Metadata model and application for folktale resource discovery

Kate Susan Smith

Master's Program in Informatics

Degree Programs in Comprehensive Human Sciences

Graduate School of Comprehensive Human Sciences

University of Tsukuba

September 2022

Metadata model and application for folktale resource discovery

Name: Kate Susan Smith

This research proposes a metadata model, dataset and application, using Linked Open Data, for aiding with folktale discovery and selection, in order to make resource discovery of folktales easier in non-academic contexts for general users.

Folktales can be defined as traditional tales that have been passed down orally through generations, and are important for examining social issues, customs, traditions, beliefs, and the historical contexts in which the tales were created. However, despite these tales being significant for a number of reasons, certain issues with prevailing classification systems can make the discovery and selection of folktales that are relevant to user needs seem difficult, and can prevent general users from accessing tales that are suitable for their needs. Therefore, making folktales more accessible and easier for a variety of users to find is the central point of this research.

The proposed model will combine elements of existing folktale classification systems with new detailed structured categories. Folk Legends will be focused on, rather than various forms of folk narratives, as there is a gap in this area, with regards to one of the main folktale indexes, and it is a genre that is significant for users interested in local history. Folk Legends often cover topics such as superstitions, beliefs, or personal anecdotes, and can often include verifiable details, such as place names, names of historical figures, or exact dates and times.

As detailed knowledge of existing folktale classification systems will not be necessary in order for users to query this proposed dataset, the hope is that this will make discovering folktales more accessible, and lead to more knowledge about social history and local history being shared.

The building of the dataset was started in conjunction with the data model design, in order for the relevant information extracted from the tales to be used to help inform the design choices for the data model. The choice of materials for the data set were obtained from a single volume of a collection of British folktales. 109 folktales were used in the dataset. Data extracted from the collection was recorded in RDF using Turtle syntax, in order to express information about the relationships between the data. To evaluate the data model, a list of hypothetical user needs were proposed. SPARQL query language was then used in order to query the dataset for these user needs.

The results of the SPARQL queries seemed to satisfy the user needs, as tales featuring the specified search terms were returned in the results. This implies that the data model is relatively effective, as the tales can be searched and the results narrowed down to find more specific information for the users. However, a number of issues were raised that needed further examination. Location data joined with a hyphen (for example, ‘Somerset-Devon’), and keywords about the same topic but recorded using different terms did not appear in the results. This suggests that the way in which the location data and keywords are recorded needs to be examined further. There were also limitations to the research, as no user studies were carried out, with the user needs and use cases being hypothesised. Furthermore, the sample size is small, with only 109 tales across 7 folktale subjects being used in the dataset.

Main Academic Advisor: Masao TAKAKU
Secondary Academic Advisor: Toshihito WAKI

Table of Contents

1. Introduction page 1

2. Related Research page 3

    2.1 Methods of Folktale Classification page 3

    2.2 Computational Folkloristics page 4

    2.3 Tailored Schema page 5

3. Data Model page 5

    3.1 Overall Structure of Tale, Creature, Source and Collection Resource page 7

    3.2 Tale Resource page 7

    3.3 Creature Resource page 13

    3.4 Hierarchy of Creature Classes page 16

        3.4.1 Black Dogs page 17

        3.4.2 Bogies page 19

        3.4.3 Dragons page 20

        3.4.4 Fairies page 21

        3.4.5 Ghosts page 22

    3.5 Source Resource page 23

    3.6 Collection Resource and Hierarchy page 25

4. Dataset page 27

    4.1 Conversion of dataset into RDF page 30

5. Evaluation page 31

    5.1 User Needs page 31

    5.2 SPARQL Results page 33

6. Discussion page 42

    6.1 Results Interpretation page 42

6.2 Limitations page 47

7. Conclusion page 49

7.1 Future Work page 50

8. Bibliography page 50

1. Introduction

This research focuses on creating a metadata model, dataset and application for aiding with folktale discovery and selection, in order to make resource discovery of folktales easier in non-academic contexts for general users.

Current methods of classifying folktales are advantageous for the comparative study of folktales. Using the ATU index [1], tales are classified into groups of stories that have similarities. If a user wants to find similar tales, then they are able to search for tales with the same type index. However, indexes such as the ATU are usually used in academic contexts, and can be difficult for general users to navigate [2]. Therefore, it is necessary to examine other methods of enabling users to find tales relevant to their needs.

Since 1989, UNESCO has considered folklore to be 'intangible cultural heritage' [3], with folklore referring to a variety of customs and art forms, including "verbal art forms such as tales, legends, proverbs, songs, etc" [4]. Folktales can be defined as traditional tales that have been passed down orally through generations, and can include many sub-genres such as fairy tales, fables, jokes, and folk legends [5]. Folktales are significant as they are important for examining local and social history, culture, and are often used as teaching aids in schools. For example, there are often tales that give explanations for why certain geographical features in the landscape have come to be, or the history and incidents related to particular places or named individuals. Folktales can also give us information about social issues, customs, traditions, and beliefs and ideologies that people in the past held, giving us insights into what daily life might have been like, how people thought about certain issues, or the historical contexts in which the tales were created [6]. They are also used in schools to educate children on issues such as story structure, bias and point of view, and learning about the world.

However, despite these tales being significant for a number of reasons, certain issues with current classification systems can make the discovery and selection of folktales that are relevant to specific user needs seem difficult, and can prevent general users from accessing tales that are relevant to their needs. During my undergraduate dissertation project, it was necessary that I find examples of folktales with specific elements to examine. However, finding such tales to meet my requirements proved difficult. Also, after talking with various individuals engaged in educational professions, it seemed that there were also issues arising for them, in not being able to find tales that fit their specific needs for educational purposes. Therefore, making folktales more accessible and easier for a variety of users to find is the central point of this research.

Current well-known methods of classifying folktales include the ATU index [1] and the Motif-Index [7]. The ATU index is an index of topics that appear in folktales, classifying tales according to their tale type. The Motif-Index identifies common motifs, or smaller elements, that appear across a number of tales. These classification schemes are advantageous for the comparative study of folktales, and for examining the relationships between motifs and tales. They are well suited to dealing with sorting tales

according to their ‘etic’ units [8], such as the type of animals that appear, or the magical objects that are involved in a tale.

However, these ‘etic’ units are more changeable and have more variation than the ‘emic’ units [8], which consider the narrative structure and plot points of the tales. Broadwell et al. state that this focus on the changeable ‘etic’ units can lead to tales with similar narrative structures being classified under different tale types, which can lead to difficulties with resource discovery and tales not being accessible to users [8]. Also, these classification schemes are more commonly used for Folk Narratives such as fairy tales, rather than Folk Legends, and so the ATU categories are not always relevant or applicable to these Folk Legends, as some tales may not fit into any of the categories. Briggs states that with regards to Folk Legends, the “Type Index ceases to be a guide, for it does not purport to deal with legendary matter” [9].

In addition, there are issues with the inelasticity of the categories. Although the ATU classification method only assigns one label to a tale, some tales can be difficult or unable to be placed, as they could potentially belong to multiple categories [8]. The Tale Type Index also “leaves little room for the insertion of newly collected types”, and Briggs also states that the Motif-Index needs to be expanded significantly in order to cover more subject matter [9].

Furthermore, the ATU categories can be quite large, and do not provide more than one level of sorting the tales. For example, there are over 345 different versions of cinderella [10], but they can be categorised under the same tale types (510A and 510B). There can be hundreds of tales in just one category, and it can be difficult to search through these in order to find tales with specific aspects or similar plot points that may be relevant to user needs. As such, the classifications can be difficult to navigate, and are not generally used outside of academia.

Perspectives on analysing folktales have shifted over time. There has been a shift from analysing tale variants in order to examine how tales may have migrated across locations, to analysing tales from a socio-historical perspective, in which the tales are seen to reflect the social and historical conditions of the environment and time periods in which they were created [11]. In this way, many collections of tales have been classified using systems that were designed in accordance with earlier theories and methods of analysis [8] and so may not completely suit the needs of users today.

The research question is whether it will be possible to propose a metadata model, dataset and application to enable users to find folktales that are relevant to their needs and specific criteria. A model will be proposed which aims to combine current methods of classification with new detailed structured categories using Linked Open Data. The new categories will work to combine searching for both ‘etic’ and ‘emic’ elements of the tales. A hierarchy of classes based on the subjects that appear in the tales, the creatures, will be created, in order for users to be able to narrow down which tales to search. However, with the aim of making the proposed model more flexible with regards to classification, it will be possible to assign more than one label or class to a creature.

Other searchable functions will be included, such as location data, keywords, and named entities (such as monuments or historical figures). Also included will be a list of plot points, that will list the events or actions that take place in each of the tales. The aim is that this will enable users to search for tales based on their narrative structure in addition to the other elements mentioned above. Folk Legends will be focused on, as there is a gap in this area where the Tale Type Index is concerned, and it is a genre that is significant for users interested in local history. As detailed knowledge of the ATU and Motif-Index will not be necessary in order for users to query this proposed dataset, the hope is that this will make discovering folktales more accessible, and lead to more knowledge about social history and local history being shared.

2. Related Research

The aim of this research is to propose a metadata model, dataset and application for aiding with folktale discovery and selection, in order to make resource discovery of folktales easier in non-academic contexts for general users. This is related to: 2.1 Methods of Folktale Classification, such as the ATU index and the Motif-Index; 2.2 Computational Folkloristics, using computational methods to aid in the study of folklore; and 2.3 Tailored Schema, creating various schema related to folklore, tailored for specific audiences. Therefore, an overview of the related research will be divided and organised into these separate areas.

2.1 Methods of Folktale Classification

Current well-known methods of classifying folktales include the ATU index [1] and the Motif-Index [7]. These have developed from a combination of different researchers' work in the field, beginning in 1910 with Antti Aarne [12] and revised in 1928 by Stith Thompson [13], and revised again in 2004 by Hans-Jörg Uther [1]. The ATU index classifies tales according to their tale type, which is the model form, or template of a tale. Tales with the same tale type are often retold or reimagined slightly differently from each other, but will still be recognisably similar to the model form. These recognisably similar tales are described as variants. The Motif-Index [7] identifies common motifs that appear across a number of tales. Motifs can be described as the small parts of the tales that are recognisable and memorable enough to continue reappearing in subsequent retellings and reimagining of tales, being "the smallest element in a tale having a power to persist in tradition" [14] . These classification systems are advantageous for the comparative study of folktales and motifs, and are well suited to dealing with sorting tales according to their 'etic' units rather than the 'emic' units [8]. Other classification schemes, such as Christiansen's 'The Migratory Legends' [15] from 1958, deal with the classification of Folk Legends from Norway. Christiansen created this classification scheme for Norwegian Folk Legends (or Sägen) in order to be able to include a large number of additional types without disturbing the order and structure of the existing Aarne-Thompson index, and because Folk Legends were regarded as having an "essential difference" from the folktales contained in the Aarne-Thompson index [15]. Baughman's work 'Type and Motif-Index of the Folktales of England and North America' [16] from 1966 supplies additional types for tale and

motif classifications that did not appear in Thompson's earlier work [7]. Propp's work 'Morphology of the Folktale' [17], although not a classification scheme, identified various structural elements (called 'functions') within Russian folktales, that can be combined together to depict the structure of whole tales. He identified a list of 31 functions that appeared in the selected tales he examined. Each tale contained a combination of various functions, in order, from this list. The functions are numbered, so a record of the functions that appear in each tale can be created, thereby describing a tale's narrative structure.

