SHARE.TEC PORTAL:
HOW THE METADATA REPOSITORIES OF
EDUCATIONAL RESOURCES
MEET THE USERS' EXPECTATIONS[*]

E. Stefanova, K. Stefanov, P. Boytchev, N. Nikolova,
A. Georgiev, A. Grigorov

ABSTRACT. This paper presents the process of building contemporary digital libraries with metadata for educational resources in Europe. The importance of enriching digital resources with complete, descriptive and accurate metadata is discussed as well as how these metadata are supported within the European project Share.TEC. The paper presents subsequently: the results of a workshop with teachers and teacher educators, who were asked to provide feedback about specific issues related to the metadata of digital resources for the teacher education community; design of a system in response to the user requirements: its architecture, main features, their implementation; and, finally, testing and validation of the Share.TEC portal, built as a result of a join effort of researchers, developers and users.

1. Introduction. The current Web 2.0 phenomenon is focusing on sharing digital educational resources and discussing the best way to use them. This poses serious challenges to teachers' trainers—to apply the style which will be the most flexible and feasible for teachers and to prepare them to be ready to teach and learn in the current digital society.

In order to explore the knowledge and resource sharing approach in the best possible way, current teachers need to know how to solve several general problems. Available resources are scattered and not structured well, which makes them difficult to find. There is significant lack of sufficient meta-information, which can help in the process of searching for the right information at the right time and at the right place.

Teacher education, in principle, embraces slowly innovations such as these offered by Web 2.0 tools. National educational systems are culturally bound. Teacher Education (TE) communities have local locus and sharing of digital resources at international level is scarce and still in embryonic form.

Thanks to IT the traditional characteristics of the education underwent some modifications. The modern educational model includes:

- Pedagogical characteristics and classification of the TE resources, more often in the form of ontologies.
- Education directed to the individual needs and to specific competences.
- Abilities to understand different languages and different cultures.
- Abilities to search for, analyse and reuse information in digital form.

The project Share.TEC: Sharing Digital Resources in the Teaching Education Community [1] is aiming to help teachers and teacher educators to achieve these general goals by solving several specific tasks: building an advanced user-focused system; aggregating Europe-wide metadata; providing personalized, culturally-sensitive brokerage; supporting the development of a perspective among those working in and with the TE community; a powerful and flexible recommender system taking into account the user needs and personal preferences; convenient and community-oriented Web 2.0 features such as commenting, ranking, rating, community building. In this paper we show how the project achieved these objectives and present its main results.

2. Digital libraries in education. Digital libraries are organized collections of digital content made available to the public by cultural and scientific institutions (libraries, archives and museums) and publishers. They consist of all kinds of "physical" materials that have been digitalized (books, audiovisual or multimedia material, photographs, documents in archives, etc.) and materials

originally produced in digital format. Knowledge sharing is the main function of digital libraries. It can be achieved through:

- Creation and management of digital collections
- Providing free access to leading world scientific achievements
- Sharing of digital learning resources
- Increasing the visibility and widening the influence
- Displaying the best results and products available

Metadata are the key for providing the needed meaning to the original resources, making them more transparent, easy to find and use. They are additional data which describe details about the original data. These details may include different characteristics, features, links and properties of the original data. In the past metadata were used mainly to catalogue books in traditional libraries. Now metadata are the key for searching, finding and using the right data.

There is a specific metadata standard used for the classification of educational resources—LOM [2]. There exist many digital libraries storing educational resources and most of them use LOM-based metadata profiles (specific variants including different subsets of all possible LOM fields).

3. Share.TEC model and ontology. In order to complete its tasks, the Share.TEC project uses digital metadata repositories. This section clarifies our understanding of this concept.

Common metadata model and Teacher education ontology. The Share.TEC approach is based on the use of an educational digital library for storing information about a huge number of educational resources all over the world. All metadata stored in the repository follow the **Common Metadata Model** (CMM) format ([3], [4]), which is based on the Learning Object Metadata (LOM) format. The main extension concerns the pedagogical characterization of digital content.

The main innovation of our approach is the combination of the metadata with a specific ontology. In such a way we extend the digital library with additional rich semantic information. We called this new ontology **Teacher Education Ontology** (TEO) ([4], [5], [6]). The goal was to provide a more robust, flexible and powerful way for classifying TE resources in the Share.TEC system.

TEO [7] addresses the world of Teacher Education (TE) and especially TE digital resources and practices across Europe. The ontology has a multi-layered structure, with a common top level that can be instantiated at lower levels into concrete, language-specific ontologies. These gain specificity by being contextualized in particular national settings.

The purposes of TEO are to provide a pedagogical multilingual and eventually multicultural characterization of digital content, a representation of user profiles and competencies allowing personalized interaction with adaptive applications and support for recommending functions. TEO also seeks to capture areas considered crucial for describing, exchanging, sharing, and developing resources devoted specifically to TE. Its complex structure is organised in a set of ontology branches, such as digital content, competencies, knowledge domain, contexts, actors, etc.

