Harnessing Geo-tagged Resources for Web Personalization

Markus Zanker
Alpen-Adria-Universitaet
Klagenfurt
9020 Klagenfurt, Austria
markus.zanker@aau.at

Markus Jessenitschnig
Sm@rter Software
Lakeside B01
9020 Klagenfurt, Austria
markus.jessenitschnig@smarter-software.com

Martin Stromberger
FH Kärnten
9500 Villach, Austria
m.stromberger@fh-kärnten.at

ABSTRACT
The increasing plethora of information available on the Web necessitates effective personalization mechanisms that allow users to retrieve pieces of information that are more likely to be of interest to them. Fortunately, the democratization of the Web (aka, Web 2.0) that provides online users with widely available tools that allow them contribute and integrate themselves into the global information space not only increases the sheer amounts of data, but also offers opportunities to extract semantic meaning.

This paper therefore presents an integrated approach that harnesses geo-tagged web resources like tourism services or track data from bike trails to derive semantic annotations for objects from their geographic proximity. Following this, a recommendation mechanism is proposed that hybridizes collaborative mechanisms with the additional knowledge about semantic annotations to make predictions about what will be relevant to a user in a specific situation. The utility of this integrated approach is showcased by an adaptive Web-GIS scenario that supports travelers in their decision making. Finally, the proposed algorithms are evaluated using historical log data from real users who were exploring the map of a tourism destination. The results indicate that, despite very short interaction sequences, improvements compared to a collaborative filtering baseline can be achieved. An additional advantage lies in offering users more detailed options to express their search preferences that is not quantified by the presented evaluation scenario.

1. INTRODUCTION
Web 2.0 enables users to easily contribute to the online information space with what is known as user-generated content. However, this myriad of information overwhelms users unless it is processed and structured prior to presentation. Traditional information retrieval facilities like popular search engines try to reduce the information volume primarily based on content similarity with respect to a specific user query and by considering reputation heuristics like PageRank. In contrast, Semantic Web research aims to automatically augment and enrich web data with meaning by merging it with structured data like taxonomies or ontologies and formal reasoning mechanisms leading to intelligent services. Within this line of research we propose an integrated application scenario that includes an automated mechanism for geo-tagged resources that exploits spatial proximity and derives qualitative annotations for information entities. This content is then exploited by a collaborative personalization mechanism before finally being used to compute recommendations for users. The motivation for this approach stems from the need to provide personalization services for WebGIS like Google Maps. Due to the wide distribution of mobile devices with GPS functionality among consumers, (social) web platforms are accumulating more and more geo-tagged objects that may quickly overburden users when they are displayed on a map. Therefore, this paper’s contribution is a comprehensive application scenario that can thus be termed a collective knowledge system in the sense of Gruber [11] as it provides an effective information filtering and personalization mechanism by structuring and classifying online content.

This paper starts by referring to related work in Section 2 and then outlines the integrated application scenario in Section 3. Furthermore, in Section 4 we explain our technical approach by describing the extraction of semantics, the mining of frequent patterns based on these semantics and finally the collaborative exploitation of the semantic patterns for personalization and recommendation. Finally, an exemplary evaluation on historical user sessions is provided as a proof of concept and conclusions are drawn.

2. RELATED WORK
Due to the increasing popularity of GPS devices among consumers, geo-tagging or geo-coding has become popular and can now be considered common for many Web 2.0 platforms. As such, it is one of the primary phenomena that has led to the notion of the geospatial Web [24]. Geo-tags assign GPS coordinates or denominations for geographic places to an object. However, we only consider the former case in our scenario, as it allows us to directly compute distances. Based on an object’s geographic neighborhood additional semantic annotations can be derived as presented in [29]. Reeve & Han [19] classified different semantic annotation platforms based on the annotation methods used, differentiating between pattern-based, machine learning-based and...
strategies combining both. Annotation is in principle a classification task that either builds on a set of manually crafted rules/patterns or tries to learn a statistical classifier from an already annotated corpus of documents. However, to the best of the authors’ knowledge none of these approaches exploits spatial proximity for annotating objects and consequently feeds them into a personalization mechanism.

