Greek Mythology as a Knowledge Graph: From Chaos to Zeus and Beyond

Juan-Antonio Pastor-Sánchez[a], Efstratios Kontopoulos[b,*], Tomás Saorín[a], Thomas Bebis[c] and Sándor Darányi[d]
[a]*University of Murcia, Spain*
[b]*Catalink Limited, Nicosia, Cyprus*
[c]*Athens University of Economics & Business, Greece*
[d]*School of Library and Information Science, University of Borås, Sweden*

Abstract. Greek mythology has been exerting a lasting influence on Western culture, but a respective ontology has been missing from the Semantic Web until now. To remedy this deficiency, from 5377 Wikidata items with 283 properties, 34 of these properties were selected to generate a first version of an Ontology of Greek Mythology (OGM). This limited set of properties was used to define a set of classes to instantiate the descriptions of the individuals according to reification requirements. The ontology also includes the representation of contradictions between statements, a well-known symptom of classical storytelling. A retrieval tool was added to use the Wikidata Query Service through SPARQL queries in order to display and download results in various formats, thereby developing OGM into a scholarly tool. Further, as Wikidata has little information about classical sources grounding the truth of its statements, we tested a semantic enrichment workflow to extract additional statement types from source texts in the 'Theoi Project' as statement anchors. This workflow experiment proved necessary to go beyond Wikipedia to address mythological complexities in a knowledge graph, but, as discussed in the article, its scalable automation requires further development.

Keywords: Greek mythology, ontology development, knowledge graph, Wikidata, scholarly research tool

1. Introduction

Given the lasting cultural impact of Greek mythology on Western culture, it would be hard to overestimate its importance for intangible cultural heritage, making this field of study a natural point of departure for research in *Digital Humanities* (*DH*). For technical reasons more than for lack of willingness, apart from many important results on several tracks over time, a certain state of fragmentedness may have characterized related efforts in DH, but there is reason to believe the tide is turning.

The Greeks had several cosmogonies, and, according to Hesiod, Chaos was the first to emerge from nothingness, without parents. The reference to her reign is on one level a metaphor for the obscure nature of knowledge about Greek mythology. This can be illustrated in texts recalling Zeus' birth and upbringing, with various facts contradicting one another from the very beginning, both with regard to the place of birth, place of concealment, and names of caregivers to the infant. However, in such a vast archive of stories about more than 5000 deities, demigods, heroes and mere mortals, the often meandering nature of genealogy, the events unfolding, and their circumstances is typical. On a next level though, as much as the reign of Chaos led ultimately to the rule of Zeus, we hope to show that building a *Knowledge Graph* (*KG*) in this realm is holding the promise of unprecedented clarity, and the full-scale handling of otherwise frightening complexities.

Our goal is to demonstrate how existing (structured and unstructured) resources can be integrated into a workflow, towards building a rich and evolving KG to act as a scholarly reference tool. Below we report on respective work in progress, which revolves around (a) building a KG anchored in an OWL ontology and SKOS vocabulary from Wikidata state-

*Corresponding author. E-mail: e.kontopoulos@catalink.eu

ments by means of advanced SPARQL queries; and (b) linking in a sample of texts available in English from a particular external source, the 'Theoi Project', on a proof-of-concept level. The statements in the thus completed KG express semantic content supported by reified knowledge of the respective sources to justify them, contradictory as they may be, representative of text variation in the source documents.

The structure of the rest of the article is as follows: Section 2 gives a brief overview of related research; Section 3 presents our multidisciplinary methodology applied to experiment design. Section 4 lists the results of different steps, followed by their evaluation in Section 5. Section 6 explains why our contribution is relevant, with Section 7 indicating potential research directions for the future.

2. Related research

We consider the research track we are introducing here as a blend of semantic technologies applied to folk narrative studies, prominently Greek mythology and religion in its archaic, classical, Hellenistic and folk incarnations. Starting with this latter component, our endeavour took inspiration from Martin P. Nilsson who argued for the abundance of folktale motifs in Minoan-Mycenaean religion and its successive layers, such as classical Greek religion [27], [28]; and from Walter Burkert, who suggested that specific activities called functions, i.e., units of plot action typical for Russian fairy tales about the hero's quest, appear much earlier, already in the Hittite *Illuyankas* myth as contrasted with Apollodorus' rendering of the story of Typhon [3]. Further parallels between the respective sky gods, Zeus and Kumarbi can be drawn, but apart from these, his book is a plethora of Greek myths worth reconsidering for formal approaches. Edmund Leach also proposed the applicability of structural analysis to certain cycles of Greek myths [19], an encouragement leading to our own earlier work in this area [8]. Central to the current effort however was Lévi-Strauss' tenet that, for the structural analysis of myth, all of its variants must be taken into consideration [20], a challenge for traditional comparative methodologies but easy to adopt for ontology building.

For a background to the above, Stith Thompson's monumental *Motif-Index of Folk Literature* [35], a six-volume catalogue, was one of the most important works in folklore studies produced during the twentieth century. Some may have viewed it merely as a taxonomic endeavour, but as Thompson stated, "Before it can become an object of serious and well-considered study, every branch of knowledge needs to be classified. There was a time when geology and botany consisted of random collections of facts and hastily constructed theories. It was only when this anecdotal stage gave way to systematic classification that real progress was made toward a thorough method of study" [36].

As to the former ingredient, semantic technologies have been available for Proppian structural analysis applied to folktales and narrative research [30], including their conversion to an ontology [9]-[11]. However, with a single exception, neither structural nor other approaches to the study of Greek myths resulted in an ontology to the best of our knowledge, so this field falls short of experimentation with semantic reasoning. The aforementioned exception was a first initial study [34], which we are going to extend below through combination with a rich source of structured raw material in Wikidata [37]. Related in scope is [4], where network theory is used to show the directed scale-free nature of Greek and Roman mythology, while [26] used topology to demonstrate that Homer's *Odyssey* as a social network is of a highly clustered small-world nature. The authors of [17], in their interesting effort to apply network theory to different classical mythologies, argued that networks of characters can be the basis of classification and comparison, and applied statistical physics viewpoints and techniques on an experimental basis, limited to three epic texts.

