a common claim in the field. Poor results of content-based recommendation approaches highlight the need for more complete learning meta-data. Furthermore, by demonstrating the substantial differences between big-data learning environments and other typically sparse learning settings, it contributes to practices of TEL evaluation, which previously recommended using MOOC datasets for the evaluation of learning in TEL environments. Previous studies show the suitability of the BLL-based approach to predict the availability of words in memory (Kowald, Kopeinik, Seitlinger, et al., 2015). The tag recommender experiment corroborates these results on sparse TEL data sets. Findings also reveal the suitability of MP for recommending tags in computationally sensitive environments.

iii) **real-life tag recommender evaluation in IBL:** This thesis suggests tagging to create learning resource meta-data and to promote deeper learning. Furthermore, it recommends the application of cognitive-inspired tag recommendation algorithms to motivate and assist learners in the tagging process, and with this, reduce the sparsity of content description in learning data. This is expected to enhance the organization, finding and recommendation of learning content. Results of the conducted online study suggest the selection of algorithms for learning settings that either prioritise recommender accuracy, the development of a shared vocabulary (semantic stabilisation) or low computational cost. These findings i) contribute to the yet, poorly explored field of tag recommender systems in TEL and ii) inspire the design and development of future TEL environments.

8.3. Open Questions and Limitations

This Section reflects on the presented research contributions identifying weak spots and open questions.

8.3.1. Significance of Evaluation Results

The implemented evaluations either cover user satisfaction (RQ1) or performance measurements of resource and tag recommendation algorithms (RQ2 and RQ3). Other relevant indicators, such as task support, learning performance and learning motivation (Erdt, Fernandez, and Rensing, 2015) have not been investigated so far:

The CbKST-based approach (RQ1) has been evaluated in regard to its usability and perceived usefulness. Offline data evaluations are rather difficult to pursue, since they would require comprehensive TEL datasets that provide information about the learning domain and related assessment and learning items, as well as user's explicitly assessed knowledge states. Thus, it remains unknown whether the suggested approach presents recommendations accurate to a learner's needs and furthermore if it leads to improved learning performance.

The SUSTAIN-based recommender has been extensively studied on offline data (RQ2 and RQ3). However, online evaluations are less prone to error and misinterpretation. They provide direct user feedback in comparison to offline studies, where wrong predictions could be the result of a user's poor searching abilities. Thus, it remains to be seen whether the algorithm can provide additional benefits in cold-start and sparse data environments when applied in real-life learning environments.

The real-life tag recommendation study (RQ3) was restricted to a three algorithm setup, due to the sparsity of evaluation data in the school setting. An online comparison with other state-of-the-art algorithms is still pending. Furthermore, clear evidence of the tagging process itself or the degree of semantic stabilisation influencing the quality of the learning process would significantly increase the impact of this work.

8.3.2. Applying the CbKST

The CbKST is theoretically well-founded and shows potential to model closed learning domains well. By shifting the responsibility of ontology building to an acknowledged domain expert (e.g., teacher), well-suited and precise models can be developed. Nevertheless, the application of the approach revealed some issues: i) the high cost of domain modelling and its dependence on domain experts, which may lead to bottlenecks, ii) the dependence of the quality of the domain model on the competences and the motivation of the involved expert, and iii) the static nature of ontologies, which makes an application in informal learning environments difficult. Consequently, the challenge remains to provide ontology-based approaches that can be derived by automatic domain modelling services.

8.3.3. Development of SUSTAIN

The SUSTAIN model has been investigated in RQ2. In its current implementation, the recommendation model is applied on a pre-selected resource candidate set, which is obtained by CF_U. On the one hand, this causes additional computational expense for calculating CF_U, and on the other hand it entails CF specific issues such as the exclusion of new or diverse items. Furthermore, in respect to the learner model, the clustering algorithm bears potential for a stronger consideration of the aspect of time. When the learner engages with a new learning resource, this updates only the cluster that accommodates the resource in the learner model and leaves other clusters of learning categories unchanged. However, it can be expected that topics a learner explored more recently are of greater importance to current and future learning episodes. This is not depicted in the model so far. Therefore, it could be relevant to investigate how the factor time can be incorporated in the SUSTAIN model.

9. Future Work

Based on limitations that have been identified, this Chapter outlines directions of future work.

9.1. Evaluation Studies

To improve the understanding and impact of here proposed models, the implementation of extensive user studies to test the effects in real-life settings is warranted. A four-step evaluation procedure as suggested by Manouselis et al., 2013 would be an ideal approach. However, the evaluation of TEL recommender systems and its educational value is a very demanding task and experiments show that objective standard measurements such as knowledge gain hardly lead to significant results (Manouselis, Drachsler, Vuorikari, et al., 2011). Also, the challenge remains to gain access to learning environments, with frequent learners who are willing to participate in such studies.

Yet, further evaluation studies are planned. The first one will take place as part of a one-week Erasmus+ Seminar[1]. The plug-ins presented in Section 6.1 will be integrated in a Moodle training course, which imparts knowledge about the CbKST to attending PhD students. The evaluation will consider the user's perceived usefulness of the system, their learning performance and their motivation. Because the one week seminar setting supports the monitoring of learners over a reasonable amount of time, more significant results are expected.

Moreover, additional studies in informal or open social learning environments are planned i) to investigate the potential of SUSTAIN in real-life settings and ii) to re-enact the tag recommendation study with a broader variety of algorithms. Ideally, these experiments would include pre- and post- knowledge assessments

[1] https://tquant.eu/

that may lead further to insights on the learning process.

9.2. Development of SUSTAIN

First, to realize a more dynamic recommendation logic in terms of novelty or diversity, the network's sensitivity towards a user's mental state (i.e. curiosity) could be explored. In particular, based on creative cognition research (e.g., Finke, Ward, and S. M. Smith (1992)) and in line with findings of the evaluation studies in Kopeinik, Kowald, Hasani-Mavriqi, et al. (2016), it can be assumed that a broader attentional focus (i.e., higher curiosity) is associated with a stronger orientation toward novel or more diverse resources. If the algorithm integrates this association, depending on the user model, recommendations should become either more accurate or diverse. Second, further studies could help find a variant of the model that is independent of a resource candidate set obtained by CF_U and thus, searches for learner-specific recommendations only by means of the correspondingly trained SUSTAIN network. For instance by integrating the factor time into the clustering approach, the algorithm could concentrate on the identification of resources that relate only to most recent learning episodes.

9.3. Exploring Other Approaches

An alternative approach for domain modelling is the Formal Concept Analysis (FCA) (Wille, 2005), which hierarchically structures a domain according to concepts consisting of objects, attributes and binary relation between them. The hierarchy provided by the FCA is in line with theoretical assumptions from information retrieval research (Cole, Kennedy, and Carter, 1996) that describe the information need of undergraduate students as rather generic in the primary phase of engagement with a topic growing to more specific demands as their knowledge representation of the field evolves. The FCA-based domain and open learner model has already been integrated in the weSPOT IBL environment and is currently used in a school in Graz. Concept and design of the approach and how it can be used for learning recommendations is presented in Bedek et al. (2015). Moreover, the FCA has been investigated for automatic ontology construction in the past (e.g., Cimiano, Hotho, and Staab (2005), Cho, Richards, et al. (2006)). An ambitious endeavour would therefore be constituted in the combination of data-driven methodologies to create an FCA-based ontology, with the theory-driven generation of learning recommendations based on the resulting domain model. This would cater to the needs of formal learning environments, which typically operate with a restricted number of users, i.e. on sparse data.

9.4. Closing Remarks

Recommender systems have not yet made the breakthrough in TEL environments, but it remains one of the most prominent research areas in the field (Herder, Sosnovsky, and Dimitrova, 2017). I would like to conclude this work with the thought that the translation of psychologically plausible theoretical models into technology, bears advantages for both technical and conceptual aspects of recommender system research. It is my contention that a deeper understanding of the mechanisms of human memory in information retrieval and learning will lead to practical insights, various technologies can benefit from. This thesis contributes to this goal. It shows that the application of computational models of human cognition holds advantage for the design of recommender mechanisms and, at the same time, for gaining a deeper understanding of interaction dynamics in virtual learning systems.

Bibliography

Abbas, Assad, Limin Zhang, and Samee U. Khan (2015). "A survey on context-aware recommender systems based on computational intelligence techniques". In: *Computing* 97.7, pp. 667–690 (cit. on p. 22).

