Преглед НИД 30 (2017), 40–45

Aleksandra Zdravkovic, Sinisa Tomovic, Milan Todorovic, Vladisav Jelisavcic, Aleksandra Arsic, Bojan Marinkovic, Milica Knezevic
Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

Miroljub Stojanovic
National library of Serbia, Belgrade, Serbia

Ivana Djurkovic Djordjevic
Dr Archibald Reiss Primary school, Belgrade, Serbia

IMPROVING THE DIGITAL NATIONAL LIBRARY OF SERBIA[1]

Abstract. Digital National Library of Serbia includes a large number of digitized collections that are continuously enriched with new contents. Some of the collections are Newspapers, Magazines, Books, Handwritten Periodicals, Cyrillic Manuscripts, Audiobooks, Old Printed Books, Photographic documents, Posters and documentary material. Collection of Handwritten Periodicals includes 15 titles that were created in the period from 1874 to 1995. 14 out of 15 titles represent periodicals which has at least one edition that is entirelyhand-written, while the remaining title is address typing – machine typed, yet supplemented by hand to a lesser extent.

From these 15 titles, 14 represents periodicals which are at least oneedition hand-printed in its entirety, while the remaining address typing - typed on the machine, but to a lesser extent, supplemented by hand. Collection of cyrillic manuscripts NBS contains digital copies of selected pages from 180 cyrillic manuscripts of the National Library of Serbia.The collection of audiobooks contains pieces of Bora Stankovic, Branislav Nusic, as well as publications from the portal "Great War" in audio format, available to users for direct playback using the embedded player. Collection of old printed books contains a selection of 16 library units, among which are mainly represented edition of Cetinje and the Venetian editions of Bozidar Vicenca Vukovic. The collection of photographic documents contains digital copies of old photos and postcards from the collection of photographic documents of special funds of National Library of Serbia. Collection of posters and documentary material is a set of digital copies of old posters and ephemeral material from the Graphic collection and Collection of documentary material of special funds of National Library of Serbia. This collection contains collection of theater posters which have been printed in the territory of the Kingdom of Serbia and Kingdom of Yugoslavia in the period from the mid-19th century to 1945.

The technical backbone of the Digital National Library of Serbia is Serbia Forum[2], which is based on the principles of electronic encyclopedias, such as Wikipedia. The main part of the system is a tree-structure of pages, which represent collections, subcollections or digital objects themselves. Serbia Forum is a modular system that supports easy creation and installation of new modules with new functionalities. Some of the most important modules are: an electronic book reader, audio player, system metadata of digitized objects, metadata search system and a module that supports the Open Access Interface protocol. In this paper, authors will present the improvement of existing modules, and will feature some of the new functionality of the system.

Keywords. SerbiaForum, digitization, collections

Introduction

Digitization is a process that contributes to the preservation of knowledge and spirituality of a nation for future generations.[3]We can notice that millions of artifacts

that are in museums are behind the wall and still waiting the discovery in museum archives. Special aspect of digitization is preserving act that arise but also disappear in time, like languages. According to an article in The Guardian in March 2007, if all spoken language since the dawn of time were digitized, it would consume five exabytes of storage space.[4] Natural disasters (such as collapse or fire) or war (during the NATO bombing the Federal Republic of Yugoslavia (Republic of Serbia and Republic of Montenegro) for three months in 1999, a great deal of cultural heritage in Serbia (objects, buildings, cultural landscapes) was damaged or destroyed as a result of direct hits, or as collateral damage) reminds us of the importance of digital preservation of our cultural heritage. Having that on mind we developed, in collaboration with University of Graz, Austria, an application of that kind, named Serbia Forum, that contributes to the storage, preservation and presentation of culturally relevant data. From the beginning of its development in March of 2012, the web application was built to serve two specific purposes. Firstly, it serves as an encyclopedic collage of articles written by credible authors. The credibility of authors is checked, along with the sources of information within the articles. If an article is duplicated, that is if it is written twice by two different authors, then both articles are saved for the sake of recording article progress. This allows users, i.e. readers of these articles to track certain changes and to observe different points of view of authors. Secondly, the application serves as a hub or centralized node for collecting quality controlled, digitized units of cultural heritage content from other distributed collections, i.e. archives and museums.

As we wereworking on improvingof the software at the same time the collaboration with local institutions in Serbia is being enforced. Particularly, One institution in particular has proved to be pivotal player in the fast development of the Serbia-Forum, the National Library of Serbia (NBS). Since collaboration was established between NBS and MISANU, the Serbia-Forum received its first major content contribution of approximately 1,800,000 pages of digitized material. Within that collaboration, we developed new portal, Digital National Library of Serbia, for representing of the collections of special funds of National Library of Serbia.Serbia Forum is a software platform for Digital National Library of Serbia.

