KNOWLEDGE TECHNOLOGIES FOR DESCRIPTION OF THE SEMANTICS
OF THE BULGARIAN FOLKLORE HERITAGE[∗]

Desislava Paneva, Konstantin Rangochev, Detelin Luchev

Abstract: *Preserving and presenting the Bulgarian folklore heritage is a long-term commitment of scholars and researchers working in many areas. This article presents ontological model of the Bulgarian folklore knowledge, exploring knowledge technologies for presenting the semantics of the phenomena of our traditional culture. This model is a step to the development of the digital library for the "Bulgarian Folklore Heritage" virtual exposition which is a part of the "Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage" project.*

Keywords: *Knowledge Technologies, Ontology, Digital Libraries, Bulgarian Folklore, Ethnology.*

ACM Classification Keywords: *I.2.4 Knowledge Representation Formalisms and Methods, H.3.7 Digital Libraries – Collection, Dissemination, System issues.*

Introduction

In the first ICT Work Programme under the Seven Framework Programme of the European Community for Research and Technological Development (FP7), which defines the research priorities for 2007-2008, cultural heritage research is part of Challenge 4, named "Digital Libraries and Content". Its main objective is the development of "large-scale European-wide digital libraries of cultural and scientific multi-format and multi-source digital objects, assisting communities of practice in the creative use of content in multilingual and multidisciplinary contexts, and based on robust and scalable environments, cost-effective digitisation processes, semantic-based search facilities and tools for digital preservation" [ICT, '07].

The "Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage" (FolkKnow) project follows this idea and aims to build a multimedia digital library with a set of objects/collections, selected from the fund of the Institute for Folklore of the Bulgarian Academy of Science, which corresponds to the European and world requirements for such activities, and is consistent with the specifics of the presented artefacts. The project will use the knowledge technologies and digital libraries as they are the most suitable tools for semantic description and virtual multimedia presentation of cultural historical artefacts.

This paper presents the first stage of the work done on module 3 of the project, named "Development of Digital Libraries and Information Portal with Virtual Exposition "Bulgarian Folklore Heritage"". It tracks out the creation of Bulgarian folklore ontology, describing the knowledge about Bulgarian folklore objects and their main features, technical data or context. This ontology is the backbone of the subsequent work of folklore digital library development. Section 2 of the paper is a short description of the main issues of the FolkKnow project. In section 3, a brief outlook of the project's digital library is included. Section 4 summarises the current state of ontology

development. Sections 5 and 6 deal with different aspects of the Bulgarian folklore ontology including its scoping, conceptions, relations, and its implementation and utilization in the project.

Knowledge technologies for creation of digital presentation and significant repositories of folklore heritage

The aim of the project "Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage" is to build a multimedia digital library with a set of objects/collections, selected from the fund of the Institute for Folklore of the Bulgarian Academy of Science, which corresponds to the European and world requirements for such activities, and is consistent with the specifics of the presented artefacts. The complex structure and the multi-layer characters of the folklore objects require innovative approach for knowledge representation. The rich-in-content Web-presenting of the Bulgarian folklore knowledge defines the usage of modern methods and technologies for digital archive developing, which will be used not only for preservation and access to the information, but as a tool for scientific research analysis development. The main tasks is to create a digital library and information artery using knowledge-based technologies and Semantic web approach, in order to present in virtual form the valuable phenomena of the Bulgarian folklore heritage. The realization of the project gives possibility for wide social applications of the multimedia collections for the purposes of Interactive distance learning/self-learning, research activities in the field of Bulgarian traditional culture, and for the cultural and ethno-tourism in Bulgaria.

We assume that when Bulgarian folklore heritage is digitalized and presented virtually there will be a need of contemporary information technologies that allow complex multimedia presentations and descriptions, as well as broad and flexible access method. We believe that the digital libraries and Semantic web meet those requirements because they are powerful technologies for digitalization, semantic description, access provisioning, preservation, and virtual representation of cultural and historic values and especially of Bulgarian folklore heritage. The approach for building the module is formed as a result of the research experience of the team of Institute of Mathematics and Informatics and its know-how in multimedia applications gained in numerous European Information Society projects. It includes analytical research, choice and usage of suitable methods, tools and environments for digital representation and preservation of significant cultural and historical artefacts and their exposure into the global information space. This approach allows the integration of the idea of the traditional Bulgarian culture and folklore in the European culture space, while completely preserving its identity and diversity [Bogdanova et al. '06].