Adomavicius, Gediminas and Alexander Tuzhilin (2005). "Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions". In: *IEEE Transactions on Knowledge and Data Engineering* 17.6, pp. 734–749 (cit. on pp. 22, 23).

Albert, Dietrich, Alexander Nussbaumer, and Christina Steiner (2008). "Using visual guidance and feedback based on competence structures for personalising e-learning experience". In: *Proceedings of the 16th International Conference on Computers in Education (ICCE 2008)*, pp. 27–31 (cit. on pp. 86, 158).

Aleven, Vincent, Bruce Mclaren, Ido Roll, and Kenneth Koedinger (2006). "Toward meta-cognitive tutoring: A model of help seeking with a Cognitive Tutor". In: *International Journal of Artificial Intelligence in Education* 16.2, pp. 101–128 (cit. on p. 34).

Anderson, John R., Daniel Bothell, Michael D. Byrne, Scott Douglass, Christian Lebiere, and Yulin Qin (2004). "An integrated theory of the mind." In: *Psychological review* 111.4, p. 1036 (cit. on pp. 6, 49, 56, 57).

Anderson, John R, C Franklin Boyle, Albert T Corbett, and Matthew W Lewis (1990). "Cognitive modeling and intelligent tutoring". In: *Artificial intelligence* 42.1, pp. 7–49 (cit. on p. 34).

Anderson, John R. and Lael J. Schooler (1991). "Reflections of the environment in memory". In: *Psychological science* 2.6, pp. 396–408 (cit. on pp. 57–59, 133).

Anjorin, Mojisola, Ivan Dackiewicz, Alejandro Fernández, Christoph Rensing, et al. (2012). "A framework for cross-platform graph-based recommendations for TEL". In: *Proceedings of the 2nd workshop on recommender systems in technology enhanced learning*, pp. 83–88 (cit. on p. 1).

Augustin, Thomas, Cord Hockemeyer, Michael D Kickmeier-Rust, Patrick Podbregar, Reinhard Suck, and Dietrich Albert (2013). "The simplified updating rule in the formalization of digital educational games". In: *Journal of Computational Science* 4.4, pp. 293–303 (cit. on pp. 4, 54, 82).

Aydin, Cansu Cigdem and Guzin Tirkes (2010). "Open source learning management systems in e-learning and Moodle". In: *Education engineering (EDUCON)*, pp. 593–600 (cit. on p. 86).

Balby Marinho, L., A. Hotho, R. Jõschke, A. Nanopoulos, S. Rendle, L. Schmidt-Thieme, G. Stumme, and P. Symeonidis (2012). *Recommender Systems for Social Tagging Systems*. SpringerBriefs in Electrical and Computer Engineering. Springer. ISBN: 978-1-4614-1893-1. DOI: 10.1007/978-1-4614-1894-8 (cit. on p. 20).

Bar, Ariel, Lior Rokach, Guy Shani, Bracha Shapira, and Alon Schclar (2013). "Improving simple collaborative filtering models using ensemble methods". In: *Multiple Classifier Systems*. Springer, pp. 1–12 (cit. on p. 98).

Basilico, Justin and Thomas Hofmann (2004). "Unifying collaborative and content-based filtering". In: *Proceedings of the twenty-first international conference on Machine learning*. ACM, p. 9 (cit. on p. 74).

Bateman, Scott, Christopher Brooks, Gordon Mccalla, and Peter Brusilovsky (2007). "Applying collaborative tagging to e-learning". In: *Proceedings of the 16th international World Wide Web conference (WWW2007)* (cit. on p. 41).

Bechtel, William, George Graham, and David A Balota (1998). *A companion to cognitive science*. Blackwell Oxford (cit. on p. 49).

Bedek, Michael, Simone Kopeinik, Bernd Prünster, and Dietrich Albert (2015). "Applying the Formal Concept Analysis to Introduce Guidance in an Inquiry-based Learning Environment". In: (cit. on pp. 134, 137, 164).

Beham, Guenter, Hermann Stern, and Stefanie Lindstaedt (2010). "APOSDLE-DS A Dataset from the APOSDLE Workintegrated Learning System". In: *1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)* (cit. on p. 70).

Benchmark Folksonomy Data from BibSonomy (2013/2015). Available from http://www.kde.cs.uni-kassel.de/bibsonomy/dumps. Knowledge and Data Engineering Group, University of Kassel (cit. on p. 68).

Biswas, Gautam, Hogyeong Jeong, John S Kinnebrew, Brian Sulcer, and ROD ROSCOE (2010). "Measuring self-regulated learning skills through social interactions in a teachable agent environment". In: *Research and Practice in Technology Enhanced Learning* 5.02, pp. 123–152 (cit. on p. 34).

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003). "Latent dirichlet allocation". In: *The Journal of machine Learning research* 3, pp. 993–1022 (cit. on pp. 41, 66).

Bloom, Benjamin S, Max D Engelhart, Edward J Furst, Walker H Hill, and David R Krathwohl (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. Vol. 19. New York: David McKay Co Inc (cit. on p. 33).

Bobadilla, Jesús, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez (2013). "Recommender systems survey". In: *Knowledge-based systems* 46, pp. 109–132 (cit. on p. 17).

Bobadilla, Jesus, Francisco Serradilla, Antonio Hernando, et al. (2009). "Collaborative filtering adapted to recommender systems of e-learning". In: *Knowledge-Based Systems* 22.4, pp. 261–265 (cit. on p. 39).

Bourkoukou, Outmane and Essaid El Bachari (2016). "E-LEARNING PERSONALIZATION BASED ON COLLABORATIVE FILTERING AND LEARNER'S PREFERENCE". In: *Journal of Engineering Science and Technology* 11.11, pp. 1565–1581 (cit. on p. 40).

Breese, John S., David Heckerman, and Carl Kadie (1998). "Empirical analysis of predictive algorithms for collaborative filtering". In: *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 43–52 (cit. on pp. 19, 24).

Brusilovsky, Peter and Eva Millán (2007). "User models for adaptive hypermedia and adaptive educational systems". In: *The adaptive web*. Springer-Verlag, pp. 3–53 (cit. on pp. 32, 33, 35, 80).

Brusilovsky, Peter and Carlo Tasso (2004). "Preface to special issue on user modeling for web information retrieval". In: *User Modeling and User-Adapted Interaction* 14.2, pp. 147–157 (cit. on p. 37).

Buder, Jürgen and Christina Schwind (2012). "Learning with personalized recommender systems: A psychological view". In: *Computers in Human Behavior* 28.1, pp. 207–216 (cit. on pp. 2, 29).

Burke, Robin (2007). "Hybrid web recommender systems". In: *The adaptive web*. Springer, pp. 377–408 (cit. on p. 19).

Campos, Pedro G., Fernando Diez, and Ivan Cantador (2013). "Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols". In: *User Modeling and User-Adapted Interaction*, pp. 1–53 (cit. on p. 65).

Cantador, Iván, Peter Brusilovsky, and Tsvi Kuflik (2011). "2nd Workshop on Information Heterogeneity and Fusion in Recommender Systems (HetRec 2011)". In: *Proceedings of the fifth ACM conference on Recommender Systems*. RecSys 2011. Chicago, IL, USA: ACM (cit. on p. 112).

Carr, Brian and Ira P. Goldstein (1977). *Overlays: A theory of modelling for computer aided instruction*. Tech. rep. MASSACHUSETTS INST OF TECH CAMBRIDGE ARTIFICIAL INTELLIGENCE LAB (cit. on p. 35).

Cazella, Sílvio, Eliseo Reategui, and Patrícia Behar (2010). "Recommendation of learning objects applying collaborative filtering and competencies". In: *Key Competencies in the Knowledge Society*, pp. 35–43 (cit. on p. 40).

Chang, Ting-Wen, Jeffrey Kurcz, Moushir M El-Bishouty, Sabine Graf, et al. (2015). "Adaptive and personalized learning based on students' cognitive characteristics". In: *Ubiquitous Learning Environments and Technologies*. Springer, pp. 77–97 (cit. on p. 34).

Cho, Woo Chul, Debbie Richards, et al. (2006). "Automatic construction of a concept hierarchy to assist Web document classification". In: *Proceedings of 2nd International Conference on Information Management and Business (IMB. 2006)*, pp. 13–16 (cit. on p. 165).