Critical for the success of such research directions are databases and/or datasets with source texts to support facts and underpin argumentation. Here, development with regard to other disciplines has been slow: folklore databases of different narrative genres exist (e.g., [15], [18], [24]), but datasets for machine learning do not. The situation is somewhat similar when it comes to text collections about Greek and Roman mythology. The Perseus Digital Library at Tufts University [33] has many of the relevant texts partly in their original languages, partly in English, whereas the remaining classical texts are being collected up by, e.g., the Open Greek and Latin Project (OGL – http://bit.ly/3dClznR) at the University of Leipzig, and the First Thousand Years of Greek project (FF1KG – http://bit.ly/3bxUvDC) as one of its pilots. English translations of many classical myths are also available, e.g., at the 'Theoi Project' website (www.theoi.com). Thereby, material to ground classification in philology is in principle there, although, for our particular case, this requires a kind of reverse engineering, proceeding from ontology to its sources. Further, for workflow development, natural language processing (NLP) tools for modern and classical languages are increasingly available, see, e.g., the Classical Language Toolkit (CLTK – `http://bit.ly/2ZSXp00`), and so are text analytics tools from data science, while the same holds for standard semantic reasoning tools for the querying and visualization of knowledge graphs.

3. Methodology

3.1. Creating an ontology of Greek mythology from Wikidata

Wikidata is a collaboratively edited knowledge graph developed by the Wikimedia foundation. There are different editions of Wikipedia for each language, but there is only one Wikidata edition. Therefore, all pages from different editions of Wikipedia about the same concept correspond to a Wikidata item. In Wikidata, such items are described by structured data statements. Each statement includes a property and a value. Statements with values that refer to other items establish relationships between the item described, and the item the property refers to. Statements may also include references to sources of information in order to verify their truth. It is also possible to add qualifiers to describe detailed aspects of statements. Both references and qualifiers work as statements' reifications.

Wikidata items are identified by a "`Q`" code. Item "`Zeus`" corresponds to code `Q34201` and to URL `http://www.wikidata.org/entity/Q34201`. On the other hand, Wikidata properties are identified with a "`P`" code. Property "`father`" corresponds to code `P22` whose URL is `https://www.wikidata.org/prop/P22`.

Wikidata uses its own data model that is mapped to an RDF representation. In this way it is possible to construct queries to retrieve data from the knowledge graph using a SPARQL query service (WDQS) [22].

An important aspect of Wikidata is its knowledge organization dynamics. The data model offers some properties to this end. Property `P279 (subclass of)` allows relating two items, defining one as a superclass and the other as its subclass. The `P31 (instance of)` property allows one item to be related as an instance of another one that has been defined as a class. The property `P361 (part of)` establishes a part-whole relationship between two items.

One of the key weaknesses of Wikidata is the curation of its data. Large amounts of valuable data can be found together with other ones that are inconsistent, incomplete, or incorrect. Furthermore, statements with references to information sources are scant in relation to the size of the Wikidata knowledge graph. These problems do not affect all knowledge domains equally and make it necessary to design specific strategies for Wikidata-based ontology development [2], [16].

The representation of data extracted from the Wikidata knowledge graph through an ontology would necessitate a greater degree of formalization and adaptation to more specific applications and contexts. Defining such an ontology for Greek mythology from Wikidata requires exploring its knowledge graph. This allows one to have an overview of how data are organized and the problems that can emerge from their retrieval. As a starting point to develop our model, the work in [34] was analysed, where classes and properties were proposed that enable the creation of a simple ontology to fundamentally represent kinship relations and organize individuals in a basic taxonomy.

One of the essential objectives of this work is that statements should have references to sources of information. This could be accomplished through an RDF reification process, a complex mechanism though with highly verbose results difficult to visualize. Therefore, in addition to representing the information obtained from Wikidata, our ontology should also allow for enriching statements with references in a simple way.

It is also convenient to consider the use of controlled vocabularies for concepts, characters, places and other types of individuals who may contribute to the description of the Wikidata items. The use of SKOS vocabularies permits these elements to be separated in a way that facilitates their reuse and reduces modifications in the ontology schema.

3.1.1. Analysing the taxonomy of classes

The first step towards developing an ontology from Wikidata is the analysis of items related to the knowledge domain. For this, it was necessary to examine the taxonomy of classes from which individuals are instantiated. As a working hypothesis, "`Greek mythology (Q34726)`" was chosen as the item that should be the upper class of the taxonomy of classes.

The taxonomy of classes is composed of all the items between which there is a relationship defined by means of the "`Subclass Of (P279)`" property. However, some Wikidata editors do not correctly distinguish between the concepts of class, instance and part. For this reason, it is possible to find class-subclass relationships using the "`Instance of (P31)`" and "`Part of (P361)`" properties to this end. Considering all of the above, the following SPARQL query was used to retrieve all the items related to `Q34726` from WDQS, and also show the classes of which they are instances:

```
SELECT ?subject ?subjectLabel ?proper-
ty ?class ?classLabel ?superclass ?su-
perclassLabel
WHERE {
 ?subject ?property wd:Q34726 ;
      wdt:P31 ?class .
 ?class wdt:P279|wdt:P31|wdt:P361 ?su
      perclass .
 SERVICE wikibase:label {
      bd:serviceParam wikibase:language
      "en","de","it","el","es","fr" .
 }
}
ORDER BY ?classLabel ?superclassLabel
```

The analysis of the query results showed that item `Q34726` is not used in Wikidata to organize items about Greek mythology. At the same time, two items located at the beginning of the class taxonomy were identified: "`mythological Greek character (Q22988604)`" and "`group of Greek mythical characters (Q28061975)`".

Wikidata includes in its data model the number of Wikimedia sites (sitelinks) in which items have equivalences. It should be considered that some items, classes when referring to taxonomy, do not have any sitelink because they are created natively in Wikidata. At the same time some of these classes have been created unnecessarily, since they have no instance items. These classes are not relevant to the ontology, so the following SPARQL query was formulated to filter them out. The query also includes the total number of sitelinks for each class, and the total number of instance items.

```
SELECT ?num-
ber_of_sitelinks ?class ?classLabel
(lang(?classLabel) AS ?langClass)
(count(?item) as ?number_of_items) ?su-
perclass ?superclassLabel
WHERE {
 {?class wdt:P279* wd:Q22988604 .} UN
      ION {?class wdt:P279*
      wd:Q28061975 .}
 OPTIONAL{?class
      wdt:P279|wdt:P31|wdt:P361 ?super-
class}
 OPTIONAL{?item
      wdt:P31|wdt:P361 ?class .}
 ?class wikibase:sitelinks ?num
      ber_of_sitelinks .
 FILTER(BOUND(?item) || (!BOUND(?item)
      && ?number_of_sitelinks>0))
 SERVICE wikibase:label {
      bd:serviceParam wikibase:language
      "en","de","it","el","es","fr" .
 }
}
GROUP BY ?num-
 ber_of_sitelinks ?class ?classLabel ?l
 angClass ?superclass ?superclassLabel
ORDER BY desc (?number_of_sitelinks)
```

The results were revised to exclude those superclasses that are not relevant to this research.