The collaborative recommendation paradigm itself can be seen as one of the pioneer applications of what has now come to be known as 'Web 2.0' where users share opinions on items and benefit from the ratings and recommendations of other users from the community [14]. The concept dates back to the early 90-ies when the GroupLens system was developed for recommending netnews [20]. Soon afterwards recommending movies became a popular application domain for collaborative filtering (CF) [12] and remain so due to the availability of datasets for experimental offline evaluation and competitions like Netflix [5]. However, the CF technique suffers from ramp-up problems when new users or new items without any associated ratings are introduced to the system. This is particularly the case for our proposed scenario of personalizing WebGIS interactions where anonymous user sessions are rather short and initially no ratings are known. Therefore, the standard CF mechanism needs improvements for this application scenario. As a consequence hybrid recommendation algorithm designs that incorporate additional content have been proposed. Burke provides a coherent survey on how additional knowledge can be inserted to a CF system [4]. Web 2.0 supports such efforts from two perspectives. First, users are free to provide additional data about entities that could be recommended by tagging them or writing even more detailed reviews. Second, they disclose information about themselves like demographics and social relationships. The feature combination approach proposed by Basu et al. [2] addresses the first aspect. Their approach combines a user’s likes and dislikes with content features of catalog items, such as user Alice liked mostly horror movies if nearly all of her positive ratings refer to movies of that specific genre. Content-boosted collaborative filtering [17] is another example of how knowledge about items can lead to CF algorithm improvements. Melville et al. computed similarities between rated objects and derived additional artificial ratings for items that are similar to the actually rated ones. A meta-level approach was first introduced by the Fab system [1]. This hybridization approach uses one method to build a model that is enacted by a second one. Fab employs a content-based recommender that builds user models based on a vector of term categories and the users’ degrees of interest in them. The recommendation step, however, does not propose items that are similar to the user model but employs a collaborative technique. The latter determines the user’s nearest neighbors based on content models and recommends items that similar peers have liked. A similar principle is also employed by the collaborative constraint-based meta-level recommender presented here. CF identifies which association rules have the highest support among a user’s nearest neighbors and a constraint-based recommender [30] evaluates the personalized constraint sets. The CHIP demonstrator [25], however, is a recent example of a recommender that solely relies on (semantically enriched) content information about museum collections. Although, this is advantageous at the beginning of a system’s life cycle with only few users, once system adoption increases, additional collaborative us-

age data helps to improve the quality of recommendations.

Knowing more about users themselves can also lead to more accurate recommendation systems. Pazzani [18], for instance, is bootstrapping a collaborative recommender system with demographic user characteristics when not enough item ratings were known. Another approach for feature combination was proposed by [28] that dynamically determines a set of different types of rating feedback based on their predictive accuracy and availability for computing recommendations. Alternatively, trust relationships that are encoded in social network applications could be exploited to determine user neighborhoods. Goldbeck and Hendler [9], for instance, researched how trust relationships may be inferred in the Web 2.0 context. However, in our application scenario we chose not to exploit trust-based mechanisms and instead focused on the similarity based approach for neighborhood determination of traditional CF.

The adoption of maps or WebGIS in recommendation applications has only recently been explored. Ricci et al. [21] used maps in combination with an interactive critiquing recommender in a mobile environment. Users were able to give qualitative feedback on nearby restaurants and the system iteratively proposed alternatives. However, in contrast to our work no implicitly collected user feedback was exploited.

3. APPLICATION SCENARIO

In order to illustrate our proposed approach in a comprehensible way we sketch the application scenario in Figure 1. On the bottom left a typical excerpt from a WebGIS is depicted that is overloaded with geotagged entities that might represent different attractions, restaurants or user generated travel blogs and pictures. Obviously, one needs to zoom deeper into such an area to explore all of these pieces of information at the risk of possibly loosing oversight of the overall picture. Therefore, effective filtering mechanisms that restrict the number of visible objects based on their presumed relevance for the current user are required. However, clues about the semantics of these objects are required to offer explicit filtering options to users or to learn about patterns from user transactions.