Cimiano, Philipp, Andreas Hotho, and Steffen Staab (2005). "Learning concept hierarchies from text corpora using formal concept analysis." In: *Journal of Artificial Intelligence Research (JAIR)* 24.1, pp. 305–339 (cit. on p. 165).

Colardyn, Danielle and Jens Bjornavold (2004). "Validation of formal, non-formal and informal learning: Policy and practices in EU member states". In: *European journal of education* 39.1, pp. 69–89 (cit. on pp. 2, 3, 30, 154).

Cole, Charles, Lynn Kennedy, and Susan Carter (1996). "The optimization of online searches through the labelling of a dynamic, situation-dependent information need: The reference interview and online searching for undergraduates doing a social-science assignment". In: *Information processing & management* 32.6, pp. 709–717 (cit. on p. 164).

Corbett, Albert T. and John R. Anderson (1994). "Knowledge tracing: Modeling the acquisition of procedural knowledge". In: *User modeling and user-adapted interaction* 4.4, pp. 253–278 (cit. on p. 34).

Csikszentmihalyi, Mihaly (2014). "Toward a psychology of optimal experience". In: *Flow and the foundations of positive psychology*. Springer, pp. 209–226 (cit. on p. 154).

Davidson, James, Benjamin Liebald, Junning Liu, Palash Nandy, Taylor Van Vleet, Ullas Gargi, Sujoy Gupta, Yu He, Mike Lambert, Blake Livingston, et al. (2010). "The YouTube video recommendation system". In: *Proceedings of the fourth ACM conference on Recommender systems*. ACM, pp. 293–296 (cit. on p. 1).

De Bra, Paul, Ad Aerts, Bart Berden, Barend De Lange, Brendan Rousseau, Tomi Santic, David Smits, and Natalia Stash (2003). "AHA! The adaptive hypermedia architecture". In: *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*. ACM, pp. 81–84 (cit. on p. 36).

Dellschaft, Klaas and Steffen Staab (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*. Milwaukee, Wisconsin, USA: ACM, pp. 73–82. ISBN: 978-1-4503-1335-3. DOI: 10.1145/2309996.2310009 (cit. on p. 20).

Denaux, Ronald, Vania Dimitrova, and Lora Aroyo (2004). "Interactive ontology-based user modeling for personalized learning content management". In: *Proceedings of AH'04 Semantic Web for E-Learning Workshop* (cit. on pp. 31, 36).

Desmarais, Michel C. and Ryan S. Baker (2012). "A review of recent advances in learner and skill modeling in intelligent learning environments". In: *User Modeling and User-Adapted Interaction* 22.1-2, pp. 9–38 (cit. on p. 34).

Despotović-Zrakić, Marijana, Aleksandar Marković, Zorica Bogdanović, Dušan Barać, and Srdjan Krčo (2012). "Providing adaptivity in Moodle LMS courses". In: *Educational Technology & Society* 15.1, pp. 326–338 (cit. on p. 86).

Diaz-Aviles, Ernesto, Marco Fisichella, Ricardo Kawase, Wolfgang Nejdl, and Avaré Stewart (2011). "Unsupervised auto-tagging for learning object enrichment". In: *Towards Ubiquitous Learning*. Springer, pp. 83–96 (cit. on p. 41).

Dimache, Aurora, Simone Kopeinik, Attracta Brennan, Thomas Roche, Lisa C. Winter, and Dietrich Albert (2014). "Innovative online vocational training of renewable energy technologies (INNOVRET)". In: *International Journal of Information and Education Technology* 4.1, p. 127 (cit. on pp. 12, 80).

Dimache, Aurora, Thomas Roche, Simone Kopeinik, Lisa C. Winter, Alexander Nussbaumer, and Dietrich Albert (2015). "Suitability of Adaptive Self-Regulated e-Learning to Vocational Training: A Pilot Study in Heat Pump System Installation". In: *International Journal of Online Pedagogy and Course Design (IJOPCD)* 5.3, pp. 31–46 (cit. on pp. 12, 80).

Doignon, Jean-Paul (1994). "Probabilistic assessment of knowledge". In: *Knowledge structures*. Springer, pp. 1–57 (cit. on pp. 51, 52, 54).

Doignon, Jean-Paul and Jean-Claude Falmagne (1985). "Spaces for the assessment of knowledge". In: *International journal of man-machine studies* 23.2, pp. 175–196 (cit. on p. 50).

Drachsler, Hendrik, Hans G. K. Hummel, and Rob Koper (2009). "Identifying the goal, user model and conditions of recommender systems for formal and informal learning". In: *Journal of Digital Information* 10.2 (cit. on pp. 1–4, 28–32, 39, 41, 42, 44, 45, 79, 155).

Drachsler, Hendrik, Hans GK Hummel, and Rob Koper (2008). "Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model". In: *International Journal of Learning Technology* 3.4, pp. 404–423 (cit. on p. 28).

Drachsler, Hendrik, Hans Hummel, and Rob Koper (2007). "Recommendations for learners are different: Applying memory-based recommender system techniques to lifelong learning". In: (cit. on pp. 1, 4, 28, 29, 35, 43).

Drachsler, Hendrik, Katrien Verbert, Olga C. Santos, and Nikos Manouselis (2015). "Panorama of recommender systems to support learning". In: *Recommender systems handbook*. Springer, pp. 421–451 (cit. on pp. 2, 37, 38, 43, 158).

Drasgow, Fritz and Charles L. Hulin (1990). "Item response theory". In: *Handbook of industrial and organizational psychology* 1, pp. 577–636 (cit. on p. 34).

Duval, Erik (2011). "Attention please!: learning analytics for visualization and recommendation". In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. ACM, pp. 9–17 (cit. on p. 1).

Embarak, Ossama H (2011). "A method for solving the cold start problem in recommendation systems". In: *Innovations in Information Technology (IIT), 2011 International Conference on*. IEEE, pp. 238–243 (cit. on p. 36).

Erdt, Mojisola, Alejandro Fernandez, and Christoph Rensing (2015). "Evaluating Recommender Systems for Technology Enhanced Learning: A Quantitative Survey". In: *Learning Technologies, IEEE Transactions on* 8.4, pp. 326–344 (cit. on pp. 18, 27, 42, 128, 160).

Erdt, Mojisola and Christoph Rensing (2014). "Evaluating recommender algorithms for learning using crowdsourcing". In: *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*. IEEE, pp. 513–517 (cit. on p. 43).

Falmagne, Jean-Claude, Dietrich Albert, Christopher Doble, David Eppstein, and Xiangen Hu (2013). *Knowledge spaces: Applications in education*. Springer Science & Business Media (cit. on pp. 50, 51, 81).

Falmagne, Jean-Claude and Jean-Paul Doignon (1999). *Knowledge spaces*. Springer (cit. on pp. 34, 38, 50, 51).

Fazeli, Soude, Babak Loni, Hendrik Drachsler, and Peter Sloep (2014). "Which recommender system can best fit social learning platforms?" In: *Open Learning and Teaching in Educational Communities*. Springer, pp. 84–97 (cit. on pp. 39, 124, 156).

Finke, Ronald A., Thomas B. Ward, and Steven M. Smith (1992). "Creative cognition: Theory, research, and applications". In: (cit. on p. 164).

Font, Frederic, Joan Serrà, and Xavier Serra (2015). "Analysis of the Impact of a Tag Recommendation System in a Real-World Folksonomy". In: *ACM Trans. Intell. Syst. Technol.* 7.1, 6:1–6:27. ISSN: 2157-6904. DOI: `10.1145/2743026`. URL: `http://doi.acm.org/10.1145/2743026` (cit. on pp. 41, 138).

Friedrich, Martin, Katja Niemann, Maren Scheffel, Hans-Christian Schmitz, and Martin Wolpers (2007). "Object Recommendation based on Usage Context". In: *Educational Technology & Society* 10.3, pp. 106–121 (cit. on p. 75).

Ge, Mouzhi, Carla Delgado-Battenfeld, and Dietmar Jannach (2010). "Beyond accuracy: evaluating recommender systems by coverage and serendipity". In: *Proceedings of the fourth ACM conference on Recommender systems*. ACM, pp. 257–260 (cit. on pp. 24, 25).