3.1.2. Constructing a SKOS vocabulary for the representation of the taxonomy of classes

Next, a SKOS vocabulary was built where classes were transformed into hierarchical categories. The reason underpinning this decision is the low level of logical formalization of the Wikidata class and subclass organization mechanism. On the other hand, it was also necessary to consider that Wikidata classes are really items and, therefore, may be subject to substantial changes in the future. In this way, a modular design was preferred in which the ontology schema would be subject to the fewest possible changes, while the data sets of the statements and SKOS vocabularies could be updated more frequently. Details about the implementation of the SKOS taxonomy are given in Subsection 4.1.

3.1.3. Analysing the properties

To develop an ontology, it is necessary to define a set of properties appropriate to the domain of knowledge to be represented. In the case of this research, object properties were considered for representing the relationships between Wikidata items. The properties used in the instances of the classes of the taxonomy were examined. For this purpose, the following SPARQL query was used:

```
SELECT ?p ?pLabel
WHERE {
 {?class wdt:P279* wd:Q22988604 .} UN
      ION {?class wdt:P279*
      wd:Q28061975 .}
 ?item
      wdt:P31|wdt:P361 ?class ; ?prop ?
      object .
```

```
 FILTER(contains(str(?prop),"/direct/
      "))
 SERVICE wikibase:label {
      bd:serviceParam wikibase:language
      "en","de","it","el","es","fr" .
 }
 ?p wikibase:directClaim ?prop .
}
GROUP BY ?p ?pLabel
ORDER BY ?p
```

3.1.4. Defining object properties for the representation of relationships

Not all properties retrieved in the previous step were relevant. It was thus necessary to make a selection of those that fit the scope of our design. For this reason, those relationships proposed by [34] were reconsidered. The results of property analysis also offered other relevant attributes that were selected for inclusion.

3.1.5. Ontology design

The ontology had to meet a series of requirements which were not only related to the representation of the knowledge domain but also relevant to the functionality of referencing information sources of the different statements. Both OWL and RDF natively provide a mechanism for defining reification properties, which, however, can make maintaining and querying datasets very complex.

Another essential aspect, mentioned above, was the decision to represent the classes as SKOS categories. This same decision was made to represent the items, concepts, and sources used in the statements.

The implementation of the ontology is reported in Subsection 4.1.

3.1.6. Query tool design.

SPARQL has been used extensively to analyse the domain of knowledge about Greek mythology in Wikidata. However, it was necessary to have a mechanism to retrieve data obtained from Wikidata as well for their viewing and downloading, towards populating the ontology, and formulating a KG. Therefore, data had to be represented according to the scheme of the designed ontology. Finally, it was considered of interest to have some additional function that allows comparing the results obtained from different items. For this purpose, a simple web application was developed, presented in Subsection 4.2.

3.2. Semantic enrichment of the KG

We then shifted our attention to investigating how the Wikidata-derived KG presented above could be enriched by additional information from an external source, towards eventually developing a scholarly reference tool. More specifically, we wished to extract statements for the ontology from that particular website accompanied by their respective references to the original documents, to add them to the Wikidata-based KG, and to establish interrelations such as contradiction and complementarity. An example of a typical target statement in natural language would be: "According to Pseudo-Apollodorus (Bibliotheca 1. 4 – 5 - trans. Aldrich), Zeus was born on Mount Dikte (Dicte) in the island of Crete".

3.2.1. Data source

Created by Aaron J. Atsma from Auckland, New Zealand, the 'Theoi Project' is a reference website dedicated to Greek mythology and religion and its representations in classical literature and ancient Greek art. It contains rich information about gods, spirits, heroes, and fantastic creatures with more than 1,500 pages and 1,200 images.

3.2.2. Semantic enrichment workflow

In order to extract the aforementioned statements from 'Theoi' and perform semantic enrichment of the available KG, we adopted the following workflow:

- Content extraction from 'Theoi';
- NLP of the extracted content, which included: (a) substitution of synonyms by concept-like umbrella terms; (b) coreference resolution to find all expressions that refer to the same entity; (c) relation extraction to identify semantic relationships between two or more entities;
- Conversion of extracted relations into reified RDF statements;
- Addition of the statements into the KG;
- Establishment of relationships (e.g., contradictions) between the added statements.

While attempting to implement the above workflow, we came across several challenges. The implementation process, as well as the problems we experienced, are reported in Subsection 4.3.

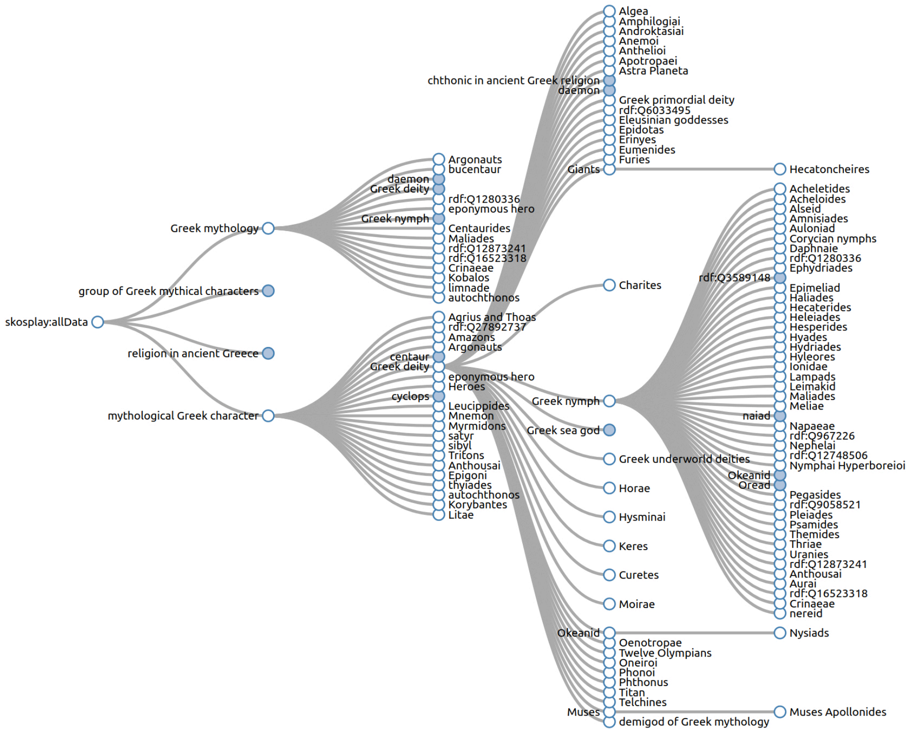

Fig. 1. SKOS Vocabulary Taxonomy. Visualization obtained with SKOS Play! (https://skos-play.sparna.fr).

