

Design and implementation of a social semantic digital library

Maria Nisheva-Pavlova ^{a,b,*}, Dicho Shukerov ^a and Pavel Pavlov ^a

^a *Faculty of Mathematics and Informatics, Sofia University, 1164 Sofia, Bulgaria*

^b *Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, 1113 Sofia, Bulgaria*

Abstract. The paper analyzes some current trends of research and development in the field of digital libraries. The presentation is focused on the main features of two new generations of digital libraries – the so-called semantic digital libraries and social semantic digital libraries. The design characteristics, principles of functioning and some implementation details of a particular academic digital library have been discussed as an illustration of the suggested ideas.

Keywords: Digital library, semantic technology, ontology, semantic interoperability, search engine, information retrieval, sentiment analysis

1. Introduction

During the last 2–3 decades, digital libraries are one of the most rapidly developing areas of research and considerable practical results. According to the IFLA/UNESCO Manifesto for Digital Libraries [7], “a digital library is an online collection of digital objects, of assured quality, that are created or collected and managed according to internationally accepted principles for collection development and made accessible in a coherent and sustainable manner, supported by services necessary to allow users to retrieve and exploit the resources”. Digital libraries contain electronic copies of valuable books, periodicals, documents, maps, audio archives, etc., and provide convenient tools for comparatively inexpensive access to them. A digital library is an information retrieval system that maintains collections of digital objects along with means for organizing, storing, and retrieving the resources contained in its collections.

Digital libraries should play the role of environments supporting the full life cycle and the best practices of creation, preservation and use of rich digital content. Interoperability and sustainability are the most important principles of building digital libraries able to communicate with each other.

The so-called semantic digital libraries have been based on design and implementation standards that are a significant step to providing interoperability at the semantic level.

Social semantic libraries reflect the changes in users’ expectations resulting from the wide penetration of social networks in everyday life of a continuously widening variety of communities.

This paper discusses the main features of a particular semantic digital library called DjDL [12] and the results of our recent activities directed to its growing into a social semantic digital library. It is an extended version of the one published in The ELPUB 2015 Conference Proceedings [13].

*Corresponding author. E-mail: marian@fmi.uni-sofia.bg.

2. Semantic digital libraries

The considerable results in the area of digital libraries during the last two decades played a determinant role in the development of the Digital Library Reference Model [2]. This model has the aim to lead to an agreement between experts with respect to the main concepts, structures and activities in digital libraries. Three types of systems play a central and distinct role in the corresponding digital library reference architecture [2]:

- Digital Library – “an organization, which might be virtual, that comprehensively collects, manages and preserves for the long term rich digital content, and offers to its user communities specialized functionality on that content, of measurable quality and according to codified policies”;
- Digital Library System – “a software system that is based on a defined (possibly distributed) architecture and provides all functionality required by a particular digital library”;
- Digital Library Management System – “a software system that provides the necessary software infrastructure both (i) to produce and administer a digital library system incorporating the suite of functionality considered fundamental for digital libraries and (ii) to integrate additional software offering more refined, specialized or advanced functionality”.

The next generation of digital libraries – *semantic digital libraries* – may be considered as digital library systems that apply semantic technologies to achieve their specific goals [10,14]. They provide new search paradigms for the information space – *intelligent search* (also known as semantic or ontology-based search) [6,11] and *community-enabled browsing*. The specific technologies of semantic digital libraries make it possible to integrate metadata from various heterogeneous sources. In this way they support the interconnection of different digital library systems.

The utilization of proper *ontologies* is one of the main characteristics of semantic digital libraries. In particular, ontologies play a key role in semantic search. Three types of ontologies have been identified as a support for this type of search [9]: bibliographic ontologies, subject ontologies, and community-aware ontologies. Bibliographic ontologies describe metadata standards. Subject ontologies are useful as knowledge sources which define the meaning of most domain concepts, their hierarchy, properties and relationships. Community-aware ontologies are oriented to the description of the different types of users, their requirements and interactions.

Ontologies are also one of the well-accepted types of resources for achieving semantic interoperability of digital libraries. According to [5], semantic interoperability depends mainly on the existence and use of well-formed and accepted upper and core ontologies, in which the basic concepts and relationships are defined. In addition, the concepts defined in the upper and core ontologies, should be extended by appropriate domain ontologies.

3. Social semantic digital libraries

The critical study of the experience in development and use of digital libraries shows that current semantic digital libraries are not enough from the point of view of their typical end users because [10]:

- digital libraries should not be for scholars and librarians only but mostly for average people;
- they concentrate on delivering content, not on knowledge and opinion sharing within a community of users;
- digital libraries have lost the human part of their predecessors.