Gemmell, Jonathan, Thomas Schimoler, Maryam Ramezani, Laura Christiansen, and Bamshad Mobasher (2009). "Improving folkrank with item-based collaborative filtering". In: *Recommender Systems & the Social Web* (cit. on pp. 67, 74, 105).

Ghahramani, Zoubin (2001). "An introduction to hidden Markov models and Bayesian networks". In: *International journal of pattern recognition and artificial intelligence* 15.01, pp. 9–42 (cit. on p. 34).

Goldberg, David, David Nichols, Brian M. Oki, and Douglas Terry (1992). "Using collaborative filtering to weave an information tapestry". In: *Communications of the ACM* 35.12, pp. 61–70 (cit. on p. 17).

Goldstein, Ira P. (1979). "The genetic graph: a representation for the evolution of procedural knowledge". In: *International Journal of Man-Machine Studies* 11.1, pp. 51–77 (cit. on p. 35).

Gomez-Uribe, Carlos A. and Neil Hunt (2016). "The netflix recommender system: Algorithms, business value, and innovation". In: *ACM Transactions on Management Information Systems (TMIS)* 6.4, p. 13 (cit. on p. 1).

Goodwin, Gregory A (2017). "Learner Models in the Generalized Intelligent Framework for Tutoring: Current Work and Future Directions". In: *Proceedings of the 5th Annual Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym5)*. Robert Sottilare, p. 89 (cit. on pp. 33, 34, 45).

Griffiths, Thomas L., Mark Steyvers, Joshua B. Tenenbaum, et al. (2007). "Topics in semantic representation". In: *Psychological review* 114.2, p. 211 (cit. on pp. 5, 101).

Hale, Gordon A (1977). "On use of ANOVA in developmental research". In: *Child Development*, pp. 1101–1106 (cit. on p. 145).

Hämäläinen, Wilhelmiina and Mikko Vinni (2010). "Classifiers for educational data mining". In: *Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, pp. 57–71 (cit. on p. 32).

Helic, Denis, Christian Körner, Michael Granitzer, Markus Strohmaier, and Christoph Trattner (2012). "Navigational efficiency of broad vs. narrow folksonomies". In: *Proceedings of the 23rd ACM conference on Hypertext and social media*. ACM, pp. 63–72 (cit. on p. 21).

Heller, Jürgen, Christina Steiner, Cord Hockemeyer, and Albert Dietrich (2006). "Competence-based knowledge structures for personalised learning". In: *International Journal on ELearning* 5.1, p. 75 (cit. on pp. 4, 49, 51–54).

Herder, Eelco, Sergey Sosnovsky, and Vania Dimitrova (2017). "Adaptive Intelligent Learning Environments". In: *Technology Enhanced Learning*. Springer, pp. 109–114 (cit. on pp. 34, 35, 44, 165).

Herlocker, Jonathan L., Joseph A. Konstan, Loren G. Terveen, and John T. Riedl (2004). "Evaluating collaborative filtering recommender systems". In: *ACM Trans. Inf. Syst.* 22.1, pp. 5–53. ISSN: 1046-8188. DOI: 10.1145/963770.963772 (cit. on pp. 25, 66, 107).

Hetzner, Sonia, Christina M. Steiner, Vania Dimitrova, Paul Brna, and Owen Conlan (2011). "Adult self-regulated learning through linking experience in simulated and real world: a holistic approach". In: *European Conference on Technology Enhanced Learning*. Springer, pp. 166–180 (cit. on p. 82).

Heymann, Paul, Daniel Ramage, and Hector Garcia-Molina (2008). "Social tag prediction". In: *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, pp. 531–538 (cit. on pp. 20, 40).

Hidalgo, J. Ignacio, Esther Maqueda, José L Risco-Mart'ın, Alfredo Cuesta-Infante, J. Manuel Colmenar, and Javier Nobel (2014). "glUCModel: A monitoring and modeling system for chronic diseases applied to diabetes". In: *Journal of biomedical informatics* 48, pp. 183–192 (cit. on p. 18).

Hintzman, Douglas L. (1984). "MINERVA 2: A simulation model of human memory". In: *Behavior Research Methods, Instruments, & Computers* 16.2, pp. 96–101 (cit. on pp. 6, 49, 59, 60, 133).

Hockemeyer, Cord, Owen Conlan, Vincent P. Wade, and Dietrich Albert (2003). "Applying competence prerequisite structures for eLearning and skill management". In: *Journal of Universal Computer Science* 9.12, pp. 1428–1436 (cit. on p. 52).

Hockemeyer, Cord, Theo Held, and Dietrich Albert (1997). "RATH-A relational adaptive tutoring hypertext WWW-environment based on knowledge space theory". In: (cit. on p. 39).

Howe, Jeff (2006). "The rise of crowdsourcing". In: *Wired magazine* 14.6, pp. 1–4 (cit. on p. 43).

Hu, Yifan, Yehuda Koren, and Chris Volinsky (2008). "Collaborative filtering for implicit feedback datasets". In: *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, pp. 263–272 (cit. on pp. 74, 98).

Jannach, Dietmar, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich (2010). *Recommender systems: an introduction*. Cambridge University Press (cit. on pp. 17, 18, 21, 23, 24, 43).

Jäschke, Robert, Leandro Marinho, Andreas Hotho, Lars Schmidt-Thieme, and Gerd Stumme (2007). "Tag recommendations in folksonomies". In: *Knowledge Discovery in Databases: PKDD 2007*, pp. 506–514 (cit. on pp. 73, 132).

Jäschke, Robert, Leandro Marinho, Andreas Hotho, Lars Schmidt-Thieme, and Gerd Stumme (2008). "Tag recommendations in social bookmarking systems". In: *Ai Communications* 21.4, pp. 231–247 (cit. on p. 20).

Keeney, Ralph L. and Howard Raiffa (1993). *Decisions with multiple objectives: preferences and value trade-offs*. Cambridge university press (cit. on pp. 2, 28).

Khribi, Mohamed Koutheaïr, Mohamed Jemni, and Olfa Nasraoui (2015). "Recommendation Systems for Personalized Technology-Enhanced Learning". In: *Ubiquitous Learning Environments and Technologies*. Springer, pp. 159–180 (cit. on pp. 2, 41, 44).

Kickmeier-Rust, Michael D., Dietrich Albert, Cord Hockemeyer, and Thomas Augustin (2007). "Not breaking the narrative: Individualized competence assessment in educational games". In: *Proceedings of the European Conference on Games-based Learning (ECGBL)*, pp. 25–26 (cit. on p. 3).

Kintsch, Walter and Praful Mangalath (2011). "The construction of meaning". In: *Topics in Cognitive Science* 3.2, pp. 346–370 (cit. on pp. 66, 106).

Klašnja-Milićević, Aleksandra, Mirjana Ivanović, and Alexandros Nanopoulos (2015). "Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions". In: *Artificial Intelligence Review* 44.4, pp. 571–604 (cit. on pp. 2, 38, 41).

Kohavi, Ron, Alex Deng, Roger Longbotham, and Ya Xu (2014). "Seven rules of thumb for web site experimenters". In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 1857–1866 (cit. on p. 72).

Kohavi, Ronand and Roger Longbotham (2016). "Online Controlled Experiments and A/B Testing". In: *Encyclopedia of Machine Learning and Data Mining*. Boston, MA: Springer US, pp. 1–8 (cit. on p. 72).

"Introduction to Recommender Systems: Algorithms and Evaluation" (2004). In: *ACM Trans. Inf. Syst.* 22.1. Ed. by Joseph A. Konstan, pp. 1–4. ISSN: 1046-8188. DOI: 10.1145/963770.963771. URL: http://doi.acm.org/10.1145/963770.963771 (cit. on pp. 1, 17).

Kopeinik, Simone, Michael Bedek, Olga Firssova, Jürgen Mack, and Dietrich Albert (2015). "INTRODUCING TECHNOLOGY-ENHANCED INQUIRY-BASED LEARNING TO SUPPORT SCIENCE EDUCATION IN SECONDARY SCHOOLS: A TEACHER PERSPECTIVE". In: *EDULEARN15 Proceedings*. 7th International Conference on Education and New Learning Technologies. Barcelona, Spain: IATED, pp. 6035–6045. ISBN: 978-84-606-8243-1 (cit. on pp. 12, 132).