4. Results

On a general level, statistics about the number of statements etc. are discussed in Subsection 4.1 and establish the current constraints to searching the ontology. However, due to semantic enrichment, we also identified interrelations not dealt with in Wikidata but key to correctly capturing domain-specific knowledge.

4.1. OGM: Ontology of Greek Mythology

The development of the steps indicated in the methodology made it possible to retrieve 132 Wikidata items identified as classes. Among these items, 239 relations were identified. Both the items and their relations were represented as a SKOS Vocabulary (see Fig. 1). The items were defined as SKOS concepts (`skos:Concept`). The class-subclass relationships were represented as SKOS hierarchical relationships (`skos:broader`/ `skos:narrower`).

For the analysis of properties, 5377 items were retrieved defined as instances or parts of the recovered classes. In the item descriptions, 283 different properties were identified. Based on these properties, and considering the scope of the investigation, 34 properties were identified. Some of these properties were based on those identified in [34]. The properties were used to define a set of classes to instantiate the descriptions of the individuals. This was done according to the reification requirements of the ontology. The ontology also includes a set of classes (e.g., “Location”, “Gender”, “Event”, “DeityDomain”, etc.) to define as individuals those items used in the description of properties. Fig. 2 shows the class taxonomy of OGM.

The `Statement` subclasses include the necessary integrity constraints to check the consistency of OGM-based datasets.

The ontology also includes the properties needed to link the subject and object of a statement. OGM possesses the necessary properties to link sources to statements, as well as to define contradictions between them. A specific data property has also been defined to link to the ‘Theoi Project’ pages since some Wikidata items include this property. Fig. 3 shows the defined properties.

The OGM ontology is publicly available at http://purl.org/umu/ogm and the link for visualizing it via WebVOWL [21] is available at http://bit.ly/38pvUjB.

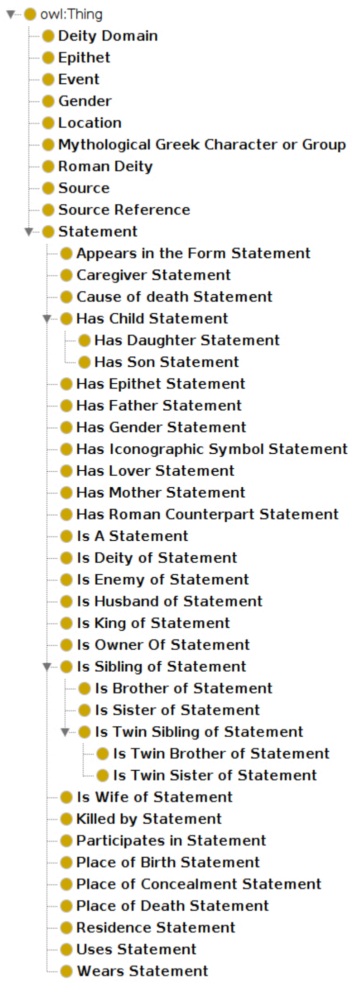

Fig. 2. OGM class taxonomy.

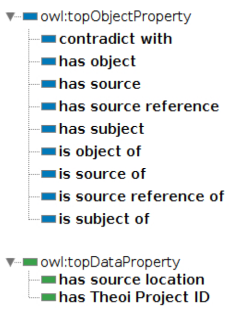

Fig. 3. OGM properties.

4.2. Wikidata Ontologizer

To represent the data obtained from Wikidata according to the OGM model, it is necessary to carry out an adaptation process. This can be done automatically using a simple web application called "*Ontologizer*". The tool has been developed to use WDQS through SPARQL queries and process the results to display and download in various formats.

Ontologizer allows for entering one or more Wikidata items in a web form and displaying the statements adapted to OGM. It is also possible to retrieve unique (different) vs. shared (similar) statements by at least two specified items. Data is retrieved and transformed through a SPARQL query, and the displayed tables can be downloaded in CSV format.

The RDF representation of the complete dataset can be downloaded in TTL format. This representation includes:

1. The taxonomy of categories represented using SKOS, with preferred labels (`skos:prefLabel`), alternative labels (`skos:altLabel`), descriptions represented as definitions (`skos:definition`) and hierarchical relationships (`skos:broader` and `skos:narrower`). The categories have been defined as SKOS concepts (`skos:Concept`) and as instances of the OGM class `Mythological Greek Character or Group`.
2. Wikidata statements represented as instances of the corresponding subclasses of OGM `Statement`. Each OGM statement is related to its corresponding subject and object items using the corresponding properties (`ogm:hasSubject` and `ogm:hasObject`).
3. The items of the statements represented as SKOS concepts and using the preferred labels, alternative labels and descriptions. The items have also been defined as instances of the appropriate OGM classes according to the constraints defined in the ontology.

To understand how the Ontologizer generates the RDF representation, an example with the following Wikidata statement will be typical:

```
Zeus (Q34201) / official residence
(P263) / Mount Olympus (Q80344)
```

Enter item(s) separated with spaces: wd:Q35500 Select filter: None (get all statements) Ontologize

Downloads: Tables (csv format) - RDF dataset (ttl format) - Only SKOS Taxonomy (ttl format)

has Subject	Subject Label	Statement Class	has Object	Object Label	Object Sitelinks
Q35500	Aphrodite	has gender	Q6581072	female	4
Q35500	Aphrodite	has father	Q79999	Uranus	81
Q35500	Aphrodite	has father	Q34201	Zeus	130
Q35500	Aphrodite	has father	Q44204	Cronus	86
Q35500	Aphrodite	has mother	Q199923	Dione	43
Q35500	Aphrodite	has mother	Q1374029	Evonyme	2
Q35500	Aphrodite	is wife of	Q44384	Hephaestus	85
Q35500	Aphrodite	is a	Q205985	goddess	70
Q35500	Aphrodite	is a	Q22989102	Greek deity	0
Q35500	Aphrodite	is a	Q23015914	fertility deity	4
Q35500	Aphrodite	has child	Q184353	Charites	50
Q35500	Aphrodite	has daughter	Q213440	Tyche	44
Q35500	Aphrodite	has daughter	Q611171	Peitho	27
Q35500	Aphrodite	has daughter	Q641286	Rhodos	26

Fig. 4. Example of contradictory statements in OGM about Aphrodite's parentage.