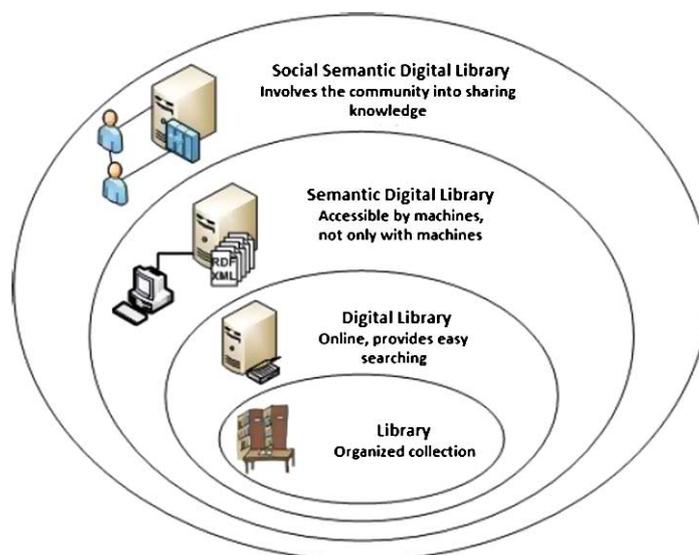


Fig. 1. Evolution of libraries.

The so-called *social semantic digital libraries* [10] are suggested as a solution of these problems. They have the aim to make users (in other words, readers) involved in the content annotation process. They also allow users/readers to share their knowledge within a community. Social semantic digital libraries provide better communication between users in and across communities than the traditional ones.

Figure 1 illustrates the main stages in evolution of libraries. The development of the concept of social semantic digital library is a natural consequence of the extending importance of social networks in communication and information sharing within and between various communities of users.

4. Main characteristics of DjDL – A digital library with Bulgarian folk songs

DjDL preserves a collection of over 1000 digital objects representing folk songs from the Thrace region of Bulgaria. This collection constitutes a considerable part of the digitized archive of the distinguished Bulgarian folklorist Prof. Todor Dzhidzhev published in [3]. In particular, the files with metadata and lyrics of songs (in LaTeX format) and the files with the encoded musical notations of songs (in LilyPond format) have been used as original resources in building the repository of DjDL.

The development of the prototype of DjDL was supported by the Bulgarian National Science Fund within a project titled “Information technologies for the presentation of Bulgarian folk songs with music, notes and text in a digital library” [8]. The main characteristics of this prototype were presented at the ELPUB 2012 conference [12]. In this paper we discuss an entirely new version of DjDL that has some essential features of a social semantic digital library.

DjDL has the typical architecture of an academic digital library. Its functional structure is shown in Fig. 2. It includes five main components: metadata catalogue, repository, search engine, module implementing the library functionality, interface module. A subject ontology was especially created and has been used to support the full functionality of the search engine.

The folk songs treasured in the repository of DjDL have been presented with their notes (musical notations), text (lyrics) and music (digitized versions of their authentic performances).

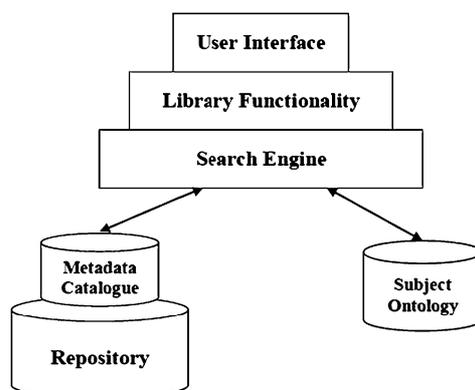


Fig. 2. Functional structure of DjDL.

The subject ontology describes a proper amount of knowledge in several domains, relevant to the content of Bulgarian folk songs. It contains definitions of the main domain concepts, descriptions of their properties and some kinds of relationships between them, as well as a selected set of their representative instances. This subject ontology consists of a set of interrelated subontologies needed by the search engine of DjDL and developed especially for the occasion:

- ontology of folk songs which includes various genre classifications of folk songs (by their thematic focus, by the context of performance, by their cultural functions, etc.);
- ontology of manner of life and family (professions, instruments, clothing, ties of relationship, feasts, traditions and rites, etc.);
- ontology of impressive natural phenomena;
- ontology of social phenomena and relationships;
- ontology of historic events;
- ontology of administrative division – combines the current administrative division of Bulgaria with the one from the beginning of the 20th century.

In addition, a set of natural language-dependent patterns of typical stylistic or thematic constructs, called *concept search patterns*, have been defined and used as domain knowledge aimed at providing satisfactory precision and recall of the search engine.

