RECOMMENDER SYSTEMS AND REPOSITORY SEARCH: THE SHARE.TEC PROPOSAL

P. Boytchev[1], A. Grigorov[2], L. Sarti[3],
A. Georgiev[4], K. Stefanov[5] and M. Chechev[6]

[1,2,4,5,6] Faculty of Mathematics and Informatics, Sofia University, Bulgaria
[3] Institute for Educational Technology, National Research Council, Italy
[1,2,4,5,6] {boytchev, Alexander.Grigorov, atanas, krassen, milen.chechev}@fmi.uni-sofia.bg
[3] luigi.sarti@itd.cnr.it

Abstract: *This paper presents a general overview of possible approaches for implementing recommender systems and describes a specific implementation in the context of the Share.TEC project, which aims to foster the sharing of digital resources in the Teacher Education field. An outline of the main functionalities is given, together with a brief technical description of how these have been implemented.*

Keywords: *recommender system, adaptability, digital library, user profile, user interface, search portal, education*

СИСТЕМИ ЗА ПРЕПОРЪЧВАНЕ В ЦИФРОВИ БИБЛИОТЕКИ: ПРИМЕР ОТ ПРОЕКТА SHARE.TEC

доц. П. Бойчев[1], доц. А. Григоров[2], проф. Л. Сарти[3],
ас. А. Георгиев[4], доц. К. Стефанов[5] и ас. М. Чечев[6]

[1,2,4,5,6] Факултет Математика и Информатика, СУ „Св. Кл. Охридски", България
[3] Институт по Образователни Науки, Национален Изследователски Център, Италия
[1,2,4,5,6] {boytchev, Alexander.Grigorov, atanas, krassen, milen.chechev}@fmi.uni-sofia.bg
[3] luigi.sarti@itd.cnr.it

Анотация: *В тази статия се прави общ преглед на възможните подходи за реализация на системи за препоръчване и се представя конкретната разработка на система за препоръчване в рамките на проект Share.TEC, който цели подпомагане на процесите на споделяне и съвместно използване на дигитални ресурси в контекста на обучение на учители. Описани са основните функции на системата за препоръчване, и е дадено кратко обяснение как те са реализирани на практика.*

Ключови думи: *системи за препоръчване, приспособимост, цифрова библиотека, профил на потребител, потребителски интерфейс, портал за търсене, обучение*

1. Introduction

The Share.TEC[1] project is building an advanced user-focused system designed to foster a stronger digital culture in the field of Teacher Education (TE). The aim is to help users across Europe search for, learn about and exchange digital resources of various kinds, and also to share experience about the use of those resources.

Key aspects of Share.TEC's mission are that the system should have (a) adaptability capabilities to support diversity and (b) recommending functions to facilitate, support and enrich the user experience. The latter are the prime focus of this paper.

---

[1] Share.TEC - SHAring Digital REsources in the Teaching Education Community, eContentplus programme (ECP 2007 EDU 427015), http://www.sharetecproject.eu/.

Share.TEC is co-funded under the EU's eContentplus programme, which stipulates that, wherever possible, projects should develop services by reusing existing (preferably open source) technological solutions. This policy is the foundation for many of the implementation choices adopted in Share.TEC and described in this paper.

A key element of the Share.TEC architecture is the Share.TEC web portal, which allows users to query and annotate a common repository of metadata referring to a shared corpus of TE related digital resources.

The Share.TEC portal[2] has been developed to allow automatic personalization of the interface so as to match the user's language, area of knowledge, preferences, community role, and history. All queries can be performed in the project partners' native language (Bulgarian, Dutch, English, Italian, Spanish, and Swedish) and in reference to the user's own context via:

- simple and advanced range of parameters for query filtering;
- key parameter values automatically set in accordance with user profile (override available);
- allowance for multiple values to be attributed to each parameter.