Tags that are associated with these geo-positioned objects constitute a valuable starting point. Knowing that an entity is denoted with ’hotel’ or ’restaurant’ alongside with metadata like a tourism ontology or a folksonomy would allow at least some categorical filtering.
Additional qualitative information about objects can be extracted if their geographic position and their spatial proximity to other entities is considered. For instance, comparable to a density based clustering approach the degree of crowdedness for a specific place can be determined and objects can be classified based on whether they are part of a hot spot area or not. In addition, when taking tag values into account more interesting properties can be derived such as if an entity tagged as a hotel lies close to objects tagged with terms like beach it will probably be suitable for tourists that enjoy water-related activities. The application of such heuristics for generating qualitative profiles of tourism regions is described in [27, 29]. Obviously, this principle becomes even more powerful when it also incorporates sentiments from geo-tagged travel blogs, for instance, that could be used to derive degrees of service quality or cost/value ratios. Compare, for instance, the work in Graebner et al. [10], where quantitative ratings are derived by sentiment analysis and problem-specific lexica.

However, item semantics are a necessary prerequisite for building personalization models of high quality. Collaborative filtering is a very popular paradigm for the personalization of displayed items due to its effectiveness and simplicity. It identifies similar users and deduces recommendable items for the current user from what worked well for her/his most similar peers (see the upper and lower right depictions in Figure 1). Users that interact with maps leave digital traces like selecting objects to read their associated tags and entries, zooming or sliding a map’s pane which may all be abstractly denoted as transactions. Alternatively, reputation based mechanisms that are derived from social network analysis can be exploited for forming user neighborhoods. [5] is an example of such a foafing-based recommendation approach.

Finally, we propose learning personalization models using a frequent pattern mining approach that is capable of processing the diverse situational characteristics of user sessions like terms from search queries, map panes or semantics of identified objects in order to determine association rules that can be exploited for making predictions. Examples include for instance that users who used the terms A and B in their search terms showed interest in items that are annotated with tags C and D. or users who select an Indian restaurant also look for alternative restaurants offering multicultural food. Thus, the proposed recommendation approach for this scenario is a collaborative constraint-based meta-level recommender that applies association rules that have been mined from the most similar peers of the current user to make recommendations. Consequently, it does not assume that a single set of personalization rules serves the whole user community, but rather using the collaborative filtering principle exploits clusters of users showing similar behavior and/or trusting each other. In Zanker [26] it has been shown that such an approach in traditional e-commerce shopping scenarios reaches a higher predictive accuracy than applying exactly the same set of personalization knowledge to all users.

In the next three sections we will present how this scenario integrates user generated content from Web 2.0 with semantic technologies and data mining techniques in more detail.

4. MAP-BASED RECOMMENDATION APPROACH

The basic collaborative filtering mechanism itself does not require any knowledge about objects that can be proposed except a unique identifier. However, hybridizing it by exploiting additional content information typically leads to better results [4]. In map-based recommendation scenarios the number of objects that should be visualized on a map is typically rather large. In our evaluation scenario, for instance, there were more than 16,000 different tourism objects. In addition, user profiles are typically very sparse and locally biased, which means that given an entry point on a map, users will probably show interest in items that are within a short distance of or at least close to that point. Therefore, pure collaborative filtering can only identify similar peers for a user that showed interest in approximately the initial region and learns relationships of the form users who were interested in object A also showed interest in objects B and C. However, the basic idea of the proposed approach is to also learn more topical association rules like users who were interested in objects related to water sports also showed interest in objects related to nightlife. Hence additional semantics about objects are required that may stem from tagging activity, available data repositories or information extraction efforts like the one presented in next subsection.

