Towards Online Accessibility of Valuable Phenomena of the Bulgarian Folklore Heritage

Radoslav Pavlov, Desislava Paneva-Marinova, Konstantin Rangochev, Maxim Goynov, Detelin Luchev

Abstract: *Bulgarian Folklore Digital Library (BFDL) represents a complete web-based environment for the registration, documentation, and access to a wide range of Bulgarian folklore objects. A very important task during the BFDL development was the provision of the strictly designed functionalities. A special attention was pay to content creation, search and administration, trying to cover a wide range of possible solutions such as easy content annotation and indexing, content structuring, storage, maintenance, access, search, selection, group, management, tracking, data export, etc. This paper presents the main functionality, services specification and implementation of the BFDL.*

Key words: *Digital Libraries, Collection, Systems Issues, User Issues, Online Information Services.*

INTRODUCTION

The ethnological knowledge and therefore the ethnological research have systemic character [8]. Since the early period of Bulgarian ethnology until the present day, scholars describe, investigate, and analyze different subsystems of this type of knowledge. As a rule, scholars study a certain area of knowledge in a particular topos of the Bulgarian ethnical territory and find out an algorithm (where there is a process) or a structural description and afterwards the procedure is repeated in another topos, *etc*. Finally, a summarized algorithm or a structural description is achieved which is – as a matter of fact – the research abstraction[1]. All that means the ethnological studies are hierarchically organized. Leaving the particular topos (a village, for instance), the scholar focuses a bigger entities (such as a region, ethnographical region, or an ethnical territory) and thus he deals with a model of the studied area of ethnological knowledge. A danger in this hierarchical approach (modeling) could be the possibility to neglect important systematic links of knowledge[2]. Another danger comes from the specifics of fieldwork investigations. As a rule, the scholar extracts parts of ethnological knowledge by the means of interview with the interviewees. Therefore, the ethnologists study phenomena which are not person specific but characterize the community but they use for this purpose the memories and opinions of particular people (persons).

An important problem is the specifics of the ethnological research: these types of studies are mostly abstract – due to several historical, objective and subjective reasons (technology of recording, ethical, ideological, and scholarly prejudices, *etc.*), the records of samples of Bulgarian folklore which are studied by scholars, in practice contain partial information: for instance, songs have been recorded only as texts without notation; or there ids no information for the discourse practices convoying the oral narratives; or in more cases the records are made by means of structured interviews and not by inclusive interviews. Therefore, the conclusions of scholars are usually based on partial information.

[1] For instance, the "full" description of the "Bulgarian koleduvane" (Christmans rites) is an algorithm which does not coincide with its local variants.

[2] For instance, if we consider some folklore paradigms of kinship, it can turn out that the same person is involved in several systems of kinship: 1. by blood: grandson- son-brother-uncle; 2. by rite: brother-in-law; 3. by profession.

All these problems require new flexible methods for representation of knowledge in formal and single structures for securing manners for access and management of this knowledge. Thus, the use of interdisciplinary solutions with active involvement of modern semantic-based and digital libraries and technologies promises a good basis for innovative presentation and creative usage of the Bulgarian folklore knowledge [6][7][8]. The work on the research project "Technologies, Based on Knowledge of the Bulgarian Folklore Heritage" (FolkKnow) concentrated on solving the problems enumerated above. In order to formally represent the folklore knowledge it was produced *Ontology of Bulgarian folklore (BFO):* its detailed description was made in [2][3][4][5]. The present articles aims at representing the functionality, services specification and implementation of the Folklore Digital Library produced in the framework of the project. Its main goal is to secure access and management of folklore objects which are semantically annotated through BFO.

FUNCTIONAL SPECIFICATION OF THE BULGARIAN FOLKLORE DIGITAL LIBRARY

The main architectures and functional specifications for digital libraries are described in [9][10]. In BFDL we follow this services specification and the requirements of experts in the area of Bulgarian folklore. Following them the base BFDL functional modules are:

- A module for adding and editing folklore objects;
- A module for viewing the content of folklore objects (according to their base type and rubric to which they belong or by different descriptive characteristics);
- A module for searching by: signature and archive number; keywords of the following categories: name, language, annotation, type of the folklore object/rubric; file type; record information (simultaneously or one by one): by situation, by reporter name, by recorder name, by record date and by recording location; extended search – it provides the option for searching through all the object characteristics;
- A module for managing the user data;
- A module for monitoring the user actions, which keeps track of the following actions:
    - Actions, related to working with the system: registration, logging in the system, unsuccessful log-in attempts, logging out, changing of the user password, e-mail address change, *etc*.
    - Actions, related to the object manipulation: adding an object, editing an object, deleting an object, adding a file, deleting a file.
    - Actions, related to the content viewing: review of objects by their characteristics, view of a single object, searching for objects by characteristics;
    - Other administrative actions: changing the user's level, deleting a user, generation of an XML copy of the data in the system;
- A module for file format conversion;
- A module for generation of XML copies of the objects in the system.