On the basis of the personal profile, analysis of system history and user behaviour, the Share.TEC system offers a rich set of adaptation features, including a recommender system. This provides: (1) automatic ranking of search results in accordance with user profile characteristics; (2) comprehensive recommender hints, as well as individual user tagged bookmarks and annotations; (3) personalized profile-specific form for adding feedback (rating and experiential annotation).

The recommender systems (RS) provides support for the effective selection of interesting content from a huge set of elements, drawing on the opinions of a community of users (Herlocker *et al.* 2004). Recommender systems have been developed in the field of e-commerce to assist users who, when looking for interesting products or useful information, receive hints and advice together with the bare results of their queries. The suggestions may rely on product features or on community behaviour. Some RS incorporate explanations, i.e. on request they can explain the reasons for a particular suggestion, in order to increase their level of trustworthiness.

2. Simple theoretical background

RS rely on a number of techniques (Santos & Boticario 2008) including *collaborative filtering* and *information-based techniques*. Collaborative filtering can be *user-based,* where users who chose the same items are assumed to have the same preferences; *item-based,* which assumes that items rated similarly by users are probably similar; and *stereotype-based,* where users with similar attitudes are matched. Information-based techniques comprise *case-base reasoning* (e.g. if a user likes certain items, she will probably like items that share the same attributes) and *attribute-based techniques* that match item attributes to the user profile. The best results are usually obtained by *hybrid* techniques, where two or more of the above approaches are combined.

RS are often evaluated on the *quality* (*accuracy*) of their suggestions. By the same token, it can also be useful to measure how often the system leads its users to make unsuitable choices. Users can express the perceived usefulness of the RS indications by answering explicit questions ("was this item worth exploring?"). The system can provide accuracy-related information by counting the number of times a suggested resource has actually been downloaded or visited. In any case, the bottom-line measure of RS success should be user

---

[2] Share.TEC Web Portal, available at http://sharetec.it.fmi.uni-sofia.bg, visited May 2010.

satisfaction (Herlocker *et al.* 2004). An error rate of about 15% is generally considered satisfactory performance for a well-tuned RS.

RS base their functionality on a *user model* (UM) aimed at representing user characteristics and *needs*. A *need* is a constraint posed on the value of the resources that the user is likely to be seeking. It can be used as a default value for querying the repository and identifying those resources that fulfil that need.

A widely used approach is to represent user stereotypes, i.e. a set of relevant profiles (defined *a-priori*) that represent "typical" users: individual, concrete instances will be compared against these stereotypes to identify the (approximate) best match.

The main use of the recommender system is to provide support to the query system: the RS can identify and suggest resources that, although not exactly matching the specific query conditions, are nevertheless "close" to the profile of both the user and the community she belongs to. This might be obtained by loosening the query, and is especially useful when strict adherence to the query filters would provide few or no results. This approach can also be used to sharpen the focus of an underspecified query that would provide too many results: in this case the RS influences the ranking of the resources, as they are ordered according to their closeness to the requirements of the user/group profile.

The RS can also be used when the user enters the system by displaying various kinds of information blocks, like:

- resources recommended according to the user profile;
- notifications (another user has annotated the current user's resources, or a new version of a resource that the current user previously visualized has become available, etc.);
- resources that are recommended to all the users or to specific user categories (e.g. content providers) by the system administrators.

Finally, recommending can also take place when the user shows some interest in a specific resource: along with the full description of the selected resource (including the link to the contents) the user is notified about similar resources ("people who viewed this resource also viewed …").

RS usually rely on representations of individual user profiles that also incorporate data from previous interactions with the system (queries, tagging, annotations, etc.). In this respect, the user model should take into account factors like:

- user characteristics;
- user history;
- user query behaviour and preferences;
- simple and advanced range of parameters for query filtering;
- social bookmarking;
- existing groups.

3. Main functions of the Share.TEC recommender system

The Share.TEC recommender system provides a set of functions in order to help the user locate the most appropriate Teacher Education (TE) resources according to the user profile and history of activity.