In order to enable users to search for tales based on various criteria, this research aims to combine these already existing classification schemes with new detailed structured categories. These structured categories will include a list of plot points, which will be actions or events that take place in the tales, that combine to make up the narrative structure. This will differ from the ATU index, and aims to enable users to search for the 'emic' units of the tales in addition to the 'etic' units [8]. This research will also differ in that it will focus on Folk Legends, rather than other types of Folk Narratives that the ATU focusses on. Unlike Christiansen's work [15], this research will deal with a collection of folktales from Britain, rather than Scandinavian folktales. Unlike Propp's work, the number of plot points will not be restricted to the 31 functions, and the actions will be more specific than the broad functions. Also, the characters and subjects that appear in the Folk Legends will not be assigned labels such as 'villain', 'hero' or 'victim' like the dramatis personae in Propp's work.

2.2 Computational Folkloristics

Computational folkloristics uses digital tools in order to analyse folklore from different perspectives, such as using text mining or data visualisation, in order to gain new insights [18]. It can also enable data analysis of folklore on a larger scale. The research that Broadwell et al. [8] conducted was to examine how to classify folktales that researchers were previously unsure what category they should be assigned to, as the tales could potentially belong to multiple categories. They utilised computational methods in order to identify common words that appeared in tales that already had clear classifications, then used these word lists to discover if the unclassified tales had any similarities to these lists. This could then help pinpoint which categories the unclassified tales appeared to have more similarities with, or demonstrate how various categories overlapped. The ambiguous nature of deciding where some tales should be placed was emphasised, and the liminal tales where uncertainty arises were shown to be the "richest hunting ground" [8] for research. Returning agency to the researcher through exploring these differing classifications was discussed. Meder et al. [19] discuss their investigations into how machine learning techniques can be used in order to automate the process of assigning metadata to tales in the Dutch Folktale Database [20]. Due to the amount of folktales waiting to be added to the database, and the time consuming nature of assigning the metadata manually, experiments were carried out to ascertain whether metadata such as sub-genre and tale type identification could be carried out automatically. Supervised machine learning was carried out, with the computer learning from examples of stories that had been correctly annotated manually. They discovered

that the computer was more effective in correctly assigning the sub-genres of jokes, riddles and legends to the tales.

In contrast, my research will not be focusing on classifying the tales into preexisting categories, but aims to be situated from a user's perspective, enabling users to find appropriate Folk Legends with specific criteria, by combining existing indexes with new detailed structured categories. Also, the sample of tales that will be used in the dataset will not be large enough to use machine learning techniques.

2.3 Tailored Schema

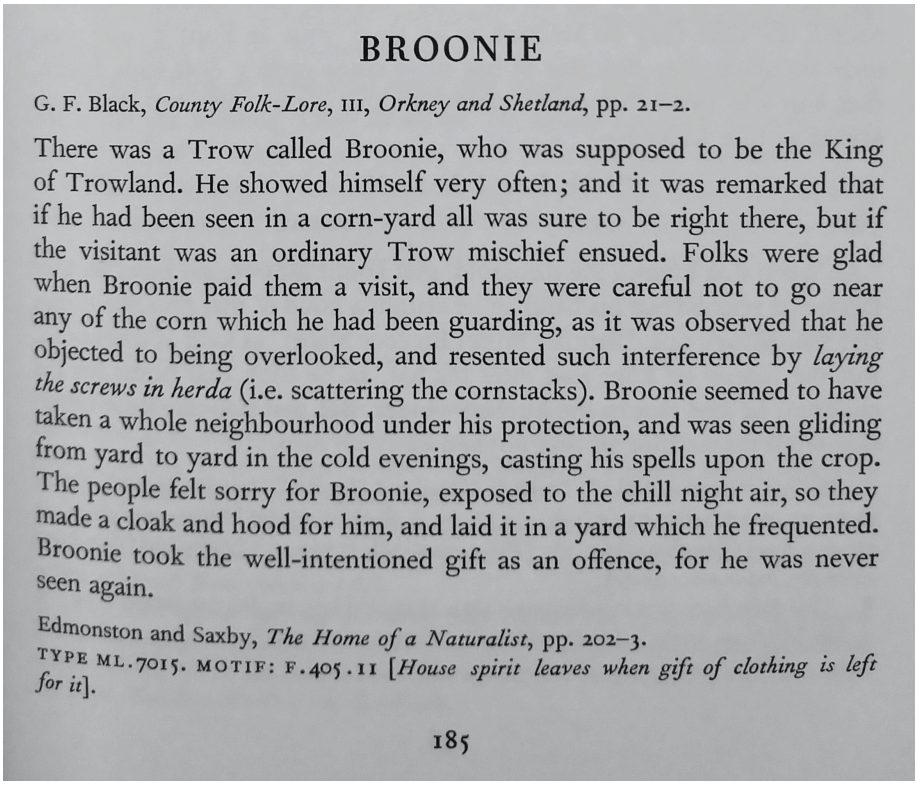
Crevasse [2] conducted research in order to create a tailored descriptive schema that would be able to be used as a foundation for a fairytale and folklore database for extra-academic professionals. The schema would be beneficial for extra-academic professionals wanting to find relevant fairy tales that would be appropriate for their jobs, such as professional storytellers and school librarians finding suitable stories to share with children. Interviews were conducted in order to gather information on the views of the extra-academic professionals, examining what they may find useful and necessary. Attention was paid to the user needs, as existing classification systems used in academic settings may not have been useful. A hybridised schema was proposed, that incorporated aspects of existing classification systems with the extra-academic professionals' needs. Giannoulakis et al. [21] carried out research into a proposed metadata schema for intangible cultural heritage in the form of folk dances. The needs of many different users were considered when proposing the framework, such as dance professionals, teachers, researchers and the general public. Also, how the users would utilise the information was considered, such as using the information to learn how to perform a particular dance. They proposed various metadata elements to describe the dance information, creating a digital choreographic model.

Although the intended audience is different, and Crevasse's [2] researched focused on literary fairytales and folk narratives, and Giannoulakis et al. [21] focused on folk dance (as apposed to this research, which will be focused on Folk Legends), aspects of this research will have similarities to the above mentioned works. For example, the focus on a user perspective, and paying attention to how the intended audience will use the schema. Also, the hybridisation of existing classification systems with new elements.

3. Data Model

Firstly, a data model for folk legends and the creature subjects that appear in them was proposed. The data model was designed to focus on information that may be of assistance when searching for folktales with specific criteria, accommodating the user's needs. The data model will be constructed to demonstrate the links and properties between the different resources, and the hierarchical relationships between them and the proposed classes. The building of the dataset was also started in conjunction with the data model design, in order for the relevant information extracted from the tales to be used to help inform the design choices for the data model. The choice of materials for

the data set will be obtained from Volume I of the collection 'A Dictionary of British Folk-Tales, Part B Folk Legends' by Katharine M. Briggs [9].



BROONIE

G. F. Black, *County Folk-Lore*, III, *Orkney and Shetland*, pp. 21–2.

There was a Trow called Broonie, who was supposed to be the King of Trowland. He showed himself very often; and it was remarked that if he had been seen in a corn-yard all was sure to be right there, but if the visitant was an ordinary Trow mischief ensued. Folks were glad when Broonie paid them a visit, and they were careful not to go near any of the corn which he had been guarding, as it was observed that he objected to being overlooked, and resented such interference by *laying the screws in herda* (i.e. scattering the cornstacks). Broonie seemed to have taken a whole neighbourhood under his protection, and was seen gliding from yard to yard in the cold evenings, casting his spells upon the crop. The people felt sorry for Broonie, exposed to the chill night air, so they made a cloak and hood for him, and laid it in a yard which he frequented. Broonie took the well-intentioned gift as an offence, for he was never seen again.

Edmonston and Saxby, *The Home of a Naturalist*, pp. 202–3.
TYPE ML.7015. MOTIF: F.405.11 [*House spirit leaves when gift of clothing is left for it*].

185

Figure 1. An example of a tale from the Fairies chapter of Volume I of 'A Dictionary of British Folk-Tales, Part B Folk Legends' by Katharine M. Briggs [9].

In order to follow the design principles of Linked Data [22] for interconnecting structured data, URIs were used to refer to the relevant parts of the tale, creature, source and collection resources. example.org was used to provide example domains of the resources. For instance, the example HTTP URI ex:BlackDog or <https://example.org/BlackDog>. RDF was used in order to express information about the relationships between the data recorded from the Folk Legends, and how this data relates to other resources. RDF was used so that the data could be discovered more easily by "RDF-aware applications" and so become more usable, as it can connect to other datasets and sources [23]. RDF was ideal for expressing the relationships in the folktale dataset, as it was able to depict information that is "expressed using different schemata in a single graph" [23]. The description of the RDF triples was written using Turtle syntax. SPARQL query language was used in order to query the dataset. Links to other URIs, such as data on locations featured in the tales, information on people or places mentioned, or further information about the Motif-Index, were also added to enrich the data and enable the users to discover more relevant information.

3.1 Overall Structure of Tale, Creature, Source and Collection Resource

The data model connects the Tale Resource, Creature Resource, Source Resource, and Collection Resource together.

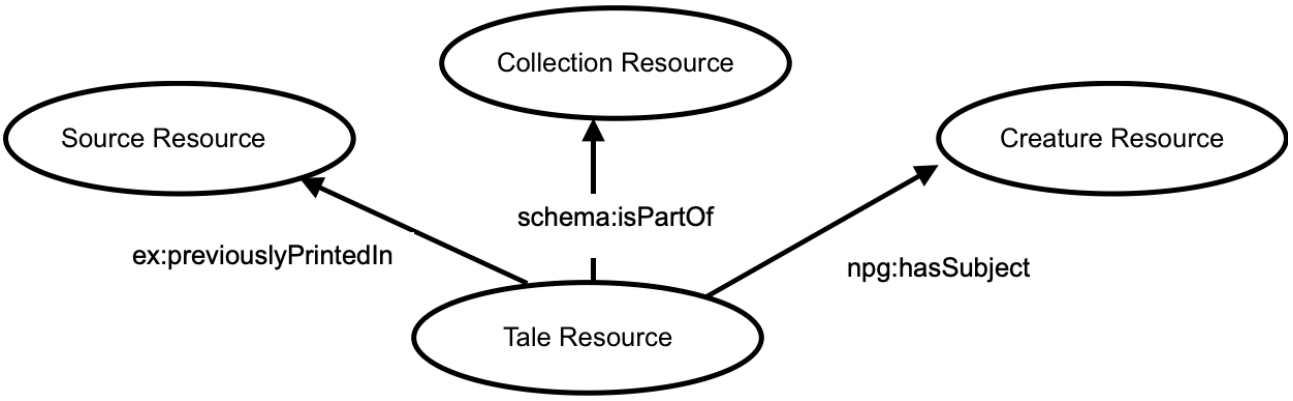


Figure 2. Diagram of how the Tale, Creature, Collection and Source Resources are connected together with properties.

Each individual tale in the collection will be a resource, and the properties and objects connected to the Tale Resource will demonstrate information about each tale. This information will be details that correspond with the user needs, assisting users to find tales with specific criteria.