4.1 Extracting semantics of geotagged items

Exploiting semantics from geospatial proximity is analogous to density-based clustering in spatial databases like GDBSCAN [22]. There, data points are classified to be within a cluster, at the cluster’s border or noise depending on existence of other points within their e-neighborhood. Similarly, we propose to derive semantic annotations from the ensemble of other objects in an object’s neighborhood. The approach assumes a distance measure, some domain knowledge such as tags and an optional set of annotation patterns to limit the search space. Figure 2 sketches a simple example based on the notion of formal contexts [8]. On the left the small circles denote arbitrary objects in a planar space that possess the property of being white or red. Now, one might be interested to know if a red object is located close to white objects. For instance, if red objects represent hotels and white objects constitute tourism infrastructure like sights, restaurants, beaches or ski schools it would be useful to know if the accommodation is centrally located or farther away. In [27] 10 different properties like aptness for water sports activities, proximity to golf courses or the nightlife factor were computed to assist the rather complex decision processes of prospective tourists.
The larger circles with dotted lines represent the area within which white objects are considered to be close to the red object in order to make binary decisions. The table on the right in Figure 2 sketches the formal binary contexts, denoting that only objects $o_1$ and $o_2$ possess the extracted property close to white (abbrev. by ctw in the following).

Formal Concept Analysis [8], a popular formalization for mining and discovery problems in Web 2.0, like the mining of frequent tri-concepts to identify shared conceptualizations in folksomnies presented by [15], may be used to define the extraction process in a generic fashion. In general, a formal context is defined by a triple $\Gamma(O,P,R)$, where $O$ denotes the set of objects, $P$ represents the formal properties or attributes of $O$ and $R$ is a binary relation between $O$ and $P$ such that $R \subseteq O \times P$. Given a pair of elements $o \in O$ and $p \in P$ where $(o,p) \in R$ means that $o$ has the property $p$. The intent of a set of objects $O_k \subseteq O$ signifies its properties and is denoted as $O_k^P$ while the extent of a set of properties $P_k \subseteq P$ denoted as $P_k^O$ stands for the objects possessing that property, i.e. $\{o_1\}_R = \{ \text{red}, \text{ctw} \}$ and $\{ \text{ctw} \}_O = \{ o_1, o_2 \}$. In addition, we introduce the operator $\epsilon$ that determines the sets of objects that are in its $\epsilon$-neighborhood $o^\epsilon$ for an object $o \in O$, i.e. $\{o_1, o_2\}_\epsilon$. Note, that the value for the threshold distance $\epsilon$ needs to be chosen depending on the semantics that should be derived, for instance, proximity is more important for après ski activities than for skiing in general. In addition, we need an extent operator $\rho$ that decides if an object possesses a new property. The outcome of the binary classification decision depends on an object’s $\epsilon$-neighborhood. Here, a variety of different predicates implementing this oracle would be possible. An initial simple approach would assume a threshold number of objects (minsum). However, neighboring objects could also be weighted based on their actual distances or an individual relevance score. For instance, in Paris Tour Eiffel and Louvre are obviously more relevant for a sightseeing factor than an architectural site. Furthermore, the distance measures themselves could be varied. We used linear Euclidean distance in our implementation, however logarithmic distances or considering actual routing distances could lead to more accurate results. In cases where objects are tagged with geographical names an additional resolution step could derive latitude and longitude information where necessary. Furthermore, the threshold value minsum itself could be reflexively defined. For instance, the average or the upper quartile of the number/weights of neighboring objects among all objects could be used as a threshold, for instance. Obviously, a variety of empirical studies is necessary to compare these different operator design variants.

Consequently, when applying this geospatial annotation approach to derive a new set of properties $\Pi$ the formal context $\Gamma(O,P,R)$ is transformed into $\Gamma'(O; \hat{P}; R')$, such that $\hat{P}' = P \cup \Pi$ and $R' = R \cup \bigcup_{o \in O} \bigcup_{p \in \Pi} (o, p, \pi)^\rho$ with

\[
(o, p, \pi)^\rho = \begin{cases} 
(o, \pi) & : \text{if } |o^\epsilon| \text{ exceeds predefined threshold} \\
\emptyset & : \text{else}.
\end{cases}
\]

Finally, a formal context constitutes a pair $\langle O_k, P_k \rangle$ with $O_k \subseteq O$ and $P_k \subseteq P$ such that $O_k^P = P_k$ and $P_k^O = O_k$. For instance $\{\{o_1, o_2\}, \{\text{red}, \text{ctw}\}\}$ forms the concept red objects that are close to white objects (see Figure 2).

