ONTOLOGY OF THE BULGARIAN CULTURAL HERITAGE. BULGARIAN FOLKLORE

Tsvetomira Ivanova[1], **Iva Nedeleva**[2]

Abstract: This article presents the creation of an ontology describing Bulgarian folklore. The ontology has been developed based on the CCO (Cataloging Cultural Objects) standard. CCO leads us to provide data content standards. The set of ontologies is a knowledge base used by intelligent agents. One of the main problems with systems and developments describing different cultural and/or historical sites is the difficulty in sharing information between individual societies or systems. The use of standards greatly facilitates the dissemination of data. The CCO standard provides clear and precise definitions of the individual attributes that an object must possess. The ontologies developed by Protégé are designed in such a way as to meet the requirements of the standard. The created ontologies describe 7 folklore areas with specific ones being: music, dances and musical instruments. The work on creating this ontology directs us to create web applications of the developed ontologies.

Introduction

Ontology and Cataloging Cultural Objects(CCO)

Semantics in the field of information technology explores the meaning of things. In computer science, its analysis and interpretation usually use formal systems based on mathematical logic. "Semantic computing" refers to attempts to define the logical systems of computers, i.e. to define a formal language, grammar and inference apparatus as a format, data and computer program.

The term "ontology" comes from philosophy and expresses the nature of existence: the nature of things that actually exist and how to describe them.

Their connection is between the concepts of Cyber-Physical Spaces (CPS) and the Internet of Things (IoT). The connection between them is based on the integration of the virtual and physical worlds. When the person is at the center of such spaces, they become Cyber-Physical-Social Spaces (CPSS). Professor Stoyanov et al. (2019) names the CPSS ecosystem as the Virtual-Physical Space (ViPS). This ecosystem uses an agent-oriented approach that focuses on providing architectural and semantic modeling services. Agents that represent architectural components provide dynamism, flexibility and intelligence. Semantic models and ontologies in particular are an effective complement to enhance the intelligence of the space.

The ontology is not only a standard classification or a hierarchical representation of concepts. The Web Ontology Language (OWL) ontology is based on Description Logics (DLs) – a class of formal languages used to represent knowledge. OWL seeks maximal expressiveness in logic description. This technology is popular in different areas of science such as medicine, biology, cultural-historic heritage, art, etc.

The semantic models for data normalization are wide-spread relational databases and their utilization in different areas proves that the relational model of data is successfully applied in modeling, but creating a database using the normalization of data is a complicated process and this needs the use of more-convenient and powerful tools for modeling the respective data and requires the usage of semantic models of the data. The main advantage of semantic models is making the visualization of semantics for the data possible. This is how important features of objects are defined, their attributes and the link between them.

Throughout the years, semantic models are developed in two main streams, which were established in the 1970s by Cod and Chen. The development of semantic models leads to a better expression of the meaning of data in those models, making them simpler to learn and fit them for usage by Case, with more realism and detailed descriptions when digitalizing real world objects. The first semantic model, Entity-Relationship (ER) was developed by Dr. Chen in 1976.

Modeling the object is based on using graphical diagrams, containing a set of different components: objects, attributes and the link between objects. The most prominent tool for data modeling is the

[1] Plovdiv University "Paisii Hilendarski", Faculty of Mathematics and Informatics, Computer Systems Dept., Bulgaria, tsvetomira_kazashka@abv.bg

[2] Plovdiv University „Paisii Hilendarski", Faculty of Mathematics and Informatics, Computer Systems Dept., Bulgaria, iva.nedeleva@gmail.com

diagrams of the entity-relationship type. ER diagrams help us define the entities, their attributes and the relationship with other entities. These diagrams are used when developing relational databases. Chen's model is also further developed by Barker where Barker's model uses Case by Oracle.

Objects by Cod can be associated with their analogues in the ER model in the following way: The Code matches the Object, the Characteristic matches the Attribute, the Association matches the Relationship (only for links of the type m-n).

Allemang and Hendler (2011) discuss the capabilities of Semantic Web modeling languages such as RDFS (Resource Description Framework Schema) and OWL (Web Ontology Language). The authors used the life and works of William Shakespeare to demonstrate some of the most basic capabilities of the Semantic Web. They discuss semantic modeling and how it can support the development from chaotic information gathering to one characterized by information sharing, cooperation, and collaboration.

The Standard Cataloging of Cultural Objects (CCO) is of great importance for the functioning of cultural institutions, libraries, museums, and institutions that deal with the description and documentation of works of art, architectural monuments, artifacts, or photographs. It is also very important for the specialists who work in these institutions like museum documentation experts, curators, archivists, and librarians. The CCO standard is not a guide for system design, but it could also be useful by system designers who need to understand the nature and form of information of a cultural object. In this sense, based on the SSO standard, our work is unified.