Kopeinik, Simone, Dominik Kowald, Ilire Hasani-Mavriqi, and Elisabeth Lex (2016). "Improving Collaborative Filtering Using a Cognitive Model of Human Category Learning". In: *The Journal of Web Science* 2.1 (cit. on pp. 13, 65, 99, 125, 155, 164).

Kopeinik, Simone, Dominik Kowald, and Elisabeth Lex (2016). "Which Algorithms Suit Which Learning Environments? A Comparative Study of Recommender Systems in TEL". In: *European Conference on Technology Enhanced Learning*. Springer, pp. 124–138 (cit. on pp. 65, 124, 132, 133, 139, 145).

Kopeinik, Simone, Elisabeth Lex, Paul Seitlinger, Dietrich Albert, and Tobias Ley (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. ACM*, pp. 409–418 (cit. on pp. 12, 13, 132).

Kopeinik, Simone, Alexander Nussbaumer, Michael Bedek, and Dietrich Albert (2012). "Using CbKST for Learning Path Recommendation in Game-based Learning". In: *Proceedings of the 20th International Conference on Computers in Education*, pp. 26–30 (cit. on pp. 39, 81).

Kopeinik, Simone, Alexander Nussbaumer, Lisa C. Winter, Dietrich Albert, Aurora Dimache, and Thomas Roche (2014). "Combining self-regulation and competence-based guidance to personalise the learning experience in moodle". In: *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on*. IEEE, pp. 62–64 (cit. on pp. 12, 80).

Körner, Christian, Dominik Benz, Andreas Hotho, Markus Strohmaier, and Gerd Stumme (2010). "Stop thinking, start tagging: tag semantics emerge from collaborative verbosity". In: *Proceedings of the 19th international conference on World Wide Web*. WWW '10. Raleigh, North Carolina, USA: ACM, pp. 521–530. ISBN: 978-1-60558-799-8. DOI: 10.1145/1772690.1772744 (cit. on p. 20).

Korossy, Klaus (1997). "Extending the theory of knowledge spaces: A competence-performance approach". In: *Zeitschrift fur Psychologie* 205.1, pp. 53–82 (cit. on pp. 37, 38, 50, 53).

Korossy, Klaus (1999). "Modeling knowledge as competence and performance". In: pp. 103–132 (cit. on p. 54).

Kowald, Dominik, Simone Kopeinik, and Elisabeth Lex (2017). "The TagRec Framework as a Toolkit for the Development of Tag-Based Recommender Systems". In: (cit. on pp. 104, 125).

Kowald, Dominik, Simone Kopeinik, Paul Seitlinger, Tobias Ley, Dietrich Albert, and Christoph Trattner (2015). "Refining Frequency-Based Tag Reuse Predictions by Means of Time and Semantic Context". In: *Mining, Modeling, and Recommending'Things' in Social Media*. Springer, pp. 55–74 (cit. on pp. 13, 59, 125, 159).

Kowald, Dominik and Elisabeth Lex (2015). "Evaluating Tag Recommender Algorithms in Real-World Folksonomies: A Comparative Study". In: *Proceedings of the 9th ACM Conference on Recommender Systems*. ACM, pp. 265–268 (cit. on pp. 65, 66, 124, 129).

Kowald, Dominik, Paul Seitlinger, Simone Kopeinik, Tobias Ley, and Christoph Trattner (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*. Springer, pp. 75–95 (cit. on p. 133).

Kowald, Dominik, Paul Seitlinger, Christoph Trattner, and Tobias Ley (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*. New York, NY, USA: ACM (cit. on pp. 49, 57, 58).

Krestel, Ralf, Peter Fankhauser, and Wolfgang Nejdl (2009). "Latent dirichlet allocation for tag recommendation". In: *Proceedings of the third ACM conference on Recommender systems*. ACM, pp. 61–68 (cit. on p. 66).

Kuhn, Alex, Brenna McNally, Shannon Schmoll, Clara Cahill, Wan-Tzu Lo, Chris Quintana, and Ibrahim Delen (2012). "How students find, evaluate and utilize peer-collected annotated multimedia data in science inquiry with zydeco". In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, pp. 3061–3070 (cit. on pp. 6, 40, 147).

Ley, Tobias, Barbara Kump, and Cornelia Gerdenitsch (2010). "Scaffolding self-directed learning with personalized learning goal recommendations". In: *User modeling, adaptation, and personalization*, pp. 75–86 (cit. on p. 3).

Ley, Tobias and Paul Seitlinger (2015). "Dynamics of human categorization in a collaborative tagging system: How social processes of semantic stabilization shape individual sensemaking". In: *Computers in human behavior* 51, pp. 140–151 (cit. on pp. 6, 40, 157).

Ley, Tobias, Armin Ulbrich, Peter Scheir, Stefanie Lindstaedt, Barbara Kump, and Dietrich Albert (2008). "Modeling competencies for supporting work-integrated learning in knowledge work". In: *Journal of knowledge management* 12.6, pp. 31–47 (cit. on p. 37).

Lin, Nan, Daifeng Li, Ying Ding, Bing He, Zheng Qin, Jie Tang, Juanzi Li, and Tianxi Dong (2012). "The dynamic features of Delicious, Flickr, and YouTube". In: *Journal of the American Society for Information Science and Technology* 63.1, pp. 139–162 (cit. on pp. 21, 140).

Linden, Greg, Brent Smith, and Jeremy York (2003). "Amazon. com recommendations: Item-to-item collaborative filtering". In: *IEEE Internet computing* 7.1, pp. 76–80 (cit. on p. 1).

Lindstaedt, Stefanie, Günter Beham, Barbara Kump, and Tobias Ley (2009). "Getting to know your user–Unobtrusive user model maintenance within work-integrated learning environments". In: *Learning in the synergy of multiple disciplines*, pp. 73–87 (cit. on pp. 36, 37).

Lindstaedt, Stefanie, Peter Scheir, Robert Lokaiczyk, Barbara Kump, Günter Beham, and Viktoria Pammer (2008). "Knowledge services for work-integrated learning". In: *Times of Convergence. Technologies Across Learning Contexts*, pp. 234–244 (cit. on pp. 28, 39).

Ling, Charles X and Chenghui Li (1998). "Data mining for direct marketing: Problems and solutions." In: *KDD*. Vol. 98, pp. 73–79 (cit. on p. 24).

Lohmann, Steffen, Stefan Thalmann, Andreas Harrer, and Ronald Maier (2007). "Learner-Generated Annotation of Learning Resources–Lessons from Experiments on Tagging". In: *Journal of Universal Computer Science* 304, p. 312 (cit. on p. 41).

Love, Bradley C. and Douglas L. Medin (1998). "SUSTAIN: A model of human category learning". In: *AAAI/IAAI*, pp. 671–676 (cit. on pp. 49, 54).

Love, Bradley C., Douglas L. Medin, and Todd M. Gureckis (2004). "SUSTAIN: a network model of category learning." In: *Psychological review* 111.2, p. 309 (cit. on pp. 4, 54, 99, 101, 102, 107, 108, 114, 121).

Lu, Jie, Dianshuang Wu, Mingsong Mao, Wei Wang, and Guangquan Zhang (2015). "Recommender system application developments: a survey". In: *Decision Support Systems* 74, pp. 12–32 (cit. on pp. 17, 19).

Macgregor, George and Emma McCulloch (2006). "Collaborative tagging as a knowledge organisation and resource discovery tool". In: *Library review* 55.5, pp. 291–300 (cit. on p. 21).

Manouselis, Nikos, Hendrik Drachsler, Katrien Verbert, and Erik Duval (2012). *Recommender systems for learning*. Springer Science & Business Media (cit. on pp. 37, 38).

Manouselis, Nikos, Hendrik Drachsler, Katrien Verbert, and Erik Duval (2013). "Challenges and Outlook". In: *Recommender Systems for Learning*. Springer, pp. 63–76 (cit. on p. 163).

Manouselis, Nikos, Hendrik Drachsler, Riina Vuorikari, Hans Hummel, and Rob Koper (2011). "Recommender systems in technology enhanced learning". In: *Recommender systems handbook*. Springer, pp. 387–415 (cit. on pp. 27, 28, 37, 42, 44, 163).

Manouselis, Nikos, Riina Vuorikari, and Frans Van Assche (2010). "Collaborative recommendation of e-learning resources: an experimental investigation". In: *Journal of Computer Assisted Learning* 26.4, pp. 227–242 (cit. on pp. 124, 129, 156).