The *ranking* function provides personalized ranking of search results. It can use several different ranking schemes, e.g. numeric values in some range (say 1-100), a star ranking system, bars for ranking, etc. The user is able to select the type of ranking to be used. The desired ranking scheme can be selected in the user profile or chosen explicitly for any given search.

The *relevance indicator* function is attached to each result and shows how relevant it is to the query. The indicator can be expressed in the interface as a percentage, a "progress bar", or a colour legend. Optionally, the user can vote on whether the relevance rating is accurate.

The *recommended tag* function accompanies each search result, denoting that it is recommended by other users with similar interests. Several options are available:

- recommended because clicked by many users;
- recommended because clicked by similar users;
- recommended because of closeness to the user profile.

The recommendation can also be a combination of the above items.

The *recommendations* function is related to the capacity of the recommender system to recommend different types of resources, which may be displayed on the user's home page, e.g.:

- the 10 resources ranked most highly by users of the same stereotype;
- the 5 most viewed resources;
- the 10 users who are closest to the current user based on her preferences, or on the history of searches/used resources, or on some other criterion.

Explanations are descriptions of how the system motivates its automatic adaptive decisions, e.g. broadening search criteria because of membership in a specific stereotype or similarity group; or explaining why a piece of news is shown to the user. This functionality relies on collaborative filtering, information-based techniques and on hybrid techniques.

Feedback is explicitly provided by end users and is used for evaluating the quality (accuracy) of the suggestions given by the RS. When it is taken into account together with the user history, a clear picture emerges of how the system is used.

The recommender system provides and supports several social network functions. The *Find similar users* function is the process of finding users with similar interests and behaviour. This uses the stereotypes (or similarity rings, which group users according to their proximity to one another), although the interface will hide all the complexity of this process and provide a user-friendly way of searching. The system also provides capabilities for searching for other users according to various parameters, depending on who is looking and what initial information is available. For example, the user may search for a specific user by her username or e-mail address, or for a group of users in the same country, city, or school type, or users with similar areas of interests.

Find items by similar users is a function enabling search for digital content that is considered valuable by users with similar interests and behaviour.

And, finally, the *news about users* function announces new users who have similar interests to a given user. It is as if a user can set a trigger to "notify me if someone similar to me has joined Share.TEC".

4. Implementation of the Share.TEC RS

The recommender system is composed of several modules (or services) that control different aspects of adaptability. The *customization* module is responsible for adapting the user interface. This includes (but is not limited to) language preferences and layout/colour schemes. The module for *smart filtering* provides additional filters depending on user interests (explicitly stated in her profile and implicitly deduced from her activities). This module also provides default or initial values for some filtering criteria. The *ranking* module affects the sorting order of the retrieved results, promoting records that are likely to reflect the user's intentions.

This section briefly describes the implementation of the Share.TEC recommender system. It is based on the Apache Mahout recommender and uses the Solr search engine.

Using Mahout to build a recommender system

Apache Mahout[3] is an open source project by the Apache Software Foundation whose primary goal is to create a scalable machine-learning library. Mahout contains implementations of algorithms for recommender engines, clustering and classification. Furthermore, it uses the Apache Hadoop library to enable Mahout to scale effectively in the cloud[4] (Owen & Anil 2009). Mahout's primary features are:

- Taste, which is an open source project for Collaborative Filtering (CF) started by Sean Owen on SourceForge and donated to Mahout in 2008;
- several Map-Reduce enabled clustering implementations, including k-Means, fuzzy k-Means, Canopy, Dirichlet, and Mean-Shift;
- distributed Naive Bayes and Complementary Naive Bayes classification implementations;
- distributed fitness function capabilities for evolutionary programming;
- matrix and vector libraries.