Each Tale Resource will have at least one creature which appears as the main subject of the tale, which will be the Creature Resource. The Creature Resource connects to the Tale Resource and demonstrates what creatures appear in the tale and information about those creatures. There is also a hierarchy of classes for the creatures.

Each Tale Resource was also printed in other sources before being collected in Briggs' folktale collection. The sources that the tales were previously printed in will be separate resources that connect to the Tale Resource. The Source Resource will contain information about the book or material that each Tale Resource was previously printed in.

There will also be the Collection Resource, which will contain information about Briggs' collection, and where in the collection each Tale Resource is located.

3.2 Tale Resource

It was decided that the central node would be the Tale Resource. The properties related to the Tale Resource were established as: ex:hasPlotPoints; ex:TypeIndex; ex:typeOfTale; schema:identifier; schema:name; schema:pagination; ex:motifIndex; ex:motifInfo; schema:location; ex:locationAltName; ex:placesMentioned;

ex:locationLinks; ex:locationInfo; schema:keywords; schema:mentions; ex:mentionedInfo; and ex:dateMentioned.

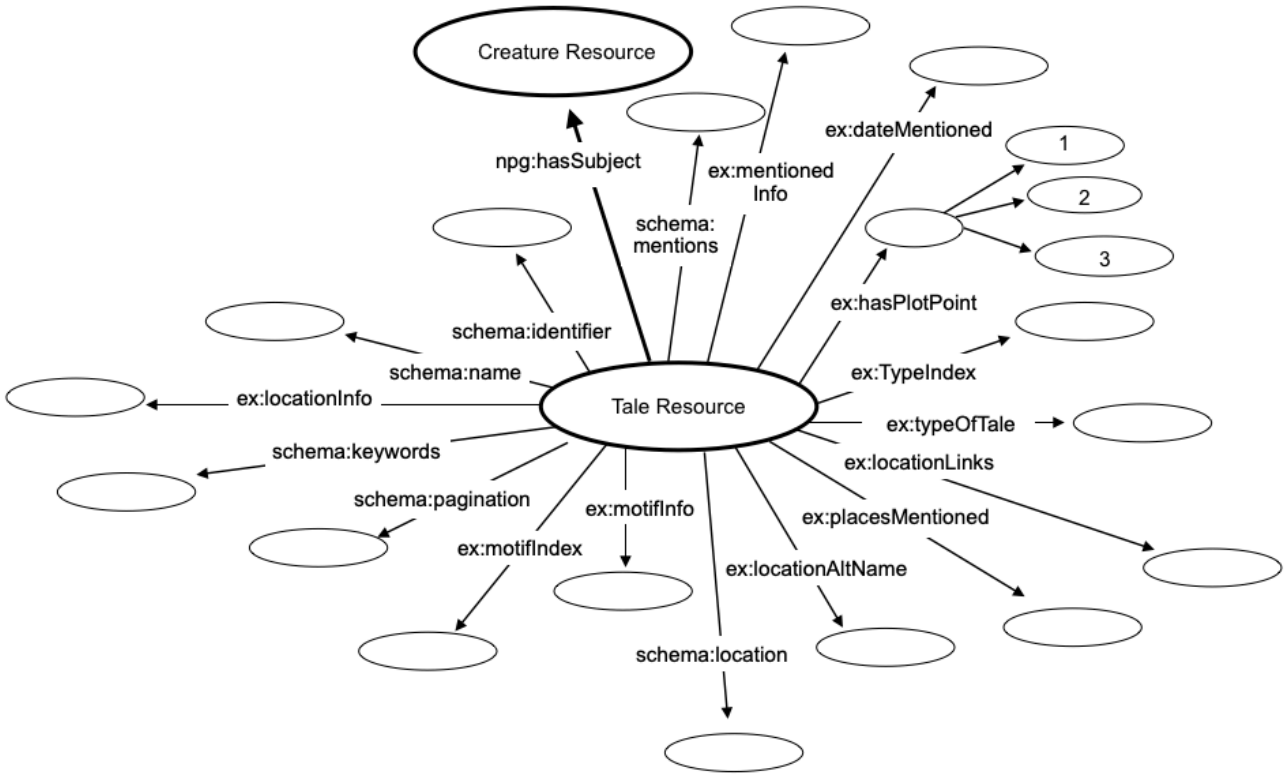


Figure 3. Diagram of the structure of the central node Tale Resource with the properties that connect the Tale Resource to the objects.

The property ex:hasPlotPoints describes the relationship between the Tale Resource and the plot points that the tale contains. The narrative structure of the tale is analysed and then separated into small phrases that describe the events of the tale. The property ex:hasPlotPoints is connected to a blank node, then the list of plot points are numbered and ordered in the sequence in which they appear in the tale. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. The property is mandatory and repeatable.

It was decided to include this property in order to document the structure of the folktales, suggesting a way in which to sort the tales by examining the ‘emic’ units of their narrative structure, rather than the changeable ‘etic’ units [8]. Many of the topics in the Motif Index [7] also correspond to certain plot points that happen in the tales. However, these motifs often deal with specific ‘etic’ units, such as the kind of creature or subject the event is based around, and so it can be difficult to match these events across tales dealing with different subject matters. For instance, two separate motifs listed in the index are “F405.11 House spirit leaves when gift of clothing is left for it” and “F381.3 Fairy leaves when he is given clothes” [7]. These motifs both deal with the

event of a creature leaving a dwelling because it has been given clothes. However, because the motifs include the changeable elements of ‘fairy’ and ‘house spirit’, they are listed under separate sections, and so searching for tales with the motif F381.3 will not include tales with motif F405.11, despite them being so similar. Therefore, it was decided that by removing the changeable element of which creature or subject is involved in the event, this would make it possible to search across the whole dataset for the tales that contain the same plot point. In this instance, ‘clothes given to subject’ would be one plot point, and later on in the tale ‘subject leaves due to gift’ is listed as another plot point. This separates the motif into small actions that can be combined together in various ways to create a whole tale. This enables users to search across all the tales for a single plot point, regardless of the changeable units surrounding them, such as the different creatures or subjects. It also enables users to search for multiple plot points in combination with each other, in order to examine plot points that often exist together. For example, ‘clothes given to subject’ and ‘subject leaves due to gift’ appear together multiple times.

The property of ex:TypeIndex was used in order to describe, when applicable, the number and name of the tale type index that Briggs had assigned to the tales. For example, multiple tales were assigned “Type 1090 Contest Between Man and Ogre: Mowing Contest” [13], as they have a similar story. As previously stated, the types in the Tale Type Index are not always applicable to this collection of tales, so not all of them have been assigned a type. The name of the property does not specifically mention the ATU index, as this version of the index did not exist when Briggs’ collection was published in 1971. The Tale Type Index was originally published by Aarne in 1910 [12] and then translated into English and built upon by Thompson in 1928 [13]. It was then further built upon by Uther in 2004 [1], becoming the ATU index that is used today. Therefore, Briggs had assigned tales using the Aarne and Thompson version of the index. Briggs also used Christiansen’s “The migratory legends” [15] to assign types for some of the tales. Due to this, no specific indexing system was mentioned in the property, so that it could be used to refer to more than one index. This property of the previously assigned tale types was included in order for existing classification methods to be combined with the other tale information in the data model. The number and name of the tale type, as well as the index that each classification is from, is written as a literal in the object node for this property. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. The property is optional and can be repeated if necessary.

The property of ex:motifIndex was used in order to describe, when applicable, the identifier and name of the motifs that Briggs had assigned to the tales. For example, “F381.3 Fairy leaves when he is given clothes” [7]. Thompson’s Motif Index [7] and Baughman’s Type and Motif-Index [16] were both used. This property of the previously assigned motifs was included in order for existing classification methods to be combined with the other tale information in the data model. Not all of the tales have assigned motifs and some have multiple motifs assigned to them, so the property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property of ex:motifInfo was used in order to link other sources of information about the tale motifs to the tale resources. The object nodes of this property will have links to a website listing various motif categories [24] applicable to the tale resources. For example, if a tale is assigned “F381.3 Fairy leaves when he is given clothes” [7], then it will be linked to the webpage listing all the other motifs designated as “F. Marvels” [7]. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. Not all of the tales have assigned motifs and some have multiple motifs assigned to them, so the property is optional and can be repeated if necessary.

The property ex:typeOfTale was used in order to describe what type of tale the Tale Resource is. For example, the type of tale for “The Dragon of Kingston” is a ‘Folk Legend’. This property may seem redundant, as it was previously stated that the dataset would be comprised specifically of tales from Part B Folk Legends, so theoretically all of the tales should be listed as ‘Folk Legend’. However, some of the tales in Part B are also included in the Part A Folk Narratives section, and could be considered both a Folk Narrative and Folk Legend. Therefore, it was decided that this property would be included, as it may be useful for those tales that cross between the two types. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. The property is mandatory and can be repeated if necessary.

The property schema:location was used in order to describe where the tales took place. This property was significant as Folk Legends can often be particular to one area, or related to local legends specifically about a certain place or landmark [9]. It can also be interesting to discover patterns about where certain tales or creatures appear. For instance, there are numerous tales about giants in Cornwall, and the characteristics of Black Dogs often vary according to the geographical region [25]. This information was also important for fulfilling the user needs, as searching for tales based around a particular area is useful for users interested in the history or culture of an area. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. Not all of the tale resources mention locations, and some take place across multiple locations, so the property is optional and can be repeated if necessary.

The property ex:locationAltName was used in order to note alternate place names that may have been used in the tale resources. For example, some place names have local variations that are not used on official documents but are well known locally. Also, due to the age of some of the tales, the place names have changed over time, so there may be a need to refer to the location with the name written in the text, but also use it’s official current name, so that users can find the location on a current map. Also, some locations have names in different languages. For example, some locations have names that are used currently, but are also sometimes still referred to using the Cornish language. Not all of the tale resources mention locations or have alternate names, and some take place across multiple locations, so the property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property ex:placesMentioned was used to describe any other locations that were mentioned in the tales, but at which the tale did not actually take place. This was

included as although a tale may not take place at a certain location, the information given about a location, or the way in which it was mentioned, can still be useful for users interested in local culture or history. For example, a tale may mention a nearby town and give some information about it, such as the state of its market or landscape, even though the tale does not take place there. Not all of the tale resources mention locations, and some take place across multiple locations, so the property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property ex:locationLinks was used in order to link other sources of information about the tale locations to the tale resources. The object nodes of this property will have links to webpages that have various information about the places listed, such as Wikipedia pages of the locations. For example, if a tale takes place in a village in Cornwall called St Teath, then it will be linked to the Wikipedia page for that village, which contains more information about the location and it's history. Not all of the locations have their own webpages, and some tales take place across multiple locations, so this property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property ex:locationInfo was used in order to provide information on the tale locations either from GeoNames [26] or by providing the latitude and longitude for locations or landscape features that were too small or rural to be listed on GeoNames. Some locations had the latitude and longitude listed on maps or information for hiking on ordnance survey websites [27]. Some locations without GeoNames information also had listings on Historic England [28], due to the locations being of historical significance, such as ancient burial mounds. These were also listed, where applicable. Not all of the tale resources mention locations, and some take place across multiple locations, so the property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource

The property schema:keywords was used to list any tale elements or words that were not described in the ex:hasPlotPoints but were significant to the tale resources. For example, the animal ‘hare’ is mentioned in a number of tales and this is significant because hares were often linked to the appearance of witches [9]. The appearance of the hare may not be integral to the plot of the tale, but its appearance does signify a particular supernatural element. By listing notable words like this, it may help users looking for tales with particular elements not searchable elsewhere in the ex:hasPlotPoints section. The property is optional and can be repeated. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property schema:mentions was used to describe any people, well-known or named creatures, or events that are mentioned in the Tale Resource. For example, the name of a fair that has taken place in an area for a long period of time, and so has become historically significant. Also, people mentioned by name in the tale resources, or specific creatures that appear often or are so distinct that they have their own local or historical

name. The property is optional and can be repeated. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property ex:mentionedInfo was used to link any other sources of information about the people, creatures or events to the tale resources. The object nodes of this property will have links to webpages that have various information about the subjects listed, such as Wikipedia pages if a person or family has one. For example, one tale mentions the ghost of a 'Lady Ann'. The tale does not mention her last name, but by discovering the family who owned the building in which the tale takes place, and the dates of when the tale was supposed to have happened, we can find information on a man who appears to have been married to two different Anns. Finding information on their children also gave the detail that one Ann was supposed to have been murdered, which corresponds with the story in the Tale Resource of her haunting the building. So it can be assumed that the 'Lady Ann' is possibly 'Lady Ann Strode, neé Wyndham'. In this case, as she doesn't have her own wikipedia page, the pages of her husband [29] and son [30] are linked to the Tale Resource, as they both give some small details about her. Not all of the tales mention specific people or events, so this property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property ex:dateMentioned was used to listed any dates or times that were mentioned in the tale resources. Because many of the tales are old and passed down over generations, the dates are not usually very specific, and are sometimes quite vague, maybe listing just the century in which they happened. On the other hand, some have very specific dates, such as 'Christmas Eve 1483', or 'Easter Monday 1484'. For this reason, a property which uses more specific date information was not used. Not all of the tales mention specific dates, so this property is optional and can be repeated if necessary. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource.

The property schema:identifier was used in order to identify each tale used in the dataset. Each tale was given a sequence of letters and numbers relating to their placement in Briggs' collection. For example, the first tale in the selection of tales about dragons was given the identifier ColPBVolN1dr1 in order to indicate its location in Collection Part B Volume Number 1, dragon tale 1. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. The property is mandatory and is not repeatable.

The property schema:name was used to describe the name of each tale as listed in Briggs' collection. For example, the name of the first dragon tale in the collection is "The Dragon of Kingston". This was used so users can easily recognise the tale as it appears in the collection.The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Tale Resource. The property is mandatory and is not repeatable.

The property schema:pagination was utilised so that users who have access to the printed material would be able to easily find the location of the tales in the collection.

As this information did not need to be searchable to apply to the users' needs, single tales that are printed across multiple pages were contained in one node. For example, pages '8-9', rather than having a separate node for page 8 and page 9 independently. The rdfs:range of the property is xsd:integer and the rdfs:domain is the Tale Resource. The property is mandatory and is not repeatable.

3.3 Creature Resource

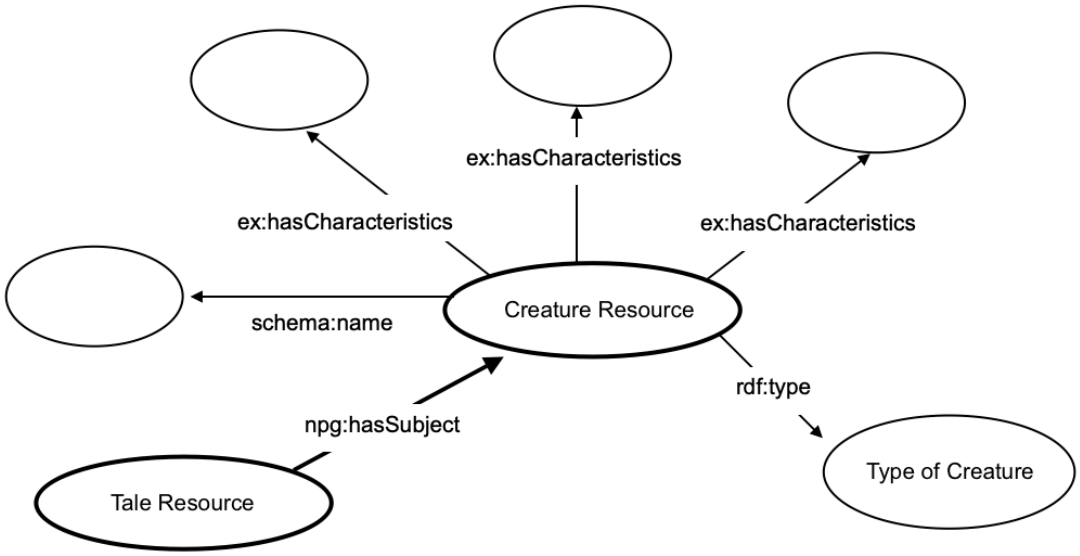


Figure 4. Diagram of the structure of the Creature Resource node, with its properties and how it connect to the Tale Resource and the Type of Creature.

The properties related to the creature resources were established as ex:hasCharacteristics and also the type of the creature. A hierarchy of creature classes was designed, to demonstrate the different relationships between the types of creatures. The node showing the type of creature will link to the hierarchy of creature classes.

The property ex:hasCharacteristics was used to describe the different characteristics and features that each particular creature has. Each Tale Resource has at least one creature appearing in them, and each creature is unique and has different characteristics. However, certain types of creatures may have similar characteristics, so this property was included as it may be useful for users wanting to search for patterns in the creature descriptions.

For example, certain types of fairies have particular characteristics that are common among them, such as Brownies (a type of Fairy) that often disappear after having been given clothes or gifts.

Furthermore, when examining tales about Black Dogs, there are often descriptions and characteristics that appear quite frequently. For instance, the dangerous Black Dogs are often described as having 'eyes the size of saucers' or 'eyes bright as flame'. On the

other hand, Black Dogs that are helpful or act as guardians often look like normal dogs and appear and disappear mysteriously without leaving any trace. Being able to search a folklore database for these specific characteristics may be helpful for users, as certain Black Dogs have different characteristics depending on what area of the country they appear in [25], illuminating how variations of the creatures differ across regions. For example, in folktales from the areas of Cambridge, Essex, Suffolk and Norfolk, it seems that the Black Dogs are often described as just having one eye. Whereas, in the Westmorlands, the Black Dogs often walk around without a head, and on the Isle of Man the dog sometimes shape-shifts into a tall man without a head [25].

A list of specific characteristics can also be useful because some of the characteristics described in the tales are also linked to local customs and rituals, enabling readers to gain some insight into how people thought about society and life at the time. For example, there is a characteristic of a Black Dog appearing from out of the mouth of a fish. In the folktale, a person catches a fish, and then a Black Dog leaps out of the fish's mouth and runs around causing trouble. According to Briggs, this is linked to the custom of not fishing on a Sunday. As the Christian Church decided that Sunday was a day to rest on, people were not supposed to go fishing. Therefore, the characteristic of a Black Dog appearing out of a fish's mouth might have been used to persuade people that fishing on a Sunday was a bad idea, because it could lead to trouble with the supernatural [9].

In this way, it is hoped that by enabling users to search for creatures based on their characteristics, this will be helpful for users wanting to examine aspects of local customs and variations of creatures across different regions. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Creature Resource. The property is mandatory and is repeatable, as the creatures often have multiple characteristics.

The property schema:name is also used to show the individual names of the creatures, if they have been given one. Many of the creatures do not have specific names, and they are just known as a 'Fairy' or a 'Devil'. However, some of them are given specific names in the tales, perhaps due to their notoriety or being well known in a particular area, such as the fairy 'Blue Burches', which seemed to appear often in a number of guises to the villagers in the tale, and was well known for causing mischief. The creatures in the Ghost section also often have specific names, as they are usually the ghost of a deceased person, and their name is sometimes recorded. For example, the ghost of 'Archbishop Richard Scrope', who appears in tale ColPBVolN1gh5 'The Bishopsthorpe Ghost', was supposed to be the ghost of an actual historical figure. As many of the creatures do not have given names, the property is optional, and repeatable. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Creature Resource.

There can be more than one Creature Resource connected to a Tale Resource, as some tales have more than one creature appearing in them, and each of these creatures may have their own class, characteristics, or name. For example, tale ColPBVolN1gh11 'The Chantmarle (Cattistock) Ghost' actually has two ghosts appearing in the tale, the ghost

of ‘Lady Ann’ and the ghost of ‘Wat Perkins’. Each ghost has their own name and characteristics, so it was decided that a Tale Resource could have multiple creature resources linked to it, in order that information could be recorded about each subject separately.

The property rdf:type is also used in order to illustrate what kind of subject is being described in the Creature Resource. The object of this property will be one of the classes from the hierarchy of creature classes (please refer to section 3.4). The rdfs:range of the property is rdf:Class and the rdfs:domain is the Creature Resource. This property is mandatory, as all the creatures in the data set belong to at least one class. The property is also repeatable, as there are some instances where a creature could possibly belong to more than one class if the description of them in the tale was vague or could fit multiple categories.

For example, in tale ColPBVolN1bo1 ‘Black Devil and White Devil’, it is unclear whether the creature is a ‘Bogie’ or a ‘Devil’. The tale appears in the collection under the chapter about Bogies, but the sinister nature of the creature, the fact that it is referred to as ‘the Evil One’ in the text, and that there is a note at the end of the tale that “In most of the English versions the ghost or devil is real”, means there is some ambiguity over which creature class it should belong to. As the aim of this research is to help users find the folktales that are specific to their needs, rather than focusing solely on the classification of the subjects, it was decided that this creature would have an rdf:type of ex:Bogie and ex:Devil. Therefore, in this case, if a user searches for tales containing Devils, this tale would be included in the results, and likewise, if a user searches for tales containing Bogies, the tale would also be included in the results. In this event, the aim would be to help users get access to the tales that are relevant to their needs, and they can then make the decision whether this tale will be of use to them.

Another example of this ambiguity in creature classification are tales ColPBVolN1dr12 and ColPBVolN1dr13. These tales are included in the Dragon section of the collection, however, it is actually specified in the text that the dragons are the devil transformed into a dragon. In this case, although it seems that this creature is not an actual dragon, but a shapeshifting devil, it was decided that these creatures would be classified as both ex:Dragon and ex:Devil, as they are included in the Dragon chapter of the collection.

There are also examples of tales that are present in the Ghost section of the collection, but the creatures in them are very similar to Brownies, a type of Fairy. For example, in tale ColPBVolN1gh9, ‘The Cauld Lad O’Hylton: I’, it states in the text that the creature is a “*Ghost* or *domestic spirit,* of the brownie or Robin Goodfellow genus” [9]. Briggs’ footnotes on that tale also state that this creature is “equally poised between that of a ghost and a brownie” as it performed tasks and disappeared when given gifts (like a Brownie), and because brownies and fairies are also “supposed to be ghosts of the dead” [9]. Tale ColPBVolN1gh10 ‘‘The Cauld Lad O’Hylton: II’ is also similar, in that the text states that the creature “was said to be the ghost of a stable boy” and that “He worked like a Brownie”. As there were so many similarities between the two ghosts and the characteristics of brownies, and the endnotes made by Briggs for tale ColPBVolN1gh10,

it was decided that these two creatures would also belong to two classes, ex:Ghost and ex:Fairy.