The module for viewing the content of folklore objects is available to all the users of the library, except the administrators. The reason is that the administrators of such systems are often people, who don't have any relation to their content. They only do support tasks. The module itself was implemented similar to the Windows OS file browser and KDE, so that it is closer to the familiar user interfaces for viewing hierarchical information. The left side shows a tree of all classes, which inherit "Type of folklore object" and the right side shows a list of objects of the selected class in the tree.

The module for creating and editing folklore objects is used for adding new objects and modifying the information of already created objects. Through it, one can add more

multimedia files to an object or delete existing ones.

Searching for information is the most frequent search and therefore the most important operation in a digital library. This is why there are several different modules for searching by different criteria:

- Searching by a signature or archive number – This search module is useful for finding objects by their archive number (for example, AIF No 200, folder 1, page 57). In general, there is only one search result. In case of incorrect data, it is possible to have several objects as a result.
- Search by a keyword in the object properties – by name, language, annotation and type of the folklore object – Searching is performed simultaneously over all these properties. It is expected that this module is the most frequently used one. This is why special attention has been paid to its optimization.
- Searching by record information – This module is used to find all the objects, which cover some of the following conditions: all the objects, recorded in a given situation, for example an interview, chat/conversation, *etc*.; all the objects, recorded by a given person; all the objects, recorded by a given informer; all the objects, recorded in a given period of time; all the objects, recorded in a given location.
- Searching by file type – This module allows getting a list of all the objects, to which there is a multimedia file attached – audio, video or images. This type of searching uses the database, in which the administrative information is stored instead of the OWL file that contains the ontology.

Most types of searching use SPARQL (SPARQL Protocol and RDF Query Language). This is a language for requests to the RDF and OWL ontologies. The language is in a standardization process by RDF Data Access Working Group as an official recommendation of the World Wide Web Consortium. The SPARQL syntax is similar to the most widespread language for database requests – SQL.

The module for monitoring the user actions is intended to keep logs for the modified and deleted objects by the users, so that in case of deleted data by mistake or wrong entered data, the responsible user can be found. There is also a log for search requests, whose purpose is to enable statistical reports about the search types that are used least and most often. It would allow the removal of the rarely used search types and the priority optimization of the ones that are used most often.

The module for file format conversion was developed to provide the ability to present every file, which is unsuitable for internet preview in a "light" and convenient form for web preview. The module recognizes the "inconvenient" files, it tries to covert them and on success it replaces the original file with the new "lighter" file. On failure, the module keeps the original file in the library. The module for generating an XML copy of the data is available only to the system administrators. The purpose behind this module is creating a copy, which can be used as an archive copy on one hand and on the other hand it may serve as raw data for other systems, using information from the library.

IMPLEMENTATION OF FUNCTIONAL COMPONENTS OF THE ARCHITECTURE OF THE BULGARIAN FOLKLORE DIGITAL LIBRARY

A module for adding objects to the BFDL – Adding objects is implemented through filling and sending a form to the web server. Because the great number of fields to fill, the form is not generated thoroughly. Only the necessary fields for the creation of the objects are generated, following the semantic descriptions, presented in the BFO, built in the first stage of module 3 of the project. The technology, used for the implementation is AJAX. The user interface passes a request to the server, in which it requires only that part of the form, which according to the user is necessary to create the object. The server processes the request and returns the required fields as a result, which is visualized in the user

interface. After all the fields are filled, the user submits the form. The server validates the data and if everything is correct, it adds the object to the data storage. If there is something wrong, it returns a message to the user, relative with the error (usually, an empty field or unacceptable field value). After the server adds the information from the form to the data storage, there follows a check for attached files in the user request. If there are attached files, the server checks if there are file formats, which are unsuitable for web presentation (for example, wav, .doc, .mpg, .avi, .mpeg, *etc*.) and if it finds such files, the system refers to the module for file format conversion to formats, suitable for web preview. For each of these files, the module for file format conversion tries to convert them. Upon success, it adds the converted file to the library. On failure (which can occur if the added file has some specifics, which the system cannot recognize), it adds the original file to the library. At the end of the object adding procedure, the system refers to the module for monitoring the user actions, where it adds an "object added" event and writes the author (the user, who created the object) and the event date.

