ABSTRACT
In the field of personal learning environment (PLE) research is focusing on the generation and provision of recommendations. Amongst others, approaches reach from decision making tools based on psycho-pedagogical principles over specialized social recommender functionality up to general community or context-aware recommendations. The variety of the solutions results from the fact that pure collaborative filtering (CF) techniques are not sufficient for PLE-based scenarios. In this paper we propose utilizing learner interaction recordings for generating PLE recommendations fitting the educational and social context of a learner. Besides pointing out how we have realized this approach as part of a research prototype, we evaluate and discuss such recommendations generated from data captured in former studies.

Categories and Subject Descriptors
H.3.3 [Information Storage and Retrieval]: information search and retrieval – information filtering

General Terms

Keywords
Personal learning environments, recommender strategy, user interaction data, collaborative filtering, clustering.

1. INTRODUCTION
Personal learning environments (PLEs) deal with supporting learners in their everyday activities, e.g. by empowering them to design and use their environments so that they can connect to learner networks and collaborate on shared artifacts to achieve their goals [1]. One important instrument of PLEs is recommender technology which is applied according to very different paradigms and techniques. Amongst others, recommendations can guide through the learning process on the basis of psycho-pedagogical principles [2], help to identify relevant artifacts, experts or learning events at the workplace [3], or provide community and context-aware information to learners [4, 5].

Moreover, recommender systems for PLEs are often based on a collaborative filtering although CF techniques produce unsatisfying results as the underlying learner interaction model is more complex than simple learner-task pairs (‘each learner performs a task one time’) [6]. An overview of possible interactions of learners with their PLEs is given in [7]. According to this simplified model interactions with the following PLE entities can be identified: (a) activities, (b) collection of learning resources (repositories), (c) single artifacts, (d) communities, and (e) agents, i.e. humans and software systems.

Consequently we propose to capture interactions with these PLE-related entities and to utilize this data for generating recommendations. Therefore, the upcoming section gives an overview of selected recommender approaches and shows how to overcome the restrictions of CF techniques for real-world situations we are facing in PLE scenarios. Then we propose a strategy for creating recommendations for different PLE entities (activities, artifacts, tools, and peers) and for different situations of using PLE technology. Finally we describe data-sets we have captured in various studies and indicate the quality of our PLE recommendations on the basis of concrete scenarios.

2. LIMITATIONS OF COLLABORATIVE FILTERING AND POSSIBLE WAYS OUT
Collaborative filtering (CF) techniques aim at predicting appropriate items on the basis of interaction data of many users within a community [8]. In the field of PLEs CF can be applied on the basis of learner interaction recordings which are shared voluntarily [7]. Such a PLE recommender could suggest activity patterns (i.e. shared PLE designs) as well as single entities (i.e. peers, artifacts or tools) which are useful for a specific context. However, pure CF techniques are not sufficient for PLE-based activities due to three important reasons.

First of all [6] state that CF is based on the assumption that each user rates each item once, which is not the case for educational environments. Normally learners perform tasks several times and continuously interact with the different PLE entities (activities, actors, artifacts, tools). Secondly, CF techniques tend to provide recommendations of the most popular items, i.e. the top-n of the overall data-set or the so-called ‘global top-n’. In PLE-based activities, however, learners rather require recommendations which are appropriate for a specific situation. For instance, suggesting the most successful search engine (e.g. Google) to a learner who has specific information need (e.g. on mathematical equations or on documents within a corporate repository) could be counterproductive. So, in many cases the global top-n is irrelevant thus decreasing the accuracy of a recommender while popular items of local data-sets (e.g. PLE interaction recordings of one’s clique or retrieved according to a search term) might be more useful. Thirdly, CF suffers from the data sparsity problem, and the quality of recommendations is low if only a small set of a large database of items is rated by users [5].
To overcome these limitations of CF techniques for PLEs, literature provides a few and rather diverse solution approaches. Amongst others, [6] proposes context-aware factorization techniques to generate recommendations utilizing all interactions (performances) of students. Consequently authors show that this approach slightly outperforms CF techniques like the k-nearest neighbors method. Furthermore, [5] applies clustering techniques to overcome sparsity problems and create local top-n at the same time. In practice model-based CF techniques seem to be appropriate for a PLE recommender. Such a semantic model can not only be used to cluster usage data and create recommendations on the level of these clusters but also for other techniques which increases the accuracy of a recommender. For instance, [4] propose a PageRank-like approach to rank recommendations according to actor-artifact-activity networks. Others, like [7], utilize such a semantic model and suggest exploiting user feedback to improve recommendations.

