Digital Libraries as a Social Media

Krassen Stefanov

Sofia University St. Kliment Ohridski, Faculty of Mathematics and Informatics
krassen@fmi.uni-sofia.bg

Abstract. This paper presents an innovative approach for enhancing digital libraries functionalities. An innovative distributed architecture involving digital libraries for effective and efficient knowledge sharing was developed. In the frame of this architecture semantic services were implemented, offering multi language and multi culture support, adaptability and knowledge resources recommendation, based on the use of ontologies, metadata and user modeling. New methods for teacher education using digital libraries and knowledge sharing were developed. These new methods were successfully applied in more than 15 pilot experiments in seven European countries, with more than 3000 teachers trained.

Keywords: adaptability, adaptive systems, user modeling, recommender systems, ontology, metadata.

1 Introduction

Lifelong competence development is one of the main topics of the current society. Current research is trying to find the best ways to support it using various technological and organizational frameworks based on current standards and innovative technologies. Such frameworks demand a kind of knowledge management system for searching, sharing and exchange of needed information resources. In such a way individual users can share their own information resources with other users and form different social groups. Different solutions exist to solve this problem. On one end we can see big digital libraries with their centralized services, while at the other end of the spectrum are decentralized networks like Peer-to-Peer networks (P2P), which enable users to share resources without losing control on them. In the middle we see various Web 2.0 applications like Facebook, Flickr, YouTube and many others, enabling users to share their information resources with high level of trust and motivation, making them very strong competitors to big institutional digital libraries. The integration of all such systems at the moment is rather mission impossible, because of the lack of suitable standards, which is a serious obstacle for the free exchange and sharing of knowledge and information resources.

The ideal solution demands development of new information models extending digital libraries with new services and software components and following new standards, enabling free development, storage and sharing of knowledge between all users. This new model should enable the unification of information and knowledge in all

existing systems: digital libraries, learning management systems, P2P networks, and social networks.

In this paper we will describe such an information model and its implementation, and present experiments with real users. We will discuss the main results obtained and will formulate ideas for further improvement.

The paper is organized as follow. First we show some similar ideas and implementations. Then we describe the model in details. Later on we stress on the implementation of the model, and present the experiments made with the implemented system. At the end we analyze the main results obtained, and formulate the main conclusions and ideas for improvement of the system.

2 Similar Research

The main features of the proposed model are adaptability, user and domain models, recommendations, social interaction. We will investigate shortly how these features were implemented in systems closed to digital libraries.

Adaptability is a key component of the Web 2.0 phenomenon. The possibility of the systems to adapt to the user needs and wishes is the key characteristic of the adaptability. Adaptive systems are best known in educational settings.

Adaptive learning proposes the learning process to be oriented to the current needs of each particular learner, and the learning system to be able to support each individual learner differently, depending of her/his abilities, performance, drawbacks. According to [1], adaptive learning involves three components: a model of the content to be learned (a content model), a means of understanding student abilities (a learner model), and a method of matching the content to the student (an instructional model).

Adaptive learning systems are web-based [1] and they implement adaptability in relation to content presentation, user interface, learning activities and problem solving. They take into account the context in order to provide the needed adaptation.

User modeling is one of the key dimensions for providing adaptability. User modeling is also best known in learning management systems [2]. The first and classical meth-od is known as “overlay user model”, and it is considering the particular individual user model as a subset of another ideal super-user model. In learning management systems various approaches for automatic construction and support of such model exist [2]. Depending on how we collect information in order to construct the model, we can distinguish explicit methods (when usually the user enters the information) and implicit methods (systems are trying to collect information needed from various existing sources). Implicit methods are more widespread, because they are not demanding extra efforts from users. They involve watching users’ activities (what resources the user is accessing), collecting statistical data about user’s history and how the user is using the system, and analysis of all collected statistical data in order to guess how the regular user will react in each particular situation.