The purpose of the search engine is to provide adequate access to the variety of resources stored in DjDL. It provides two main types of search: keywords-based and semantic (ontology-based) search. The semantic search tool provides a set of facilities for augmentation and refinement (automatic reformulation according to the available explicit domain knowledge) of the queries for keywords-based search. The augmentation of the user queries is based on proper utilization of the two forms of conceptual knowledge maintained in DjDL – the subject ontology and the set of concept search patterns based on this ontology.

The search engine realizes some additional functionalities enabling the user to combine the search and retrieval of documents kept in the repository of DjDL with a kind of sentiment analysis of their texts. For this purpose some of the subject ontology classes are associated with proper positive or negative numbers which play the role of sentiment estimates of the corresponding concepts. The sentiment estimates of the ontology concepts have been used as default values for their specializations and forms.

The library functionality and the user interface of DjDL are designed in accordance with the expected requirements of its typical users. Three levels of access to the library and the corresponding differenti-

ated roles of users have been defined: *librarian*, *author* and *reader*. Librarians maintain the user accounts and their associated roles. They also may add and register new library resources and upload new versions of the subject ontology. Authors may use a specialized authoring tool developed especially for the purpose. It allows one to create and edit catalogue descriptions and texts of songs (their lyrics and musical notation). Users registered as readers may examine the texts, musical notations and sound recordings of the available folk songs, define and send queries to the search engine, write comments that can be read and replied by others. In this way users are enabled to participate in the content annotation process, to communicate and to share their impressions and opinions.

5. Semantic search and sentiment analysis of the lyrics of songs in DjDL

The semantic search tool of DjDL is aimed at making some kind of pre-processing of the user queries in order to provide better precision and recall of their accomplishment. When the user defines his/her particular query and indicates the search sources (the lyrics of songs or specific metadata), the search engine augments the query so much as possible, in accordance with the available explicit domain knowledge.

The most significant knowledge source for augmentation of the user queries is the taxonomy (the “is-a” hierarchy) of concepts that forms the core of the subject ontology. During the augmentation of the user query, first of all an exhaustive breadth-first search in the graph representing the “is-a” concept hierarchy is performed, starting from the node which corresponds to the original user query. The names of the visited nodes that are in fact the respective more specific concepts described by the ontology, are added to the one formulated by the user. The resulting list of concepts if properly visualized and placed at the user’s disposal for optional refinement.

During the next step of query expansion, the search engine adds to the newly constructed set of concepts some derivatives and synonyms of the main terms found as values of their “form” and “synonym” properties in the subject ontology. The corresponding property values from the definitions of all concepts included by that time in the expanded user query and the existing instances of these concepts are added to the query as well. Finally, the values of some properties of the newly included instances that have been explicitly specified as significant for their classes with respect to search purposes, are included in the resulting augmented query. If the search activities have been realized in the lyrics of songs and there is a concept in the augmented query provided with appropriate search patterns, the pattern matching module performs an additional search for each of these patterns.

Thereby the user query is augmented as far as possible in terms of the subject ontology and in fact it has the shape of a disjunction of all included forms of concepts and instance names. In this form the resulting query is ready for further refinement and processing.

As example queries for semantic search being of interest for folklorists, that can be executed by the search engine of DjDL, we could indicate the queries for search and retrieval of:

- songs praising or mentioning significant historic events;
- songs in which typical folk beliefs or rites are described;
- songs in which elements of country work and life are described or mentioned;
- songs in which significant family events are mentioned.

The search engine provides also some facilities for processing of user queries presuming examination of equality or inequality. For example, it is possible to formulate and execute queries for search and retrieval of:

- songs performed alone or in a group;
- songs performed by men or women only;
- songs performed by one and the same singer;
- songs performed by singers, born in one and the same settlement or region;
- songs performed in a specific region (grouped by regions of performance).

Let us suppose for example that the user defines a query for semantic search in the lyrics of songs which concerns the concept “historic event” (“историческо събитие” in Bulgarian). During the execution of this query, first of all it is augmented and refined with the assistance of the user (see Fig. 3). Then a consecutive search in the lyrics of songs follows. As a result, all documents with texts of songs containing phrases that are juxtaposed with at least one element of the augmented query, are extracted. A list with the titles of the discovered songs is properly visualized on the user screen, as shown in Fig. 4.

When the user clicks on the name of a chosen song satisfying the augmented query, the text of this song is displayed in a new window along with the corresponding metadata. The discovered words and phrases that match the query, are highlighted (see e.g. Fig. 5).