Mahout provides tools for building a recommendation engine through the Taste library – a fast and flexible engine for collaborative filtering. Taste supports both user-based and item-based recommendations and comes with many options for implementing recommendation services, as well as interfaces that make it possible to define your own recommending system. Taste consists of five primary components that work with Users, Items and Preferences:

- *DataModel* – storage for Users, Items, and Preferences;
- *UserSimilarity* – interface defining the similarity between two users;
- *ItemSimilarity* – interface defining the similarity between two items;
- *Recommender* – interface for providing recommendations;
- *UserNeighborhood* – interface for computing a neighborhood of similar users that can then be used by the recommenders.

Figure 1 shows a diagram of the relationships between various Taste components in a user-based recommender for the Share.TEC system.

Making a user-based recommender involves a number of steps: create a DataModel, define UserSimilarity, get a neighborhood of users, create a Recommender and get recommended items. Similar steps are needed to create an item-based or custom recommender.

---

[3] Mahout project home page, available at: http://lucene.apache.org/mahout/, visited May 2010.
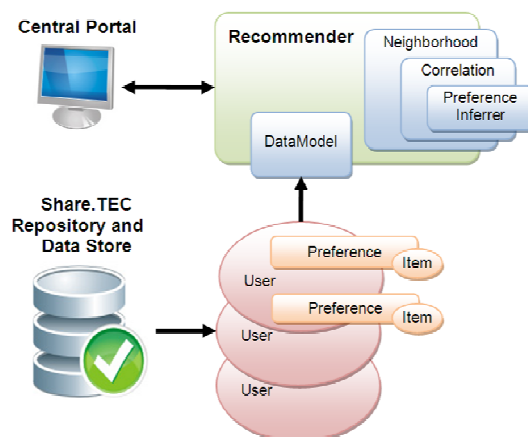[4] Ibid.

Figure 1 Relationships between Taste components in a user-based recommender

Data collection, indexing and processing

In the Share.TEC system the recommender system and the search engine use data from various sources, like the Common Metadata Model (CMM) (Alvino *et al.* 2009a) records in the central repository, the Teacher Education Ontology (TEO) (Alvino *et al.* 2009a; Alvino *et al.* 2009b), and explicit and implicit data drawn from user profiles. The CMM records and TEO are stored in the Share.TEC central repository, while the user profiles, resource comments, ratings and tags are stored in the database of the Share.TEC web portal.

The Solr search engine[5] is configured to use indices on CMM fields and TEO. The indices are built incrementally and updated whenever a CMM record is added to the repository.

The Share.TEC web portal keeps track of user activities – resources viewed, rated, tagged and commented, search queries, etc. These logs need to be processed in order to build the implicit user profiles. This is implemented by creating a special job in Hadoop that works asynchronously with the database of the web portal. Other jobs in Hadoop are defined to process user profiles, compute the distance between users applying the metrics (Tosato 2009) and compute the user stereotypes/rings.

Recommending

Recommending to all users

Some resources are recommended to all users, including those for whom no specific UM exists in the system, e.g. anonymous or not authenticated users. For example, the most frequently viewed overall metadata descriptions are presented, and the system can also display the top-rated resources.

A content-based resource recommender has been also implemented and is available to all users. When a user views a metadata description of a resource, she can click a "See More Like This" link and the system shows a list of similar resources. This functionality is implemented by using the MoreLikeThisHandler component of the Solr search engine, based on text similarity between documents. For this purpose, the current implementation uses the *title* and *description* metadata fields, but it can be configured to use other CMM fields as well, for instance *keywords*.

Recommending based on user profile

---

[5] Apache Solr home page, available at: http://lucene.apache.org/solr/, visited May 2010.

Some resources are recommended only to authenticated users, namely those with user profiles in the Share.TEC system.