Another key function of adaptive systems is their ability to recommend to each individual user the best matching to her/his profile knowledge, information, resources, activities and other users and social groups. Detailed analysis of various learning rec-

ommendation systems can be seen in [3]. Santos and Boticario [4] offer semantic model for recommending, based on the learning management system context. Here we should stress that learning recommendation systems differ from standard recommendation systems, which first were implemented for e-commerce systems domain. The main difference is related to the specifics of learning systems, where each individual user has own specific learning goals, which the recommender system should support. Development of suitable software tools for specifying the relevant recommendations in different settings is one of the key characteristics of the recommendation systems.

3 Combined Adaptability Model for Digital Libraries

Before we give the description of the model, we will make some short introduction to digital libraries domain. According to different definitions [14], the current digital library includes the following components:

- digital content (scientific, cultural, educational, etc.) and other relevant knowledge objects in digital form, accessible with information technologies;
- objects in the digital library are organized in thematic hierarchies, classifications or ontological schemes;
- all objects are seamlessly described with proper metadata standard, according to the family OAI-PMH (The Open Archives Initiative Protocol for Metadata Harvesting).

Digital content may include any forms of digital information corresponding to the respective physical objects, obtained through some kind of digitalization. The main functions of the current digital libraries are to share knowledge through:

- digital collection creation and management;
- free access to leading scientific results;
- sharing of learning resources;
- widespread dissemination of knowledge;
- best products and results demonstration;
- fast search and access to relevant information.

Metadata are the key for the implementation of all these functions, providing relevant information for the meaning of the original knowledge objects, making them more transparent and easy to find and use. These metadata provide additional data characteristics, features, relations and properties. Combining metadata with more complex knowledge representation structures like ontologies we can obtain huge information stores with abilities to classify and catalogue all the relevant knowledge resources in a given domain.

The proposed model is based on the combined use of metadata and set of specific ontologies. In particular, they were applied in the teacher education domain. The Teacher Education Ontology - TEO [9, 10, 11, 12, 13, 14] has a complex structure with several levels related to:

- digital content;
- competencies related to the user domain;
- domain knowledge area;
- activities and contexts related to the domain;
- user interaction modes and activities related to different contexts.

In order to provide all the main services mentioned in previous chapter, we designed the model [8, 9] to involve three main components – adaptability, user modeling and user interface modeling. The model is shown on Fig. 1 and is structured around these three key components, taking into account and presenting in more details all the specific processes and data they involve. Adaptability is a constant iterative process, continuing during the whole life cycle of the system.

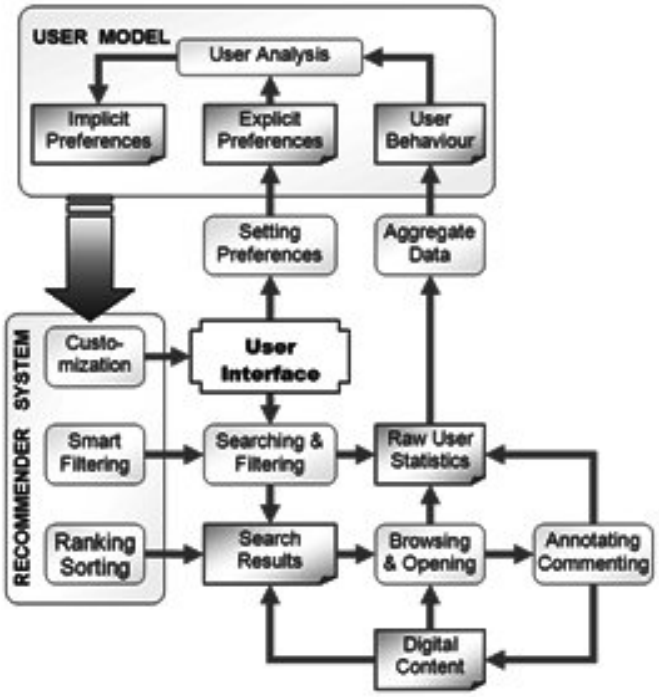

Fig. 1.