4. Teachers expectations and requirements analysis. In the beginning of February 2010 a workshop was held at Sofia University. It was a part of a series of European workshops dedicated to the research on users' expectations about metadata repository functionalities. The paper presents this particular workshop's context and results. The article reflects especially Bulgarian users' expectations and requirements, but similar results are summarised from other European workshops.

As the Share.TEC system is designed to serve especially teacher education, there were invited (as volunteers) teachers' educators and teachers in different areas. Nine teachers and eight teachers' educators (university professors engaged in pre- and in-service teachers training) accepted the invitation and were involved in the workshop.

The workshop had two main goals:
- To identify users' expectations for main functionalities of the Share.TEC portal.
- To test the teachers' education ontology (TEO) and common meta-data model (CMM) with practitioners in the field.

The workshop started with filling in a questionnaire through which we would like to evaluate what the expectations of potential users are in respect to digital repositories with metadata for teacher education resources. The questionnaire contained three groups of questions. The first one was on users' searching behaviour. The role of the social networks and community of practices in the teachers' continuing development were in the focus of second part. The last part was dedicated to criteria for determining appropriateness and quality of teacher training Internet resources.

Keeping in mind that there are many Web instruments for searching Internet resources, it was interesting to see which of them are the most popular and whether there is a need for new tools. The participants' responses on the question *What type of Web tools do you use when searching in Internet?* were various (Figure 1).
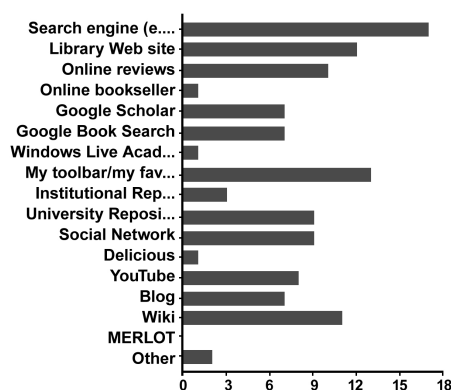
Fig. 1. Types of searching engines used

As we expected, most of the workshop participants usually use well-known searching engines.

Being asked *What types of barriers, if any, have you found while searching for digital materials?* the teachers and trainers responded:

- *Most of the digital resources are without description and I have to open the resource in order to decide is it useful for me.*
- *The information found is not classified.*
- *In many cases you have to read/view the whole material to understand whether it is applicable in your teaching situation.*

The problems discussed by participants confirm our initial belief that there is a need for classification of the resources based on an appropriate ontology. These resources should be also described through additional data—metadata.

Our accumulated impressions and experience gained during the work with teachers and their trainers gave us the ground to state the hypothesis that teachers' and trainers' Internet communities of practice have an important role in their lifelong self-education. In order to prove or reject this hypothesis we pose the second group of questions related to the use of social networks and communities of practice.

Eleven of the respondents answered "yes" to the question *Do you use Facebook or any other social network (LinkedIn, Plaxo, Xing)?*, six persons answered "no".

The participants who answered positively were asked: *Which is the feature that you prefer in your social network? Why?* The majority of the answers were:

- *They connect me with people with common interests.*

- *They help me to communicate with people all over the world and to discuss similar problems in my professional area.*
- *Sharing the amazing and useful content, communication.*

It is interesting to notice that although social networks are considered mainly for communication, the participants accept them also as a tool for sharing content and professional experience.

To the next question *Do you know what a "Community of Practice" is?* we received 12 positive and 5 negative answers.

The positive answers to the question *Have you ever been a member of a "Community of Practice"?* were 8, the negative ones were 9. Approximately the same was the distribution of answers to the question *Are you a member of a community of teachers at national or international level?*—9 positive, 8 negative.

After completing the questionnaire we were surprised that some of the participants asked about the meaning of the term *Community of Practice*. When we clarified it, the same teachers told us that they had marked the negative answer to the question, but actually they had participated in such communities.

The people who indicated positive answers to the previous two questions were asked: *Why are you a member of a professional community?* The main reasons pointed out were:

- *To share my experience and to learn from colleagues' experience.*
- *To find and share educational resources.*
- *To share and exchange experience and resources.*

Their responses showed that teachers are led by the willingness to exchange experience and resources in professional communities. All participants but one would like to work online with colleagues for problem solving they are interested in. Most, but not all of the participants would like to receive information from the community or automatic services, judging from the 13 positive and 4 negative answers of the question *Would you like to receive proposals or automatic services from the community?*

Only 12% of the people do not like to participate in professional communities, based on sharing of resources and comments. The other 88% prefer such networks and add:

- *For me as a teacher it is very important to read comments about some resources and to have a vision about their quality and possible application.*
- *It is time- and effort-consuming, or just impossible to develop/create my own materials. The colleagues' opinion is very important for me.*