It is represented according to OGM as a declaration of the OGM class `Residence Statement`, defined in the OWL ontology as follows:

```
ogm:ResidenceStatement rdf:type
owl:Class ;
rdfs:subClassOf [ owl:intersectionOf
( ogm:Statement
[ rdf:type owl:Restriction ;
 owl:onProperty ogm:hasObject ;
 owl:someValuesFrom ogm:Location ]
[ rdf:type owl:Restriction ;
owl:onProperty ogm:hasSubject ;
owl:someValuesFrom ogm:MythCharOrGroup
] ) ; rdf:type owl:Class ] ;
rdfs:label "Residence Statement"@en .
```

The class definition establishes that the subject items of the `ogm:ResidenceStatement` class declaration must be instances of the OGM class `Mythological Greek Character or Group`, while the object items must be of the OGM class `Location`.Following this restriction, the previous example would be represented (abbreviated) as follows:

```
<#Q34201-residence-Q80344> rdf:type
ogm:ResidenceStatement ;
  ogm:hasSubject ogm-data:Q34201 ;
  ogm:hasObject ogm-data:Q80344 .

<#Q34201> rdf:type skos:Concept ;
  rdf:type ogm:MythCharOrGroup ;
  skos:prefLabel "Zeus"@en .
```

```
<#Q80344> rdf:type skos:Concept ;
  rdf:type ogm:Location ;
  skos:prefLabel "Mount Olympus"@en .
```

Some subclasses of `ogm:Statement` should have a single occurrence with a subject. This is the case for classes representing father, mother, place of birth or place of death. Ontologizer indicates as contradictions those statements of any of these classes with the same subject. This indication is represented in the visual interface by a color indication and in the RDF dataset by the `ogm:contradictWith` property. Fig. 4 shows in orange contradicting statements in OGM concerning Aphrodite, one of the Olympian deities.

Our Ontologizer, which is available at https://skos.um.es/gmwdo, also offers the ability to download the SKOS taxonomy separately.

4.3. Deploying NLP for semantically enriching the KG

The practical implementation of the semantic enrichment of the KG adhered to the workflow presented in Subsection 3.1, with its steps described below.

4.3.1. Content extraction

The first step of the workflow involved content extracted from 'Theoi'. Despite the website's popularity among scholars and researchers with an interest in

Greek mythology, we came across critical challenges when attempting to deploy automated processes for extracting some of its content. Those challenges were mainly due to the website's inconsistent structure and the variable ways information is distributed among the relevant webpages. As a consequence, considerable effort was required for manually curating the results of the extraction process.

For the purposes of this work, we extracted 40 texts relevant to Zeus's birth from 'Theoi', along with their respective source references.

4.3.2. Natural Language Processing

A major challenge was due to the source texts themselves, which are English translations of classical authors. More specifically, we faced the following obstacles: (a) subject, predicate and object occurred as different sets of lexemes (e.g., names, synsets); (b) the often floral, poetic language used by the authors was highly NLP-unfriendly; (c) anaphora and coreference resolution did not perform adequately.

To cite an example, stories about Zeus' birth from Rhea and Kronos share the recurrent elements of being born somewhere, being handed over after birth to be hidden at different locations and being nurtured by caregiving guardians whose persons can be regionally different. Such partly conflicting, partly overlapping information characterizes the plot until the day when the future king of the gods overthrew his father. To ground our argument for semantic enrichment, having departed from an 834x40 term-document matrix, we manually created a 126x40 synset-document matrix with six conceptual categories as umbrella terms (Zeus synonyms, birth synset, place of birth, concealment synset, caregiver synonyms, place of concealment). This in turn was summed up by a 6x40 category-document matrix for ground truth and a means of semantic anchoring of ontology statements in philology. The observations below relate to the partial success of this enrichment process.

To face the first challenge discussed above, we substituted word forms, i.e., lexemes and synonyms based on a manually curated list of synsets from WordNet [12], informed by Indo-European semantic fields (http://bit.ly/37NhQjA) and their concept-like umbrella terms. For instance, "king of the gods" or "Kronides" would be substituted by the umbrella term "Zeus".

Regarding the remaining two challenges that affected the rest of the NLP process, we examined various tools towards achieving our goals. For extracting statements from the texts, we initially considered PIKES [7] and FRED [14], both of which extract RDF triples from free text. Unfortunately, the nature of the source text in 'Theoi' did not allow those tools to perform in a satisfactory manner. We thus resorted to investigating "pure" NLP-focused tools, specifically Stanford's CoreNLP [23] and expert.ai's NL API (http://bit.ly/2ZROZq2). After some experimentation, we concluded that CoreNLP's neural coreference system [5], [6] was very well suited to the task, and expert.ai's NL API worked in a satisfactory manner for relation extraction.

For instance, coreference resolution on the following excerpt from Pseudo-Apollodorus (Bibliotheca 1. 4-5, trans. Aldrich) regarding Zeus's birth:

"Rhea, when she was heavy with Zeus, went off to Crete and gave birth to him there in a cave on Mount Dikte."

deduced that "she" refers to "Rhea" and "him" refers to "Zeus". After substituting "she" with "Rhea" and "him" with "Zeus", the resulting excerpt became:

"Rhea, when Rhea was heavy with Zeus, went off to Crete and gave birth to Zeus there in a cave on Mount Dikte."

Submitting the above sentence to expert.ai generated the set of relations illustrated in Fig. 5.

Fig. 5. Set of relations generated by expert.ai.

4.3.3. Conversion of extracted relations and addition to the KG

The extracted relations, like those presented in Fig. 5, were then converted into reified RDF statements adhering to the reification scheme proposed in [29], and complying with the OGM ontology presented in

Subsection 4.1. Fig. 6 illustrates an example of two reified statements concerning the birthplace of Zeus; according to two different classical authors, he was born in two different locations, Arcadia and Crete.

The diagram is based on the Graffoo notation [12], where the circles indicate instances, while their captions are written in the form of "`instance_XYZ::Class_ABC`", specifying the name and the type (class) of each instance correspondingly.