Each user is connected with her/his own model, containing three clusters of data:

- Explicit data preferences provided by the user, used for adjust the system to the user;
- Implicit user data preferences extracted from the user interaction data with the system and automatically generated;

- General implicit data for the behavior of all users, generated by the system on the base of the analysis of all the individual implicit data accumulated.

The recommender system includes several modules for managing different adaptability aspects. The fine tuning module is responsible for the adaptation of the user interface. This includes the preferred language, color scheme and interface blocks customization. The intelligent filtering module provides additional filters for fine tuning the system (either stated explicitly or derived from the user behavior). This module also initializes the values for the filtering criteria. The rating module influences the order of data to be shown to the user after different operations like searching, browsing, etc. (moving forward the items that match better the user preferences). The recommender system designed is offering a set of functions and services to help the user to find the most appropriate for her/him knowledge and information resources, according to the user profile and history of her/his interactions with the system.

4 Implementation of the Model

The proposed model was implemented in the Share.TEC project [10]. This implementation allows the Share.TEC system architecture to provide innovative semantic services for multilingual and multi-cultural access with maximum adaptability features. Using innovative knowledge presentation tools like metadata, ontologies and user models allows the system to provide powerful recommender system services as well. These services also use the analysis of individual user interactions with the system as well as the generalized statistical data about overall system performance in different contexts. As a result the Share.TEC system provides:

- Automatic ordering of search results according to the user profile
- Detailed user recommendations for information resources, as well as for social interactions (user groups, relevant users, forums, discussions, etc.)
- Ability for the user to provide additional information and preferences about information resources, user interactions, social interactions and others.

The implementation is based on the use of the open source framework Apache Mahout, in combination with the open source search engine Solr [5]. The Apache Mahout system [6] offers several important recommendation algorithms and for the clustering and classification of big amounts of data. It uses another open source system - Apache Hadoop – for maximum effectiveness of applications of these algorithms on big data sources in distributed networking environment [15]. The Mahout system includes also several clustering algorithms based on the Map-Reduce model (k-Means, fuzzy k-Means, Canopy, Dirichlet, and Mean-Shift), classification algorithms (Distributed Naive Bayes и Complementary Naive Bayes), evolutionary programming algorithms and libraries for matrix and vector computations.

The Share.TEC system and the recommender sub-system use various data sources like metadata records in the digital library, various ontologies for modeling users, information domains, system operations and user interactions. While metadata records

and ontologies are kept in the digital library, all the rest of the data are kept in the independent data base stored in the Share.TEC portal. This data base stores information about each individual system interaction like access to information sources, evaluation, rating and commenting of sources, search command, browsing commands, etc. This is maintained almost automatically through initializing different system tasks through the Hadoop system, which are working asynchronously with the system portal data base.

5 Pilot Experiments with the System

The Share.TEC system was tested successfully [13, 14] during multiple massive online sessions with users from all over Europe. All these massive pilot experiments were implemented in seven different European countries. The Share.TEC system was presented and demonstrated during many international conferences and events as well as to some specific large forums involving teachers and learning scientists from all over Europe. The system is also widely used in Bulgaria. Several large national events were organized with the help of the Ministry of Education. These experiments show that the proposed model can provide huge perspectives for the massive adoption of digital libraries.

The users tested five different scenarios for the use of the system. For user evaluation and validation standard questionnaires were used combined with detailed analysis and documenting of user activities by means of the Think Aloud protocol. Detailed observation of the user activities confirmed that users have no significant problems to work with the system.

