An Approach to Data Representation and Processing of Knowledge from the Bulgarian Folklore

Veneta Tabakova-Komsalova[1, 2[0000−0002−0617−9844]], Asya Stoyanova-Doycheva[1, 2[0000−0002−0129−5002]], Alexander Petrov[1], Laska Kostadinova-Tzankova[1]

[1] University of Plovdiv "Paisii Hilendarski", Plovdiv, Bulgaria
[2] Institute of Information and Communication Technologies, Bulgarian Academy of Sciences, Sofia, Bulgaria
v.komsalova@uni-plovdiv.bg, astoyanova@uni-plovdiv.bg, apetrov@uni-plovdiv.bg, laska_kos@abv.bg

Abstract. In the context of global and high-tech development, cultural fields are often overlooked during these transitions. Therefore, it is essential to pay close attention to their inclusion. Cultural-historical heritage, which has undergone centuries of changes, deserves to be introduced into the technological world and made as accessible as possible to society, without territorial or border restrictions. Digitizing a part of this heritage, including folklore, can make it high-tech while preserving its character, peculiarity, and identity. This article examines the digitalization of Bulgarian cultural and historical heritage. It explores existing international standards for presenting cultural heritage and proposes a formalism for an adequate presentation of knowledge in the field of Bulgarian folklore. This article presents the first version of a model for the formal description of objects from Bulgarian folklore. We provide examples of the formal presentation of objects from Bulgarian folklore and their corresponding computer implementation.

Keywords: Cultural-Historical Heritage, Modelling and Knowledge Processing, Formal Models, Prolog, Ontologies.

1 Introduction

Cultural and historical heritage is an invaluable asset for humanity. It is a reminder of our past, our traditions and our cultural identity. However, these treasures are vulnerable to damage, loss and destruction over time. To prevent their destruction and preserve them for future generations, digitisation of cultural and historical heritage has become a crucial tool. Through digitisation, heritage can be made accessible to a wider audience and protected from the ravages of time.

Digitisation of cultural and historical heritage has many advantages. First, it makes these treasures more accessible to a wider audience. Digitised objects can be shared online, making them accessible to people who would not otherwise have access to them.

This is particularly important for people living in remote areas or in countries with limited access to cultural and historical artefacts. Through digitisation, they can learn about and appreciate the cultural wealth of other parts of the world. Secondly, digitisation improves the preservation of cultural and historical heritage. Digital copies of fragile or damaged objects can be made, allowing their restoration and preservation. This ensures that these treasures are protected so that they can be appreciated by future generations. Thirdly, digitisation allows for easier and more efficient research and study. Digitised objects can be catalogued and indexed in a way that facilitates access to them. This is particularly important for researchers and scholars, who can now access vast numbers of materials from the comfort of their own home. In addition, digitisation enables the creation of virtual exhibitions and other educational resources, making it easier to study and appreciate cultural and historical heritage.

To digitisation of cultural and historical objects, various international standards are used that provide a precise structure for how objects should be described to facilitate the exchange of information between different preservation and cataloguing systems. Usually, these standards deliver partially formalized representations of cultural and historical objects, so further formalization is needed when using them to create computer applications. Moreover, they are partially usable for representing the specificity and national specificity of our cultural and historical heritage (e.g. Bulgarian folklore). For example in (Madanska, 2022) is presented a semantic model of the Bulgarian revival architecture. Ontology contains information about old houses in the suggested period, their functions, materials and architecture in accordance with CCO standard. In (Paneva-Marinova, Pavlov, & Rangochev, 2010) authors present the library of Bulgarian traditional culture and folklore. They represent knowledge in ontologies. Also in (Ivanova & Nedeleva, 2021) authors develop an ontology of Bulgarian folklore dances and it is organised also in accordance with CCO standard.

In this paper, we present a different structural approach for representing and describing objects of Bulgarian folklore. The idea is that one can exploit the structure of folklore motifs for their formal representation, which is relatively easy to transform into a computer model. With examples we demonstrate the applicability of the approach.