Digital library of Bulgarian folklore heritage

Digital libraries are a contemporary conceptual solution for access to information archives. They contain diverse hypertext-organized collections of information (digital objects such as text, images, and media objects) that are organized thematically and are managed by complex specialized services such as semantic-based search, multi-layer and personalized search, context-based search, relevance feedback, resource and collection management, metadata management, indexing, semantic annotation of digital resources and collection, grouping and presenting digital information, extracted from a number of locations, services for digital information protection and preservation, *etc.* [Pavlov&Paneva, '06].

Besides that the flexibility, the automatic adaptation, the access anywhere and anytime, the decentralization, the wide variety of digital objects and collections, the information security, *etc.* are already key requirements for the advanced multimedia digital libraries [Pavlova-Draganova et al., '07] [Paneva et al., '05] [Pavlov et al., '06a] [Pavlov et al., '06b].

Information about the actual state of the research of the architecture of digital libraries, informational access to audio-visual and non-traditional objects and semantic interoperability is contained in the FP6 project DELOS "A Network of Excellence on Digital Libraries" (http://www.delos.info).

Having in mind this variety of useful properties and characteristics of the large-scale repositories of digitized knowledge their use for presentation of the valuable phenomena of the Bulgarian folklore is not casual. There are some national investigations and projects concerning the virtual existence and the digitalization of ethnographic and folklore artefacts, for example, experimental digital archive "Bulgarian Ethnographic Treasury" (http://mdl.cc.bas.bg/ethnography/) [Luchev, '05] [Luchev, '06], project "Yuper" (http://yuper.hit.bg/), project "Folklore Motives and Anthologies" (http://liternet.bg/), project "WebFolk Bulgaria" (http://musicart.imbm.bas.bg/EN/Default.htm), project "Living Human Treasures" (http://www.treasures.eubcc.bg/main.php), project "Virtual Encyclopaedia of Bulgarian Iconography" (http://mdl.cc.bas.bg/), *etc.*

The FolkKnow multimedia digital library can be similar valuable gallery of artefacts and knowledge for Bulgarian culture, art and folklore that will present a relatively limited number of specimens of different folklore narrative types (songs, rituals, faith, knowledge, proverbs, magic, *etc.*) and their audio-visual documentation. Until now, the Bulgarian folklore is always shown partially only with text, sound or image, but the authors' demand is for joint unities of words, music and motions. This possibility can be provided by contemporary multimedia environments. The ambitions of the authors are the demonstration of unique music dialects from different local folklore areas and advanced approaches for folklore content prescription representation through authentic sounds, videos, and photos of live rituals. Part of the Bulgarian folklore specimens will be presented from asynchronous point of view; other will be in their diachrony – unique materials, saved for years. Another task is the different record technique demonstration – inquiry, interview, inclusive observation, *etc.*

Multimedia digital library of Bulgarian folklore expects a wide range of potential users – professionals and scientists, non-professionals, connoisseurs and viewers, *etc.*

Ontological presentation of folklore knowledge

Originally, the term ontology comes from philosophy where it is employed to describe the existence of beings in the world. In 1993, Gruber's definition becomes the most referenced on the knowledge technologies literature: "an ontology is a formal, explicit specification of a shared conceptualization" [Gruber, '93]. Conceptualization refers to an abstract model of phenomena in the world by having identified the relevant concepts of those phenomena. Explicit means that the type of concepts used and the constraints on their use are explicitly defined. Formal refers to the fact that the ontology should be machine readable, which excludes natural language. Shared reflects the notion that an ontology captures consensual knowledge, that is, it is not private to some individual, but accepted by a group.