On the other hand, there are sometimes instances where the same tale will appear twice in the collection, but under different chapters and with different motifs assigned to it. For example, the tale 'Billy Gray and the Phantom Dog' appears in the collection both in the chapter about Ghosts as ColPBVolN1gh4, and in the chapter about Devils as ColPBVolN1dev4. This may be due to the description of the creature seeming to fit the pattern of a tale about hauntings, with the protagonist being haunted after a family member has died, and also fitting the pattern of a tale about the devil punishing someone for deeds seen as wicked. Although they are both the same tale, and previously printed in the same source, different motifs have been assigned to the tale in each chapter. Due to this, rather than the tale only appearing in the dataset once, and the creature type being listed as two different classes, it was decided that this tale would appear twice in the dataset, in the same way that it appears twice in the printed collection. The first appearance in the dataset will be with the creature classified as a ex:Devil, and the second time with the creature classified as ex:Ghost. In this way, if a user searched to find tales about ghosts, this tale would still appear in the results, likewise with users searching for tales about devils.

3.4 Hierarchy of Creature Classes

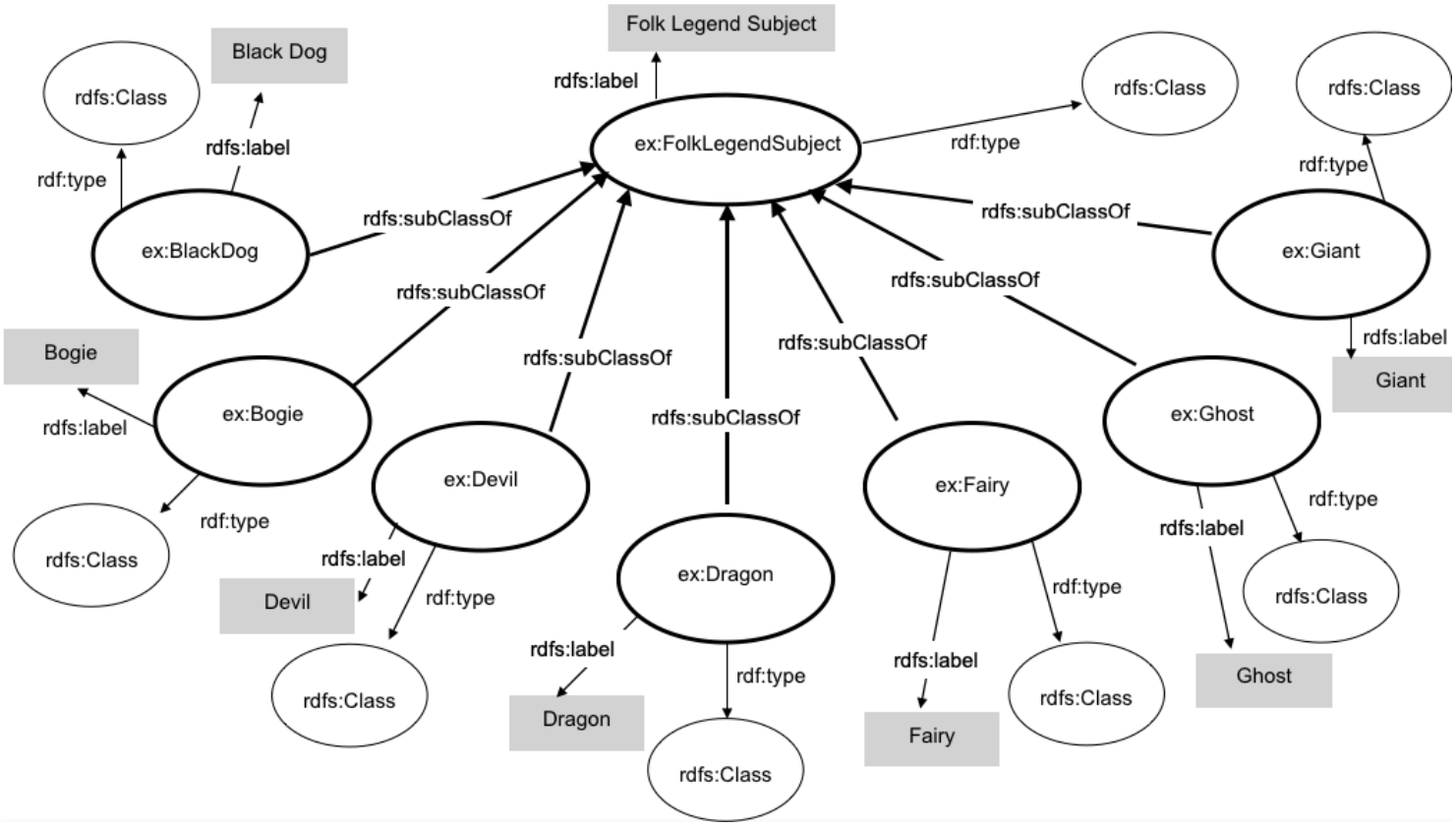


Figure 5. Diagram of one level of the hierarchy of classes for the creatures. The different creatures are all subclasses of the superclass Folk Legend Subject.

The hierarchy of classes for the creatures was designed to indicate how the different creatures are related, and aid in resource discovery for the users. If a user wanted to search for tales specifically about a particular creature, or a subtype of a particular creature, then the hierarchy of classes would enable this, allowing the results to be narrowed to what a user requires.

In this dataset, the subjects that appear in Part B Folk Legends Volume 1 of the collection are: Black Dogs; Bogies; Devils; Dragons; Fairies; Ghosts; and Giants. Each of these subjects has an rdf:type of rdfs:Class. They are also all ex:subClassOf the superclass of ex:FolkLegendSubject.

The ex:BlackDog, ex:Bogie, ex:Dragon, ex:Fairy and ex:Ghost subject classes also have subclasses of their own. However, it was decided that the ex:Devil and ex:Giant classes would not have any subclasses of their own, as a way to further divide the classes was not determined. The differences between the instances of creatures belonging to the ex:Devil and ex:Giant classes did not seem great enough to warrant whole separate subclasses, and no preexisting groups or types (such as Brownies and Hobgoblins being different types of fairies) were found. Rather, the differences between the instances of the ex:Devil and ex:Giant classes seemed to be slight differences in characteristics, rather than distinct classes. The Giants in the collection range in size from being ten feet tall to having “a stride of six miles” and are often written as the creators of various features in the landscape [9]. The Devils in the collection are described as “formidable”, and although it is possible to fool them, often the stories have more sinister aspects, with the devil hunting souls, fighting against magicians and various religious people, seducing women to claim their souls, and shape changing [9].

3.4.1 Black Dogs

There are also more subclasses within the ex:BlackDog class. Black Dogs are found across the UK, and are usually dangerous or malevolent creatures that people in the tales generally do not want to come across. However, they can also be helpful, as there are instances where the Black Dog has helped travellers or people who are in danger or have lost their way [31].

Brown has identified three different types of Black Dog, those belonging to Group A, Group B, and Group C [25]. Group A dogs have many local names (such as Bargest or Barguest, Padfoot, Shuck, Trash and Striker), and have the ability to change their shape, are ominous and malevolent, and many are often connected to burial sites or graveyards [25]. Briggs states that this connection to cemeteries was due to a custom where a black dog was buried at a graveyard that was newly consecrated, so that the dog would have the responsibility of guarding the graveyard. This responsibility would have otherwise fallen on the first person to be buried there. This means that the dog may appear after death as a ‘Grim’, haunting the graveyard and keeping passersby away [31]. Group B dogs are known simply as a Black Dogs, are often mistaken for real dogs, and are associated with specific places or families. They can also be linked to witches [25].

Group C dogs appear at certain times of the year, with a “calendar cycle” and are also associated with specific places. Group C is the rarest group [25].

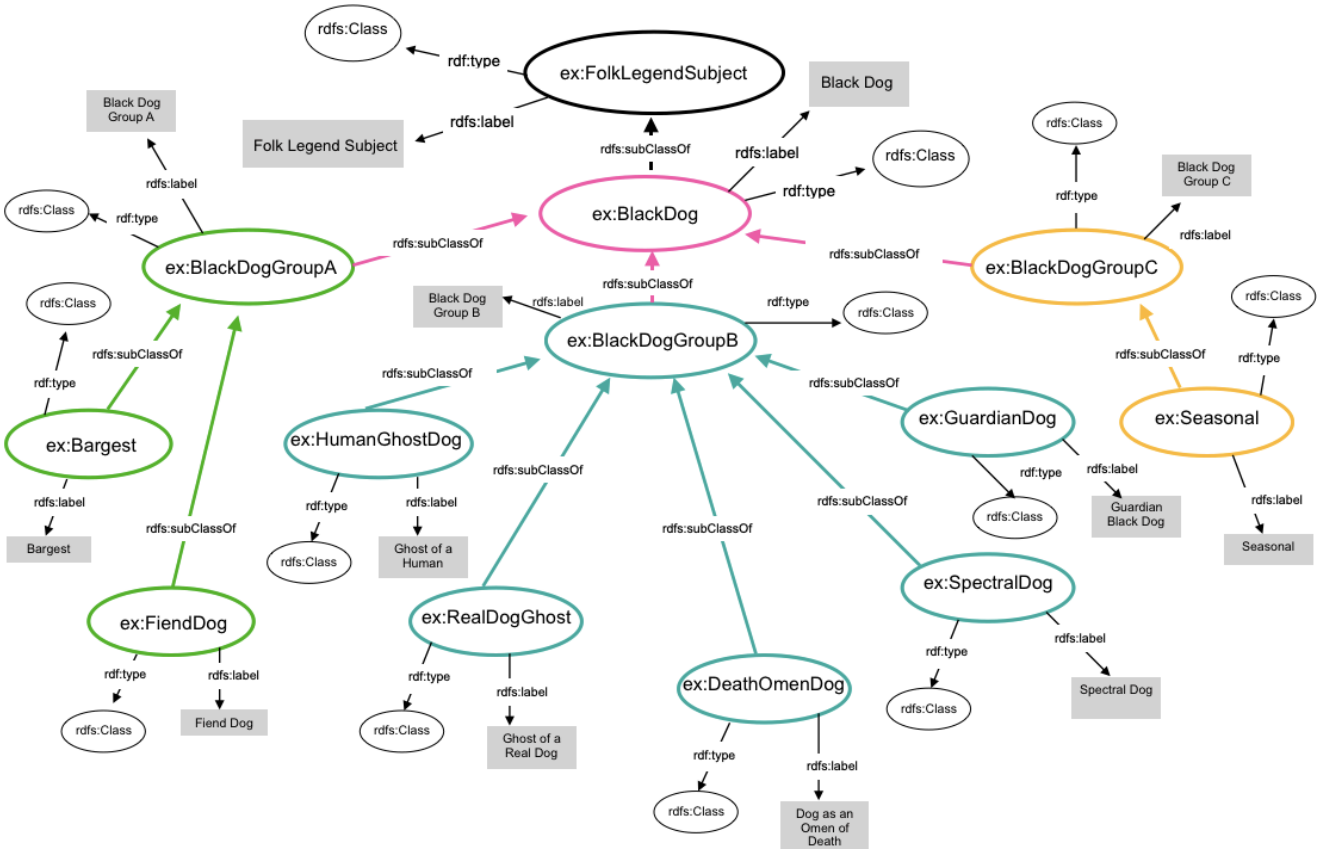


Figure 6. Diagram of the different Black Dog Groups and the subclasses of Black Dogs that belong to each group.