The last part of the questionnaire was focused on the selection of educational materials. The workshop participants answered to its first question *What*

kind of resource information will convince you to use it? as shown in Figure 2. For them the most important are resource annotation and area of knowledge. The type of resource, author and rating are also significant.
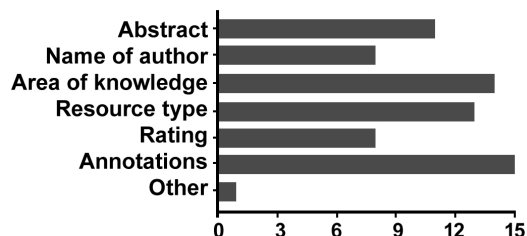

Fig. 2. Level of significance of resource selection

We received similar results (Figure 3) for the question *What kind of information contribute for understanding of resource quality?* In this case users think the most important is resource annotation as well as community rating and resource author.
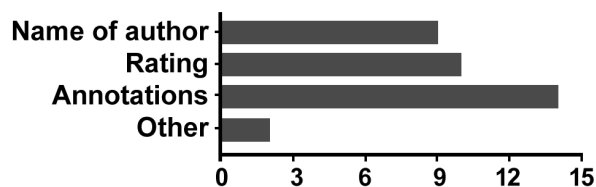

Fig. 3. Level of significance of resource assessment

The answers (Table 1) to the question *Would you like to see the choices made on the resources from your colleagues? What in particular?* show that people are influenced by others' actions and choices. The users would prefer to see a selection of people with similar interests.

Table 1

What do they see?	29%	What do they buy?	35%
What do they download?	76%	Other	12%

The analyses prove the hypothesis that teachers would like to share experience and resources and their colleagues' opinion on selection of learning materials is very important for them.

The requirements collected through the questionnaire were analyzed. The main functionalities of the Share.TEC portal were identified. With them in mind the Share.TEC system development started.

5. Share.TEC system architecture.

In order to respond to the user requirements, the Share.TEC system architecture was developed (Figure 4).
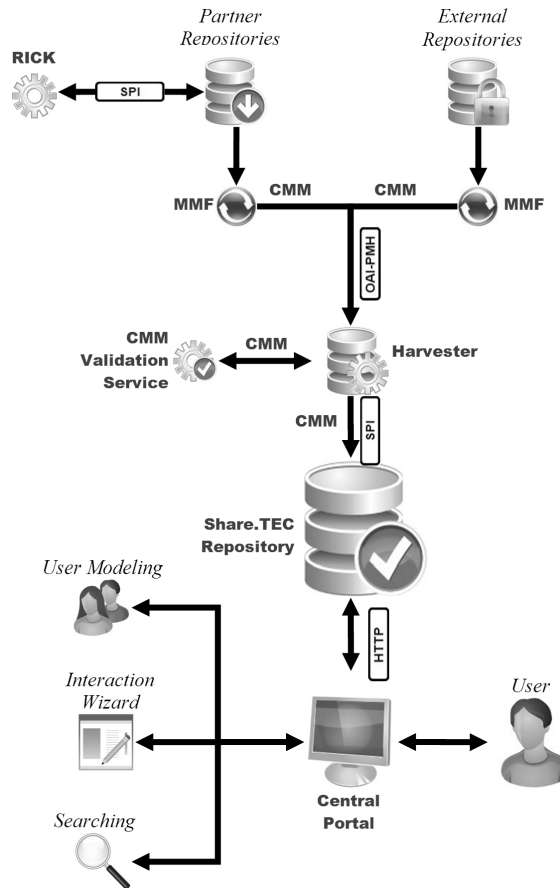

Fig. 4. Main Share.TEC architecture

The main component of the Share.TEC system is the Share.TEC **Repository**, which was also named *central repository* or *repository cache.* It contains metadata representing resources from the TE domain. These metadata are represented in the CMM format and are collected from various sources: partners'

local metadata repositories and external repositories or additional sources containing such metadata. The central repository also contains the representation of the TEO. The ontology was initially developed using the Protégé editor and later imported to the central repository.

We choose to use the Fedora Commons Repository system as a central repository (cache). The main reasons for choosing Fedora were explained in detail in [8], following a similar analysis [9, 10].

TEO is stored in the central Fedora repository using the Enhanced Content Model, which enables support for OWL Lite. The OWL description of the TEO is extracted from the Protégé tool, which was used for the implementation of the ontology, and after that imported to Fedora.

The other main component of the system prototype is the **central portal**. The central portal is the place where the user of the Share.TEC system can communicate with all of its components and to specify what services she/he needs from the system. The role of the user interface is manifold: to represent the main system services and components to the user in a natural way; to enable the user to express what she/he wants from the system; to formulate the right set of queries to the various components and services involved; to collect the results and to represent them in a natural way.

The Share.TEC Web Portal allows the automatic personalization of the portal interface to match the user's language, community role, and history. All queries can be performed in the project partner's native language and in reference to the user's own context by:

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

Automatic ranking of search results in accordance with user profile characteristics (weighting) will be also supported.

All metadata descriptions found allow previewing the corresponding resources. Comprehensive recommender hints will be offered, to which tagged individual user favourites will be aggregated.

A specific form, based on the personalized user profile, will be used for adding user feedback, such as rating and experiential annotation of resources.