Ontologies can be used for many different purposes. The literature on knowledge representation contains research on the use of ontologies for data-interchange, for data-integration, for data-querying or for data visualization. In general, visualization of information can be seen as a two-step process. In a first step, information is transformed into some intermediate semantic structure. This structure organizes the raw information into a meaningful structure. In a second step, this semantic structure is used as the basis for a formal visual representation. We will use this approach in our work on the Bulgarian folklore ontology development.

Tools for building ontologies usually provide a graphical user interface that allows ontologists to create ontologies without using directly a specific ontology specification language. Some tools have been created for merging and integrating ontologies [Fensel, '04].

Recently, many ontology languages have been developed in the context of the World Wide Web: Resource Description Language (RDF), RDF Schema, Simple HTML Ontology Extensions (SHOE), Ontology Exchange Language (XOL), Ontology Markup Language (OML), Web Ontology Language (OWL), Ontology Inference Layer (OIL), DAML+OIL, *etc.* Their syntax is based on XML, which has been widely adopted as a 'standard' language for exchanging information on the web [Fensel, '04].

To efficiently represent the folklore annotation framework and to integrate all the existing data representations into a standardized data specification, the folklore ontology need to be represented in a format (language) that not enforce semantic constraints on folklore data, but can also facilitate reasoning tasks on folklore data using semantic query algebra. This motivates the representation of Bulgarian folklore ontology model in Web Ontology Language (OWL). OWL facilitates greater machine interpretability of Web content than that supported by XML, RDF, and RDF Schema by providing additional vocabulary along with a formal semantics. Knowledge captured from folklore data using OWL is classified in a rich hierarchy of concepts and their inter-relationships. OWL is compositional and dynamic, relying on notions of classification, reasoning, consistency, retrieval and querying. We investigated the use of OWL for making Bulgarian folklore ontology using Protégé OWL Plug-in.

Ontology of Bulgarian folklore

Since one of the targets of the FolkKnow project is to present the valuable phenomena of the Bulgarian folklore in suitable virtual form using knowledge technologies, we have to observe and specify the experience that has been gained in the last 500 years in the area of traditional folklore *i.e.* to construct Bulgarian folklore domain ontology.

FolkKnow annotator/indexers using this ontology will semantically describe and index the raw audiovisual content in order to create and maintain reusable digital objects.

The ontology will be used also to realize semantic-based access to concrete digital objects, representing folklore objects, described by their main features, technical data or context. All this information is included within the Folklore Ontology Concept – the root concept for the ontology.

The process of building of the Bulgarian folklore ontology for the FolkKnow project is necessarily iterative. The first activity is the definition of the scope of the ontology. Scoping has been mainly based on several brainstorming sessions with folklorists and content providers. Having these brainstorming sessions allowed the production of most of the potentially relevant terms. At this stage, the terms alone represented the concept, thus concealing significant ambiguities and differences of opinion.

A clear issue that arose during these sessions was the difficulty in discovering of definite number of concepts and relations between these concepts. The concepts listed during the brainstorming sessions were grouped in areas of work corresponding naturally arising sub-groups. Most of the important concepts and many terms were identified. The main work of building the ontology was then to produce accurate definitions.

Description of the conceptions

The scientific classification and documentation of folklore objects provide folklorists and content generators with a rich knowledge background with plenty of multidimensional data and metadata. There is a special relation among the metadata, which reveals all the knowledge concerning the folklore object obtained from the classification procedure.

The folklore object is related to three levels of knowledge, enriched with a set of sub-levels of the data classification. All these levels of knowledge or "thematic entities" in the ontology conception are supported by the scientific diagnosis results and the related documentation.

The entity "Identification and description" consists of general historical data, identifying aspects such as title, language, archival signature, period, current location of the folklore object, annotation, first level description, second level description, *etc.*,

The entity "Technical" includes technical information both revealing the technologies used for folklore object capturing and recording, record situation, record type, record place, record date, main participants in the process (record maker and informant), *etc.*

These main entities and their metadata are supported, documented and provided by the scientific diagnosis, which has been applied to the folklore objects.

Ontological model

We will present the Bulgarian folklore ontological model using classes of concepts, organized in taxonomy and table with properties.