These groups, ex:BlackDogGroupA, ex:BlackDogGroupB, and ex:BlackDogGroupC, were created as subclasses of the ex:BlackDog class. As there were various different kinds of Black Dogs within these groups, subclasses of these groups were also created to further distinguish the different types of Black Dogs. These are: ex:Bargest and ex:FiendDog, which are subclasses of Group A; ex:HumanGhostDog, ex:RealGhostDog, ex:DeathOmenDog, ex:SpectralDog, and ex:Guardian Dog, which are subclasses of Group B; and ex:Seasonal, which is a subclass of Group C.

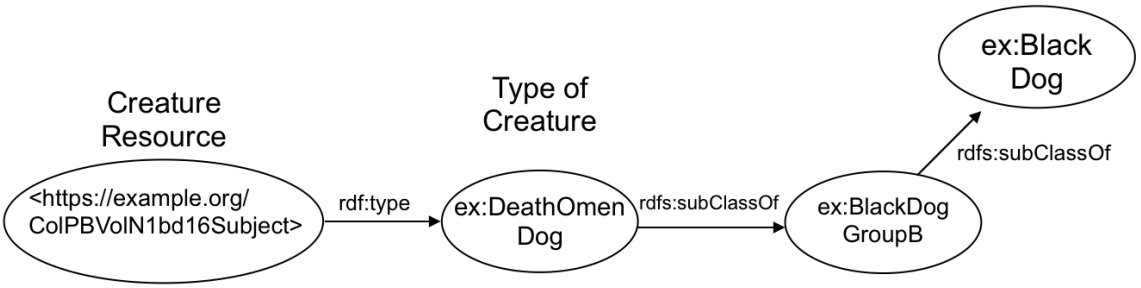


Figure 7. Diagram of the ex:BlackDog subclasses of ex:BlackDogGroupB and ex:DeathOmenDog and how they connect to the Creature Resource.

These subclasses can connect to the creature resource through the triple of the Creature Resource, property rdf:type, and the relevant ex:BlackDog subclass in the 'Type of Creature' node.

3.4.2 Bogies

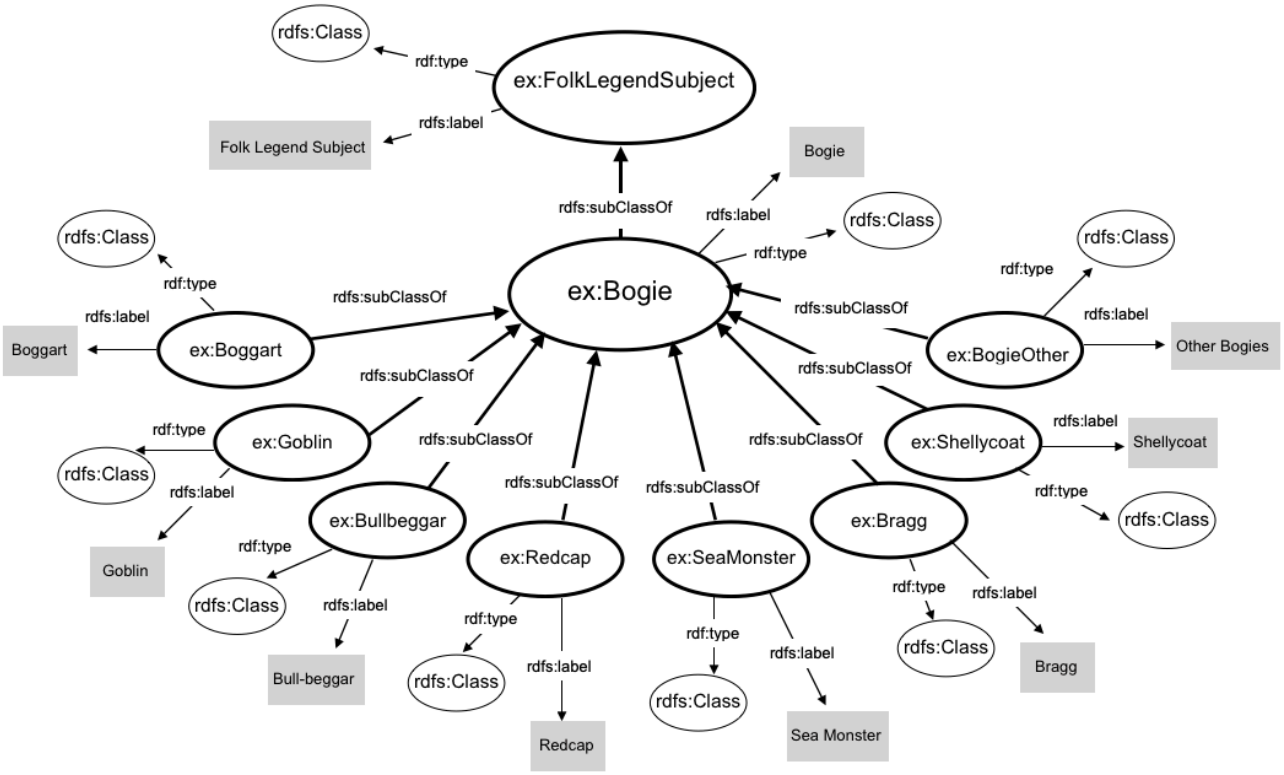


Figure 8. Diagram of the ex:Bogie class and its subclasses, and how they connect to the superclass ex:FolkLegendSubject

Bogies are spirits that enjoy causing mischief and tormenting people and, with some exceptions, are generally not particularly harmful [31]. They can be considered "minor devils, if they are devils at all" [9]. There are many different types of bogies, so for the purpose of this research, only the bogies that will appear in the tales contained within the dataset will be in the designed data model, as otherwise there would be too many different classes with no instances of them within the dataset.

Within the ex:Bogie class, the subclasses that will be used to distinguish the different types of Bogies in this research will be: ex:Boggart; ex:Goblin; ex:Bullbeggar; ex:Redcap; ex:SeaMonster; ex:Brag; and ex:Shellycoat. There is also the subclass of ex:BogieOther, as some tales do not specify which type of Bogie is appearing, or the description is vague. The ex:BogieOther subclass was created so users could search separately for any unspecified Bogie creatures, with the other types of Bogies not being included in the results.

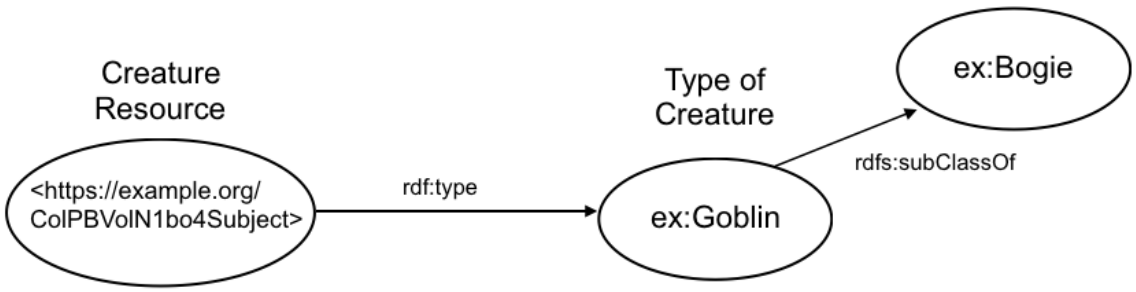


Figure 9. Diagram of how the ex:Bogie subclass of ex:Goblin connects to the Creature Resource.

These subclasses can connect to the creature resource through the triple of the Creature Resource, property rdf:type, and the relevant ex:Bogie subclass in the 'Type of Creature' node.

3.4.3 Dragons

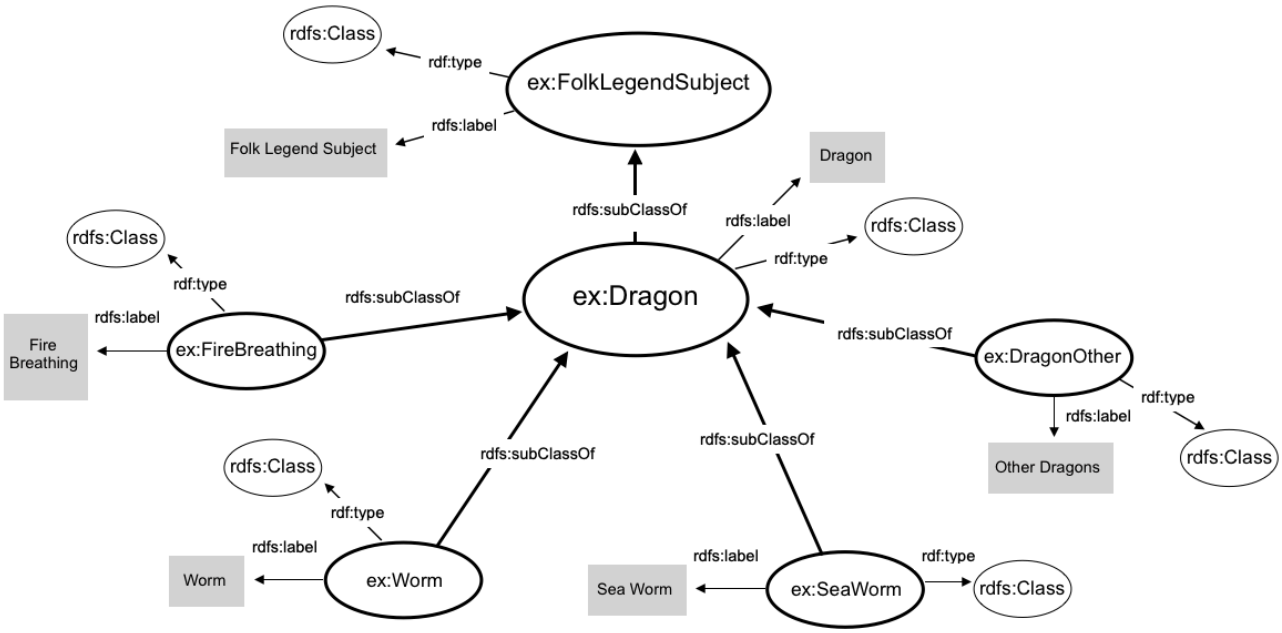


Figure 10. Diagram of the different subclasses that belong to the ex:Dragon class, and how this is connected to the ex:FolkLegendSubject superclass.

In Folk Legends, it seems that dragons that resemble snakes or worms are more common than dragons that have wings and breath fire [9]. Therefore, the subclasses within the ex:Dragon class are: ex:FireBreathing; ex:Worm; ex:SeaWorm; and ex:DragonOther. The ex:DragonOther was decided on as some tales do not specify which type of Dragon is appearing, and the physical description of the creatures is not detailed enough to conclude which type it may be. This ex:DragonOther subclass enables users to be able to search separately for any unspecified Dragon creatures, with the other types of Dragons not being included in the results.