Marinho, Leandro Balby, Andreas Hotho, Robert Jäschke, Alexandros Nanopoulos, Steffen Rendle, Lars Schmidt-Thieme, Gerd Stumme, and Panagiotis

Symeonidis (2012). *Recommender systems for social tagging systems*. Springer Science & Business Media (cit. on p. 75).

Marinho, Leandro Balby and Lars Schmidt-Thieme (2008). "Collaborative tag recommendations". In: *Data Analysis, Machine Learning and Applications*. Springer, pp. 533–540 (cit. on p. 74).

Marino, Olga and Gilbert Paquette (2010). "A competency—driven advisor system for multi-actor learning environments". In: *Procedia Computer Science* 1.2, pp. 2871–2876 (cit. on p. 38).

Mathes, Adam (2004). *Folksonomies-cooperative classification and communication through shared metadata* (cit. on p. 20).

McNee, Sean M., John Riedl, and Joseph A. Konstan (2006). "Being accurate is not enough: how accuracy metrics have hurt recommender systems". In: *CHI'06 extended abstracts on Human factors in computing systems*. ACM, pp. 1097–1101 (cit. on pp. 23, 25, 44, 72).

Miller, George A (1994). "The magical number seven, plus or minus two: Some limits on our capacity for processing information." In: *Psychological review* 101.2, p. 343 (cit. on p. 34).

Mozer, Michael C. and Robert V. Lindsey (2016). *Predicting and improving memory retention: Psychological theory matters in the big data era* (cit. on p. 57).

Niemann, Katja (2015). "Automatic Tagging of Learning Objects Based on Their Usage in Web Portals". In: *Design for Teaching and Learning in a Networked World*. Springer, pp. 240–253 (cit. on pp. 6, 41, 122).

Niemann, Katja and Martin Wolpers (2013). "Usage context-boosted filtering for recommender systems in TEL". In: *Scaling up Learning for Sustained Impact*. Springer, pp. 246–259 (cit. on pp. 41, 75, 124, 156).

Niemann, Katja and Martin Wolpers (2015). "Creating usage context-based object similarities to boost recommender systems in technology enhanced learning". In: *Learning Technologies, IEEE Transactions on* 8.3, pp. 274–285 (cit. on p. 75).

Nussbaumer, Alexander, Christian Gütl, and Cord Hockemeyer (2007). "A generic solution approach for integrating adaptivity into web-based e-learning platforms". In: *International Conference on Interactive Mobile and Computer Aided Learning 2007 (IMCL'07)* (cit. on p. 83).

Nussbaumer, Alexander, Eva-Catherine Hillemann, C Gutl, and Dietrich Albert (2015). "A competence-based service for supporting self-regulated learning in virtual environments". In: *Journal of Learning Analytics* 2.1, pp. 101–133 (cit. on pp. 92, 158).

Oard, Douglas W. and Jinmook Kim (1998). "Implicit feedback for recommender systems". In: *Proceedings of the AAAI workshop on recommender systems*. Menlo Park, CA: AAAI Press, pp. 81–83 (cit. on p. 35).

Paas, Fred and John Sweller (2014). "Implications of cognitive load theory for multimedia learning". In: *The Cambridge handbook of multimedia learning* 27, pp. 27–42 (cit. on p. 34).

Papagelis, Manos, Dimitris Plexousakis, and Themistoklis Kutsuras (2005). "Alleviating the sparsity problem of collaborative filtering using trust inferences". In: *Trust management*, pp. 125–140 (cit. on p. 21).

Parra, Denis and Shaghayegh Sahebi (2013). "Recommender Systems : Sources of Knowledge and Evaluation Metrics". In: *Advanced Techniques in Web Intelligence-2: Web User Browsing Behaviour and Preference Analysis*. Springer-Verlag, pp. 149–175 (cit. on pp. 66, 73, 107).

Parra-Santander, Denis and Peter Brusilovsky (2010). "Improving collaborative filtering in social tagging systems for the recommendation of scientific articles". In: *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*. Vol. 1. IEEE, pp. 136–142 (cit. on p. 105).

Pierce, Glenn L. and Paul F. Cleary (2016). "The K-12 educational technology value chain: Apps for kids, tools for teachers and levers for reform". In: *Education and Information Technologies* 21.4, pp. 863–880 (cit. on pp. 3, 153).

Protopsaltis, Aristidis, Paul Seitlinger, Fotini Chaimala, Olga Firssova, Sonja Hetzner, Kitty Kikis-Papadakis, and Pavel Boytchev (2013). "Working environment with social and personal open tools for inquiry based learning: Pedagogic and Diagnostic Frameworks". In: (cit. on p. 134).

Pu, Pearl, Li Chen, and Rong Hu (2011). "A user-centric evaluation framework for recommender systems". In: *Proceedings of the fifth ACM conference on Recommender systems*. ACM, pp. 157–164 (cit. on p. 42).

Rashid, Al Mamunur, Istvan Albert, Dan Cosley, Shyong K. Lam, Sean M. McNee, Joseph A. Konstan, and John Riedl (2002). "Getting to know you: learning new user preferences in recommender systems". In: *Proceedings of the 7th* *international conference on Intelligent user interfaces*. ACM, pp. 127–134 (cit. on p. 22).

Rawashdeh, Majdi, Heung-Nam Kim, Jihad Mohamad Alja'am, and Abdulmotaleb El Saddik (2012). "Folksonomy link prediction based on a tripartite graph for tag recommendation". In: *Journal of Intelligent Information Systems*, pp. 1–19 (cit. on p. 77).

Reimann, Peter, Michael Kickmeier-Rust, and Dietrich Albert (2013). "Problem solving learning environments and assessment: A knowledge space theory approach". In: *Computers & Education* 64, pp. 183–193 (cit. on p. 35).

Resnick, Paul, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl (1994). "GroupLens: an open architecture for collaborative filtering of netnews". In: *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. ACM, pp. 175–186 (cit. on p. 17).

Resnick, Paul and Hal R. Varian (1997). "Recommender systems". In: *Communications of the ACM* 40.3, pp. 56–58 (cit. on pp. 1, 17).

Sakai, Tetsuya (2007). "On the reliability of information retrieval metrics based on graded relevance". In: *Information processing & management* 43.2, pp. 531–548 (cit. on pp. 75, 76).

Salehi, Mojtaba (2013). "Application of implicit and explicit attribute based collaborative filtering and BIDE for learning resource recommendation". In: *Data & Knowledge Engineering* 87, pp. 130–145 (cit. on p. 40).

Santos, Olga C. and Jesus G. Boticario (2015). "Practical guidelines for designing and evaluating educationally oriented recommendations". In: *Computers & Education* 81, pp. 354–374 (cit. on pp. 27, 42, 43).

Sarwar, Badrul, George Karypis, Joseph A. Konstan, and John Riedl (2001). "Item-based collaborative filtering recommendation algorithms". In: *Proceedings of the 10th international conference on World Wide Web*. ACM, pp. 285–295 (cit. on p. 74).

Schafer, J. Ben, Dan Frankowski, Jon Herlocker, and Shilad Sen (2007). "Collaborative filtering recommender systems". In: *The adaptive web*. Springer, pp. 291–324 (cit. on pp. 4, 67, 73).

Schafer, J. Ben, Joseph A. Konstan, and John Riedl (1999). "Recommender systems in e-commerce". In: *Proceedings of the 1st ACM conference on Electronic commerce*. ACM, pp. 158–166 (cit. on pp. 17, 18).

Schmidt, Andreas (2004). "Context-steered learning: The learning in process approach". In: *Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on*. IEEE, pp. 684–686 (cit. on p. 38).

Seitlinger, Paul, Dominik Kowald, Simone Kopeinik, Ilire Hasani-Mavriqi, Tobias Ley, and Elisabeth Lex (2015). "Attention Please! A Hybrid Resource Recommender Mimicking Attention-Interpretation Dynamics". In: *arXiv preprint arXiv:1501.07716* (cit. on pp. 13, 65, 66, 99, 101, 107, 124, 125).

Seitlinger, Paul, Tobias Ley, and Dietrich Albert (2013). "An implicit-semantic tag recommendation mechanism for socio-semantic learning systems". In: *Open and Social Technologies for Networked Learning*. Springer, pp. 41–46 (cit. on pp. 49, 60, 61, 133, 134).