Taxonomies are used to organize ontological knowledge using generalization and specialization relationships through which simple and multiple inheritances could be applied. Properties are an interaction between individuals of the domain-classes and the range-classes.

Figure 1 depicts the main concepts and properties in the Bulgarian folklore ontological model.
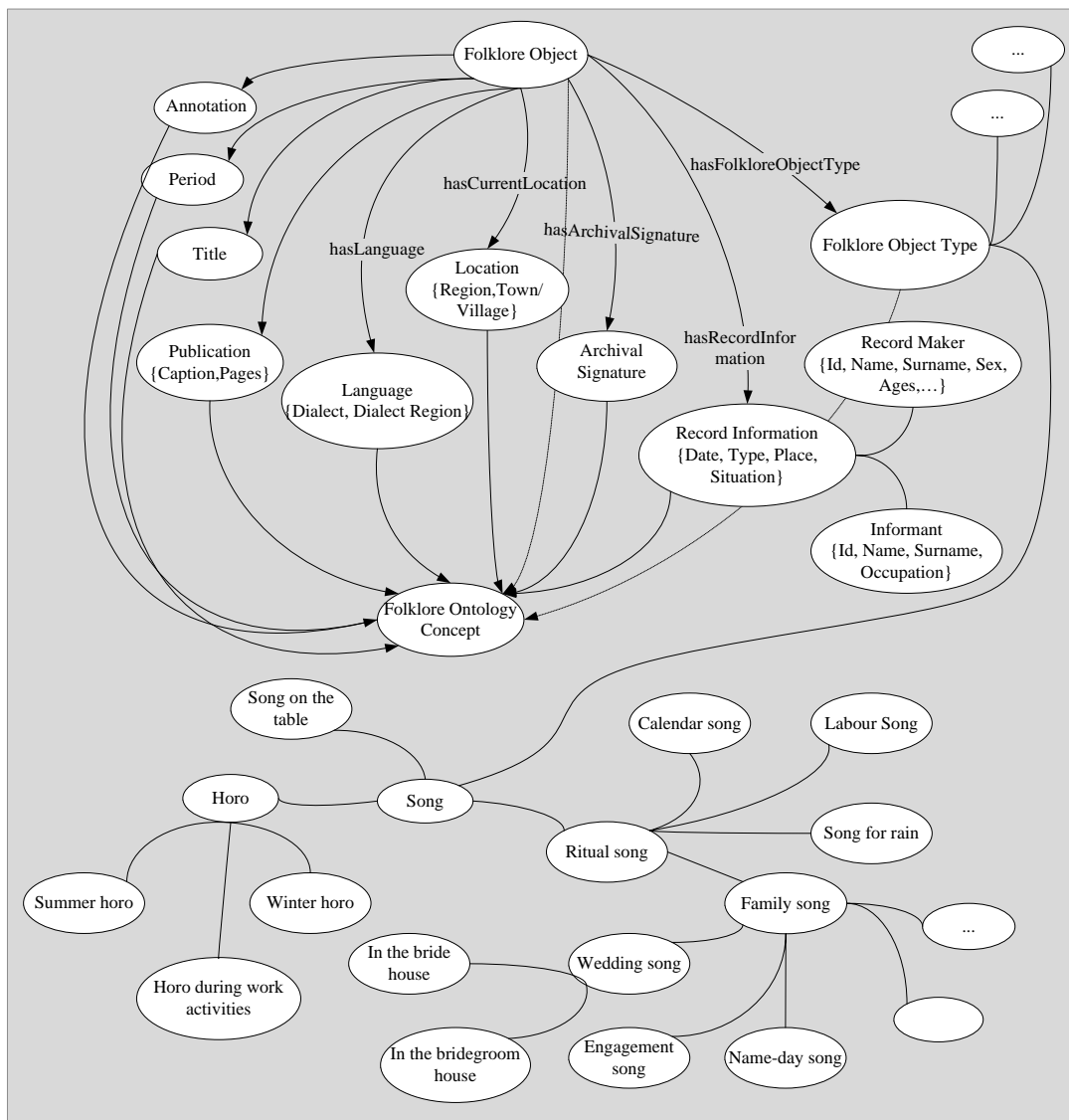
Figure 1: Part of the concepts and properties in Bulgarian folklore ontology

The most representative concepts have been defined first and then they have been specified appropriately in order to get a representation of the knowledge stored in the databases. The Bulgarian folklore ontology is composed of 70 concepts and 82 properties.