The paper is structured as follows: the next section provides a brief overview of standards for cataloguing cultural objects. The third section briefly describes the digitization of Bulgarian folklore. The fourth section presents the approach that we use and describes the formal model we have presented. It demonstrates and gives examples from Bulgarian folklore. In the fifth section, we conclude by discussing the potential of the presented formal model and future tasks.

2 Culture Cataloguing Standards

One of the first and most widespread is the CCO standard. The Cataloguing Cultural Objects (CCO) standard (Baca, Harpring, Lanzi, McRae, & Whiteside, 2006) is a set of guidelines developed by the Visual Resources Association (VRA) and the Art Libraries Society of North America (ARLIS/NA) for cataloguing and describing

cultural objects in digital format. CCO standard describes cultural objects in a structured and consistent manner, using a set of guidelines and rules for cataloguing and describing various aspects of cultural objects. The standard includes instructions on identifying cultural objects, describing their physical attributes, content, and contextual information, and managing digital images and associated metadata.

Authority files (the main elements in the standard) are standardized lists of names, titles, and other data used to describe cultural objects, people, organizations, and other entities in a consistent and controlled manner. Authority files help to ensure accuracy, consistency, and interoperability in the description of cultural objects, and they are widely used in cultural heritage institutions such as libraries, archives, and museums. There are several types of authority files commonly used in the cultural heritage field, including:

- Name authority files: These contain standardized names and variants of names for individuals, groups, and corporate bodies. Name authority files help to avoid confusion and ambiguity when referring to people and organizations in the description of cultural objects.
- Subject authority files: These contain standardized subject terms and keywords used to describe the content of cultural objects. Subject authority files help to ensure consistency in describing the topics, themes, and concepts represented in cultural objects.
- Geographic authority files: These contain standardized place names and geographic coordinates used to describe the location and context of cultural objects. Geographic authority files help to ensure consistency in describing the places and regions depicted in cultural objects.
- Genre/form authority files: These contain standardized terms and categories used to describe the type and format of cultural objects, such as photographs, paintings, sculptures, and manuscripts. Genre/form authority files help to ensure consistency in describing the physical and stylistic characteristics of cultural objects.
- Authority files for controlled vocabulary lists: These contain standardized terms and phrases used to describe the content of cultural objects, such as the Getty Art and Architecture Thesaurus or Library of Congress Subject Headings.

By using authority files, cultural institutions can ensure that the description of cultural objects is accurate, consistent, and interoperable across different platforms and institutions. This enables greater access to cultural heritage collections and promotes wider use and understanding of cultural objects.

In recent years, a large number of other standards have emerged, taking into account the specific features of different groups of cultural and historical artefacts. We will briefly mention the most used of them.

CDWA (GETTY, 2023): The Categories for the Description of Works of Art describes the content of art databases by articulating a conceptual framework for describing and accessing information about works of art, architecture, other material culture, groups and collections of works, and related images. The CDWA includes over 500 categories and subcategories. A small subset of categories is considered core in that they represent the minimum information necessary to identify and describe a work.

Cultural Objects Name Authority (GETTY, 2023) is a structured dictionary containing authoritative entries for cultural works, including architecture and movable works such as paintings, sculptures, prints, etc., whether existing, destroyed, or conceptual. CONA is linked to the other Getty dictionaries (GETTY, 2023), such as the Union List of Artist Names Online, TGN, AAT, and IA, to control values and provide more efficient access.

CIDOC CRM (GETTY, 2023): The International Committee for Documentation is a committee of the of the International Council of Museums. The CIDOC Conceptual Reference Model (CRM) provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.

DACS (GETTY, 2023): Describing Archives: A Content Standard is an output-neutral set of rules for describing archives.

Object ID (GETTY, 2023): Object ID is an international standard for describing cultural objects developed from a subset of the CDWA and in collaboration with the museum community, police and customs agencies, the art trade, insurance industry, and valuers of art and antiques.

CIMI (GETTY, 2023): The Consortium for the Computer Interchange of Museum Information attribute set is part of a Z39.50 Profile for cultural heritage information.

From the analysis of existing standards, we can conclude that they are mostly automated notations. The semantic descriptions of objects and the relations between them presented by them are partially or hardly processed with the help of a computer. For this reason, we try to create a formalism that lends itself to more complete computer processing.