A module for editing objects in a BFDL – The module for editing objects works almost in the same way as the module for adding objects. The difference is that the system doesn't add information about a new object, but replaces the existing information about an object with the new information, provided by the module for editing. Again, the system checks the form for errors, it processes the files (if there are new files added), it changes its data and finally adds an event for modified object through the module for monitoring the user's activity.

A module for viewing the content of folklore objects – This module takes a request from a user, in which the user specifies the property, by which folklore objects must be found. The module refers to the data storage and makes a request for selecting and sorting the objects by this property. The module for monitoring the users' actions records the "view objects by" event and adds data about the date, the user and the property, by which objects are listed. The storage processes the request and returns a result, which the system processes and sends to the user. The user interface visualizes the result in a proper manner.

A module for searching – This module allows the user to set a property/properties, by which objects are found.

The algorithm for searching by a single property – The user interface sends a request to the data server specifying the property and its needed value. The module for searching refers to the data storage of semantic metadata with a query for selection and sorting the objects with the needed value of the specified property. The module for monitoring the user actions records the "search" event with the provided search parameters, the date and the user, who performs the search. The storage processes the request and returns a result, which is then processed by the search module and displayed in a proper manner by the user interface. *The algorithm for searching by more than a single property* – The algorithm is analogical to the one, described above with the only difference that the query to the data storage is more complicated – there are multiple selections of objects for each search property and the result is a sorted section of the selection results.

Testing the functional components of the architecture of the BFDL

After an analysis of the means and standards in the technological implementation of the library environment and the functional modules, the following software was chosen: Operating system: Microsoft Windows Server 2008 x64 Standard; Web server: Apache HTTP Server v 2.2, PHP v 2.2.9; Database management system: MySQL v 5.1 Standard; Tools for the additional modules: FFMPEG, vwWare, HTML, JavaScript, AJAX; Database query language: SPARQL

The functional components of the architecture of the BFDL, described in [6] and [7] were implemented and tested for errors and speed on a server platform with the following hardware configuration: CPU: 2 x Intel QuadCore 2.8 GHz; RAM: 8GB DDR3; HDD: 4 x

500GB, RAID 10 SATA II; LAN: 2 x 1000Mbit.

Testing the functional module for adding/editing a folklore object – Server response time (average of 50 attempts): 0.0058 s, *i.e.* in theory, the functional module for adding/editing an object can process about 172 requests per second for each processor core, which makes 172*8=1376 requests.

Testing the module for viewing folklore objects – Time for server response: 0.009 seconds per request, *i.e.* 888 requests per second.

Testing the module for searching by a single property – Time for server response: 0.008 seconds per query, *i.e.* 1000 requests per second.

Testing the module for searching by several properties – The test was performed with 25 properties (it will happen very rarely). Time for server response: 0.01 seconds per query, *i.e.* 800 requests per second.

Testing the module for file format conversion – Converting video files: the server sends a response before it converts the video file, because the process is relatively slow. The average time of processing a video file is about 30 seconds, *i.e.* you can add about 16 video objects per minute. In this way, after adding a video object, its actual recording in the BFDL happens in 30 seconds.

Converting audio files: The server responds before the file is actually processed. The average time for processing an audio file is about 10 seconds, *i.e.* in theory, a system with such a configuration can process about 48 audio files per minute.

Converting MS Word (.doc) files: the conversion takes place in real time. The average server response time is 0.04 seconds per request, which are about 200 requests per second.