Although the personalized filtering of WebGIS objects serves as the guiding application domain, this approach based on proximity-based annotation should be seen as a preprocessing step for further analysis of the relationships between semantic concepts as well as for geo-sensitive search in general. While geo-sensitive search mainly focuses on extracting information about the geographic relevance of a document, this approach can help to answer queries like “Where can I attend an interesting conference in my field closest to a sunny beach?” that according to Ceri [6] cannot currently be answered.

### 4.2 Collaborative meta-level recommending

A meta-level hybrid recommender consists of at least two systems, where the first generates a model that is exploited as input by the second and so on. Association rule mining is a popular method for discovering relationships in large data sets. Here, users show interest in different objects that possess several properties and are in a spatial context. Hence, the basic idea is to identify popular dependencies that have a high support (i.e. occur for a large share of users) and use them to personalize the system’s interaction with new users. In [26] this approach was combined with collaborative filtering, where the set of applicable rules (i.e. logical implications) was not derived from the whole user population, but only from those that showed similar behavior (i.e. the nearest neighbors). Figure 3 sketches the recommendation process, starting in the lower left quadrant and proceeding clockwise. Collaborative filtering determines a new user’s most similar peers and recommends a set of association rules that have been mined from transaction data and related item knowledge (i.e. the formal contexts of objects in her/his neighborhood as outlined in Subsection 4.1). This personalized rule base is subsequently exploited by a constraint-based recommender which computes which objects should be visualized.

Formally, the set of user models constitutes a formal context $\Gamma_{UM}(O_{UM}, P_{UM}, R_{UM})$, where $O_{UM}$ denotes the set of users, $P_{UM}$ represents their formal properties and $R_{UM}$ relates $O_{UM}$’s elements to elements of $P_{UM}$.

Properties in $P_{UM}$ identify objects that were of interest to users, together with some taxonomical classification data denoting the type of an object like attraction, restaurant or hotel also including more detailed categorization labels like Italian or Asian restaurant or *** or **** accommodation. In
addition, properties derived from the objects’ spatial relationships were used.

Again, the operator $\epsilon$ determines the pairs of similar users for a user $u \in O_{UM}$ and their degree of similarity. To determine collaborative user similarity a set distance measure is required. Here, the Dice coefficient [7] is used. The similarity measure (function $\text{sim}(x,y)$) has to exceed a threshold in order for user $y$ to be in the neighborhood of $x$. The threshold $t$ can be either static or reflexively set to the similarity value of the $n$-th most similar user.

$$u' = \{(u_i, \text{sim}(u, u_i))|u_i \in O_{UM} \land \text{sim}(u, u_i) \geq t\}, \text{ where}$$

$$\text{sim}(u, u_i) = \frac{2 \times |u_i^{R_{UM}} \cap u^{R_{UM}}|}{|u_i^{R_{UM}}| + |u^{R_{UM}}|}$$

Thus, rule mining is applied only to a user’s nearest neighbors and rules are weighted according to the determined user similarities.

$$\text{rules}(u', u) = \{A \rightarrow B | A, B \subseteq P_{UM}\}$$

$$\text{weight}(A \rightarrow B, u) = \sum_{i \in 1..|u'|} \text{weight}(A \rightarrow B, u, u_i)$$

$$\text{weight}(A \rightarrow B, u, u_i) = \begin{cases} \text{sim}(u, u_i) & \text{if } A, B \subseteq u_i^{R_{UM}} \\ 0 & \text{else.} \end{cases}$$

Furthermore, implication rules will usually be restricted to reasonable patterns like for instance restricting the condition part to properties that can appear in user models of new users like their entry points to the map or their search requirements. If rules are maximally restricted to consider only object identifiers, the approach degrades to a traditional neighbourhood-based collaborative filtering algorithm.