3.4.4 Fairies

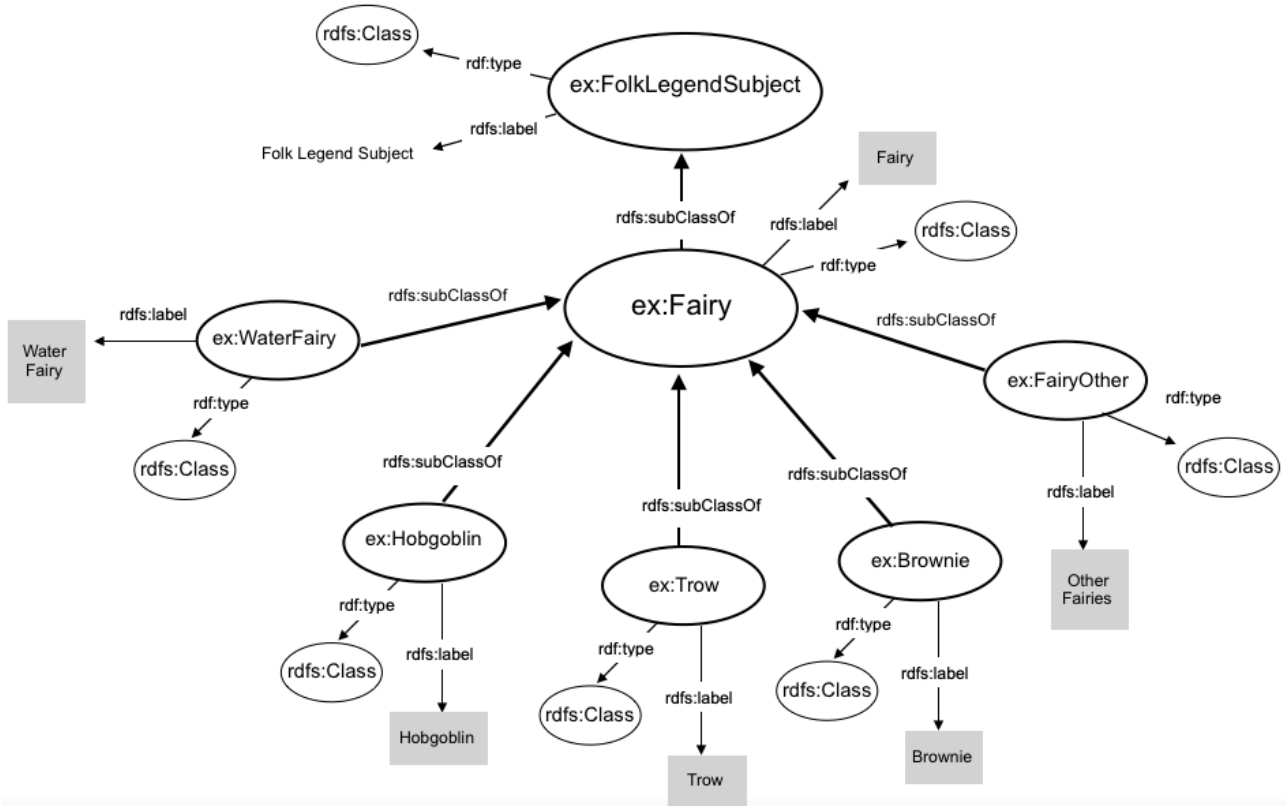


Figure 11. Diagram of the different subclasses that belong to the ex:Fairy class, and how this is connected to the ex:FolkLegendSubject superclass.

Fairies are supernatural creatures that appear in many folktales, and have various abilities and characteristics. In the collection used for this research, Briggs states that the chapter on Fairies "has been made a kind of hold-all for all supernatural beings who are not devils, bogies or angels", as distinguishing between various types of fairies and other nature spirits can be difficult [9].

Only the fairy types that have been used in the dataset set are recorded in the data model, as otherwise, there would have been too many types that would not have had any instances. The chapter on fairies was one of the biggest in the collection, spanning 240 pages, so only a small selection of tales from this chapter were able to be included in the dataset. If this research was to be expanded in the future, then more examples from this chapter would necessitate that the data model be edited in order to provide for more subclasses of different fairies.

The subclasses within the ex:Fairy class for this research are: ex:WaterFairy; ex:Hobgoblin; ex:Trow: ex:Brownie; and ex:FairyOther, which covers all other examples of fairies that were not clearly classified.

3.4.5 Ghosts

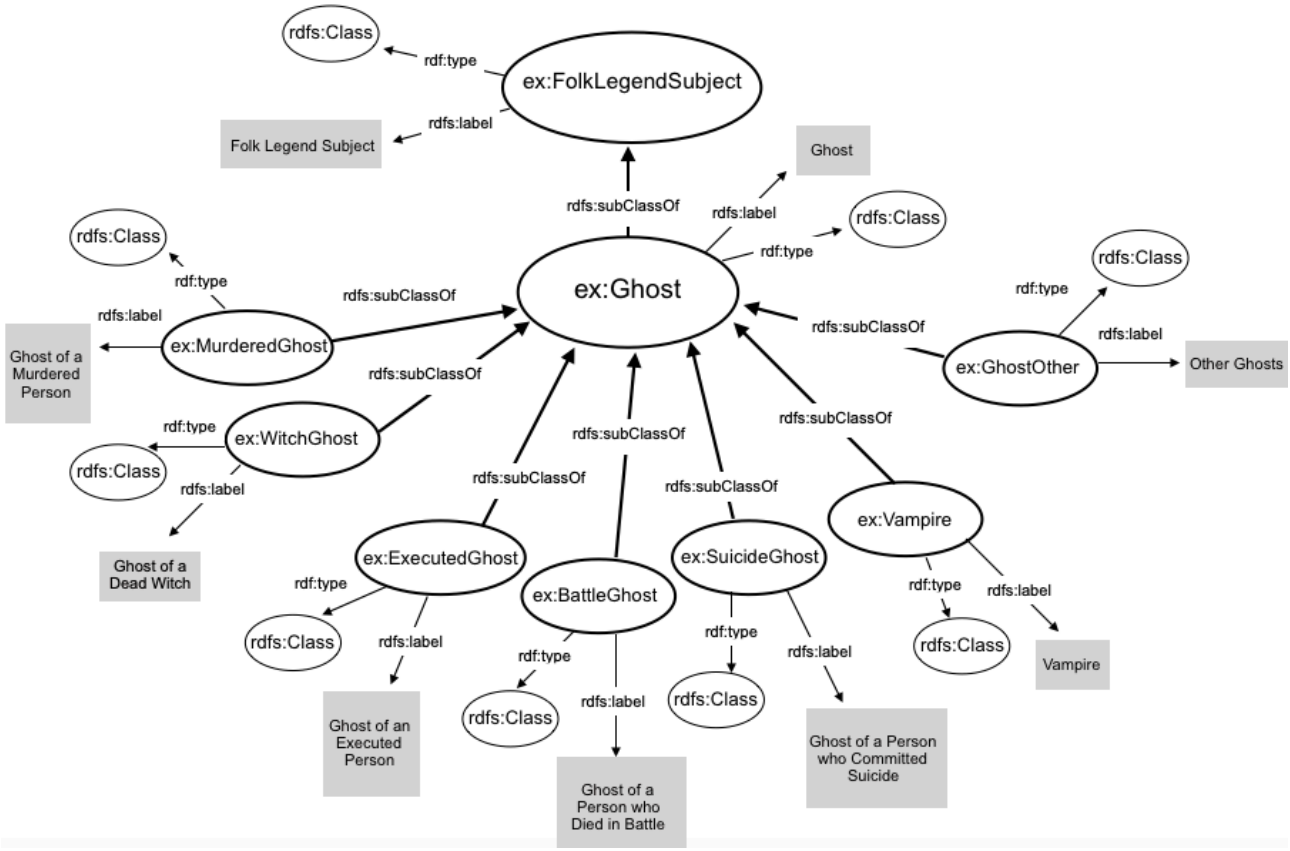


Figure 12. Diagram of the different subclasses that belong to the ex:Ghost class, and how this is connected to the ex:FolkLegendSubject superclass.

The section on ghosts in the collection is also one of the larger chapters, with 192 pages. As such, the different ghost subclasses for the data model have only been created from the tales which are used in the dataset, as the tales were too numerous to use all of them in this research.

Briggs states that “Hauntings can generally be divided into those of people who have done wrong and of people who have suffered it” [9]. Therefore, it was decided to use this idea as the basis for creating the different subclasses, separating the tales into different sections based on the manner of the ghost’s death. Other subclasses such as ex:Vampire were included, as vampires are considered a distinctly different type of ghost, or reanimated corpse. They are considered rare in England, where the collection of tales used in the dataset come from [9]. The subclasses in the ex:Ghost class used for this research are: ex:MurderedGhost; ex:WitchGhost; ex:ExecutedGhost; ex:BattleGhost; ex:SuicideGhost; ex:Vampire; and ex:GhostOther.

3.5 Source Resource

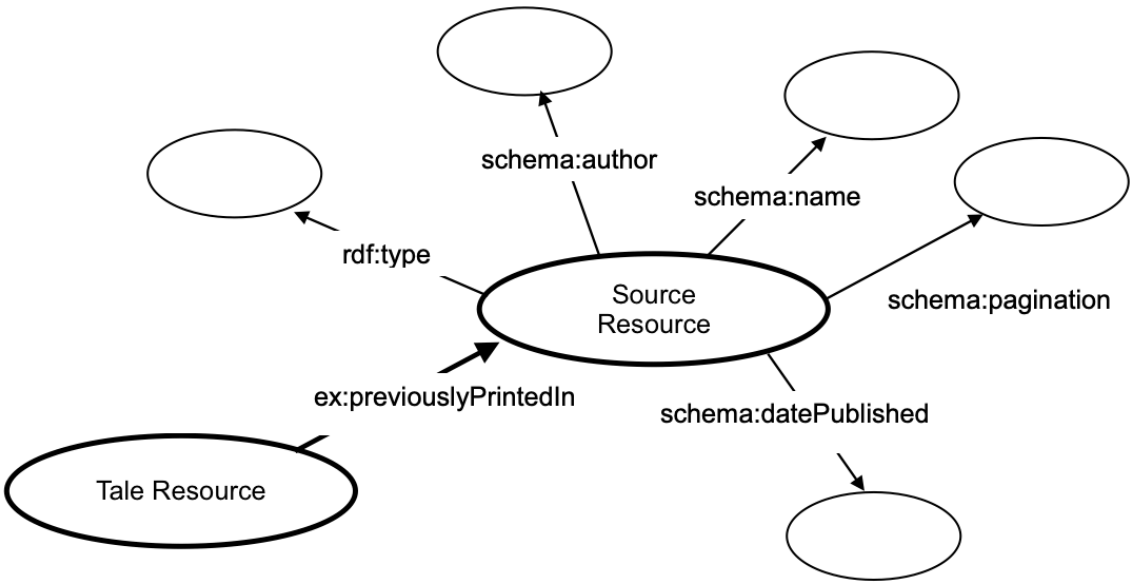


Figure13. Diagram of the structure of the Source Resource node, with its properties and how it connects to the Tale Resource.