CONCLUSIONS AND FUTURE WORK

The formal presentation of the folklore knowledge and the attempts at digitalizing the folklore archive (with the purpose of creating of a digitalized content) describe a group of problems which the folklorists (ethnologists and ethnographers who usually have a humanitarian education) haven't even suspected of their existing. Furthermore when they are working with the diffusion system of the Bulgarian folklore, they create not very clear qualification systems. On the project FolkKnow it is working actively in direction of formalization of the folklore knowledge by creating of computer workable structures. There are created resources and services for treatment of these structures in the digital library with the purpose of optimal content creation, storage, retrieval, filtering, search, classification, mining, *etc*. From this point of view the creation of digital library of the Bulgarian folklore gave us an unexpected and significant advantage – it puts correctness, clearness, *etc*. in the qualification systems of the folklore study/ethnology and it secured the services for an approach and using of the folklore knowledge in the virtual space.

ACKNOWLEDGEMENTS

This work is supported by NSF of the Bulgarian Ministry of Education and Science under grant No IO-03-03/2006 "Development of Digital Libraries and Information Portal with Virtual Exposition "Bulgarian Folklore Heritage"" from the project "Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage".

This work is partially supported by the ETN TRICE project 142399-LLP-1-2008-1-BG-ERASMUS-ENW "Training, Research and Innovations in Computing Education".

REFERENCES

[1] Bogdanova, G., R. Pavlov, G. Todorov, V. Mateeva. Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage, "Advances in Bulgarian Science", 2006, pp.7-15.

[2] Luchev, D., D. Paneva, K. Rangochev. Approaches for Utilization of the Semantic Web Technologies for Semantic Presentation of the Bulgarian Folklore Heritage. Bulgarian Museums under the Conditions of EU Membership of the Country. Sliven, 2009, pp. 371-381.

[3] Luchev, D., D. Paneva, K. Rangochev. Use of Knowledge Technologies for Presentation of Bulgarian Folklore Heritage Semantics. „Information Technologies and Knowledge", 2(4), 2008, pp. 307-313.

[4] Paneva, D., K. Rangochev, D. Luchev. Knowledge Technologies for Description of the Semantics of the Bulgarian Folklore Heritage. Proceedings of the Fifth International Conference „Information Research and Applications" – i.Tech 2007, July, 2007, Varna, Bulgaria, vol. 1, pp. 19-25.

[5] Paneva, D., K. Rangochev, D. Luchev. Ontological Model of the Knowledge in Folklore Digital Library. Proceedings of the Open Workshop „Knowledge Technologies and Applications", 31 May – 1 June, 2007, Kosice, Slovakia, pp. 47-55.

[6] Rangochev, K., D. Paneva, D. Luchev. Bulgarian Folklore Digital Library. Proceedings of the International Conference on Mathematical and Computational Linguistics „30 years Department of Mathematical Linguistics", 6 July, 2007, Sofia, Bulgaria, pp. 119-124.

[7] Rangochev, K., D. Paneva, D. Luchev. Data and Functionality Management in a Folklore Digital Library. Proceedings of the National Conference Slovo: Towards a Digital Library of South Slavic Manuscripts, 21-26 February, 2008, Sofia, "Boian Penev" Publishing Centre, pp. 246-250.

[8] Rangochev, K., Structural specifications of the folklore historical knowledge. Problems of the Bulgarian folklore, Sofia, vol. 10, pp. 444-449.

[9] Digital library architecture (WP1), DELOS project, Cluster home page, Available at: http://delos-old.isti.cnr.it/WP1.html (Accessed on May, 2010).

[10] DILIGENT project "Digital Library Infrastructure on Grid Enable Technology", Available at: http://www.diligentproject.org/ (Accessed on May, 2010).

ABOUT THE AUTHORS

Radoslav Pavlov – PhD in Mathematics, Associated Professor, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev Str., bl. 8, Sofia 1113, Bulgaria; e-mail: radko@cc.bas.bg

Desislava Paneva-Marinova – PhD in Informatics, Assistant Professor, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev Str., bl. 8, Sofia 1113, Bulgaria; e-mail: dessi@cc.bas.bg

Konstantin Rangochev – PhD in Philology, Assistant Professor, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev Str., bl. 8, Sofia 1113, Bulgaria; e-mail: krangochev@yahoo.com

Maxim Goynov – Programmer, Institute of Mathematics and Informatics, BAS, Acad. G. Bonchev Str., bl. 8, Sofia 1113, Bulgaria; e-mail: maxfm@abv.bg

Detelin Luchev – PhD in Ethnography, MA in Informatics, MA in History; Assistant Professor, Ethnographic Institute and Museum, BAS; 6A, Moskovska Str., Sofia 1000, Bulgaria; e-mail: luchev_detelin@abv.bg