3 Digitalization of Bulgarian Folklore

The digitization of folklore cultural heritage is an innovative process, the essence of which differs from the simple transfer, for example, of music or images on an electronic medium. Rather, the project we are proposing offers a form for structuring and modelling specialized knowledge from the field of Bulgarian folklore.

Digital technologies support the transformation of a wide audience of cultural institutions – tourists, library visitors, museums and galleries - into an actively practicing culture by building engaging empathy with cultural resources. In this context, one of the goals of the project is to examine some issues and find solutions in the field of "technology - education - cultural heritage" cooperation, to present basic technological aspects related to the expansion of people's opportunities to rediscover folklore wealth, interacting with modern technology. The main advantage of the sought-after formal presentation of knowledge specific to Bulgarian folklore is the possibility of their computer modelling and processing, including their semantics. With the help of such formalisms, the type objects (entities) important for the subject area, their properties (attributes) and the relations between them or the established connections between them can be determined.

On the one hand, over the years the development of formalisms (semantic models) for knowledge has led to a better reflection of the meaning of the data in the model,

greater ease of research and suitability of the model for using software tools, greater realism and activity in transferring material from the real world to the computer world. Different standards were created for cataloguing cultural monuments such as CCO (Cataloguing Cultural Objects).

On the other hand, one of the reasons and motivations for the emergence of artificial intelligence (AI) is the so-called common-sense knowledge. This is knowledge that, due to its nature (and its volume), is difficult to formalize, and there are no mathematical models for it so that it can be represented and processed by computer systems. Knowledge in many sciences (philosophy, history, literature, psychology, sociology, etc.) can be attributed to this type of knowledge.

One of the main goals of AI is to find suitable formalisms for representing such knowledge. At a certain period in the development of AI, it was assumed that it was possible to develop a universal formalism for representing knowledge with common meaning. It turned out that this task (at least for now) remains unsolvable (the formalization of the semantics is a problem). Considering the failures (the idea of a universal formalism for representing knowledge with a general meaning was completely abandoned), the thesis was adopted that it is more meaningful to look for different formal forms adequate for solving certain classes of tasks. Thus, over the years some formalisms were created, such as rules, frames, semantic networks, which proved suitable for many application cases. However, new forms adequate to represent knowledge in a general sense for use in specific knowledge-based systems continue to be constantly sought. For the digitalization of Bulgarian folklore, we use different approaches, one of them is the presentation of knowledge through rules and facts, and the other is through the development of ontologies.

Formalisms (models) simplify complex things by providing general and typical categorizations or actions. Our attention turns to the criteria for building the models. The considered standards present the ideal or typical cases. Many of the specifics in cataloguing practice are not represented in models. When looking at the modelling process, it can appear to be linear or unified. Many research methods are applied to extract value from modelling. In this article, we focus on our idea of a formal model describing the Bulgarian cultural-historical heritage.

4 Our Approach

The results of our analysis on various folklore objects and motifs suggest that these elements have a certain structure that can be formally represented. Identifying such a representation would greatly facilitate the creation of computational models for digitized folklore objects and their use. A formalism that uses principles and associated rules would not depend on users' interpretation or subjective interpretation for searching, identifying, and recognizing folklore objects. Formal models can interpret symbols and manage the structures composed of them in a valid way based on specific computation rules that define the relationships between the elements. These formal models can provide inference rules that infer new knowledge from existing knowledge.

Formal models are used in many intellectual activities and endeavors where precision and objectivity are required.

Generally, our structural approach can be represented by the following scheme:

- For each folklore theme (e.g., embroidery, costumes, ...) we specify a set of elements that we call "folklore primitives" - folklore primitives are atomic elements, i.e., we define them as an items that can only be identified by name, and its structure cannot be discerned.
- For each folklore theme we can specify folklore elements from higher levels. For example, the first-level items are those that are derived from combinations of folklore primitives. Specific combinations are obtained by specific "folklore relations". Second-level elements are structures formed by combinations of first-level "folklore elements" or first-level "folklore elements" and primitives.
- Higher level 'folklore elements' can also be specified on the same principle if necessary.