Finally, the constraint-based recommendation paradigm returns those items that satisfy the largest share of all applicable personalization rules. Thus, the recommendation function $\text{rec}(u, o)$ returns a score normalized to the interval $[0 \ldots 1]$ for item $o$ and user $u$. Note, that objects $o \in O$ derive from formal contexts $\Gamma(O, F, R)$ as defined in Subsection 4.1.

$$\text{rec}(u, o) = k \times \sum_{r \in \text{rules}(u', u)} \text{weight}(r, u) \times \text{sat}(r, o, u)$$

$$k = \frac{1}{\sum_{r \in \text{rules}(u', u)} \text{weight}(r, u)}$$

$$\text{sat}(r, o, u) = \begin{cases} 1 & \text{if } r \equiv A \rightarrow B \land \neg (A \subseteq u^{R_{UM}}) \lor B \subseteq o^{R} \\ 0 & \text{else.} \end{cases}$$

Note, that $\text{sat}(r, o, u)$ denotes that the implication $r$ is satisfiable for item $o$ and user $u$. Hence, $\text{sat}(r, o, u)$ can only be true if $r$ is not applicable to $u$ or if $o$ possesses the properties in the consequent part of the propositional implication.

5. EVALUATION AND DISCUSSION

The goal of the exercise was to demonstrate the applicability of the proposed recommendation approach by conducting an evaluation using a dataset containing historical user interactions from a Google Maps application in tourism. The dataset was collected from the Carinthian destination platform kaernten.at over a period of 12 months. The dataset contains log data from almost 1000 anonymous user sessions that investigated the geographic neighborhood of an object. Users were offered the complete functionality of Google Maps like zooming, switching between different map panes and requesting details for visualized objects. Therefore, user models contain the identifiers of at least one entry object and several associated objects that users showed interest in. Furthermore, property information on visualized objects and additionally extracted properties from the spatial proximity of objects as described in Subsection 4.1 was also included.

Figure 4 depicts the computed property suitability for water sports, where intensity denotes the sum of weights of objects in the neighborhood that are tagged with water-related activities. When comparing this depiction with a topological map of the Austrian province of Carinthia one can easily determine that the emphasized areas correspond to the major lakes where most of water-related tourism takes place. In order to apply the mining and recommendation approach the property values were discretized to a binary value using the overall average, i.e., the property value was only set to true if the weight was above average. The evaluation’s goals are, first, to provide evidence that the proposed collaborative constraint-based meta-level recommender (MLCF) can be applied to data derived from real user interactions in a practical setting and, second, to demonstrate an improvement over standard collaborative filtering (CF) that does not consider any additional content information about objects except their identifiers and is therefore considered as the baseline method.

5.1 Methodology

The evaluation followed the procedure for offline experimentation as described for instance by Herlocker et al. [13]. We used a leave one out approach, where all user models other than the one being probed (i.e simulating the new user) were used to derive association rules. The properties of the tested user were split into a learning set (LS) and a testing set (TS). Essentially, the identifier and additional object properties of the entry object were used to find similar users and the identifiers of all other objects the user showed interest in constituted the testing set (TS) to be predicted. The properties of the testing set objects were not considered in this step because they are only relevant for association rule mining.

The average testing set size was found to be 2.18 and therefore we varied the recommendation size $n$ from 1 to 4, meaning that the items with the $n$ highest recommendation
scores for the probed user constituted the recommendation set (RS). Thus, successful predictions (aka. hits) were defined as follows:

\[ \text{Hits} = RS \cap TS \]

The accuracy of recommendations was computed using the popular Precision (P) and Recall (R) classification metrics [23, 13].

\[ P = \frac{|\text{Hits}|}{|RS|} \]
\[ R = \frac{|\text{Hits}|}{|TS|} \]

Note, that we are actually using Precision@n and Recall@n, where the number n equals the cardinality of the recommendation set \( n = |RS| \), i.e. the recommendations made to the user. We consider Recall@n to be more relevant in our situation as 100% Recall indicates that all items of the testing set were correctly predicted with \( n \) recommendations while the maximum achievable Precision value depends on the number of items in the testing set. Another important metric is User Coverage (Ucov) that is defined as the share of users from the overall dataset that received a non-empty set of recommendations during the trial [13]. Thus, User Coverage focuses on an algorithm’s ability to actually make recommendations.