The main focus of the CCO is descriptive metadata and power control - data designed to describe a cultural work, data used to create and catalog records of that work and its images. Administrative metadata (data used in the management and administration of information resources) and technical metadata (data to write properties of a digital image file) are excluded, except where it is appropriate to discuss descriptive metadata.

Ten key principles of CCO:

1. Establish the logical focus of each work record, whether it is a single object, a work composed of several parts, or a physical group or collection of works. Let there be a clear distinction between Work Records and Image
2. Inclusion of all necessary CCO elements.
3. Following the rules of the CCO, creating additional rules to disseminate information for efficient retrieval, redirection and exchange.
4. Use of control dictionaries.
5. Use of reliable dictionaries, with correct terminology, published as standard.
6. Use established metadata standards, such as VRA Core Categories or categories to describe works of art.
7. Presentation of classification, indexation and the display of different but related functions.
8. Consistency in establishing the relationships between the works.
9. Consistency in capital letters, punctuation and syntax. Avoid layoffs.
10. Use of data values in English when possible.

Access to systematic information is of particular importance for the cultural and creative industries. Ruseva et al. (2018) formulate the need for projects in the field of art as follows: the digitalization of cultural heritage is a necessity in the modern IT world which predisposes to permanent storage, easier use, and easier distribution. At the same time, the digitalization of material and non-material cultural heritage requires resources - time, money, and people. In a market economy, the acquisition of these resources puts cultural managers in a difficult situation and requires flexibility, activity and creative solutions. Ruseva et al. (2018) see the participation and development of projects at the national and European level as an opportunity to reduce the financial burden. Through the active use of financial instruments for financing culture and art, the necessary resources for the digitization and preservation of cultural heritage can be provided.

Standardized information is also needed to ensure quality projects. In this context, creating standard ontologies is a necessary process.

Data and methodology

The ontologies developed by Protégé are designed in such a way as to meet the requirements of the standard. Glossaries of terms are also maintained, as well as the objects themselves with their attributes.

The model for semantic modeling uses RDF. Further usage of RDFS and OWL provide additional powerful tools for modeling and ontologies.

Results and Discussion

Protégé is selected for the development of ontologies. It provides a wide range of axioms that can be created and typological groups of Bulgarian folklore be defined with as much detail as possible. For those reasons, seven ontologies were created according to the CCO standard (Baca et al., 2006). Using standards facilitates the sharing of knowledge between different societies and systems.

Seven ontologies were created (Figure 1), most of which correspond to the dictionaries of the standard. These ontologies are: *folklore region*, *Subjects, Agents, Musical instruments, Objects, Bg folk music, Bg folk dance*. Each of them describes a concept and its characteristics necessary for the presentation of Bulgarian musical folklore. An ontology was also created that interprets the idea of the object from the CCO. The object described in this ontology maintains the elements required by the standard but additional features were also added to fully represent the concepts. The first ontology contains knowledge about various landmarks in Bulgaria.