The third main component from the first system prototype is **the Harvester**. The harvesting process, as chosen for implementation in the Share.TEC system prototype [11], includes *the Harvester* (program for importing metadata records taken from a number of digital repositories into one central repository),

metadata validation service (program ensuring that all the harvested metadata records have the right data format) and *OAI-PMH target* (the point where a given digital repository provides access to its metadata records through the OAI-PMH protocol).

The central repository is searching and retrieving all the sources of CMM metadata using the OAI-PMH *Harvester,* which periodically connects to all sources of metadata and using the OAI-PMH [12] protocol collects all new instances of metadata from these sources, then writes all these new metadata into the central repository. The Harvester is also responsible for the validation of the metadata through the *validation service*—it checks that all metadata conform to the CMM format, and inserts only the successfully validated ones into the central repository.

Other very important components for the system are the existing **external well-known digital repositories**, containing metadata relevant for the domain of TE domain, as well as the **local repositories** established at each project partner site. They served as the initial databases from which the central Share.TEC repository was initially populated with metadata.

To support the initial population of CMM metadata records to the central Share.TEC repository from the local partner's repositories by automated importing of existing metadata formats (for example, providing automatic translation of existing DC or LOM related metadata formats to CMM), a specific program called **MMF (Metadata Migration Facility)** was also provided. MMF mainly plays the role of translating from well-known metadata formats to CMM. Beside populating metadata records from partners' repositories, MMF will take all the metadata records coming from the well-known TE external repositories, convert them to CMM, and feed them to the central repository through the Harvester.

In order to generate new CMM metadata records at partners' repositories, we provide a comprehensive support tool for the creation of metadata records (**Resource Integration Companion Kit—RICK**). Initially, this component extracts some of the metadata fields' values automatically (such as file type, owner, date, keywords, etc). Next, this tool offers a web form for filling in or editing all fields. At the end, RICK creates new metadata records in the user's local repository, while an opportunity will also be offered to store the resource itself either in the local repository or in some other place of choice (such as Youtube, Delicious, Flickr). For the first prototype, we offer interactive filling-in with no automatic generation, and we store metadata in the local repository.

The RICK tool gives support to individual users for *creating resource records according to the Share.TEC metadata model.* Therefore, we envision

RICK as an interactive user request-driven tool (not providing full automatic metadata record creation) that employs the SPI publishing interface between RICK and partners' CMM-compliant repositories.

6. ShareTEC Adaptivity model. Adaptability is the overall capacity of a system to adapt to the user. This adaptation can be controlled explicitly by the user (i.e., customization) and implicitly by analyzing user behavior and interaction with the system (i.e., personalization). Adaptivity encompasses the collection of statistics about the user, building a user model, analyzing knowledge emerging from this model, determining how the system can adapt better to the user and, finally, the actual adaptation.

In search portals adaptability is closely related to the user model. The **user model** is a set of data describing the user and algorithms that generate and use these data. The user model is a virtual representation of a user and is used to provide adaptivity.

According to Barra, Negro and Scarano, adaptivity is "the ability to be aware of user's behavior so that it can take into account the level of knowledge and provide the user with the right kind of documents" [13]. Other authors, like Norvig and Cohn, mention adaptive systems where "much of the searching may go on when the user is not even logged in. The application does more on behalf of the user without constant interaction, and the sophistication comes from a splitting of responsibilities between the program and the user" [14]. A user model is "a representation of a user that captures goals / tasks, knowledge / background / experience, interests, traits / cognitive styles (holist, serialist) and context of work" as described in [15], and it can be empirical and analytical.

In this context three general interconnected, mutually related issues are involved: adaptability, user modelling, and user interface.

A global adaptability model incorporating these three main concepts is represented in Figure 5. Adaptability is an iterative, continuous process. It runs throughout the lifetime of the whole system. As users interact with the system, it collects more data about their preferences and goals and fine-tunes itself. The main goal of this continuous process is to converge the virtual model of a user to a model that can provide sufficient data for adequate adaptation.

The two main user activities related to adaptability in digital libraries are *setting preferences* and *searching* for digital content. When a user sets preferences (like language or topic of interest) this is stored in the model of this user containing:

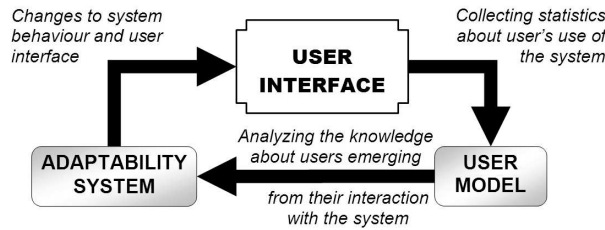- *explicit preferences* which are defined by the user and support system customization;

Fig. 5. Global adaptability model

- *implicit preferences* which are defined automatically by the system based on user behavior and preferences;
- *user behavior data* which are generated from the raw user statistics or their aggregated version.