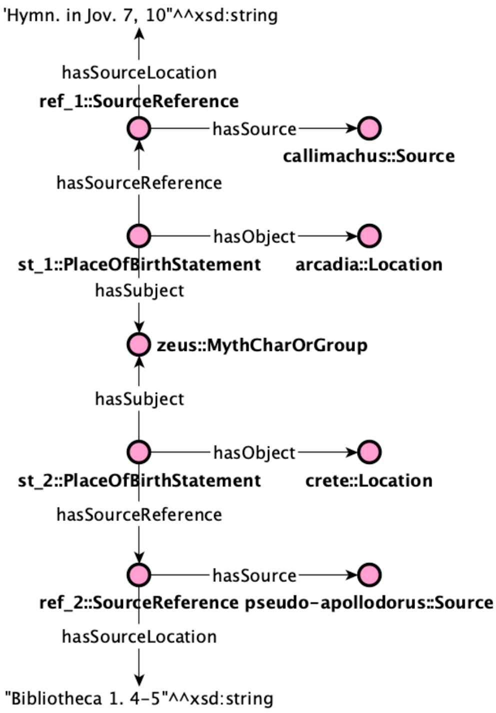

Fig. 6. Reification of two sample statements.

4.3.4. Inferring relationships between statements

A key element in such a rich variety of texts is the existence of underlying contradictory statements. For instance, in Fig. 6, the obvious contradiction is that nobody (not even an Olympian god) can have two birthplaces. In order to infer the existence of such contradictions, we conducted named entity resolution from Wikidata, enriched the KG with new information derived from Wikidata, and ran a set of SPARQL queries on top of the KG that determined whether contradictions exist.

Fig. 7 displays the update of the previous figure after running the named entity resolution and respective SPARQL query. Arcadia and Crete are now instances of `wd:Q82794 (geographic region)` and the contradiction is inferred after comparing their lat/long coordinates from Wikidata (property `P625 (coordinate location)`).

4.3.5. Addition of new statement types

One of the typical findings in this proof-of-concept stage was that Greek myths contain many facts on an episodic level, not covered by Wikidata. An example thereof was that during the process described in the previous sub-subsections, we came across two additional types of statements relevant to the birth of Zeus, which we manually added to OGM:

1. `ogm:CaregiverStatement`, which indicates the caregivers of a god during his/her upbringing;
2. `ogm:PlaceOfConcealmentStatement`, which represents their hiding place.

Over time, as more episodes will be covered in narratives about the twelve Olympians, we expect more structural peculiarities of the above kind to enrich OGM.

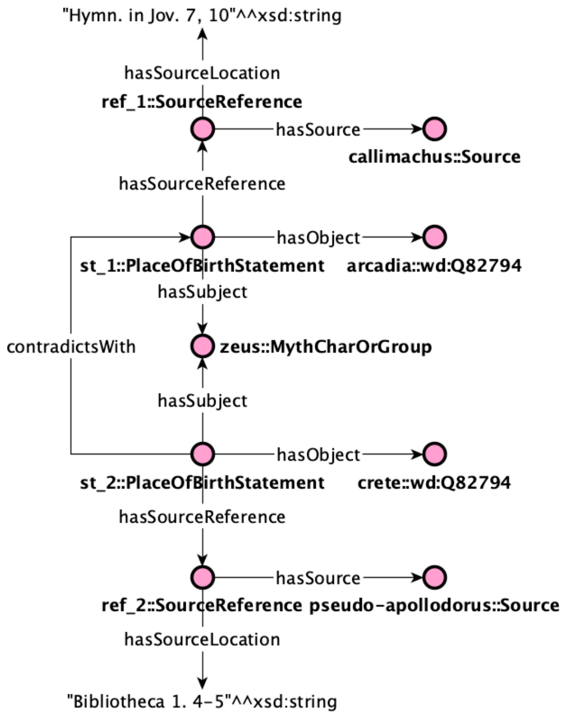

Fig. 7. Example of contradictory statements.

5. Evaluation of OGM

The design of OGM followed the guidelines in [32]. After an initial design, a first evaluation of the ontology was carried out using the OOPS service

(OntOlogy Pitfall Scanner!), which can analyse an ontology and provide a series of indicators about the different pitfalls that can be detected [31]. OOPS! identified some minor class overlap issues. Once these problems were fixed and the OGM design was finalized, OOPS! did not detect any other pitfalls.

To measure OGM's adequacy for representing the corpus used for its design (Wikidata), a dataset was created from the statements corresponding to the items of the twelve Olympian gods in Wikidata using the Ontologizer (see Subsection 4.2). The data extraction process identified a total of 888 Wikidata items that were represented as individual instances of the OGM classes (see Table 1).

Table 1

Wikidata items as individuals of OGM classes

OGM classes	*Frequency*
MythCharOrGroup	845
DeityDomain	23
Epithet	2
Event	1
Gender	2
Location	2

Furthermore, all items were also represented as SKOS concepts. The dataset also included the 136 categories used in Subsections 3.1.1 and 3.1.2. These categories were essential for defining the statements of the `ogm:IsAStatement` class that connects the items with the category to which they belong.

Other categories retrieved from Wikidata (using properties `P31`, `P361` and `P279`) that are not in the SKOS taxonomy vocabulary have also been included. These categories are not directly related to the domain of Greek mythology, but it is convenient to incorporate them to ensure the completeness of the dataset and to assess whether OGM's current design is adequate for the representation of the knowledge domain. Table 2 shows the number of `ogm:IsAStatement` class statements. The first two ("`Q101609 Twelve Olympians`" and "`Q22989102 Greek Deity`") are included in the SKOS taxonomy and the rest were obtained after retrieving the statements of the Wikidata items for the twelve Olympian gods.

Table 2

Use of Wikidata categories in `ogm:IsAStatement` class statements

Wikidata categories	*Frequency*
Q101609 (Twelve Olympians)	12
Q22989102 (Greek Deity)	11
Q1589492 (King of the Gods) *	1
Q842697 (Thunder God) *	1
Q205985 (Goddess) *	3
Q23015914 (Fertility Deity) *	2
Q511056 (Solar Deity) *	1
Q24284226 (Agricultural deity) *	1
Q95074 (Fictional character) *	1
Q41863069 (War deity) *	1
Q1916821 (Water deity) *	1
Q13405593 (Nature deity) *	1

(*) Not included in SKOS Taxonomy

It is obvious that, as the number of items belonging to the domain of knowledge about Greek Mythology analysed by Ontologizer increases, the number of Wikidata categories of this domain present in the statements of the `ogm:IsAStatement` class will also increase. The schema proposed by OGM represents the Wikidata properties as subclasses of `ogm:Statement`. The Ontologizer does not just do a transpose, it also breaks down some of the Wikidata properties into more specific subclasses. Table 3 shows how Wikidata properties `P26 (spouse)`, `P40 (child)` and `P3373 (sibling)` are represented as classes differentiated by gender. This is possible because property `P21 (gender)` of the subject and object items has been examined to determine the specific subclass. At the same time, Wikidata property groupings were defined that grouped together Wikidata properties and represented them as a subclass of `ogm:Statement`. An example of this can be seen with Wikidata properties `P607 (conflict)` and `P793 (significant event)` that have been represented via the subclass `ogm:ParticipatesInStatement`.