The main conclusion from this experiment was that each user confirmed to find and use relevant and very valuable learning resources. The most valuable were the conclusions from some chemistry teachers to find such valuable knowledge objects they were not possible to find in other ways. All experiments confirmed that the user interface completely conform to the user needs. The Share.TEC portal enables the fully feasible system for learning and sharing of knowledge resources, experience and competences for users' future lifelong competence development. The analysis of the seminars and pilot experiments show that the combined use of ontologies and metadata provides the relevant framework for the use of digital knowledge objects for lifelong competence development.

From the feedback from users we can conclude that the proposed model and its implementation can provide the feasible user interface through the web portal based on the social functionalities provided.

6 Conclusion

In this paper we presented the key and innovative solutions adopted in the Share.TEC system, including: adaptability model based on portal functionalities and user interface, multilingual and multicultural approaches, semantic layer (ontology and applica-

tion profile), user modeling features and recommender system, portal architecture and the user-oriented interactive social services.

The evaluation approach and the results from the validation activities in relation with the Share.TEC system were also provided. We show some similar research projects to compare with our research results.

The work on the Share.TEC system is continuing and includes handling secure payments, making a complete track of user actions, adding more user-interface languages, improving the graphical browsing of resources, implement user notifications for deleted/changed resources, etc.

References

1. Brusilovsky P.: Adaptive navigation support in educational hypermedia: The role of student knowledge level and the case for meta-adaptation. British Journal of Educational Technology, Vol. 34(4), pp. 487-497 (2003)
2. Brusilovsky, P., and Millán, E.: User models for adaptive hypermedia and adaptive educational systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), The Adaptive Web, Springer, pp. 3–53 (2007)
3. Manouselis, N., Drachler, H., Vuorikari, R., Hummel, H. and Koper, R.: Recommender Sys-tems in Technology Enhanced Learning, Chapter in Recommender Systems Handbook, Springer, pp. 387-415 (2010).
4. Santos, O. and Boticario, J.: Modeling recommendations for the educational domain. Procedia Computer Science, Vol. 1(2), pp.2793-2800, Elsevier (2010).
5. Apache Solr home page, http://lucene.apache.org/solr/
6. Apache Mahout system, http://lucene.apache.org/mahout/
7. Stefanova E., Nikolova N., Kovatcheva E., Stefanov K., Boytchev P.: A Study of Users Re-quirements for Metadata Repositories of Educational Resources, In proceeding of international conference S3T'2010, pp.38-46 (2010)
8. Stefanov, K., Boytchev, P., Grigorov, A., Georgiev, A., Petrov, M., Gachev, G., and Peltekov, M.: Share.TEC Repository System. In Proceedings of First International Conference on Software, Services & Semantic Technologies (S3T, 29-29 October 2009), Eds.: D. Dicheva, R. Nikolov and E. Stefanova, Sofia, Bulgaria, pp. 84-91, ISBN 978-954-9526-62-2 (2009)
9. Boytchev, P.,Bocconi, S., Stefanov, K.: An Ontological Approach for Multicultural Digital Repositories, Second International Conference S3T, September 11-12, Varna, Bulgaria, ISBN 978-954-9526-71-4. (2010)
10. Stefanov, K., Boytchev, P., Grigorov, A., Georgiev, A., Petrov, M., Gachev, G., and Peltekov, M.: Share.TEC System Architecture. In Proceedings of First International Confer-ence on Software, Services & Semantic Technologies, Eds. D. Dicheva, R. Nikolov and E. Stefanova, Sofia, Bulgaria, pp. 92-99, ISBN 978-954-9526-62-2 (2009)
11. Boytchev, A. Grigorov, J. Earp, K. Stefanov and A. Georgiev: Adaptability Approaches in Digital Libraries, Second International Conference S3T, September 11-12, Varna, Bulgaria , ISBN 978-954-9526-71-4 (2010)
12. P. Boytchev, A. Grigorov, L. Sarti, A. Georgiev, K. Stefanov and M. Chechev: Recommender systems and repository search: the Share.TEC proposal, Sofia University e-Learning Journal, 2010/3, ISSN 1314-0086.