Against this background, we have already described a recommender strategy for PLEs based on a simple activity model (cf. [7]). Additionally we have showed which data to capture in PLE-based activities and how to exploit this data for generating recommendations. In the following we briefly summarize the method of capturing learner interactions and then present a recommender approach being based on some of the before-mentioned concepts.

3. CAPTURING AND COLLECTING PLE USAGE DATA

In former research, we have developed a PLE-like prototype which is based on the Actor-Network Theory (ANT). According to [9] a learner can interact with these PLE entities:

- **Processes**: Lifelong learning activities carried out at the workplace, for educational reasons, or due to personal goals (e.g. a job task in a business process, attending a course for further education, or a spare time activity requiring the acquisition of new competences)

- **Media**: Collection of learning resources required for or used in these activities (e.g. the Wikipedia platform, learning objects repository, or simply the Internet)

- **Artifacts**: Documents and other (digital or real-world) artifacts collaboratively created and accessed by learners (e.g. Wiki articles or a joint paper)

- **Agents**: Other actors, no matter if human or systemic ones (e.g. peer learners or functionality provided via Internet)

- **Communities**: People sharing the same environment in terms of having common interests, working on the same artifacts, being connected to the same actors (e.g. a group of learners trying to achieve a course goal or a special interest group for a specific topic)

Following this model of a learner-centric ecology we implemented a client-sided PLE prototype in the form of a Firefox add-on. ‘PACMan’ – which stands for Personal Activity Manager – allows users to manage their online resources and tools according to a very simple model, the notion of a (learning) activity. Such activities serve as elements to describe and structure the learning context. Basically, users can group tagged online resources and tools (URLs) to activities and give them titles. In order to keep the model simple, we do not support other relations, like dependencies or semantic relations between activities.

Finally and displayed at the bottom of the side-bar, PACMan provides facilities to connect to a pattern repository which allows sharing PLE experiences with others. This integrated web service enables practice sharing in PLE settings, as users can publish patterns of their activities, retrieve and instantiate the patterns available on the repository, and get recommendations for different aspects within a PLE. Our prototypic pattern repository is realized as a component for the OpenACS server (http://openacs.org) and is based on the object-oriented scripting language XoTCL extending the Wiki generator XoWiki (http://openacs.org/xowiki). PACMan as well as the pattern repository component (called PLEShare) are open source and accessible via SourceForge (http://sourceforge.net/projects/rolewp7). Besides, we provide PACMan also via the Mozilla Add-on Developer Hub under the URL https://addons.mozilla.org/en-US/firefox/addon/176479.
4. COMMUNITY AND TOPIC-AWARE TOP-N RECOMMENDATIONS

So far, we have used this infrastructure – the client-sided PLE solution PAcMan and the pattern repository PLEShare – to capture and collect data-sets within various case studies. After manually filtering out patterns of low quality, we have identified 47 patterns of PAcMan activities containing 260 actions, 228 unique URLs, 151 tools (URLs clustered according to domain name), and 14 peer users. In our first case studies we have identified two interesting strategies for generating local top-n recommendations.

In a first study we captured a series of activities of one user with his colleagues. Overall, we collected 9 patterns and 70 online resources (plus user-given tags) of 6 actors collaborating with each other. Analyzing these patterns, however, led to the conclusion that activities of such a clique are far too different to identify items which are worth being recommended to other users, e.g. users who plan to join this clique. Even the patterns being associated to the same activity vary, as each actor has a specific role in this activity. Basically, only a few items (URLs, user tags and tools) occurred more than once. Yet there is no significant deviation observable in the frequency distribution of the items.

In a second study we collected patterns related to language learning (French). The data-set of 14 patterns which was created by querying the whole repository by the search term “French” showed a similar behavior. Hardly any tool or artifact appeared more than once. Anyway, both approaches demonstrate ways how to cluster the data-sets and generate local top-n recommendations, whereby we consider the first one being based on community-awareness (the patterns of a clique) and the second one topic-awareness (the patterns retrieved by a search term).