Figure 6 illustrates some search results for a user query containing the phrase “love infidelity”. A pre-defined concept search pattern matches the query.

The search engine of DjDL holds up some additional functionalities that enable the user to combine the search and retrieval of documents with a kind of sentiment analysis of their texts (see e.g. Fig. 6). The sentiment analysis tool uses the subject ontology as a source of knowledge about the emotional intensity of its concepts and computes rough estimates of the mood of songs.

More precisely, some of the subject ontology classes are associated with proper positive or negative numbers which play the role of sentiment estimates of the corresponding concepts. The sentiment of a song is currently defined in accordance with the sum of the sentiment estimates of the particular words in the lyrics of this song. Moreover, the specializations of ontology concepts and all their forms and

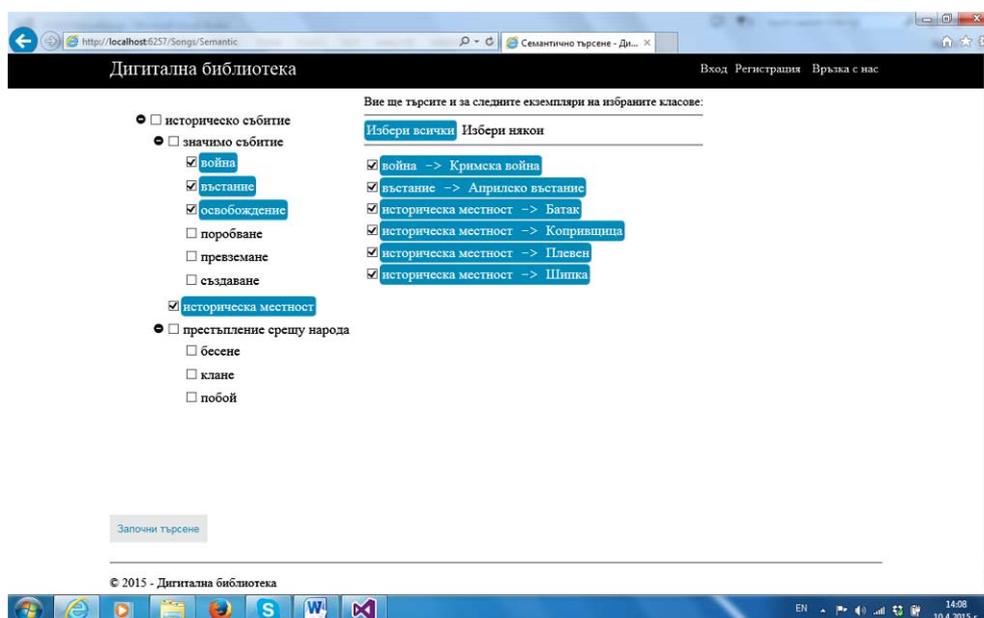


Fig. 3. Construction of a user query for semantic search.

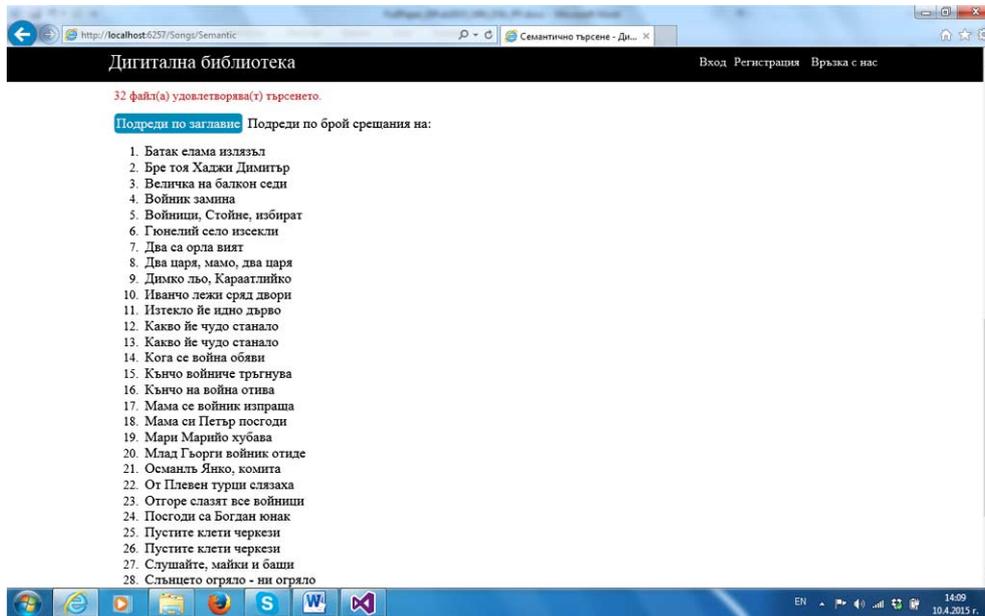


Fig. 4. Search results for a user query containing the phrase “historic event” (level 1 – document retrieval).