OWL properties represent semantic relationships between classes of objects. Below, a piece of the Bulgarian folklore ontology code, defining the property *isRecordPlaceOf* is presented. Here the declaration of the transitivity condition and the definition of property *hasRecordPlace* as its inverse can be seen.

```
<owl:ObjectProperty rdf:ID="isRecordPlaceOf">
  <rdfs:domain rdf:resource="#Record_Place"/>
  <owl:inverseOf>
   <owl:ObjectProperty rdf:ID="hasRecordPlace"/>
  </owl:inverseOf>
```

```
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Represents that the element "Record_Place" is a record place of the folklore object.
    </rdfs:comment>
    <rdfs:range rdf:resource="#Folkore_Object"/>
</owl:ObjectProperty>
<owl:ObjectProperty rdf:about="#hasRecordPlace">
    <rdfs:range rdf:resource="#Record_Place"/>
    <rdfs:domain rdf:resource="#Folkore_Object"/>
    <owl:inverseOf rdf:resource="#isRecordPlaceOf"/>
    <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
        Represents that the element "Folklore_Object" has a record place.
    </rdfs:comment>
</owl:ObjectProperty>
```
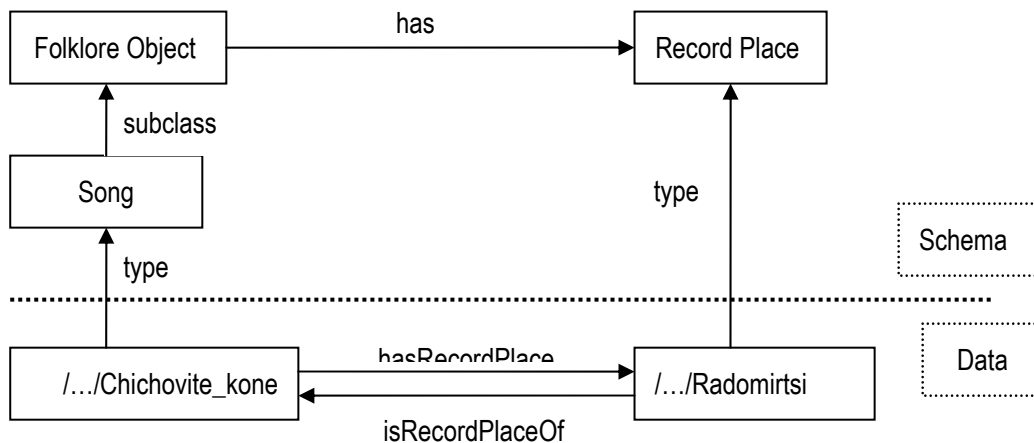

Figure 2: Scheme of relationships between classes of objects in Bulgarian folklore ontology

Implementation of Bulgarian folklore ontology

Knowledge about Bulgarian folklore heritage and oral tradition is interesting not only for the wide audience of professionals (historians, philologists, psychologists, ethnologists etc), but also for non-professionals and institutions dealing with these problems. Folklore heritage specialists will reach to organized objects and semantically structured knowledge for their investigations. For example, the searching for an object "Ritual", semantically connected with an object "Festival", can give back not only the all rituals of the festival, but also the "Songs", "Faith and Knowledge", "Magic", "Food", *etc.*, semantically bound up with both "Ritual" and "Festival". Cultural institutions and organizations, as well as non-professionals will be able to find information for semantically joined complexes of folklore objects on the base of starting points as "Location", "Period", "Language/Dialect", *etc.*

The ontology gives the ability to describe the semantics of folklore content and to use new knowledge management services such as semantic search across aggregations of varied and complex sub-classes and objects in a robust, rich and user-friendly manner, personalized search, context-based search, multi-criteria search, metadata management, *etc.*

The semantically annotated objects can also be used as a base for eLearning courseware development; for example, folklore objects can be easily discovered and grouped in learning lessons, modules or parts of them.