The formal models created in this way can be implemented as intelligent systems. We will demonstrate the applicability of our approach with two examples. The first example will be implemented using rules in the logic programming language Prolog (SWI-Prolog, 2023), and the second example will be implemented as an ontology (a specialized semantic network) in the development environment Protégé (Protégé, 2023).

4.1 Example "Bulgarian Embroidery"

The first example we have chosen is Bulgarian embroidery, which is an interesting folklore theme. As early as the primitive-communal system, embroidering clothes with symbolic images was widespread. These images were passed down from tattooed skin to clothing. Like language, embroidery uses visual images and ornaments to express aspects of our people's life and culture. Today, it is difficult to decipher the symbols from medieval times. In the age of technology, we have access to logical explanations. Once we understand a symbol, we can use ornamental sources to interpret its hidden meaning and understand its impact. The stitches could indicate a person's social or marital status, as well as which part of Bulgaria they came from.

Applied to the stitches, our approach follows the steps given here:

- Specification of the set of folklore primitives used in the various motifs of the embroideries $FP_{embroideries}$ = {triangle, …};
- First level folk elements $FE^{1}_{embroideries}$ = {makaz, rombus, wing, …};
- Second level folklore elements $FE^{2}_{embroideries}$ = {kanatica - family, pedigree, nation, nest, babitsa};
- Third-level folk elements – {World - upper and lower, …};
- Folklore elements of .... level;
- Folklore elements of .... level – {the tree of life, swastika, cross, whole paintings, …}.

We define relations with the elements mentioned above in the way indicated in Fig. 1 and Fig. 2 by obtaining more complex objects - "folklore elements from higher levels" - from "folklore primitives" and/or "folklore elements" from lower levels.

In Fig. 1a), a triangle (FP) embedded in a triangle (FP) results in FE^1 called "Makaz". By the relation of Fig. 1b) "over" or "under" executed over two triangles (FP) forms FE^1 - Rombus. Performing the relation "to the bottom right corner" of Fig. 1c) with two FP - triangles results in FE^1, called "Left wing". Similarly using the relation " to the lower left corner" - "Right wing".

When we set the FP-triangle in the "center common line" of FE^1 - "Left wing" we obtain FE^2 called "Nest" as shown in Fig.1d). "Nest" under reversed "Nest" = World (upper and lower).

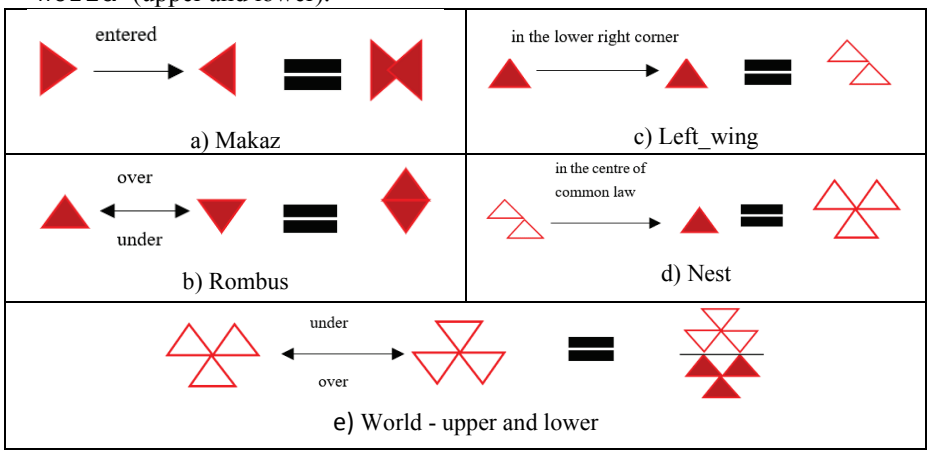

Fig. 1. Relations over Folklore Primitives and Folklore Elements from the first level.

Add a triangle (FP) to the left of the first level folklore element "Rhombus" and then add another triangle (FP) to the left of the resulting figure (FE^1) results in FE^2 - Babitsa (Fig.2a).