Table 3

Frequency of individuals for any subclass of `ogm:Statement` and equivalence with Wikidata properties

Classes statement	*Freq.*	*Wikidata prop.*
appears in the form	11	P4675
has child	20	P40
has daughter	123	
has son	394	
has father	14	P22
has mother	15	P25
has gender	12	P21
has iconographic symbol	14	P4185
has lover	250	P451
has roman counterpart	13	P460
is a	36	P31 P361 P279
is brother of	18	P3373
is sister of	13	
is deity of	24	P2925
is enemy of	1	P7047
is husband of	15	P26
is wife of	2	
is owner of	2	P1830
participates in	1	P793 P607

place of birth	1	P19
residence	6	P263 P551
uses	1	P2283
wears	1	P3828
Web link to THEOI *	12	P3545

(*) Defined as data property in OGM

It is necessary to consider that OGM defines subclasses of `ogm:Statement` that are not applicable to the case of the Twelve Olympians but to other Wikidata items in the domain of knowledge about Greek mythology. This is the case for classes such as `ogm:IsKingOfStatement`, `ogm:PlaceOfDeathStatement`, etc.

6. Discussion

We consider this first version of the proposed OGM as proof that our ontology of Greek mythology, based on Wikidata, and potentially enrichable by external sources, is a viable way to proceed with future updates. In our view, it is thereby a significant contribution to the Semantic Web in its Cultural Heritage segment.

There were two major lessons we learnt during the process:

- A key problem was the reuse of Wikidata as an exclusive source, with high- and low-quality data provided at the same time without warning or distinction. Also, the sources of its statements in our domain were ill-documented if at all, either being very few, or unreliable and possibly incorrect;
- Formalizing the knowledge extracted from Wikidata that is completed with data extracted (and referenced) from other sources can be a mechanism to improve the quality of its curation.

Besides the above overall points, more concrete conclusions are described in the following subsections.

6.1. The important role of the SKOS taxonomy

It is obvious from the figures in Table 1 that class `ogm:MythCharOrGroup` displays an excessive concentration of individuals. This is due to the fact that the Wikidata items in the Greek mythology domain analysed by the Ontologizer suffer from numerous curation problems and logical organization errors. This lack of logical formalization prohibited them from being represented as subclasses of `ogm:MythCharOrGroup`, and were instead defined as individuals of that class. This lack of specificity implies loss of knowledge after the semantic representation. For this reason, the relationship between individuals of this class and the Wikidata items that play the role of classes/subclasses has been preserved. This is the goal of the `ogm:IsAStatement` class and the SKOS taxonomy. This solution avoids the loss of information that would involve directly discarding these items from Wikidata.

Moreover, it would be convenient to analyse the SKOS taxonomy and define a more refined taxonomy of subclasses for `ogm:MythCharOrGroup`. Each of the concepts in the SKOS taxonomy would be mapped to one of the new subclasses. In this way, the items analysed by the Ontologizer could be automatically defined as individuals of one of these subclasses.

6.2. The role of NLP in semantic enrichment

As already discussed, the application of NLP in English translations of classical texts suffers deeply and, in our view, no single tool currently seems capable of handling all the inherent peculiarities. The solution proposed in this work revolves around creating a workflow of NLP tools for performing the required operations (see Subsection 4.3), while further tools (like, e.g., DBpedia Spotlight [25] or LODifier [1]) and/or custom solutions could be added to the workflow in order to augment its functionality.

Overall, NLP-based semantic enrichment of an existing KG plays a critical role, because it results in potentially identifying new relationships that do not even exist as Wikidata properties. In the case of OGM, this would lead to expanding the number of subclasses of `ogm:Statement`, and possibly defining appropriate constraints on these subclasses.

7. Future research

Our future work will focus next on defining mechanisms to validate statements obtained from Wikidata, and incorporate others obtained by means of NLP from external sources. Parallel to this, we will experiment with the semantic enrichment of existing ontology statements (statement grounding/anchoring).

As the coverage of Greek mythology on an episodic level will proceed by means of structural analysis, we expect to be able to specify additional rules for inferring interrelationships between statements. This roadmap will hopefully take us to the point where we will be able to consider narratives as sequences of statements, and apply the workflow to semantic markup in running texts.

References

[1] Augenstein, I., Padó, S., & Rudolph, S. (2012, May). Lodifier: Generating linked data from unstructured text. In *Extended Semantic Web Conference* (pp. 210-224). Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-30284-8_21.

[2] Balaraman, V., Razniewski, S., & Nutt, W. (2018). Recoin: Relative Completeness in Wikidata. In *Companion Proceedings of the The Web Conference 2018 (WWW '18).* International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1787–1792, https://doi.org/10.1145/3184558.3191641.

[3] Burkert, W. (1982). *Structure and history in Greek mythology and ritual* (No. 47). Univ of California Press, http://dx.doi.org/10.2307/1087336.

[4] Choi, Y. M., & Kim, H. J. (2007). A directed network of Greek and Roman mythology. *Physica A: statistical Mechanics and its Applications*, 382(2), 665-671, https://doi.org/10.1016/j.physa.2007.04.035.

[5] Clark, K., & Manning, C. D. (2016). Deep Reinforcement Learning for Mention-Ranking Coreference Models. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, http://dx.doi.org/10.18653/v1/d16-1245.

[6] Clark, K., & Manning, C. D. (2016). Improving Coreference Resolution by Learning Entity-Level Distributed Representations. *Proceedings of the 54th Annual Meeting of the Association for Computation-al Linguistics* (Volume 1: Long Papers), https://doi.org/10.18653/v1/p16-1061.

[7] Corcoglioniti, F., Rospocher, M., & Palmero Aprosio, A. (2015). Extracting knowledge from text with PIKES. In *14th International Semantic Web Conference (ISWC 2015)* (Vol. 1486).

[8] Darányi, S., Wittek, P., & Kitto, K. (2013, July). The Sphynx's new riddle: How to relate the canonical formula of myth to quantum interaction. In *International Symposium on Quantum Interaction* (pp. 47-58). Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-662-45912-6_5.