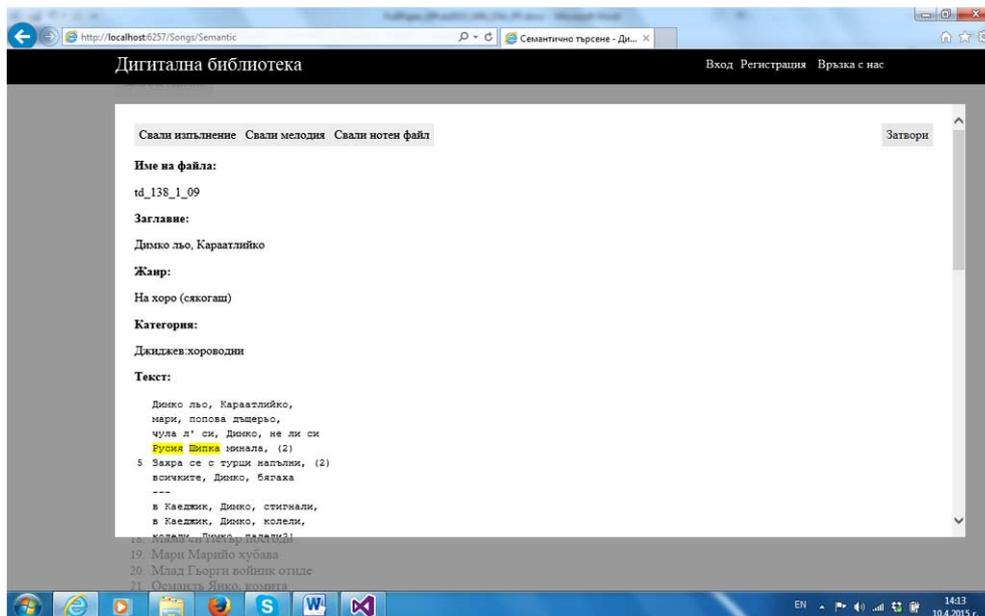


Fig. 5. Search results for a user query containing the phrase “historic event” (level 2 – display of lyrics).

synonyms inherit the sentiment estimates of the main concepts. In other words, the sentiment estimates of the ontology concepts have been used as default values for their specializations and forms.

The results of the experiments carried out with the texts of songs stored in the repository of DjDL indicate that the presented approach is not completely adequate for the domain specificity. For example,

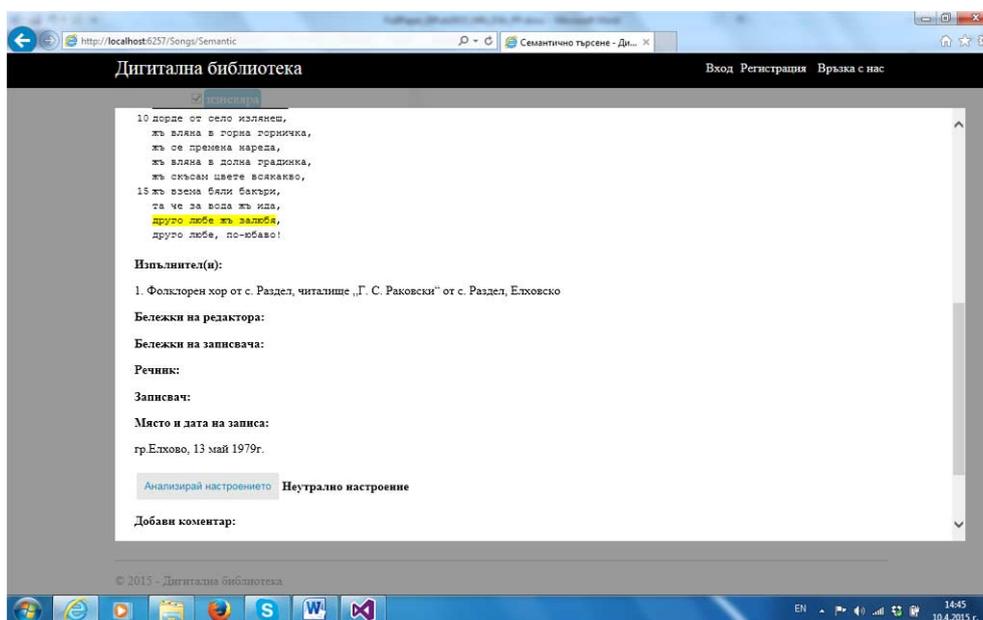


Fig. 6. Some results of semantic search in combination with sentiment analysis.