Shardanand, Upendra and Pattie Maes (1995). "Social information filtering: algorithms for automating "word of mouth"". In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press/Addison-Wesley Publishing Co., pp. 210–217 (cit. on p. 17).

Shen, Li-ping and Rui-min Shen (2004). "Learning content recommendation service based-on simple sequencing specification". In: *Advances in Web-Based Learning–ICWL 2004*, pp. 293–323 (cit. on p. 38).

Shute, V. J. (2009). "Simply Assessment". In: *International Journal of Learning and Media* (cit. on p. 36).

Shute, Valerie J and Diego Zapata-Rivera (2008). "Using an evidence-based approach to assess mental models". In: *Understanding models for learning and instruction*, pp. 23–41 (cit. on p. 35).

Specht, Marcus (2000). "ACE-adaptive courseware environment". In: *Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer, pp. 380–383 (cit. on p. 38).

Stefaner, Moritz, Elisa Dalla Vecchia, Massimiliano Condotta, Martin Wolpers, Marcus Specht, Stefan Apelt, and Erik Duval (2007). "MACE–enriching architectural learning objects for experience multiplication". In: *Creating New Learning Experiences on a Global Scale*. Springer, pp. 322–336 (cit. on p. 69).

Steiner, Christina M., Alexander Nussbaumer, and Dietrich Albert (2009). "Supporting self-regulated personalised learning through competence-based knowledge space theory". In: *Policy Futures in Education* 7.6, pp. 645–661 (cit. on p. 3).

Sun, Ron (2008). "Introduction to computational cognitive modeling". In: *Cambridge handbook of computational psychology*, pp. 3–19 (cit. on p. 49).

Trattner, Christoph, Dominik Kowald, Paul Seitlinger, Simone Kopeinik, and Tobias Ley (2016a). "Modeling Activation Processes in Human Memory to Predict the Use of Tags in Social Bookmarking Systems". In: *The Journal of Web Science* 2.1, pp. 1–16 (cit. on pp. 58, 59, 138).

Trattner, Christoph, Dominik Kowald, Paul Seitlinger, Simone Kopeinik, and Tobias Ley (2016b). "Modeling Activation Processes in Human Memory to Predict the Use of Tags in Social Bookmarking Systems". In: *The Journal of Web Science* 2.1, pp. 1–16 (cit. on p. 125).

Verbert, Katrien, Hendrik Drachsler, Nikos Manouselis, Martin Wolpers, Riina Vuorikari, and Erik Duval (2011). "Dataset-driven research for improving recommender systems for learning". In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. ACM, pp. 44–53 (cit. on pp. 4, 39, 44, 75, 124, 126, 156).

Verbert, Katrien, Nikos Manouselis, Hendrik Drachsler, and Erik Duval (2012). "Dataset-Driven Research to Support Learning and Knowledge Analytics." In: *Educational Technology & Society* 15.3, pp. 133–148 (cit. on p. 31).

Verbert, Katrien, Nikos Manouselis, Xavier Ochoa, Martin Wolpers, Hendrik Drachsler, Ivana Bosnic, and Erik Duval (2012). "Context-aware recommender systems for learning: a survey and future challenges". In: *IEEE Transactions on Learning Technologies* 5.4, pp. 318–335 (cit. on pp. 22, 26, 28, 30, 44).

Vuorikari, Riina and David Massart (2010). "dataTEL challenge: European Schoolnet's Travel well dataset". In: *1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010)* (cit. on p. 69).

Wagner, Claudia, Philipp Singer, Markus Strohmaier, and Bernardo Huberman (2014). "Semantic stability and implicit consensus in social tagging streams". In: *IEEE Transactions on Computational Social Systems* 1.1, pp. 108–120 (cit. on pp. 21, 132, 140).

Wang, Chong and David M. Blei (2011). "Collaborative topic modeling for recommending scientific articles". In: *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pp. 448–456 (cit. on p. 22).

Wille, Rudolf (2005). "Formal concept analysis as mathematical theory of concepts and concept hierarchies". In: *Formal Concept Analysis*. Springer, pp. 1–33 (cit. on p. 164).

Wilson, Scott, Oleg Liber, Mark W. Johnson, Philip Beauvoir, Paul Sharples, and Colin D. Milligan (2007). "Personal Learning Environments: Challenging the dominant design of educational systems". In: *Journal of e-Learning and Knowledge Society* 3.2, pp. 27–38 (cit. on p. 87).

Winter, Lisa-Christina, Simone Kopeinik, Dietrich Albert, Aurora Dimache, Attracta Brennan, and Thomas Roche (2013). "Applying Pedagogical Approaches to Enhance Learning: Linking Self-Regulated and Skills-Based Learning with Support from Moodle Extensions". In: *Advanced Applied Informatics (IIAIAAI), 2013 IIAI International Conference on*. IEEE, pp. 203–206 (cit. on pp. 12, 80).

Xu, Zhichen, Yun Fu, Jianchang Mao, and Difu Su (2006). "Towards the semantic web: Collaborative tag suggestions". In: *Collaborative web tagging workshop at WWW2006, Edinburgh, Scotland* (cit. on p. 20).

Zapata, Alfredo, Víctor H. Menéndez, Manuel E. Prieto, and Cristóbal Romero (2013). "A framework for recommendation in learning object repositories: An example of application in civil engineering". In: *Advances in Engineering Software* 56, pp. 1–14 (cit. on p. 40).

Zheng, Nan and Qiudan Li (2011). "A recommender system based on tag and time information for social tagging systems". In: *Expert Systems with Applications* 38.4, pp. 4575–4587 (cit. on pp. 67, 74).

Ziegler, Cai-Nicolas, Sean M McNee, Joseph A Konstan, and Georg Lausen (2005). "Improving recommendation lists through topic diversification". In: *Proceedings of the 14th international conference on World Wide Web*. ACM, pp. 22–32 (cit. on pp. 23–25).

Zimmerman, Barry J. (2000). "Self-efficacy: An essential motive to learn". In: *Contemporary educational psychology* 25.1, pp. 82–91 (cit. on p. 82).

Zimmerman, Barry J. (2002). "Becoming a self-regulated learner: An overview". In: *Theory into practice* 41.2, pp. 64–70 (cit. on pp. 39, 82).

Zimmerman, Barry J. and Dale H. Schunk (2012). *Self-regulated learning and academic achievement: Theory, research, and practice*. Springer Science & Business Media (cit. on p. 82).

Glossary

ACT-R Adaptive Control of Thought-Rational. 56, 57, 184
AI Assessment Items. 80, 184

BLL Base Level Learning Equation. 6, 7, 49, 59, 63, 123, 126, 131, 132, 137, 138, 144, 147, 155, 157, 184

CbKST Competence-based Knowledge Space Theory. 3, 13, 38, 39, 50, 81, 83, 89, 96, 97, 152, 184
CF Collaborative Filtering. 4, 13, 39, 45, 73, 152, 153, 184

FCA Formal Concept Analysis. 162, 184

GUI Graphical User Interface. 87, 184

IBL Inquiry-based Learning. 6, 40, 122, 131, 133, 134, 138, 155, 157, 162, 184

KST Knowledge Space Theory. 38, 50, 184

LDA Latent Dirichlet Allocation. 66, 106, 184
LMS Learning Management System. 86, 184
LR Learning Resources. 80, 86, 89, 92, 184
LTM Long Term Memory. 59–61, 184

MaTGM Macro Tag Growth Method. 139, 184
MOOC Massive Open Online Course. 123–125, 128, 130, 147, 184
Moodle Modular Object-Oriented Dynamic Learning Environment. 83, 84, 86, 88, 92, 93, 96, 184

RQ1 Research Question 1. 13, 14, 79, 96, 97, 158, 184
RQ2 Research Question 2. 5, 13, 79, 112, 121, 122, 155, 158, 159, 184
RQ3 Research Question 3. 13, 26, 123, 146–148, 158, 184

RS Recommender Systems. 1–3, 13, 151, 184

SRL Self-regulated Learning. 13, 83, 184

STM Short Term Memory. 59, 60, 184

STS Social Tagging Systems. 20, 184

SUR Simplified Updating Rule. 54, 82, 85, 184

SUSTAIN Supervised and Unsupervised STratified Adaptive Incremental Network. 4, 5, 8, 9, 13, 50, 54, 98, 99, 101, 104, 114–116, 118, 121, 122, 125–128, 130, 152–156, 159, 161, 162, 184

TEL Technology-Enhanced Learning. 1–6, 9, 10, 13, 18, 26–28, 30, 35–38, 44–46, 73, 75, 79, 122–126, 128–130, 148, 151, 155–158, 184

Part IV.