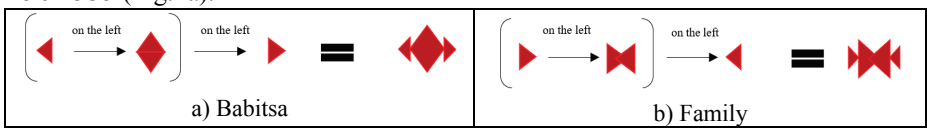

Fig. 2. Relations over Folklore Elements.

On the left side of the first level folklore element "Makaz" we add a triangle-shaped figure (FP) and then we add the resulting figure on the left side of another triangle (FP), the result is FE^2 - Kanatitsa, which symbolizes "Family" (Fig.2b). Similarly, performing the same relations over the resulting embroidery for Kanatitsa from Fig.2b), we obtain a third level folklore element (FE^3), which symbolizes "Genus". And performing the same relation over the element "Pedigree" results in a fourth level folklore element that symbolizes "Nation".

Computer Representation of "Bulgarian Embroidery", using Prolog Rules

Here we provide some of the Prolog programming code that we have used to describe folklore primitives and relations to generate folklore elements of different levels.

```
% observed basic figures in a particular embroidery
triangle(a). triangle(b).     % folklore primitives
triangle(c). triangle(d).     % folklore primitives
% relations over folklore primitives
entered(triangle(a), triangle(b)).
over(triangle(c), triangle(d)).
under(triangle(d), triangle(c)).
in_the_lower_right_corner(triangle(d), triangle(c)).
in_the_lower_left_corner(triangle(c), triangle(d)).
% specific folklore relations
center_of_common_law(left_wing,triangle(c)).
center_of_common_law(right_wing,triangle(d)).
% first level folk elements
makaz:- entered(triangle(X), triangle(Y)), X\=Y.
rombus:- over(triangle(s), triangle(d)).
rombus:- under(triangle(d), triangle(c)).
left_wing:-       in_the_lower_right_corner(triangle(d),
             triangle(c)).
right_wing:- in_the_lower_left_corner(triangle(c),
             triangle(d)).
% second level folklore elements
nest:-            center_of_common_law(X,Y),
          (X=left_wing;X=right_wing),
          Y=triangle(c).
has(embroideries, triangle(a)).
has(embroideries, rombus).
has(embroideries, triangle(b)).
fig(F):-setof(Y, setof(X,has(X,Y),S),F).
babitsa(F):-F = [triangle(b),rombus,triangle(a)].
```

4.2 Example "Bulgarian Folklore Costumes"

As a second example, we will examine the presentation of knowledge about Bulgarian folklore costumes. Typically, each Bulgarian traditional costume is defined by the types of clothing that make up the composition. Additionally, each costume has permanent parts that further characterize it. Below, we will describe the structure of a Bulgarian folklore costume.

Specification of the set of folklore primitives used in the different motifs of Bulgarian traditional costumes `FPcostume = {apron, coat, hat, pinafore, shirt, sock, trousers, …};`

- First level folk elements `FE1costume = {garment, accessories, shoes, …};`
- Second level folklore elements `FE2costume = {Two apron female costume, One apron female costume, Sayana costume,`

`Pinafore dress costume, White colored costume , Black colored costume};`

- Third-level folk elements – `{Female costume, Male costume};`
- Folklore elements of .... level.
- Folklore elements of .... level – `{folklore areas, …}.`

For each type of garment, accessory, and shoes, sub-types can be defined that are involved in the construction of a particular folk costume from different regions. Below, we will present examples of a black-clothed and a white-clothed male costume.

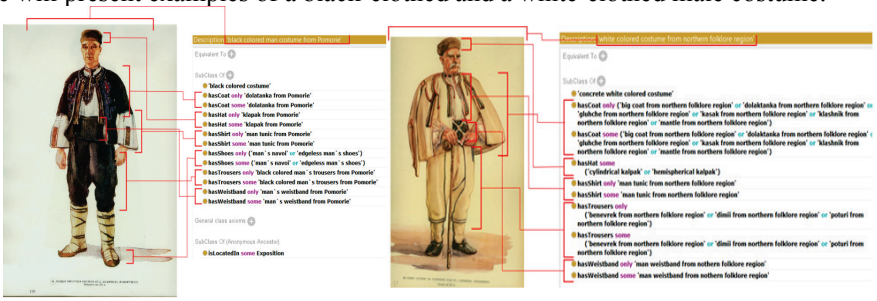

Fig. 3. White coloured man costume from Pomorie and Black coloured man custom from northern folklore region.