the sentiment of a part of the songs has been inferred as “merry” while it may be defined as “sadly” by a human reader. Because of that a new version of the sentiment analysis tool has been under development. Two basic changes have been considered in order to improve its performance. First, the sentiment symbolized by phrases that match existing concept search patterns, will be considered first of all during the sentiment analysis process. A set of new patterns will be defined for the purpose and the concept search patterns will be provided with proper sentiment estimates. Next, the sentiment estimates of some particular forms of a set of distinct ontology concepts will be revalued in accordance with their typical sense and cases of use.

6. Authoring tools and social aspects

The library functionality and the user interface of DjDL are oriented to its three basic types of users and their specific roles: librarian (or library administrator), author and reader (end user). The library administrators add and register new resources in the repository and are responsible for the maintenance of the user accounts and their corresponding roles as well as for any other security and system settings issues. Figure 7 illustrates one of the user management tools accessible to library administrators.

A set of specialized authoring tools have been developed in order to allow the users with author’s role to create and edit metadata, lyrics and musical notations of songs. Authors may also define new and edit existing concept search patterns (see e.g. Fig. 8) and define queries for creation of indexes, MIDI files with melodies of songs, etc. They can upload new versions of the subject ontology and edit the Mood dictionary indicated in Section 7.

The users of DjDL registered as readers may examine the texts, musical notations and sound recordings of the available folk songs, define and send queries for keywords-based and semantic search and sentiment analysis. They also may write comments that can be read and replied by others – by all users,

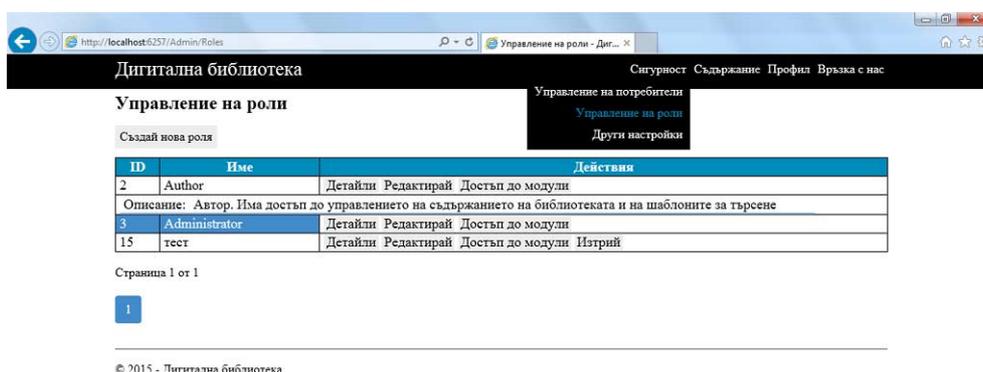


Fig. 7. User management tools: assignment of roles.

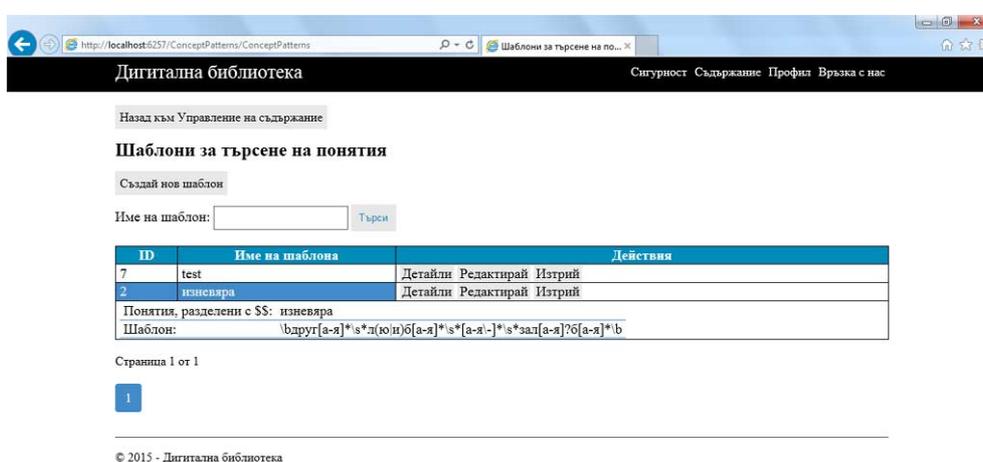


Fig. 8. Authoring tools: creating and editing concept search patterns.

by administrators or authors only, by all or specific group(s) of readers. These comments could refer specific metadata or the lyrics or music of particular songs, issues concerning the qualities of the search engine or the subject ontology, the performance of the software system of DjDL as well as any other topic of interest to the user.