Due to this motivating finding, we propose to cluster the data-set of the repository which stores activity patterns shared by PLE users voluntarily according to a user’s clique and the context (e.g. given by goals or tasks). Possible techniques to be applied here would be Markov models or simple association rules – the latter one was applied for combining the two clustering strategies (cluster overall data-set according to a user’s clique and the patterns retrieved by contextual information). In this way valuable recommendations have been generated from such clusters. As the sub-set of patterns normally is not too comprehensive (approximately 10 to 50 patterns), such local top-n recommendations could be created on-the-fly, e.g. if a user wants to start a new activity and has no experiences in the area, or if a user is involved into an activity already and requires support in the form of peer, artifact, or tool recommendations. However, this first prototypic implementation of clustering – combining clique-based clustering and query-based retrieval of patterns – is promising but not evaluated very well. Future work will focus on further experiments with community-aware (e.g. clique-based) clustering and other techniques, whereby combining these algorithms could be achieved through simple association rules or Markov models.

Finally, we also had a look at pattern usage over the last 6 months. The pattern repository keeps track of how often a pattern has been instantiated so far. Hereby, an instantiation means that a user creates an activity in the PAcMan tool on the basis of the pattern. It has to be mentioned that we analyzed all patterns on the repository (also the ‘bad’ patterns filtered out for before-mentioned analysis) as well as all their versions. Thus, we have statistical data on 409 versions of 52 patterns. As shown in Fig. 3, the frequency distribution of the pattern instantiations follows a power law, i.e. 60 patterns have never been instantiated at all, 55 patterns have been used once, 38 patterns twice etc. On the other side, a few patterns have been instantiated up to 78 times, which means that some patterns are significantly more interesting and/or relevant for users than others.

In the next step, we combined these two strategies and collected patterns of a homogenous group (researchers at the same institute) and for a specific situation given by two concrete tasks: (a) finding literature for a concrete conference paper; (b) planning the travel to the conference place. Here, we captured 17 patterns created by 8 users and including 99 URLs and tags. Analyzing this data manually, we found out that 13 resource tags were identical and even that 10 pattern titles as well as 53 tags were similar (according to a topic). Moreover, we refined 33 different tools (i.e. top-level domains) out of the 99 URLs. Fig. 2 visualizes the frequency distribution of the tags and the URLs.

![Figure 2. Tag and URL frequency distribution (identical vs. similar according to topics and top-level domains).](image)

In accordance with findings from network theory (e.g. preferential attachment and robustness considerations in scale-free networks [10]), this observation could be a characteristic of a sustainably evolving community of practice. Moreover, this usage data could be used to refine the generation of recommendations, for instance by applying a context-aware factorization technique considering that users interact with PLE-related items more than one time [6].

In sum, we identified properties of scale-free networks on two different levels, namely in clusters of PLE design decisions (our
activity patterns) and in PLE usage data itself (consumption of patterns or single PLE elements, multiple user interactions with the same item etc). According to literature, networks showing scale-free properties are useful for generating recommendations, as evidenced e.g. for music recommendation networks which have been constructed by collaborative efforts [11]. Thus, we believe that the data-sets being collected in our repository can be used to generate useful and local top-n recommendations for PLEs.

5. CONCLUSIONS AND FUTURE WORK
In this paper we have proposed a strategy for generating recommendations for PLE users. With respect to the most relevant technique, collaborative filtering, we are optimistic about creating local top-n recommendations for specific situations of learners by clustering the interaction recordings according to the social context (e.g. a user’s clique) and a specific situation (e.g. a topic formulated as query term). Although this leads to small data-sets only, we have shown that the generation of useful recommendations is possible. Related work evidence that similar approaches outperform CF techniques [5, 6]. On the other hand, our own preliminary evaluation has showed that clustering according to the social structure might not be sufficient so that we plan to examine other clustering variants (or combinations) for this purpose.

Nevertheless, we are at a very early stage with this research. Future work will try to utilize usage data (e.g. the consumption of pattern and recommended items) to improve the accuracy of our recommender or to include additional contextual information for suggesting PLE elements to users. Beside the realization of recommendation facilities for end-users and tryouts in real-world settings, we have to collect far more data in order to find and evaluate a good clustering algorithm and experiment with other ideas, like utilizing usage data e.g. through context-aware factorization.

6. ACKNOWLEDGMENTS
The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no 231396 (ROLE project).

7. REFERENCES