
Towards a Linked Data Access to Folktales classified by Thompson’s Motifs and Aarne–Thompson–Uther’s Types

Thierry Declerck
declerck@dfki.de
German Research Center for Artificial Intelligence
Germany

Antónia Košťová
akostova@coli.uni-saarland.de
Saarland University, Germany

Lisa Schäfer
lisas@coli.uni-saarland.de
Saarland University, Germany

Introduction

We present in this paper work consisting in porting to an integrated ontology two central resources for the classification of folktales: The “Motif-index of folk-literature” (Thompson, 1977) and the “Types of International Folktales” (Uther, 2004). The first resource, often called Thompson Motif Index (abbreviated as TMI) is available as an [online resource](#). The second resource is a classification system for folktale types, which was published by Hans-Jörg Uther (2004), extending former work by Antti Aarne (1961) and Stith Thompson (1977). In the following we are using the acronym ATU for referring to this classification system. Recently, a large amount of the ATU data has been made available [online](#), offering also annotation facilities for tales in multilingual versions.

Our work consisted in extracting from those resources, which are stored in different formats, classification relevant information and re-organizing them in two interrelated ontologies, using for this the W3C standards [OWL](#) (which stands for “Web Ontology Language”, RDF(s)– see the [W3C recommendations](#) for more details– and [RDF](#). The aim is to make those classification resources machine readable,

interoperable and to support by this formal representation of the metadata access to folktales annotated by those classification systems in the context of the [Linked Open Data](#) framework.

Thompson Motif Index

A folktale motif can be defined as a “repeated story element, e.g., a character, an object, an action, or an event that can be found in several stories”. In TMI motifs are organized in a tree structure, providing for a parent-child relation between the listed elements. One motif entry consists of a motif-id and a motif name. Optionally, a **motif description** and **references** are provided. Table 1 provides for an example of few motifs illustrating the tree structure and hierarchy of TMI.

Motif-id	Motif name
A	Mythological motifs
A1	Identity of creator
A1.1	Sun-god as creator
A1.2	Grandfather as creator
A1.3	Stone-woman as creator
A1.4	Brahma as creator
A2	Multiple creators

Table 1: A few motifs from Motif-index of folk-literature and their hierarchical organisation

Aarne–Thompson–Uther Folktales Types (ATU)

A folktale type can be described as a main story line that can be found in several cultures. The parts of this story line can refer to specific story elements also known as motifs. A folktale type is therefore a bigger unit than a motif. As can be seen in example 1, an entry in the ATU system consists of a type id (“6*”), a title (“Animal Captor Talks with Booty in his Mouth (previously The Wolf Catches a Goose).”) and a text summarizing the typical “storyline” of this type of folktale. Within or at the end of this “script”, links to corresponding Thompson Motif-Indices can be provided (“[K561.1]”). Finally (and optionally), similar or related types can be indicated.

(Example 1):6*~Animal Captor Talks with Booty in his Mouth (previously The Wolf Catches a Goose).~A wolf catches a goose and a fox catches a chicken. The fox asks the wolf something so that he opens his mouth and the goose is able to fly away [K561.1]. Then the wolf asks the fox, but the fox answers without losing his booty.~Cf. Types 6, 227*

Generation of the Ontology

The OWL and RDF(s) representation for the ontology was generated semi-automatically from the

html code of both TMI and ATU, responding to few design decisions we had to take. For TMI we went for a double representation: the hierarchy structure of the IDs is represented as an OWL subclass hierarchy, but all terminal nodes (leaves of the tree) are represented as both an instance of a class we call “Motif” and as an instance of the pre-terminal node in the taxonomy. This reflects our intuition that what Thompson called a motif is in most of the cases the content of the terminal nodes of the classification system, while the non-terminal nodes are more to be considered as abstraction helping in the taxonomic structures.

We compared the automatically created ATU part of the ontology to the printed version of Hans-Jörg Uther's “The Types of International Folktales”. Using the ontology editor “Protégé”, we manually added missing subclasses and individuals, rearranged generated classes and corrected errors such as typos in the electronic version of ATU or splitting errors because of inconsistent punctuation. By this step we obtained 2802 ATU classes organized in seven main subclasses, which have also subclasses, in accordance with the hierarchical structure of types proposed in (Uther, 2004). Below we display examples of the encoding of ATU data in our ontology. We first show a main class (we use the [Turtle syntax](#) for serializing the RDF code) of our ATU model “:Type”:

```
:Type
rdf:type owl:Class ;
rdfs:comment "List of all terminal nodes of the ATU Hierarchy" ;
rdfs:label "ATU Types"@en ;
```