In this way DjDL is acquiring some characteristics of a social semantic digital library. In particular, its users are enabled to participate in the content annotation process (as shown in Fig. 9), to communicate and to share their knowledge, impressions and opinions. Our further plans envisage the implementation of functionalities for searching users and creating friendships; sharing search results; building and sharing lists of personally tagged songs; sharing songs with other users or in social networks, etc.

7. Implementation details

The digital library system of DjDL is a standard client-server application built on the .NET Framework 4.5 and ASP.NET MVC 5 [1]. The tool used for its implementation is Microsoft Visual Studio Ultimate 2012 with additional packages for ASP.NET MVC 5.

Време на добавяне	Име на файла	Потребителско име	Текст	Действия
13.05.2015 17:04:48	ba_001_1_01	admin	Този запис е от...	Детайли Изтрий
07.05.2015 18:49:44	td_139_2_35	a	Много красива...	Детайли Изтрий
02.03.2015 13:29:40	ba_001_1_01	admin	Тестов коментар	Детайли Изтрий
22.02.2015 14:05:19	ba_001_1_01	admin	Чувам "предуму...	Детайли Изтрий
22.02.2015 13:27:08	td_134_2_03	admin	Навсякъде се чу...	Детайли Изтрий
22.02.2015 13:26:54	td_134_2_03	admin	Под нотите е: н...	Детайли Изтрий
22.02.2015 13:12:07	td_134_2_03	admin	Навсякъде се чу...	Детайли Изтрий
21.02.2015 14:57:28	ba_001_1_01	admin	Песента "Зъбли...	Детайли Изтрий
13.02.2015 13:08:49	td_178_2_06	admin	Този запис е от...	Детайли Изтрий
13.02.2015 13:08:44	td_178_2_06	admin	"не паят" вмест...	Детайли Изтрий
10.02.2015 11:14:42	ba_001_1_07	admin	Добавям "дето...	Детайли Изтрий
10.02.2015 11:04:39	td_140_2_49	admin	Изпълнител: Ги...	Детайли Изтрий
10.02.2015 10:52:58	ba_001_1_01	admin	Чувам "Русалма"	Детайли Изтрий
05.02.2015 13:38:52	td_140_1_27	a	"жата" вместо "с...	Детайли Изтрий
05.02.2015 13:37:18	td_094_2_19	a	"гладна и жадна...	Детайли Изтрий

Fig. 9. Comments referring the lyrics of particular songs.

A class library called RDXMLClassLibrary has been especially built for the purpose of automatic conversion of the original files with metadata and texts of songs to the RDF format. It implements the RDF 1.1 XML Syntax standard of W3C.

Another relatively new library used in the project is SignalR. It enables one to add real-time functionality to the software application. In other words, it provides the server's capability for sending data to all clients in real time.

The jQuery library v. 1.10.2 has been used for JavaScript processing.

To generate files with "standard" musical notations and MIDI files with melodies of songs from the original source files, LilyPond [8] should be installed as an external software package on the server.

The software implementation is based on Entity Framework 5 technology in combination with Code First [4]. This enables one to build first of all the data model and then to create the database. The current version of DjDL uses a local database (SqlLocalDB v. 11.0).

Figure 10 shows the software architecture of the digital library system of DjDL.

The subject ontology is created using Protégé 4.3. Most concepts of this ontology are constructed as defined OWL 2 classes, by means of necessary and sufficient conditions defined in terms of proper restrictions on certain properties. The current version of DjDL includes one more ontology (named Mood dictionary on Fig. 10) which contains some of the subject ontology classes and a "mood_estimation" property. The values of this property are integers that play the role of sentiment estimates of the respective concepts. The Mood dictionary should be considered as a part of the subject ontology and will be merged with it when the values of the "mood_estimation" property will be made sufficiently precise.