Figure 1: DateBase and CCO standard
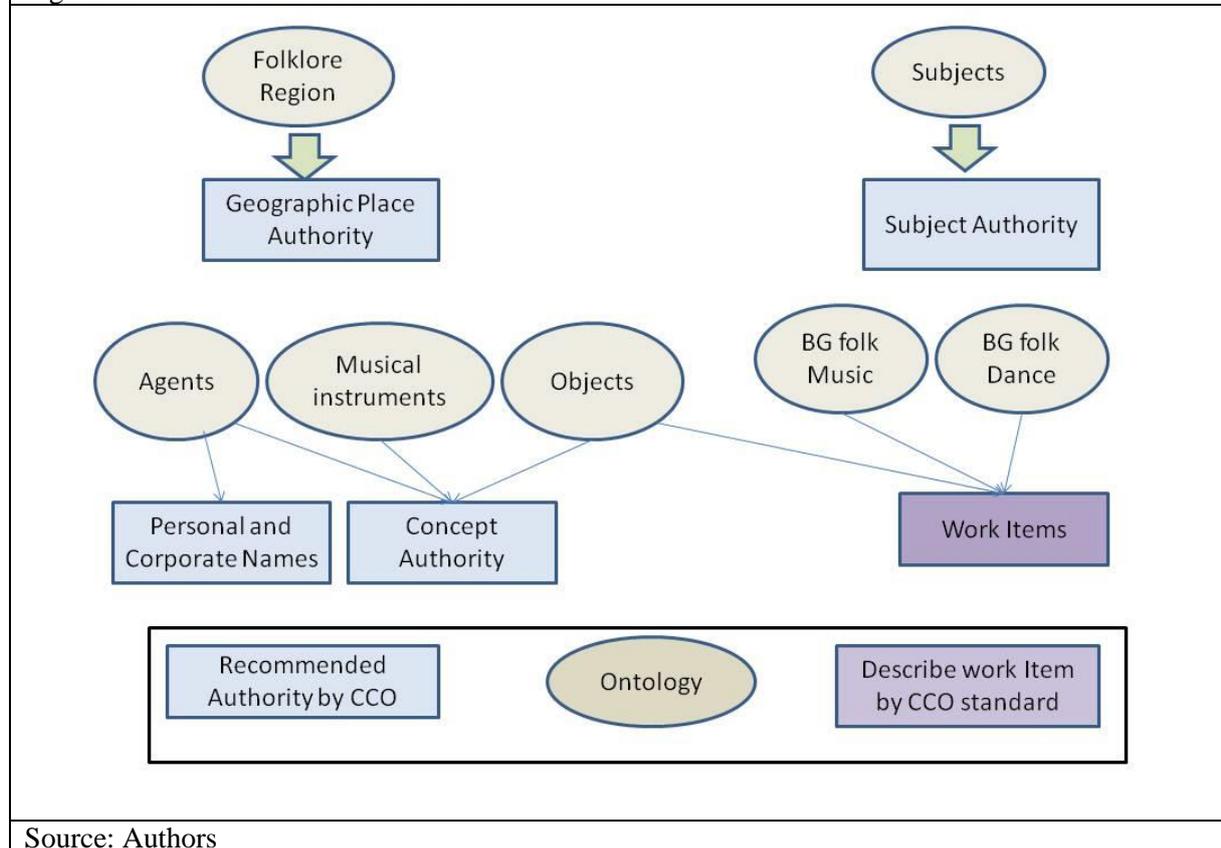
Source: Authors

The BG Folklore Ontology

One of the most interesting ontologies that were realized is the one that describes the traditional Bulgarian folklore music and dance.

Adhering to the CCO standard allows for the division of notions into separate ontologies and at the same time they are processed as a collective unit by the intelligent agents. IRI was selected to manage classes, individuals and attributes. Each defined class, individual and attribute has its unique IRI. Therefore, a class defined in one ontology can be used by another by referencing the unique IRI given when the concept was initially defined in the first ontology.

The ontologies created support two languages – English and Bulgarian. The names of the classes, attributes and individuals were created in English, using the CamelCase naming convention. Tags are used to provide easily readable text for the users, both in English and Bulgarian.

For example, a class named Bulgarian Regions has two tags added: one for Bulgarian with value "oblasti" and an additional attribute to the tag [language:bg]. The other tag is in English with value 'areas' and attribute [language:en].

In the Musical Instruments ontology, there are three types of classes: the first type encompasses those defined in the Bulgarian Folklore ontology. Notions and definitions from other ontologies are used for their description; the second type are definitions of notions already defined in other ontologies.

The main class in Bulgarian Folklore is defined in the Objects ontology where characteristics of Bulgarian folklore are defined. The other three ontologies – Bulgarian folklore regions, musical instruments and traditional dances are defined based on it. See Figure 2.