The system recommends resources that match the explicit user profile (working language, knowledge area, professional area, experience area, teacher education, teacher practice, etc.). The implementation uses the Solr search engine to find the relevant resources. A query is automatically generated and it uses the Lucene boosting mechanism to boost the importance of the terms that match the user profile. The boosting factors are selected in an appropriate way so that the results ranked at the top of the list are a better match to the explicit user profile.

Recommending most-viewed resources according to user profile

In this case the system recommends the resources most commonly viewed by users close to the current user according to the user profile. The implementation adopts a user-based recommender. First, a neighborhood of the current user is computed based on the predefined metrics (Tosato 2009). The distances between the users are already asynchronously computed by a Hadoop job. Then the recommender selects resources most viewed by users in the neighborhood and not viewed by the current user.

Recommending users according to user profile

The system recommends users that are close to the current user according to her user profile. The implementation uses the pre-computed distances between the users and finds a neighborhood of the current user in the same way as for resources.

Recommending users according to viewed resources

In this case the system recommends users that are close to the current user according to the viewed resources and the metadata of those resources.

Search and ranking of search results

Search is one of the central functionalities of the Share.TEC web portal and plays an important role in the adaptability of the system. The implementation makes use of the advanced capabilities of the Solr search engine such as full text search, faceted search, ranking of search results, etc.

Faceted search gives the user the opportunity to explore the search results using a faceted classification. The user can further filter the list of query results, for example by author, subject, language, format, etc.

Another important feature of the search implementation is the ranking of search results. By default (at basic adaptability level) the search results are sorted by relevance to the search query. At full adaptability level user preferences and interests in the explicit user profile are taken into account. The query is expanded and uses the Lucene boosting mechanism to boost the importance of the terms that match the user profile. Furthermore, based on TEO, the query can be "loosened" to return relevant results (but with a lower rank). And last but not least, the system uses a resource recommender to get a list of the most-viewed resources according to user profile. So the query is also extended to boost the rank of these resources.

5. Conclusion

In this paper we have presented some different approaches for recommending systems and we have described a specific example implementation, stressing its main functionalities and how

they were implemented. The approach adopted for developing services relies on adopting existing open source technological solutions. These have been implemented in the Share.TEC system to provide specific recommending functions capable of facilitating, supporting and enriching the user experience. The current state of the system is stable and it is being used in several international pilot experiments that are due to conclude in July 2010.

References

Herlocker *et al.* 2004: Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T. Evaluating Collaborative Filtering Recommender Systems. ACM Transactions on Information Systems, Vol. 22, No. 1, January 2004, 5–53, 2004.

Santos & Boticario 2008: Santos, O., Boticario, J.G. Recommender Systems for Lifelong Learning Inclusive Scenarios. In: Zanker, M., Felfernig, A, & Burke, R. (eds.), ECAI'08 Workshop on Recommender Systems, University of Patras, 45-48, 2008.

Owen & Anil 2009: Owen, S., Anil, R. Mahout in Action, MEAP Edition, Manning Publications, 2009.

Alvino *et al.* 2009a: Alvino, S., Bocconi, S., Earp, J., Sarti, L. (2009) Share.TEC Deliverable 2.3 "Ontology and metadata models: release versions" Available at: http://www.share-tec.eu/content/1/c6/04/41/02/DEL2_3Finalreleaseversion.pdf, visited: May 2010.

Alvino *et al.* 2009b: Alvino, S., Bocconi, S., Boytchev, P., Earp, J., Sarti, L. "Sharing digital resources in Teacher Education: an ontology-based approach", In D. Dicheva, R. Nikolov and E. Stefanova (eds.), Proceedings of First International Conference on Software, Services & Semantic Technologies (S3T, 29-29 October 2009), Sofia, Bulgaria, pp. 52-59, 2009.

Tosato 2009: Tosato, P. "Metrics for Resource-User Matching", published as deliverable D4.2 of Share.TEC project, 2009, available at http://www.share-tec.eu/content/1/c6/04/41/02/17D42Metricsforresourceusermatching.pdf, visited May 2010.