Bibliography

[Bogdanova et al. '06] Bogdanova G., Pavlov R., Todorov G., Mateeva V. (2006), Knowledge Technologies for Creation of Digital Presentation and Significant Repositories of Folklore Heritage, Advances in Bulgarian Science, pp. 7-15.

[Pavlov&Paneva, '06] Pavlov R., Paneva D. (2006), Toward Ubiquitous Learning Application of Digital Libraries with Multimedia Content, Cybernetics and Information Technologies, vol. 6, № 3, pp. 51-62.

[Pavlova-Draganova et al., '07] Pavlova-Draganova L., Georgiev V., Draganov L. (2007), Virtual Encyclopeadia of Bulgarian Iconography, International Journal "Information Technologies and Knowledge", vol.1, №3, pp. 267-271.

[Paneva et al., '05] Paneva, D., Pavlova-Draganova L., Draganov L. (2005), Digital Libraries for Presentation and Preservation of East-Christian Heritage, Proceedings of the Second HUBUSKA Open Workshop "Generic Issues of Knowledge Technologies", Budapest, Hungary, pp. 75-83.

[Pavlov et al., '06a] Pavlov R., Pavlova-Draganova L., Draganov L., Paneva D. (2006), e-Presentation of East-Christian Icon Art, Proceedings of the Fourth HUBUSKA Open Workshop "Semantic Web and Knowledge Technologies Applications", Varna, Bulgaria, pp. 42-48.

[Pavlov et al., '06b] Pavlov, R., Paneva D., Pavlova-Draganova L., Draganov L. (2006), Digital Libraries with Multimedia Content and Applications in Bulgarian Cultural Heritage (Analytical study), State Agency for Information Technologies and Communication (SAITC), by contract № 8/21.07.2005 between IMI-BAS and SAITC, 2006, Sofia, Bulgaria, Available at: http://mdl.cc.bas.bg/Digital_libraries_with_multimedia_content_and_applications_in_Bulgarian_ cultural_heritage.pdf (accessed on April 10, 2007).

[ICT, '07] Information and Communication Technologies, Work Programme 2007-08, Available online: ftp://ftp.cordis.lu/pub/fp7/ict/docs/ict-wp-2007-08_en.pdf (accessed on April 10, 2007)

[Fensel, '04] Fensel D. (2004), Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce, Second edition.

[Gruber, '93] Gruber T. (1993), Towards Principles of the Design of Oncologist Used for Knowledge Sharing. International Journal of Human Computer Studies 43 907-928, Available at: http://www.itee.uq.edu.au/~infs3101/_Readings/OntoEng.pdf (accessed on April 10, 2007).

[Luchev, '06] Luchev, D. (2006), Experimental Digital Library - Bulgarian Ethnographic Treasury, Proceedings of the Modern (e-) Learning Conference, Varna, Bulgaria, pp. 106-111.

[Luchev, '05] Luchev, D. (2005), Experimental Digital Archive "Bulgarian Ethnographic Treasury" (Analytical study) by contract № 12/08.09.2005 between SAITC and EIM-BAS, IMI-BAS (subcontractor), Sofia, Bulgaria, Available at: http://mdl.cc.bas.bg/DigLibr_BulEthnoTreasure_Analytical_Study_Luchev.pdf (accessed on April 10, 2007).

Authors' Information

Desislava Paneva - Institute of Mathematics and Informatics, BAS, Bulgaria, 8, Acad. G. Bonchev str.; e-mail: dessi@cc.bas.bg

Konstantin Rangochev – Institute of Mathematics and Informatics, BAS, Bulgaria, 8, Acad. G. Bonchev str.; e-mail: krangochev@yahoo.com

Detelin Luchev - Ethnographic Institute with Museum, BAS; Bulgaria, 1000 Sofia, Moskovska str. 6A; e-mail: luchev_detelin@abv.bg