Searching and filtering the repository produces two clusters of data: (1) a list of *search results:* the user can *browse* its elements, *open* the ones he/she is interested in, and even *comment on or annotate* resources; and (2) the *raw user statistics.* This cluster contains unprocessed data like searched keywords, found results, opened records, written comments and annotations, ratings, etc.

To make use of these data they must be *aggregated and processed* into a representation form that describes the *user behavior.*

The last set of elements in the adaptability architecture covers those related to the *recommender system.* These elements use data from the user model and decide how to accommodate the system functionality. The recommender system is composed of several modules that control different aspects of adaptability. The *customization* module is responsible for adapting the user interface. This includes language preferences and layout/color schemes. The module for *smart filtering* provides additional filters depending on user interests. The *ranking* module affects the sorting order of search results, promoting those that are likely to represent the user's intentions.

The architecture represented in Figure 6 reflects system adaptability towards *individual* users. It does not consider collaborative behavior, community interests and special interest groups.

There are various adaptability features from the user's point of view, but not all of them ought to be implemented in any digital library interface. Some of the adaptations may be inappropriate or inapplicable when factors such as the computational power, storage limitation, algorithmic complexity and data availability within a given system are taken into account.

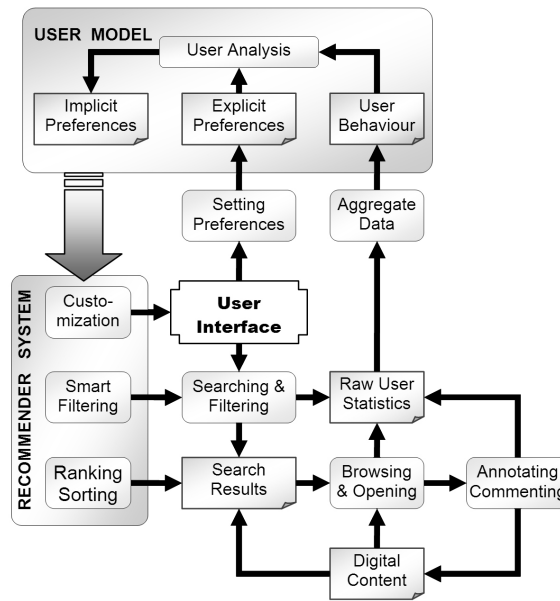The user interface for digital libraries typically uses adaptability ap-

Fig. 6. Adaptability architecture

proaches for the implementation of the following three features: (1) *adaptation of the user interface,* based on the explicit preferences given in the user profile; (2) *adaptation of search results* shown to the user and based either on explicit preferences or on the implicit preferences calculated by the system on the basis of the user behaviour and statistical data available; and (3) *specific recommendations* to the user about resources and social interactions, based on either explicit or implicit user preferences.

The adaptive behaviour of a digital library system is mainly (though not exclusively) aimed at identifying those digital resources that best suit users' needs, without asking them to enumerate the requirements in detail. Whenever possible, users are spared the trouble of explicitly expressing these needs as query parameters; the values are inferred from the user model. A tension exists between a user's general and historical characteristics and his/her momentary interests, which may vary in accordance with context shifts, either temporarily or permanently. In this perspective, it should always be possible to disregard the system suggestions, which ought to be presented in the form of *default values* for query fields that can be manually altered.

Overview of interface customization. User interface customization is an adaptability, which is explicitly controlled by the user and is stored in

the user profile. One major component of the user interface is the interface language. Many systems support different ways of changing the language by using a field *Language* or automatically depending either on the user's profile or on the language setting of the browser. Manual interface language change is volatile: it is not saved when the user logs out from the system. At the next login, the interface language will be determined by the user's personal profile setting. Automatic change in the interface language according to the profile is persistent. When the registered user specifies the working language during registration or modifies it in the personal profile settings, the selected language will be used by the system automatically for visualizing his/her personal page, regardless of the default browser language settings.

Setting the interface language automatically according to the default browser language is used by the system in cases when the user is unregistered or when the registered user has not defined her/his interface language in the profile. The system provides further adaptability opportunities by allowing the user to change personal page characteristics like text font, style, size and text and background colour. This kind of adaptability is usually called *skin adaptability*.

Adaptability regarding the user view of the information provided by the system can be done by explicit preferences. In such cases the system should have options for displaying and re-arranging different sections of the additional information on the user's home page that may be of particular personal interest.

Searching and filtering. The query system behavior of a typical digital library system can be modeled as follows:

1. A query is received, analyzed, new user model information is extracted and the query is executed in the repository, yielding a result set.
2. If the number of resources is under a pre-determined threshold, the query is transformed by relaxing some of the filters posed by the user. The additional results returned from the transformed query are ranked lower.
3. The user model is updated to take user model information into account.

The advanced query function provides users with a mask (form) to specify parameter values: these are used by the query engine to select (filter) those records that exactly match the required values. Parameters in the advanced query mask can take initial values depending on the preferences stored in the user profile, the history of the user interaction, and the information associated to the similarity ring the user belongs to (grouping of fellow-users with similar characteristics).