8. Conclusion

Semantic digital libraries integrate heterogeneous information resources based on various types of metadata. They provide interoperability at the semantic level with other digital library and information systems and deliver user friendly and adaptive search and document retrieval interfaces. The availability

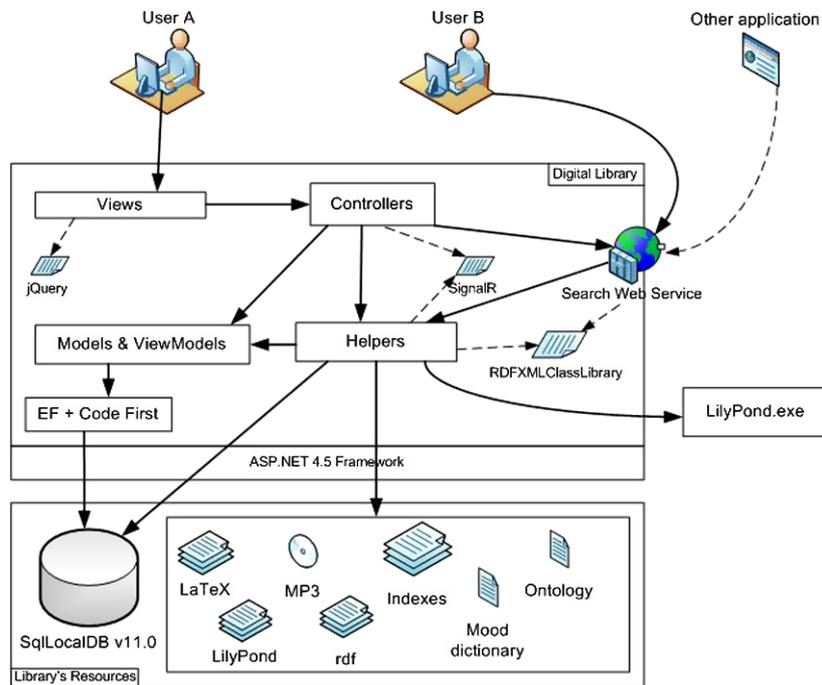


Fig. 10. Software architecture of the digital library system of DjDL.

and purposeful use of explicit conceptual knowledge at appropriate level(s) of abstraction in a digital library system may significantly improve the precision and recall of its search engine as well as give it some of the principal characteristics of a social semantic digital library.

References

- [1] ASP.NET MVC 5 official website: <http://www.asp.net/mvc/mvc5> [Accessed November 2015].
- [2] L. Candela, D. Castelli et al., *The DELOS Digital Library Reference Model: Foundations for Digital Libraries*, 2007. ISSN 1818-8044, ISTI – CNR.
- [3] T. Dzhidzhev, *Folk Songs from Thrace*, L. Peycheva, G. Grigorov and N. Kirov, eds, Prof. Marin Drinov Academic Publishing House, Sofia, 2013.
- [4] Entity Framework and Code First Documentation, available at: <https://msdn.microsoft.com/en-us/data/ee712907#codefirst> [Accessed November 2015].
- [5] N. Guarino, M. Carrara and P. Giaretta, Formalizing ontological commitment, in: *Proceedings of the 12th National Conference on Artificial Intelligence AAAI-94*, Seattle, Washington, The AAAI Press, Menlo Park, California, 1994, pp. 560–567.
- [6] R. Guha, R. McCool and E. Miller, Semantic search, in: *Proceedings of the 12th International World Wide Web Conference*, Budapest, Hungary, 2003, pp. 700–709.
- [7] IFLA/UNESCO Manifesto for Digital Libraries, available at: <http://www.ifla.org/publications/iflaunesco-manifesto-for-digital-libraries> [Accessed November 2015].
- [8] N. Kirov, Digitization of Bulgarian folk songs with music, notes and text, *Review of the National Center for Digitization* **18** (2011), 35–41.
- [9] S. Kruk et al., The role of ontologies in semantic digital libraries, in: *Proceedings of the European Networked Knowledge Organization Systems (NKOS) Workshop*, Alicante, Spain, 2006.
- [10] S. Kruk and B. McDaniel, Conclusions: The future of semantic digital libraries, in: *Semantic Digital Libraries*, S. Kruk and B. McDaniel, eds, Springer, 2009, pp. 215–222.
- [11] J. Lervik and S. Brygfjeld, Search engine technology applied in digital libraries, *ERICIM News* **66** (2006), 18–19.

- [12] M. Nisheva-Pavlova and P. Pavlov, Ontology-based search and document retrieval in a digital library with folk songs, *Information Services and Use* **31**(3,4) (2011), 157–166.
- [13] M. Nisheva-Pavlova, D. Shukerov and P. Pavlov, Building a social semantic digital library, in: *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science*, B. Schmidt and M. Dobрева, eds, IOS Press, 2015, pp. 63–72.
- [14] M. Nucci, M. Barbera, C. Morbidoni and D. Hahn, A semantic web powered distributed digital library system, in: *Proceedings of ELPUB 2008 Conference on Electronic Publishing*, Toronto, Canada, 2008, pp. 130–139.