Serbia Forum concept and technical details

Serbia-Forum is built on the foundation of Austria Forum. The web application is based on dynamic content generation and presentation delivered by the JSP Wiki[5] framework, running on the Tomcat 7[6] web server. The JSPWiki engine is the backbone of the heritage portal. It enables the user to dynamically create webpages on the server end. The pages are created by the Java Virtual Machine or JVM[7] on the web server end, creating the illusion of platform neutrality where the user need not have a JVM installed on the client machine in order to render pages. The eXist[8] database is used for content storage and organization. eXist is an open source software project for NoSQL databases built on XML technology. It is classified as both a NoSQL document-oriented database system and a native XML database (and it provides support for XML, JSON, HTML and Binary documents). Unlike most relational database management systems (RDBMS) and NoSQL databases, eXist provides XQuery and XSLT as its query and application programming languages. The database is very flexible in nature and gives a lot of moving room for the application developers in terms of content and metadata categorization, classification and other important organizational tasks necessary for the implementation of semantic content

searching mechanisms. Internet Archive BookReader[9] is used for displaying books. SoundManager 2 is used for reproducing the sound. Metadata and data input is done by user interface, we created the module for massive input metadata and data. Metadata input form is automatically generated depending on .xsdshema. Standard used for metadata is known as NCD Metadata standard [11,12] or (National Center of Digitization Standard). System can be found on different domains such as:

1. www.serbiaforum.org
2. digitalna.nb.rs
3. bosniaforum.etf.unsa.ba
4. mij.mi.sanu.ac.rs

The structureof Serbia Forum application is divided into modules. Authentication module is responsible for identifying the user. It consists of sub-module for storing information about users, as well as sub-module for identification and placement user settings. The current implementation places the users in xml database. Authorization module serves to control user access rights content. It consists of sub module for storing access rights and the rights of sub module for verification. In the forum software, access rights are assigned at the level of the digital object. Objects are hierarchically oriented, and sub object inherit the access rights from the higher level object. Access rights are stored within wiki page, with the option to be upgraded. The module for parsing wiki syntax allows the presentation of content using simple syntax, which allows the user to edit the sets and visual content.The module for filtering content allows users to add filters to the appropriate content. Examples of implemented filters are filter for automatic expansion links, dynamic insertion of images and filter for illicit things. Filter for illicit words can be used to prevent an abuse of a deliberate display of vulgar words. Module for storage wiki page is responsible for storing presentation pages (written in wiki syntax). The current implementation uses the file system to store pages, with the possibility of enlargement and the transition to a relational or NoSQL database. Module for storing extras is responsible for storing multimedia data (images, sound and video clips). The current implementation uses the file system, with the possibility of enlargement and the transition to an arbitrary database. Module for the storage of books is responsible for storing and linking digitized books. The current implementation uses the file system and a specifically defined format, with the possibility of expansion and transition to an arbitrary database.The module for storing metadata is responsible for storing metadata. The current implementation uses advanced existdb xml database. The module for generating form input is responsible for generating forms based on the scheme data. Metadata entry forms are generated automatically, based on the given scheme of data. It is possible to add a new scheme for new data types, and automatic generation of forms for their input. Search Module is in charge of searching text data. Module for renaming allows you to rename wiki pages. The module for mass data entry allows you to migrate data and metadata. Interactive book reader module allows the presentation of digitized books. The current implementation is based on the Internet Archive BookReader. It is possible to improve the system by adding more advanced readers. Module for versioning wiki page allows observation of history of the certain page. With each change author and modification date can be seen. Module for briefing page allows you to track feedback links between pages. Module for internationalization allows multilingual system. With this system it is possible to add content in different languages. It is possible to easily add new languages to the system (the only prerequisite is to deliver an adequate translation of the user interface). Module that

allows programmable additions for wiki pages allows easily addition of new functionality of the system. With this module, the authorized user (content editor) can use a wide range of tools to display multimedia content. Module for enabling flow control allows authorized users to authorize users and content. Module for managing visual templates allows different system interface for users accessing via different addresses. Example: digitalna.nb.rs is one official portal, and mij.mi.sanu.ac.rs is another portal with separately defined own appearance.

Technical improvements

In this paper we will introduce new module, wide used for exchanging metadata. OAI interface is developed in accordance with the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), a protocol developed by the Open Archives Initiative, which defines the standard for the exchange of metadata between providers on the one hand and those who provided the information collected for further use. Development of interfaces in accordance with the OAI-PMH standard allows easy exchange data with other project stakeholders, institutions and organizations, as this standard is widely accepted and supported. Remarks by OAI-PMH to interface takes place via HTTP requests. The following describes the requeststhat are defined by OAI-PMH protocol and must be supported by the OAI interface. Identify request is used to obtain information about the repository of metadata and no additional arguments are required with this request. Request ListMetadataFormats is used to obtain the available metadata formats. A parametar can be send with this request: identifier is a unique identifier of the metadata which we want to get the available formats for. The basic format required by OAI standard is oai_dc. This is the only currently supported format by implemented OAI interface. Request ListIdentifiersis used to obtain a list that contains the basic information (unique identifier and time of the last modification) of all metadata contained in the repository. We can sent additional parameters with this request such as from/until (lower and upper time limit for the selection of only those metadata changed in this time limit), set (optional parameter is specified to limit the selection of those metadata which are contained in the OAI collection only, of a given argument), metadataPrefix (parameter specified to limit the selection of only those metadata which are available in a format of agiven parametar) and resumptionToken (exclusive argument that is used to obtain the continuation of the list with basic information about the metadata). Request ListRecords is used to obtain a list that contains the metadata contained in the repository. This requirement is similar to the ListIdentifiers request, except that unlike it, ListRecords does not return the basic information (identifier and time period), but the complete metadata. Likewise ListIdentifiers, supported parameters are from, until, and metadataPrefixresumptionToken. Request GetRecord is used to obtain a particular metadata. This request is sent with mandatory parameters: identifier - a unique metadata identifier, and metadataPrefix - the desired metadata format. Request ListSets is used to obtain OAI sets that already exist in the repository. These sets should not be identified with sets described with metadata. OAI sets are used for grouping the metadata and organizing it according to the desired category (theme, time period, etc.). This request can be sent with the exclusive resumptionToken argument, which is used in the manner described for ListIdentifiers request. Metadata repositories are not yet grouped into OAI sets. OAI interface does not have implemented sets support, and therefore, in accordance with the