The user can customize the details displayed in the search results. This customization defines what fields are of interest and (only) the values of these

fields will be shown for each of the records in the result list. For example, some users may prefer to know the content provider at a glance; others may prefer to know the authors' names. Another customization possibility is to re-sort the search results according to a parameter other than relevance, e.g., date, location, etc.

The basic search field can also be personalized by placing some of the user's most commonly used search fields in a personalized basic search. Thus the basic search will be different for each user. Initially they can start with a preset list of basic filters.

Similar documents is an option that prompts the system to find other documents similar to a specific document (i.e., searching by example) by comparing only their metadata records.

Level of adaptability is a customization which allows the user to specify what level of adaptability he/she requires for the returned search results. Levels might be:

- *No adaptability*—results are returned as they are, in the fastest possible way;
- *Basic adaptability*—results are returned in accordance with metrics based only on the result metadata and the query metadata. The user's preferences and interests are not considered;
- *Full adaptability*—results are returned and ordered according to user preferences, interests and history. This option provides the most accurate ranking at the cost of performance.

7. Implementation of the Share.TEC portal. The Share.TEC portal implements these end-user functions and functionalities through the Drupal CMS portal software [16]. It allows the automatic personalization of the portal interface to match the user's language, community role, and history. All queries can be performed in the user's native language and in reference to the user's own context.

The Share.TEC portal implements the user interface to all services available. Some of these services are located and offered by the Fedora repository (central cache), others will be developed during the project and will be stored in the portal database.

The main principles of the user interface design follow the main requirements:

- The Share.TEC digital environment is tangible to users and easy to navigate.

- It is based on interactive visualizations, adapted to user profiles and current user objectives, and open up for new opportunities for collaboration and community building.
- The system provides a visual metadata search engine.
- An adaptive wizard based on the ontology model and individual user history enhances the interaction with the system and increases the effectiveness of the user's queries by making their tacit, contextualized knowledge explicit.
- The user interface is fully multilingual.

For the implementation of the user interface we use various fields, lists, hierarchies and other relevant data components from CMM and TEO. All the relations between the user interface components and the relevant CMM and TEO components were described in detail in [9].

The multi-cultural aspects are represented in the CMM and TEO models as specific multi-cultural extensions and are covered using the multilingual aspects of the user interface. These features are still in a development phase and will be further tested, evaluated and enhanced.

The adaptivity model and features described in the above sections find suitable application in the EC co-funded Share.TEC project. Share.TEC is developing a federated digital library system designed to provide culturally-aware access to resources related to the field of Teacher Education across Europe. This field is made up of people with very different backgrounds, ideas and assumptions, and with very different requirements; the community of system users is expected to reflect a variety of languages and cultures. Central to Share.TEC's mission is that the system under development should have capabilities to support diversity, and a key aspect of that support lies in adaptability.

User profile (User model). The user profile serves to provide more flexibility and adaptivity to the system. The user model is composed of three top-level components:

- *the quasi-static profile*, which includes institution, target level, language etc.; in other words, personal information about the user that is unlikely to change over a period of a year;
- *history of interaction with the system*, including ratings, query paths, annotations, resources visited, etc. This log of events grows each time the user interacts with the system;
- *counters* that synthesise quantitative elements and allow implementation of special heuristics.

After registering, the new user initially acquires "novice" status in terms of the scaffolding offered by the system; the absence of an adequate history hinders

any practical adaptability features. This status lasts only until the history size reaches a given threshold. Up to that point, the novice user inherits the characteristics of the closest fellow-users and the adaptability functions are driven accordingly.

The user model (profile) consists of some more or less static features (affiliated institution, teaching target level, language, etc.) and dynamic features such as history of interaction with the system (ratings, queries, annotations, resources visited, etc.) and some counters. For efficiency reasons the user profile is not stored in the digital library itself but in the internal user interface database. However, the user profile is closely integrated with TEO, which is stored in the digital library. User features such as Professional Area, Experience Area, Teacher Practice Context refer to corresponding TEO nodes. This enables the multi-language support of the Share.TEC system as the values associated to TEO nodes are expressed in several languages.

The dynamic part of the user profile is used mainly by the recommender system to find relevant resources and user recommendations. For details on the metric-based implementation of the recommender system see [17].

Search customization. The search component of the Share.TEC portal also employs the user profile and TEO intensively. It is based on the Solr search engine [18] and uses Lucene style queries. The search component performs semantic query expansion that consists of the following steps:

- *Initialization phase*—the query is analyzed and user profile data is extracted;
- *Expansion with user preferences*—the query is expanded with terms that match the explicit user profile (Figure 7) so that resources that match the user profile are ranked higher;
- *Ontology-based expansion*—the query references to TEO classes are moved to point to their ontological parents.
- *Multi-language expansion*—the query is analyzed to detect if it contains keywords or phrases from TEO and use their translations as synonyms to find resources in different languages.
- *Recommender-based expansion*—the resource recommender is used to get a list of the most-viewed resources according to user profile. Then the query is extended also to boost the rank of these resources.