5.2 Results

Table 1 gives an overview of the collected historic user sessions. Clearly, the average sizes of learning and testing sets were rather low. As both approaches are collaborative in nature by requiring a user neighborhood to make propositions, User Coverage remained constant for both (see Table 1).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sessions (users)</td>
<td>977</td>
</tr>
<tr>
<td>Avg. size of LS</td>
<td>1.82</td>
</tr>
<tr>
<td>Avg. size of TS</td>
<td>2.06</td>
</tr>
<tr>
<td>Ucov</td>
<td>77 %</td>
</tr>
</tbody>
</table>

Table 1: Dataset Characteristics

The accuracy of the results is presented in Figures 5 and 6. Obviously, Precision decreased with increasing recommendation set size while Recall increased. Although, the MLCF algorithm variant outperformed the standard CF algorithm by up to 2% (or 14% in relative terms) with respect to Recall@3 in this experimental setting, the corresponding Precision@3 is only marginally worse. This can be explained by the fact that MLCF adds additional items to the recommendation set by mining more general association rules that cannot be retrieved using standard CF. However, these additional recommendations are less precise. Due to the small sizes of learning and testing sets the performance increase of MLCF peaks for a recommendation set size of 3.

5.3 Discussion

Although the evaluation achieved satisfying Precision and Recall values in this map-based interaction scenario it can only be considered as an initial proof of concept. Further studies into how personalized filtering of map content is received by users will have to be conducted. In particular, the presentation of recommendation results have to be researched. Should the depiction of recommended objects on a map merely be visually emphasized or should items that are not considered to be relevant for the current user be filtered out? How should the system approach objects that are deemed to be relevant to the user but are located outside the map’s current pane? Should it recommend a navigation direction or automatically zoom in/out?

In this context some research has been conducted to research the effects of different conversational moves of travel recommender systems [16]. Mahmood & Ricci apply a reinforcement learning strategy to adapt the recommender’s interaction strategy to help users achieve their goals more efficiently. However, we expect to obtain results from a study observing users interacting with the upcoming version of the karten.at platform which includes, for instance, semantic annotations for biking and hiking GPS track data derived from several Web 2.0 tourism platforms. Currently, these annotations are exploited to adaptively rank different tourism regions based on the activities the user is interested in. Figure 7 sketches the application’s user interface. After users enter their preferred activities, they are presented with an adaptively ranked list of regions that may be further explored geographically to build their travel plan.

6. CONCLUSIONS

We presented an integrated scenario that can harness data from Web 2.0 platforms to derive semantic annotations and applied data mining and recommendation techniques in order to provide effective personalization services for users interacting with WebGIS.

The applicability of the approach was demonstrated with an experimental offline evaluation that was performed on historical log data from an e-tourism WebGIS and provided acceptable accuracy. In addition one aspect of the work,
namely the extraction of semantic annotations based on geographic proximity, has already been fielded. Therefore, live-user experiments and their evaluation will be part of the authors’ future work.

7. ACKNOWLEDGMENTS

The authors would like acknowledge the support of FFG grant MAPREC nr. 814294 - Map-based recommendation services for e-tourism applications. Logo Geoinformationssysteme GmbH, Kaernten Werbung Marketing & Innovationsmanagement GmbH and Alpen-Adria-Universitaet Klagenfurt were consortium members in this project. Furthermore, we wish to thank Alexander Seebacher, Wolfgang Schmid, Stefan Wagner, Guenther Repitsch and Andreas Wurzer for their support in system implementation and data gathering as well as Arthur Pitman for proof reading an earlier version of this publication.

8. REFERENCES