A subclass of “Type”, for example the type “6*”, has the following syntax:

```
<http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-atu-ontology#6*>
rdf:type :Type ;
rdf:type owl:Class ;
:linkToTMI <http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-atu-ontology#K561.1> ;
rdfs:comment "\"Type 6* of ATU\""@en ;
rdfs:isDefinedBy "A wolf catches a goose and a fox catches a chicken. The fox asks the wolf something so that he opens his mouth and the goose is able to fly away [K561.1]. Then the wolf asks the fox, but the fox answers without losing his booty. Cf. Types 6, 227*."@en ;
rdfs:label "\"Animal Captor Talks with Booty in his Mouth (previously The Wolf Catches a Goose)\""@en ;
rdfs:subClassOf <http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-atu-ontology#The_Clever_Fox_(Other_Animal)_1-69> ;
```

In this example, the reader can see how the type 6* is linked to a motif occurring typically in its storyline: we introduced for this a property called “linkToTMI”. Additionally, the subclass relation is expressed, using the `rdfs:subClassOf` property. The “`rdfs:label`” property stores the original short title of the ATU type in English (“@en”). We encode the original description of the type as a value to the property

“`rdfs:isDefinedBy`”. A main aspect of the ontologisation of ATU (and TMI) is that each folktale type (or motif) is now represented by a Unique Resource Identifier (URI), and thus accessible in the Linked Data framework, once our data set has been published in its cloud.

An example of a motif (“K561.1”) is given just below. We focused for the time being only on motif-ids and names. This current limitation is due to the inconsistent format of the motif descriptions and references used in the html code of the online resource, which made it difficult to be automatically extracted. We will include this information in a next version of the ontology. As pointed out earlier, elements of the TMI are encoded in a dual fashion: as belonging to the class “Motif” but also to its immediate non-terminal node (here “K561”). The `rdfs` annotation property “`rdfs:label`” is used for encoding the name of the motif (here in English, marked by “@en”). Multilingual correspondences can also be included as values of this property.

```
<http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-atu-ontology#K561.1>
rdf:type :K561 ;
rdf:type :Motif ;
rdf:type owl:Class ;
rdfs:comment "\"Index K561.1 of TMI\""@en ;
rdfs:label "\"Animal captor persuaded to talk and release victim from his mouth\""@en ;
rdfs:subClassOf :K561 ;
:linkFromTMItoATU <http://www.semanticweb.org/tonka/ontologies/2015/5/tmi-atu-ontology#6*> ;
```

In this example, we also see the property “`linkFromTMItoATU`”, which is the inverse property of the one pointing between ATU elements and motifs. Additionally, we have introduced a third “linking” property, called “`linkFromAaThToATU`”, which ensures that types of former versions of ATU are linked to the new names in the final version of ATU. By this final step of expanding our TMI-ATU ontology we ended up with the number of 14,937 classes and the number of 49,752 individuals that are interconnected by 3 object properties: “`linkFromTMItoATU`”, “`linkToTMI`” and “`linkFromAaThToATU`”. We managed thus to convert two valuable, handcrafted resources of literary knowledge consisting of more than 4000 pages into a 15.4 MB in size ontology file that can be easily accessed and searched.

Conclusions

We presented our work on the ontology generation for two widely used folktale classification systems. This ontology can be visualised and processed by standard OWL tools such as Protégé. The integrated ontology will be made openly available soon, after last quality controls. Current work is on adding as instances of the ontologies URLs of folktales that are

marked with the corresponding numbers and so to allow access to those via Linked Data mechanisms.

Bibliography

- Aarne, A.** (1961). *The Types of the Folktale: A Classification and Bibliography*. The Finnish Academy of Science and Letters, Helsinki.
- Ashliman, D. L.** (1987). *A Guide to Folktales in the English Language: Based on the Aarne-Thompson Classification System*. New York, Greenwood Press.
- Ofek, N., Darányi, S., and Rokach, L.** (2013). *Linking Motif Sequences to Tale Types by Machine Learning*. Workshop on Computational Models of Narrative, 166-182.
- Thompson, S.** (1977). *The Folktale*. Berkeley: University of California Press.
- Uther, H-J.** (2004). *The Types of International Folktales: A Classification and Bibliography. Based on the system of Antti Aarne and Stith Thompson*. FF Communications no. 284-286. Helsinki: Suomalainen Tiedeakatemia.