User statistics. To achieve a level of adaptability, the system relies on the following information: explicit user preferences and profile traits; cultural context features (when available); inferred profile and stereotype. Whenever these three sources conflict, priority is given to explicit information from the user.

Fig. 7. Explicit preferences in the user profile

The user profile consists of two logical components. The first, considered here as "static", is qualitative in nature and contains mostly explicit (i.e. not inferred) data such as name, professional profile, language, country, etc., maybe including a specification of interests by means of keywords.

The second component is dynamic and quantitative, and consists of indicators associated with specific user interests. Each indicator is incremented when the user performs an operation revealing an interest. For example, the system may have counters for each of the top-level knowledge area classes: when the user uses "mathematics" in a query or annotates a resource which is linked to a math subject, the counter associated to mathematics is incremented. Each user has a set of counters and after a period of use of the system their values provide a sketch of that individual user's interests.

Multilingual support. This is completely based on the portal software features, which easily enable the user to choose a language. The system supports a number of languages for which the proper translation for the set of system names, messages, menus, etc. is provided. The system also grants multilingual support through a multilingual metadata model, which guarantees that all concepts from TEO and CMM are translated in advance to the predefined set of languages.

Recommending. The recommender system is composed of several modules that control different aspects of adaptability. The *customization* module is responsible for adapting the user interface. This includes language preferences and layout/color schemes. The module for *smart filtering* provides additional filters depending on user interests (explicitly stated in the profile and implicitly deduced from the activities). This module also provides initial values for some

filtering criteria. The *ranking* module affects the sorting order of the results, promoting those that are likely to reflect the user's intentions.

Using Mahout to build a recommender system. Apache Mahout [19] 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. It uses the Apache Hadoop library to enable Mahout to scale effectively in the cloud [19, 20].

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. Making a user-based recommender involves creating a data model, defining user similarity, getting a neighborhood of users, creating a recommender and getting recommended items.

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 [17] and compute the user stereotypes/rings.

Recommending to all users. Some resources are recommended to all users, including anonymous or unauthenticated 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 also been 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 based on text similarity between documents.

Recommending based on user profile. 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 (language, knowledge area, professional area, experience area, 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 such a way 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 [8]. 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 (Figure 8).
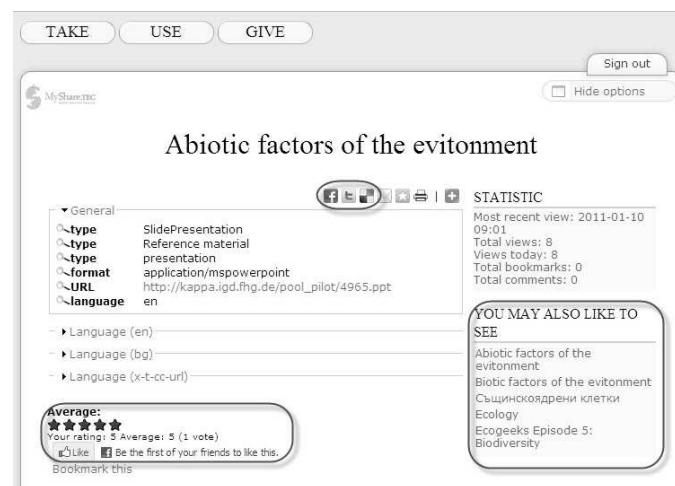
Fig. 8. Recommending according to the user profile

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. (Figure 9).
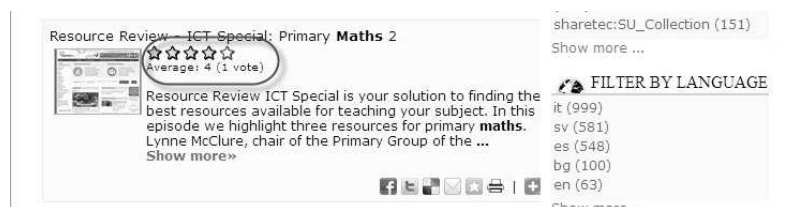

Fig. 9. Rating made by other users and filtering

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.

8. Testing and validation of the Share.TEC portal. The testing and validation of the Share.TEC portal was done in five European countries. As part of it, at the end of March and the beginning of April 2011 in Sofia 59 in-service and pre-service teachers (61% teachers), teacher's educators (8%), publishers (3%), and others (31%) tested and validated the Share.TEC portal. The participants were between 20 and 57 years old, 13 men and 56 women. They had several tasks to perform:

– Make simple search in the Share.TEC portal without registration in the system
– Register in Share.TEC portal and search with the same phase
– Use advance search in Share.TEC portal
– Rate, comment and bookmark resources chosen by them
– Create groups, join available groups and post into discussions

The feedback was collected through the Think Aloud protocol as well as through an especially designed survey. The observations of the work of the participants show that they perform the tasks without any significant problem. At the end they were happy to share with us that through searching and filtering provided by Share.TEC portal functionalities they found the resources really useful for them. For example, *Tanya*, a teacher of chemistry, was surprised that it is so easy to discover a large amount of useful digital resources, perhaps available somewhere at the end of the list, when she searched with other tools, and maybe because of that with other tools these resources are not visible. Most of the participants had the same observations. That is why it was no surprise that the question: *Would you use the Share.TEC portal in your daily life?* got more than 80% positive answers: *Yes, Share.TEC portal is easy to be use and useful*; *Yes, because it seams that shorts the path for finding the resources*; *Yes, because through it I could find a lot of information, and if I couldn't find exactly what I look for, I can ask in the community and somebody hopefully will answer me. We are happy that finally we have a possibility to collect, easily find the needed materials and to share the experience!,* one of the participants stated. The users' feedback shows that Share.TEC portal meet their requirements.

9. Conclusions. Based on the notes collected through observation on the work of the participants during the validation, the process of improvements of the user interface has begun.

Analyzing the testing and validation result, we can conclude that the Share.TEC portal provides an appropriate learning and social environment for sharing not only digital resources for teachers' education, but also knowledge and experience for self-development and building a European teacher education community. We consider the successful validation of the Share.TEC portal a result of the hard work and joint effort of researchers, developers, and users. The authors share the opinion that this is the only way to apply innovations in education in practice.

REFERENCES

[1] Share.TEC project site, `http://www.sharetecproject.eu`

[2] IEEE Standard for Learning Object Metadata 1484.12.1-2002. IEEE LTSC= `http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf`

[3] Share.TEC Project Deliverable D2.2 Common Metadata Model (CMM): version 1. `http://www.share-tec.eu/content/1/c6/04/41/02/D2_2_Common_Metadata_Model.pdf`

[4] Share.TEC Project Deliverable D2.3 Ontology and Metadata Models: release versions. `http://www.share-tec.eu/`

[5] ALVINO S., S. BOCCONI, J. EARP, L. SARTI. A Teacher Education Ontology for Sharing Digital Resources across Europe. In: Proceedings of the 5th International TENCompetence Open Workshop "Stimulating Personal Development and Knowledge Sharing", (Eds R. Koper, K. Stefanov, D. Dicheva), ISBN: 978-954-92146-5-9, 26–9.

[6] Teacher Education Ontology (TEO): version 1. `http://www.share-tec.eu/content/1/c6/04/41/02/D2_1_TEO_v1.pdf`

[7] ALVINO S., S. BOCCONI, P. BOYTCHEV, J. EARP, L. SARTI. Sharing Digital Resources in Teacher Education: an Ontology-based Approach. In: Proceedings of 1st International conference S3T'2009 Software, Services and Semantic Technologies (Eds D. Dicheva, R. Nikolov, E. Stefanova), ISBN 978-954-9526-62-2, 116–123.

[8] Share.TEC Project Deliverable D5.4 Development of the system prototype. `http://www.share-tec.eu/`

[9] Repository Software Survey. March 2009. `http://www.rsp.ac.uk/software/surveyresults`

[10] Technical Evaluation of Research Repositories. `https://eduforge.org/docman/view.php/131/1062/Repository_Evaluation_Document.pdf`

[11] Harvesting LOM. `http://ariadne.cs.kuleuven.be/lomi/index.php/Harvesting_LOM`, last access April 2011

[12] Open Archives Initiative – Protocol for Metadata Harvesting. `http://www.openarchives.org/pmh/`

[13] Barra M., A. Negro, V. Scarano. When the Teacher learns: a Model for Symmetric Adaptivity. In: Proceedings of Second Workshop on Adaptive Systems and User Modeling on the World Wide Web, Banff, Canada, 1999.

[14] Norvig P., D. Cohn. Adaptive Software. Harlequin Incorporated. `http://norvig.com/adapaper-pcai.html`, last access April 2011.

[15] Nurmi P., T. Laine. Introduction to user modeling. University of Helsinki. `http://www.cs.helsinki.fi/u/ptnurmi/teaching/UM/` `Seminar_040907%20%20Introduction.pdf`, 2011.

[16] Drupal. `http://drupal.org/`

[17] Tosato P. Metrics for Resource-User Matching, Published as deliverable D4.2 of Share.TEC project (2009). `http://www.share-tec.eu/content/1/c6/04/41/02/17D42Metrics` `forresourceusermatching.pdf`

[18] Apache Solr (home page). `http://lucene.apache.org/solr/`, April 2011

[19] Mahout (project home page). `http://lucene.apache.org/mahout/`, last access April 2011

[20] Owen S., R. Anil. Mahout in Action. MEAP Edition, Manning Publications, 2009.

E. Stefanova, K. Stefanov, P. Boytchev
N. Nikolova, A. Georgiev, A. Grigorov
Faculty of Mathematics and Informatics
St. Kl. Ohridski University of Sofia
5, J. Bourchier Blvd
1164 Sofia, Bulgaria
e-mail: eliza@fmi.uni-sofia.bg
krassen@fmi.uni-sofia.bg
boytchev@fmi.uni-sofia.bg
nnikolova@fmi.uni-sofia.bg
atanas@fmi.uni-sofia.bg
alexander.grigorov@fmi.uni-sofia.bg