OAI standard for sending information about errors and exceptions, the message that comes as a response to ListSetsrequest is: noSetHierarchy - The repository does not support sets. Purpose of OAI protocol is exchanging and collecting metadata. It is widely accepted and supported by a large number of institutions and organizations who use it for sharing/collecting metadata. Examples of such institutions and organizations are following:

- elibrary[13]– The Faculty of Mathematics virtual library provides access to metadata about dissertations through OAI service
- University of Michigen (The OAIster® database) [14] Millions of digital resources from thousands of contributors
- doiSerbiaPhD is national register of e-thesis deposited in the University repositories across the Serbia.[15]
- DOI Serbia is collection of sixty-one locally published scientific journals.[16]

OAI is standard and wide accepted protocol for exchanging metadata. The fact that digital repositories based on DSpace and Eprints solutions provide automatic support for sharing and harvesting metadata using OAI.

Conclusion

Idea for further development is that NBS represent the national aggregator for sending (meta)data into Europeana[17], which would establish a network for the exchange of knowledge and experiences between the institutions of science and culture involved in digitization. Europeana portal allows access and search through 50 million books, art, historical sources from museums, libraries, archives across Europe. Connecting with a Europeana will provide digitized materials to become more publicly available and visible.

References

[1.] See Digital National Library of Serbia, available at: www.digitalna.nb.rs (Accessed: 04.09.2016.)
[2.] See Serbia forum, available at www.serbia-forum.org (Accessed: 04.09.2016.)
[3.] See „Едукација у Дигитализацији" (Serbian edition), available at: http://forum-digitalna.nb.rs/jezik/edukacija.pdf (Accessed: 04.09.2016.)
[4.] See WhatIs.com-digitizition http://whatis.techtarget.com/definition/digitization (Accessed: 04.09.2016.)
[5.] See Apache JSPWiki, available at: https://jspwiki.apache.org (Accessed: 04.09.2016.)
[6.] See Apache Tomcat, available at: http://tomcat.apache.org (Accessed: 04.09.2016.)
[7.] See Existdb, available at: http://existdb.org (Accessed: 04.09.2016.)
[8.] See The Java® Virtual Machine Specification, available at: https://docs.oracle.com/javase/specs/jvms/se7/html/ (Accessed: 04.09.2016.)
[9.] See Internet Archive BookReader,available at: https://openlibrary.org/dev/docs/bookreader (Accessed: 04.09.2016.)
[10.] See Sound Manager 2,available at: http://www.schillmania.com/projects/soundmanager2/ (Accessed: 04.09.2016.)
[11.] Ognjanović, Z., Butigan-Vučaj, T., Marinković, B., *Predlog Nacionalnog standard opisa pokretnih kulturnih dobara*, Review of NCD, 11(2007), 1–11
[12.] Ognjanović, Z., Butigan-Vučaj, T., Marinković, B., *NCD Recommendation for the National Standard for Describing Digitised Heritage in Serbia*, Sicilia, M.-A., Lytras, M.D., Metadata and Semantics, 2009, Springer, 45–54.

[13.] See elibrary, available at: http://elibrary.matf.bg.ac.rs/ (Accessed: 04.09.2016.)
[14.] See The OAIster® database, available at: http://www.oclc.org/oaister.en.html (Accessed: 04.09.2016.)
[15.] See doiSerbiaPhD, available at: http://kobson.nb.rs/oaiphd/ (Accessed: 04.09.2016.)
[16.] See DOISerbia, available at: http://www.kobson.nb.rs/doiserboai/?verb=Identify (Accessed: 04.09.2016.)
[17.] See Europeana, available at: http://www.europeana.eu/portal/en (Accessed: 04.09.2016.)

alekzdravkovic@mi.sanu.ac.rs
sinisatom@mi.sanu.ac.rs
mtodorovic@mi.sanu.ac.rs
vladisav@mi.sanu.ac.rs
aleksandra@mi.sanu.ac.rs
bojanm@mi.sanu.ac.rs
mknezevic@mi.sanu.ac.rs
misa.stojanovic@nb.rs
djurkovic.ivana@yahoo.com