Figure 3 presents the main components of black and white coloured male costumes as parts of the Customs ontology (Miteva, Stoyanova-Doycheva, & Ivanova, 2017). The Bulgarian Folk Costumes ontology was created in compliance with the CCO standard. The black coloured male costume is from Pomorie and comprises basic male costume elements, including a `coat, hat, shirt, shoes, trousers, and a waistband.` Each element has a specific type, such as the Pomorie-style klapak hat and men's tunic shirt. Similarly, the white coloured male costume from a northern region has specific element types, such as the `dolaktanka coat` and `cylindrical hat`.

Each costume has specific elements from a particular region. For example, the shirt is a fundamental component of each costume, represented in different forms such as the male and female tunic, and `burchanka`. The aprons of women's costumes also differ and can be front or back. Additionally, they have types, such as the `brachnik` and `okrel`, which are back apron types characteristic of the tow-aproned female costume. Moreover, each shirt and apron has properties determined by decoration and embroidery, which can help identify the regions where the costumes originate from.

5 Conclusions

This article presents the first version of a model for the formal description of objects from Bulgarian folklore.

The potential for computer modelling and processing, including semantic analysis, of knowledge specific to Bulgarian folklore gives us many opportunities. With the help of this formalism, we define the important object types (entities) for the subject area,

their attributes, and the relationships or connections between them. Using a formalism for this knowledge improves the accuracy of data representation in the model and makes it easier for software tools to process. This results in greater realism when transferring material from the real world to the computer world.

Along with these extensions, practical experiments are being prepared for using the formalism we have given in the presentation of other objects from Bulgarian folklore.

Acknowledgements.

The research is supported by the project KP-06-M62/2 "Modeling of knowledge in the field of Bulgarian folklore" funded by the National Research Fund.

References

Baca, M., Harpring, P., Lanzi, E., McRae, L., & Whiteside, A. (2006). *Cataloging Cultural Objects. A Guide to Describing Cultural Works and Their Images. Chicago, United States of America: AMERICAN LIBRARY ASSOCIATION (ALA)*. Retrieved May 20, 2023, from https://vraweb.org/wp-content/uploads/2020/04/CatalogingCulturalObjectsFullv2.pdf

GETTY. (2023). *Getty Vocabularies*. Retrieved May 20, 2023, from The Getty Research Institute: https://www.getty.edu/research/tools/vocabularies/

Ivanova, T., & Nedeleva, I. (2021). Ontology of the Bulgarian culture heritage. Bulgarian folklore. *Proceedings of CBU in Social Sciences, 2*, 156-160. https://doi.org/10.12955/pss.v2.215

Madanska, S. (2022). An Ontology for Architectural Heritage: Historical Figures and Organizations. *Digital Presentation and Preservation of Cultural and Scientific Heritage, 12*, 121–130. https://doi.org/10.55630/dipp.2022.12.9

Miteva, M., Stoyanova-Doycheva, A., & Ivanova, V. (2017). Ontologiya na rodopskiya folklor [Ontology of Rhodope Folklore]. In *Education and Science - for Personal and Social Development : Proceedings of National Scientific Conference, Smolyan, 27-28 October, 2017* (pp. 49-54). http://fmi-plovdiv.org/GetResource?id=2874

Paneva-Marinova, D., Pavlov, R., & Rangochev, K. (2010). Digital Library for Bulgarian Traditional Culture and Folklore. In *Proceedings of the 3rd International Conference dedicated on Digital Heritage (EuroMed 2010)*, *8-13 November 2010, Lymassol, Cyprus* (pp. 167-172).

Protégé. (2023). Retrieved May 20, 2023, from https://protege.stanford.edu/

SWI-Prolog. (2023). Retrieved May 20, 2023, from https://www.swi-prolog.org/

Received: March 12, 2023
Reviewed: April 30, 2023
Finally Accepted: May 20, 2023