Appendix

Appendix A.

Full List of Publications

A.1. Peer Reviewed Publications

1. Kowald, D., **Kopeinik, S.**, & 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, Adapation and Personalization (UMAP'2017)*. ACM.
2. **Kopeinik, S.**, Lex, E., Seitlinger, P., Albert, D., & Ley, T. (2017), "Supporting collaborative learning with tag recommendations: a real-world study in an inquiry-based classroom project", In *Proceedings of the 7th International Learning Analytics & Knowledge Conference*, pp. 409-418. ACM Press.
3. **Kopeinik, S.**, Kowald, D., Hasani-Mavriqi I., & Lex, E. (2017). "Improving Collaborative Filtering Using a Cognitive Model of Human Category Learning", *The Journal of Web Science* 2(4), pp. 45-61.
4. **Kopeinik S.**, Kowald, D.,& Lex, E. (2016). "Which Algorithms Suit Which Learning Environments? A Comparative Study of Recommender Systems in TEL". In *European Conference on Technology Enhanced Learning*, pp. 124-138. Springer International Publishing.
5. Trattner, C., Kowald, D., Seitlinger, P., **Kopeinik, S.**, & Ley, T. (2016). "Modeling Activation Processes in Human Memory to Predict the Use of Tags in Social Bookmarking Systems". *The Journal of Web Science*, 2(1), pp. 1-16.
6. **Kopeinik, S.** (2015). "Applying Cognitive Learner Models for Recommender Systems in Small-Scale Learning Environments", Presented at *Doctoral Consortium Workshop of EC-TEL 2015*.

7. Bedek, M., Seitlinger, P., **Kopeinik, S.**, & Albert, D. (2015). "Non-Invasive Assessment of Motivation in a Digital Educational Game". In *ECGBL2015-9th European Conference on Games Based Learning: ECGBL2015*, pp. 58. Academic Conferences and publishing limited.
8. Seitlinger, P., Kowald, D., **Kopeinik, S.**, Hasani-Mavriqi, I., Ley, T., & Lex, E. (2015). "Attention Please! A Hybrid Resource Recommender Mimicking Attention-Interpretation Dynamics". In *Proceedings of the 24th International Conference on World Wide Web Companion*, pp. 339-345. International World Wide Web Conferences Steering Committee.
9. Bedek, M. A., **Kopeinik, S.**, Prünster, B. & Albert, D. (2015). "Applying the Formal Concept Analysis to introduce guidance in an inquiry-based learning environment". In *Advanced Learning Technologies (ICALT), 2015 IEEE 15th International Conference* (pp. 285-289). IEEE.
10. Kowald, D., **Kopeinik, S.**, Seitlinger, P., Ley, T., Albert, D., & 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*, pp. 55-74. Springer, Cham.
11. Kowald, D., Seitlinger, P., **Kopeinik, S.**, Ley, T., & 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*, pp. 75-95. Springer, Cham.
12. Dimache, A., Roche, T., **Kopeinik, S.**, Winter, L. C., Nussbaumer, A., & Albert, D. (2015). "Suitability of Adaptive Self-Regulated e-Learning to Vocational Training: A Pilot Study in Heat Pump System Installation". *International Journal of Online Pedagogy and Course Design (IJOPCD)*, 5(3), pp. 31-46.
13. **Kopeinik, S.**, Nussbaumer, A., Winter, L. C., Albert, D., Dimache, A., & Roche, T. (2014, July). "Combining Self-Regulation and Competence-Based Guidance to Personalise the Learning Experience in Moodle". In *Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference*, pp. 62-64. IEEE.
14. Dimache, A., **Kopeinik, S.**, Brennan, A., Roche, T., Winter, L. C., & Albert, D. (2014). "Innovative Online Vocational Training of Renewable Energy Technologies (INNOVRET)". *International Journal of Information & Education*

Technology, 4(1), pp. 127-131.

15. Winter, L. C., **Kopeinik, S.**, Albert, D., Dimache, A., Brennan, A., & Roche, T. (2013, August). "Applying Pedagogical Approaches to Enhance Learning: Linking Self-Regulated and Skills-Based Learning with Support from Moodle Extensions". In *Advanced Applied Informatics (IIAIAAI), 2013 IIAI International Conference*, pp. 203-206. IEEE.
16. Parodi, E., Fabrizio, G., Specht, M., **Kopeinik, S.**, Haimala, F., & Protopsaltis, A. (2013). "weSPOT: Working Environment with Social and Personal Open Tools for inquiry based learning". Poster in *European Conference on Technology Enhanced Learning*.
17. **Kopeinik, S.**, Bedek, M., Öttl, G., & Albert, D. (2013). "Competence Analyser: A portable GUI tool for modelling domain and learner knowledge". In *Proceedings of the 21th International Conference on Computers in Education*, pp. 133-138.
18. Bedek, M., Seitlinger, P., **Kopeinik, S.**, & Albert, D. (2012). "Inferring a Learner's Cognitive, Motivational and Emotional State in a Digital Educational Game". *Journal of e-Learning*, 10(2), pp. 172-184.
19. **Kopeinik, S.**, Nussbaumer, A., Bedek, M., & Albert, D. (2012). "Using CbKST for Learning Path Recommendation in Game-based Learning". In *Proceedings of the 20th International Conference on Computers in Education*, pp. 26-30.
20. Seitlinger, P. C., Bedek, M. A., **Kopeinik, S.**, & Albert, D. (2012). "Evaluating the validity of a non-invasive assessment procedure". In *Serious Games Development and Applications*, pp. 208-218. Springer Berlin Heidelberg.
21. **Kopeinik, S.**, Bedek, M. A., Seitlinger, P. C.,& Albert, D. (2011). "The Artificial Mentor: An assessment based approach to adaptively enhance learning processes in virtual learning environments". In *Proceedings of the 19th International Conference on Computers in Education*, pp. 106-110.
22. Bedek, M. A., Seitlinger, P., **Kopeinik, S.**, & Albert, D. (2011). "Multivariate Assessment of Motivation and Emotion in Digital Educational Games". In *Proceedings of the 5th European Conference on Games-Based Learning*, pp. 18-25.
23. Bedek, M. A., Cowley, B., Seitlinger, P., Fantato, M., **Kopeinik, S.**, Albert, D., & Ravaja, N. (2011). "Assessment of the Emotional State by Psychophysiological and Implicit Measurements". In *International Conference on*

Multimodal Interaction, Alicante, Spain.

A.2. Other publications

1. **Kopeinik, S.**, Bedek, M., Firssova, O., Mack, J., Albert, D. (2015). ”Introducing Technology-Enhanced Inquiry-Based Learning to Support Science Education in Secondary Schools: A Teacher Perspective“. In *Proceedings of the 7th International Conference on Education and New Learning Technologies*, pp. 6035-6045.

Appendix B.

Questionnaires: Evaluating the Acceptance of CbKST based Resource Recommendations in Moodle

INNOVRET Evaluation

Thank you for agreeing to evaluate this INNOVRET training course. We appreciate your time and effort.

Date:

Reviewer:

Email (optional):

Phone (optional):

Lesson:

About the Questionnaire

The questionnaire serves as a supporting instrument to help capture your feedback which helps us to understand how well the training programme worked for you and also helps to capture any comments you may have. Please complete this after you have worked through the full training course.

Evaluation Questionnaire

1. The cycle of learning, assessment, and visualisation was good for my learning*

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

*Answer this question only if you used the Learning Support Tools.

2. The system supported me to become aware about my learning process*

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

*Answer this question only if you used the Learning Support Tools.

3. The system was limiting my learning

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

4. The system provided helpful guidance for my learning

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

5. This way of learning was stressful

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

6. I enjoyed the way of learning with that system

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

7. I was successful with the learning task

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

8. The information in the user interface was easy to understand

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

9. I would like to use a system like this in the future

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

10. I am happy with the quality of the content presentation

☐ Strongly disagree ☐ Disagree ☐ Not sure ☐ Agree ☐ Strongly agree

How would you describe the learning experience?

Comments:

Have you encountered any problems with the system?

Comments: