Share.TEC System Architecture

Krassen Stefanov[1], Pavel Boytchev[2], Alexander Grigorov[3],
Atanas Georgiev[4], Milen Petrov[5], George Gachev[6], and Mihail Peltekov[7]

[1,2,3,4,5,6,7] Faculty of Mathematics and Informatics, St. Kl. Ohridski University of Sofia,
Bulgaria
[3] Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria
[1,2,3,4,5,6,7] {krassen, boytchev,alexander.grigorov, atanas, milenp, gachev, misho}
@fmi.uni-sofia.bg

Abstract. This paper presents the Share.TEC system architecture and the first system prototype. This description stresses on the main system components and underlying technologies, standards and protocols. The main system functionalities are following logically from the main use cases and user requirements. The system design allows a tight integration between data, which have been automatically harvested from various external repositories, and a representation of the ontology that describes all aspects of the teacher education domain. The proposed implementation relies on modern features such as searching a net of digital objects interconnected by custom-defined relations. A multicultural system which accommodates to the user preferences requires specific user interface. The dynamic approach for presenting the information to the user, based on the teacher education ontology, contributes to the flexibility of the system.

Keywords: Teacher Education, Ontology, User Functionality, User Interface

1 Introduction

The digital libraries initiative of the European Commission aims to make information more accessible and usable in the digital environment. There are two big efforts in the area of creating large research infrastructures from digital repositories in European Union, in the frame of the DRIVER [10] and Europeana [11] research projects. The current paper presents the results from the Share.TEC project [1]. The main objective of this project is to provide a platform, called Share.TEC system, for unified access to digital resources in the domain of Teacher Education (TE).

This paper describes the first Share.TEC system prototype. The Share.TEC system [1] has the main goal to establish a highly visible and functional portal with advanced brokerage services that will provide personalised access to a wide-range of Teacher Education (TE) content. The system prototype will be further updated and improved, based on the feedback from the first pilot tests and user feedback. One of the main innovations is the combination of metadata records and ontologies for classifying, searching and browsing through the objects in the digital library.

We start with a brief analysis of the user requirements, based on the analysis of the first Share.TEC user scenarios and use cases, which resulted in the list of main system functions needed. After that we present in details the overall

Share.TEC system architecture and describe all main components used for the system implementation. We describe also the main communication protocols, used to integrate all main components into one integrated system, as well as the main software services offered from each of the main system components.

Finally, we present the main results from the first validation session of the Share.TEC system prototype, and main directions for its future development and improvement.

2 Requirements

The Share.TEC system is based on the set of scenarios and use cases specified in Share.TEC project [1].

On the base of these use cases the main system requirements were developed, which identify the main system functionalities.

We can classify all main user functions, identified on the base of the analysis of the system requirements, into several groups:

- *Account management functions*. These are functions related to creating and removing user accounts, logging in and out of the system, and password reset. They are: register, unregister, log in, log out, password reset, profile change.
- *Profile management functions*. These functions allow the user to modify his/her profile and customize the way Share.TEC presents information. They are: personal profile, custom front page, preferences, multilingual interface.
- *Searching and navigation functions*. These functions provide the ability to browse, filter and search the data which is stored at Share. TEC repository. They are: browsing (navigation), searching, filtering and alerts.
- *Authoring functions*. These functions allow the user to contribute to the Share.TEC content by providing annotations, ranking and adding references. They are: annotation, ranking, enriching and adding references.

An important issue about the user functionalities is that they are dependent on the status of the user. Generally, there are two types of Share.TEC users – registered and non-registered (anonymous). Registered users have access to more functions and they are recognized by the system as actors. Registered users could be either logged in or logged out. A logged out user that uses the system is treated as unregistered user.

3 Share.TEC system prototype – a detailed description

The first Share.TEC system prototype was developed following the general architecture as depicted in Figure 1 (including the main system components as well as the protocols used for their communications).

The main component in the first system prototype is the Share.TEC *Repository*, which was also named *central repository* or *repository cache*. The central repository contains metadata representing resources from the Teacher

Education (TE) domain. These metadata are represented in the Common Metadata Model (CMM) format [2] and are collected from various sources: partners' local metadata repositories and external repositories or additional sources containing such metadata.
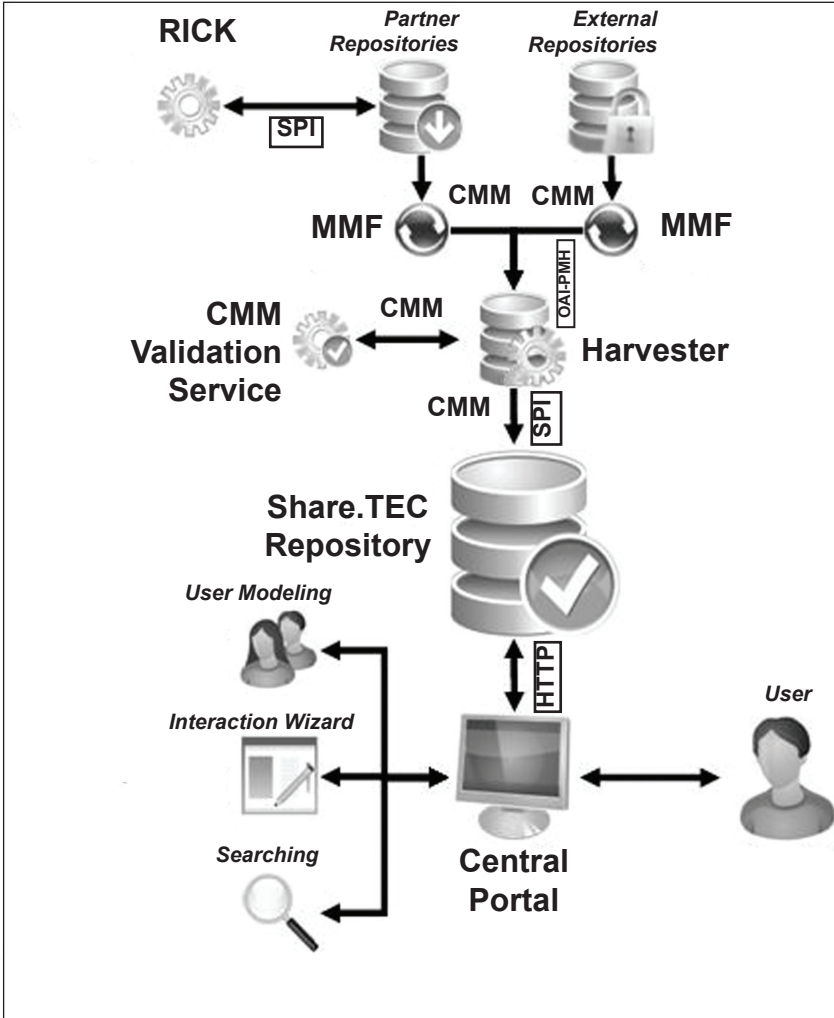

Fig. 1. Main Share.TEC architecture.

The central repository also contains the representation of the *Teacher Education Ontology* (TEO) [3]. The ontology was initially developed using the Protégé editor and later on it was imported to the central repository. The details for the representation of CMM and TEO in the repository are given in [4].

We choose to use Fedora Commons Repository system as a central repository (cache). The main reasons for choosing Fedora were explained in details in [1], following the similar analysis [5, 6].

TEO will be stored in the central Fedora repository using the Enhanced Content Model, which enables support for OWL Lite. The OWL description

of the TEO will be extracted from the Protégé tool, which was used for the implementation of the ontology, and after that imported to Fedora. When needed, additional manual editing will complement the full implementation of TEO and its features 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 multifold: 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 from these components and services, and to represent results in a natural way to the end user.

The main set of services offered to the user is related to the realization of different types of searches of varying complexity to the central repository.

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

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, like rating and experiential annotation of resources.

The third main component from the first system prototype is *the Harvester*. The harvesting process, as chosen for the implementation in the Share.TEC system prototype [7], 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 is providing 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* – special computer program which periodically connects to all sources of metadata, and using the OAI-PMH [8] protocol is collecting all new instances of metadata from these sources, and 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 whether all such metadata records conform to the CMM format, and inserts only the successfully validated metadata records to the central repository.

Other very important components for the first system prototype are the existing *external well-known digital repositories*, containing relevant for the domain of TE domain metadata, 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 is mainly playing the role of translating from well-known metadata formats to CMM. Besides 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 can extract some of the metadata fields' values automatically (such as file type, owner, date, keywords, etc). Next, this tool can offer a web form to the user for filling all of the required and missing fields, as well as any other non-obligatory field. In the end, RICK should create new metadata records in user's local repository, while an opportunity will be also 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 will offer only interactive forms filling-in (no automatic generation), and we will store only metadata records 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.

For the first system prototype we include preliminary versions of MMF and RICK.

Below we present the description of the main Share.TEC software services, grouped by the corresponding Share.TEC system component, offering them:

- Harvesting services management (accessible by system administrators)
- Adding new metadata records manually to chosen local repository via SPI
- Creating interactively metadata records to be stored in chosen local repository using the RICK tool
- Favourite service (allowing to organize digital content objects as favourites via tags from a personal user, like in del.icio.org)
- Metadata validation service (before adding metadata records to the central repository, each one should be validated)
- Searching service (search the central repository for resources based on defined queries)
- Browsing service (browsing the central repository objects and relations)

- Queries description service
- Ranking resource service (manually or automatically, based on users' profile and context)
- Presenting the search results taking into account user model, user activities (ranking etc.), history and other user interface personalization details
- Brokerage service (ensured by the User Model, built via tracking the individual user's actions and storing them for further analysis, providing the basis for weighting search results and tuning recommendations, personalising them further)
- Preview of metadata resources and (if possible) of the resource asset itself (if public)
- User authentication service
- User authorization service
- User management service
- Backup and recovery (System administration service)
- Statistics (System administration service)
- Logging (System administration service)
- Community participation Service

4 Wizard and user interface

This section describes the implementation of the user interface through the Share.TEC portal component. The Share.TEC portal implements these end-user functions and functionalities through the Drupal CMS portal software [9]. It allows the automatic personalization of portal interface to match 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 by:

- 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.

The Share.TEC portal implements the user interface to all services available. Some of these services are located and offered by 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 for the Share.TEC system follow the main requirements stated in the project proposal:

- 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 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 details in [1].

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 development phase and will be further tested, evaluated and enhanced.

5 Conclusion

In this paper we described the development of the first Share.TEC system prototype. The main URL of the system portal providing the main user interface to the system is: http://sharetec.it.fmi.uni-sofia.bg/

For the first prototype we succeed to harvest metadata from the following repositories: CENEC, OUNL, Stockholm University, SU-NIS, TCD, UVA, Teacher TV.

As a result, the main system repository contains more than 200 metadata records, describing various resources from the Teacher Education domain.

For the next system release and for the first system pilot, we expect to have also all the metadata from all the partners repositories.

During the second Share.TEC workshop held in Dublin, we performed extended testing and validation of the system prototype, in order to receive the first significant feedback from the users, and to plan all the needed system changes and enhancements for the first pilot experiment. The validation aimed to measure the reliability, availability and usability of the system, as well as to test it again several use cases designed in advance using specific think allowed protocol techniques. All details from the validation will be described in the specific project deliverable. The main preliminary results from the validation show, that the system prototype is allowing the users to perform all the planned and expected tasks with the system. Generally, the system is functional, available, reliable and easy to be used. Some specific user recommendations were recorded and will be taken into account for the development of the first system pilot version.

References

1. Share.TEC Project Deliverable D5.4 Development of the system prototype, http://www.share-tec.eu/
2. 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

3. Teacher Education Ontology (TEO): version 1, http://www.share-tec.eu/content/1/c6/04/41/02/D2_1_TEO_v1.pdf
4. Share.TEC Project Deliverable D2.3 Ontology and Metadata Models: release versions, http://www.share-tec.eu/
5. Repository Software Survey, March 2009 http://www.rsp.ac.uk/software/survey-results
6. Technical Evaluation of Research Repositories, https://eduforge.org/docman/view.php/131/1062/Repository Evaluation Document.pdf
7. Harvesting LOM, http://ariadne.cs.kuleuven.be/lomi/index.php/Harvesting_LOM, Accessed: March 10, 2009
8. Open Archives Initiative - Protocol for Metadata Harvesting, http://www.openarchives.org/pmh/
9. Drupal, http://drupal.org/
10. M. Feijen, W. Horstmann, P. Manghi, M. Robinson and R. Russell, DRIVER: Building the Network for Accessing Digital Repositories across Europe. Ariadne Issue 53 October 2007.
11. Europeana v1.0, Description of Work, http://version1.europeana.eu/c/document_library/get_file?uuid=cf720627-d351-48dc-9ef0-b10b326591d1&groupId=10602