[9] Declerck, T., Aman, A., Banzer, M., Macháček, D., & Schäfer, L. (2017). Multilingual Ontologies for the Representation and Processing of Folktales. *Proceedings of the Workshop on Language Technology for Digital Humanities in Central and (South-)Eastern Europe*, http://doi.org/10.26615/978-954-452-046-5_003.

[10] Declerck, T., Kostová, A., & Schäfer, L. (2017). Towards a Linked Data Access to Folktales Classified by Thompson's Motifs and Aarne-Thompson-Uther's Types. *Proceedings of Digital Humanities 2017*, ADHO, 2017.

[11] Declerck, T., & Schäfer, L. (2017, June). Porting past classification schemes for narratives to a linked data framework. In *Proceedings of the 2nd International Conference on Digital Access to Textual Cultural Heritage* (pp. 123-127), http://dx.doi.org/10.1145/3078081.3078105.

[12] Falco, R., Gangemi, A., Peroni, S., Shotton, D., & Vitali, F. (2014, May). Modelling OWL ontologies with Graffoo. In *European Semantic Web Conference* (pp. 320-325). Springer, Cham, https://doi.org/10.1007/978-3-319-11955-7_42.

[13] Fellbaum, C. (2010). WordNet. In *Theory and applications of ontology: computer applications* (pp. 231-243). Springer, Dordrecht, https://doi.org/10.1007/978-90-481-8847-5_10.

[14] Gangemi, A., Presutti, V., Reforgiato Recupero, D., Nuzzolese, A. G., Draicchio, F., & Mongiovì, M. (2017). Semantic web machine reading with FRED. *Semantic Web*, 8(6), 873-893, https://doi.org/10.3233/sw-160240.

[15] Ilyefalvi, E. (2018) The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics. *Acta Ethnographica Hungarica* 63(1), 209-258, https://doi.org/10.1556/022.2018.63.1.11.

[16] Kanke, T. (2020). Knowledge curation work in Wikidata WikiProject discussions. *Library Hi Tech*, Vol. 39 No. 1, pp. 64-79, https://doi.org/10.1108/LHT-04-2019-0087.

[17] Kenna, R., & Mac Carron, P. (2016). Maths meets myths: network investigations of ancient narratives. In *Journal of Physics: Conference Series* (Vol. 681, No. 1, p. 012002). IOP Publishing, https://doi.org/10.1088/1742-6596/681/1/012002.

[18] Kõiva, M., Vesik, L. (2004). LEPP: The Database and Portal of South Estonian Folklore. *Folklore: Electronic Journal of Folklore*, 27, 133−162, https://doi.org/10.7592/fejf2004.27.koivavesik.

[19] Leach, E. (1996). *Lévi-Strauss*. Fontana Press: London.

[20] Lévi-Strauss, C. (1955). The structural study of myth. *The journal of American folklore*, 68(270), 428-444, https://doi.org/10.2307/536768.

[21] Lohmann, S., Link, V., Marbach, E., & Negru, S. (2014, November). WebVOWL: Web-based visualization of ontologies. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 154-158). Springer, Cham, https://doi.org/10.1007/978-3-319-17966-7_21.

[22] Malyshev S., Krötzsch M., González L., Gonsior J., Bielefeldt A. (2018) Getting the Most Out of Wikidata: Semantic Technology Usage in Wikipedia's Knowledge Graph. In: Vrandečić D. et al. (eds) *The Semantic Web – ISWC 2018. ISWC 2018. Lecture Notes in Computer Science*, vol 11137. Springer, Cham, https://doi.org/10.1007/978-3-030-00668-6_23.

[23] Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60), https://doi.org/10.3115/v1/p14-5010.

[24] Meder, T. (2006). ISFNR Committee "Folktales and the Internet". ISFNR Newsletter, 1.

[25] Mendes, P. N., Jakob, M., García-Silva, A., & Bizer, C. (2011, September). DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on semantic systems* (pp. 1-8), https://doi.org/10.1145/2063518.2063519.

[26] Miranda, P.J., Baptista, M.S., Ely de Souza Pinto, S. (2018) The Odyssey's mythological network. PLoS ONE 13(7): e0200703. https://doi.org/10.1371/journal.pone.0200703

[27] Nilsson, M. P. (1925). *A history of Greek religion* (No. 287). Clarendon Press.

[28] Nilsson, M. P. (1972). *The Mycenaean origin of Greek mythology* (Vol. 8). Univ. of California Press., http://dx.doi.org/10.1093/nq/clxiii.oct15.287c.

[29] Noy, N., & Rector, A. (2006). Defining N-ary Relations on the Semantic Web. W3C Working Group Note 12 April 2006, available at: https://www.w3.org/TR/swbp-n-aryRelations/.

[30] Pannach, F., Sporleder, C., May, W., Krishnan, A., & Sewchurran, A. Of lions and Yakshis. *Semantic Web*, (Preprint), 1-21, https://doi.org/10.3233/sw-200417.

[31] Poveda-Villalón, M., Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2014). Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS), 10*(2), 7-34.

[32] Sabou M., Fernandez M. (2012). Ontology (Network) Evaluation. In: Suárez-Figueroa M., Gómez-Pérez A., Motta E., Gangemi A. (eds) *Ontology Engineering in a Networked World*. Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-24794-1_9.

[33] Smith, D. A., Rydberg-Cox, J. A., & Crane, G. R. (2000). The Perseus Project: A digital library for the humanities. *Literary and Linguistic Computing*, 15(1), 15-25, https://doi.org/10.1093/llc/15.1.15.

[34] Syamili, C., & Rekha, R. V. (2018). Developing an ontology for Greek mythology. *The Electronic Library*, http://dx.doi.org/10.1108/el-02-2017-0030.

[35] Thompson, S. (1935). Motif-Index of Folk Literature. A classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books and Local Legends, JSTOR, 7(3):359, http://dx.doi.org/10.2307/2919989.

[36] Thompson, S. (1946) The folktale. Holt, Rinehart and Winston: New York., cited by Garry, J and El-Shamy, H. /Eds./ (2005). Archetypes and Motifs in Folklore and Literature: A Handbook. M.E. Sharpe: Armonk, N.Y., p. xiii.

[37] Vrandečić, D., & Krötzsch, M. (2014). Wikidata: a free collaborative knowledgebase. Communications of the ACM, 57(10), 78-85, http://dx.doi.org/10.1145/2629.