Figure 2: Ontology Bulgarian folklore

Source: Authors
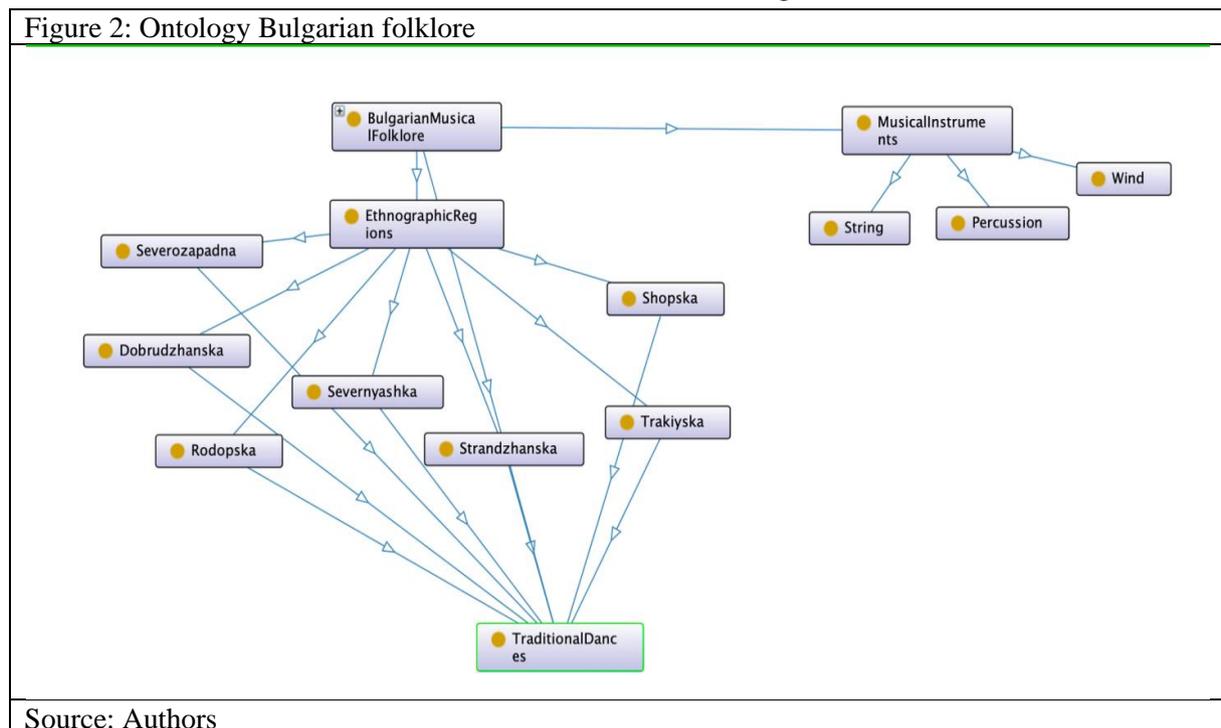
The knowledge of the folklore heritage of Bulgaria is acquired through encyclopedias in order to get a better understanding of the structure and characteristics of the separate elements included in the ontologies.

At this stage we have the following elements: the folklore regions and their geographical characteristics, musical folklore of each of the regions, dance folklore and musical instruments.

Researchers of Bulgarian folklore define seven ethnographic folklore regions: Severyashka, Thrakiyska, Strandzhanska, Shopska, Dobrudzha, Makedonska and Rhodopska (Peycheva, 2008). Each of these ethnographic regions has unique geographical borders.

The musical folklore in each region has its own characteristics and peculiarities. There are areas where a mix of folklore can be observed but this is usually due to the relocation of large groups from one region to another, or due to the geography and the close distance between those regions, but in the end, each folklore region has its own characteristic traditional dances. For example, the Strandzha folklore region's population has kept ritual tunes called "filek", which are performed during lent. Those tunes are accompanied by characteristic tunes, measured in 2 and 8 bars. Songs in the Shopska region are sang with two voices – one of the singers sings the melody while the other keeps the main line. The songs from Dobrudzha region sound rich in ornaments.

The dance folklores of the different ethnographic regions also have common characteristics and individual features. There is a major difference, for example, in the bar measurements of the tunes,

therefore dances are completely different. Similar differences are observed when it comes to the traditional folklore instruments of each region.

Conclusion

The task of collecting and organizing the knowledge of the Bulgarian Folklore Heritage into ontologies is quite a challenging and big task. This article sets the basis for further development of a complete model of realization of an intelligent agent to serve the needs of lecturers and students in the field of Bulgarian Folklore. We are aware that the development of those ontologies requires a lot of effort and resources – time, trusted sources of information, systematization of the data, etc., this is why, at this given time, we could expect that not all the ontologies will be fully developed. We plan to create an ontology related to the characteristics of Bulgarian folk dance, which will be useful for students and researchers. The ontology will describe the Bulgarian folk dances by districts and will reflect the specifics of the dance in each folklore area.

References

Allemang D., Hendler J., (2011)Semantic Web for the WorkingOntologist: Effective Modeling in RDFS and OWL, Second Edition, Elsevier Inc. DOI: https://doi.org/10.1016/C2010-0-68657-3

Baca, M., Harpring, P., Lanzi E., McRae, L., Whiteside, A. (2006).CatalogingCulturalObjects. A Guide to DescribingCultural Works and TheirImages. Chicago, United States of America.

Peycheva, L., (2008).Between the Village and the Universe: the old folk music from Bulgaria in the new times. Sofia, Academic Publishing House "Prof. MarinDrinov"

Ruseva M., Kazashka V., Kolev V. (2018).Analysis of the specific profile of Art organizations in Plovdiv,Bulgaria CBU International Conference "Innovation in Science and Education", DOI: http://dx.doi.org/10.12955/cbup.v6.1239

Stoyanov. S., et al. (2019). Cyber-Physical-Social Systems and Applications. Part I: Reference Architecture. LAP LAMBERT AcademicPublishing.