Each Tale Resource was also printed in other sources before being collected in Briggs' folktale collection. The sources that the tales were previously printed in will be recorded in a separate resource, the Source Resource, that connects to the Tale Resource. The Source Resource will contain information about the book or material that each Tale Resource was previously printed in. This source information might not be the original written sources of the tales, but they will be the versions that were used in the collection by Briggs. On occasions, the sources might be from collections of tales that were newly collected at the time, or had not yet been published by the time Briggs' collection was being printed, such as Ruth L. Tongue's collection 'Forgotten folk tales of the English counties' [32]. In this case, information such as page numbers might be missing from the collected information.

It was decided that this information would be included as it may be useful for users wanting to discover information about where a tale had previously been sourced from. As the sources seem to often be other collections of folktales, some specific to particular areas (like collections based on tales found in Somerset or the Highlands), then it was thought that this information would be relevant to users wanting to study local customs or local legends, because these sources may have other tales that may be of interest to the users about a particular area. Furthermore, it may be useful for users wanting to know more about the author of a collection that a tale was previously printed in, as the author can affect the way in which a tale has been edited or collected. For example, Briggs states that tale ColPBVolN1gh15 'Croglin Grange' about a vampire "like other stories collected by Augustus Hare, may owe something to literary reminiscence" [9].

Due to this, users might find it useful to be aware that this tale, and others collected by the same author, may be further removed from the practice of passing tales on verbally than others in the collection. Also, Briggs states that tales recorded by Robert Hunt contain much ornamentation, “he is often prolix in his style”, and so it can be difficult to tell how accurate some of the details recorded in his tales are [9]. Therefore, the information about where a tale has been selected from, and who recorded, edited or collected the previous source, might be important information for the user, so that they can be aware of the biases or literary tendencies of the author, that may also have been passed onto the tales. The properties related to the source resource were established as: rdf:type; schema:author; schema:name; schema:datePublished; and schema:pagination.

The property rdf:type describes what kind of source the Tale Resource was previously printed in, such as whether the Tale Resource was previously printed in a book or a newspaper. The rdfs:range of the property is rdf:Class and the rdfs:domain is the Source Resource. The property is not mandatory, as sometimes it is not clear whether a source was actually a book, or a pamphlet, or a series of hand written notebooks. The property is not repeatable.

The property schema:author describes the author of the source material in which the Tale Resource was previously printed in before becoming part of Briggs’ collection. This property is included so that users are able to search for the materials if they so desire. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Source Resource. The property is mandatory and is repeatable. Alternately, if the Source Resource does not have an author (for example, if the collection was not written but compiled or edited), properties such as schema:creator can be used instead.

The property schema:name was used to describe the name of each book or material that is a Source Resource. For example, the name of the book which some of the tales about Black Dogs were sourced from is called “Goblin Tales of Lancashire” [33]. This was used so users can find the books or materials themselves if they wish to examine the collections in which the tales were previously printed in. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Source Resource. The property is mandatory and is not repeatable.

The property schema:pagination was utilised so that users who have access to the printed source materials would be able to easily find the location of the tale they were examining. As this information did not need to be searchable to apply to the users’ needs, single tales that are printed across multiple pages were contained in one node. The rdfs:range of the property is xsd:integer and the rdfs:domain is the Source Resource. The property is not mandatory and is not repeatable.

The property schema:datePublished was also included so that users could use the information to find the books or materials in which the tales were previously printed in.The rdfs:range of the property is xs:date and the rdfs:domain is the Source Resource. The property is not mandatory and is not repeatable.

3.6 Collection Resource and Hierarchy

Figure 14. Diagram of the structure of the Collection Resource with its properties, and how this connects to the Collection Hierarchy and Tale Resource.

The Collection Resource is included so that users will know where the tales have come from in the collection, and to provide bibliographic information about the collection. The properties related to the Collection Resource were established as: rdf:type; schema:creator; schema:name; schema:publisher; schema:datePublished; schema:isbn;

schema:bookEdition; ex:collectionPart; schema:PublicationVolume; and schema:Chapter.

The property rdf:type will give general information about the collection, that it is a schema:Collection. The rdfs:range of the property is rdf:Class and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

The property schema:name will give general information about the name of the collection, "A Dictionary Of British Folk-Tales". The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

The property schema:publisher will give general bibliographic information about the collection. This is to aid users in searching for a copy of the collection if they wish. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

The property schema:datePublished will also give general bibliographic information about the collection. This is to aid users in searching for a copy of the collection if they wish. The rdfs:range of the property is xs:date and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

The property schema:creator will give information on who created this collection, enabling users to search for other related works by the creator. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

The property schema:isbn will give general bibliographic information about the collection. The rdfs:range of the property is xsd:string and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

The property schema:bookEdition will give general bibliographic information about the collection, and make the user aware of which edition is being used in the dataset, in case there are variations between the different editions. The rdfs:range of the property is rdfs:Literal and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

As the collection is separated into multiple sections, a hierarchy was created for the collection resource. The nodes ex:PartAFolkNarrative and ex:PartBFolkLegends use the property schema:isPartOf to connect to the main collection resource. The rdf:type ex:collectionPart was also used for these nodes, and gives information on what section of the collection they are. The rdfs:range of this property is rdf:Class and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

Each ex:collectionPart is then split into 2 sections, ex:PartAVolume1, ex:PartAVolume2, ex:PartBVolume1 and ex:PartBVolume2. These nodes use the property schema:isPartOf

to connect to the ex:collectionPart that they belong to. The rdf:type schema:PublicationVolume will give information on what section of the collection they are. The rdfs:range of the property is rdf:Class and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

ex:PartBVolume1 is then separated into 7 sections, ex:BlackDogs, ex:Bogies, ex:Devils, ex:Dragons, ex:Fairies, ex:Ghosts and ex:Giants. These are the chapters that Part B Volume 1 of the collection are divided into. These nodes use the property schema:isPartOf to connect to ex:PartBVolume1. The rdf:type schema:Chapter gives information on what section of the collection they are. The rdfs:range of the property is rdf:Class and the rdfs:domain is the Collection Resource. The property is mandatory and is not repeatable.

Only part of the whole collection was used in the dataset, so only the chapters that had tales used in the dataset were included in the data model. In the future, if further research is carried out, then the number of tales included in the dataset would be expanded to include tales from different volumes or collection parts. The data model could then be edited to include how the chapters and other parts of the collection relate to each other.

The Tale Resource connects to the appropriate chapter with the property schema:isPartOf. In Figure 14 the Tale Resource featured in the example belongs to the chapter on Black Dogs, so it is connected to ex:BlackDogs.

4. Dataset

The dataset was built from a selection of folktales from Katharine M. Briggs' collection "A Dictionary of British Folk-Tales". The collection consists of two parts, Part A Folk Narratives, and Part B Folk Legends, with each of these parts consisting of two volumes. For this research, 109 tales from Volume I of Part B Folk Legends [9] were used as the dataset.

It was decided that tales specifically from the Folk Legends section would be used, rather than the Folk Narratives section. Folk Narrative can be considered "Folk Fiction, told for edification, delight or amusement" [34], such as entertaining fairy stories told for enjoyment. On the other hand, Folk Legends are distinct in that "Folk Legend was once believed to be true" [34], such as superstitions, beliefs, or occurrences that the narrator or acquaintances of the narrator have supposedly witnessed or experienced. They can often include verifiable details, such as the names of real places or people, or exact dates and times. The tales may not be believable today, or may have been embellished greatly due to continual retellings and be "coloured by the imagination of countless storytellers" [15]. This may have happened so much that it may be difficult to decide which details were actually believable at the time, but nevertheless, "they all purport at least to recount things that actually happened, and are founded on beliefs that were actually held" [9].

It was stated previously that the ATU system is more commonly used for Folk Narratives rather than Folk Legends, as the Type Index "does not purport to deal with legendary matter" [9]. Due to this, the ATU categories may not always be relevant or applicable for these tales. Briggs also uses The Migratory Legends [15] Type Index throughout the volume to assign tale types. However, she describes this as a "rather skeletal work" that "deals chiefly with Scandinavian matter" [9] and so many of the tales in the volume are not assigned a tale type.

Therefore, as the Type Index is unable to be satisfactorily used on such tales, it seemed that examining Folk Legends, and ways to make the discovery of them easier for users, would be worthwhile. Specifically examining Folk Legends and enabling users to discover tale resources that are relevant to their needs could contribute towards making folktales related to social and local history more accessible to general users.

The selection of 109 tales were taken from each of the 7 different chapters of Part B Volume I of the collection, each covering a different subject matter: Black Dogs; Bogies; Devils; Dragons; Fairies; Ghosts; and Giants. Tales from each chapter were included in order to encompass a broad range of subject matter. The tales were analysed for information that would fit the proposed data model, such as: the location the tale takes place in; which creature subjects appear in each tale; the narrative structure and plot points that the tale is comprised of; the characteristics each of the creatures have; the motifs assigned to the tales from the Motif Index; any named entities that are mentioned, such as names of people or events; the source that the tale was previously printed in; and general information such as the name of the tales and where they are located in the collection. The aim was that the information extracted and recorded would be beneficial for users when searching for tales with specific characteristics.

An example instance of the information recorded for one of the tales in the data set is as follows.

Tale Resource: <https://example.org/ColPBVolN1bd9>
Identifier: ColPBVolN1bd9
Name: The Dog of Roslin
Pagination: 11-12
Type Of Tale: Folk Legend
Type Index: none assigned
Motif Index: Thompson, E.521.2 Ghost of a dog
Location: Roslin Castle, Midlothian, Scotland, UK
Location Alternate Name: none
Places Mentioned: none
Keywords: castle; battle; phantom; hound; soldier; dog
Plot Points: battle occurs; murder committed at location; animal killed; animal dies; subject appears; subject seeks vengeance; haunting; death occurs; death avenged; subject disappears; subject haunts location

Mentions: Mauthe Dog
Date Mentioned: February 24th, 1302 CE

Creature Resource: <https://example.org/ColPBVolN1bd9Subject>
Subject: Black Dog
Type of Creature: Black Dog Group B, Ghost of a real dog
Characteristics: baying; large; ghostly; fierce; snarls; bays
Name: Mauthe Dog

Source Resource: <https://example.org/MoreHighlandFolktales>
Type: Book
Name: More Highland Folktales
Pagination: 36
Author: R. MacDonald Robertson
Date Published: 1965

Collection Resource: <https://example.org/ADictionaryOfBritishFolk-Tales>
Name: A Dictionary of British Folk-Tales
Type: Collection
Chapter: Black Dogs
Publication Volume: Volume I
Collection Part: Part B Folk Legends
Creator: Katharine M. Briggs
Publisher: Routledge & Kagan Paul
Date Published: 1971
ISBN: 0 7100 6364 4
Book Edition: 1st

Additional information about the tales and details mentioned in them, such as locations or names, were researched. Appropriate links to more information about these details were then added to the data set, so that users could have access to further information about these points. For example, the extra information provided about the example tale given above would be:

Motif Information: https://sites.ualberta.ca/~urban/Projects/English/Content/e.htm
Location Links: https://en.wikipedia.org/wiki/Roslin_Castle
Location Information: https://www.geonames.org/maps/wikipedia_55.8526_-3.1599.html
https://www.geonames.org/2642559/midlothian.html
Mentioned Information: https://en.wikipedia.org/wiki/Moddey_Dhoo

In this way, information relevant to the user needs was gathered for each of the 109 tales in the data set.

4.1 Conversion of dataset into RDF

Firstly, the dataset was converted using Turtle syntax [35] into a format that could express and store the information gained from the Folk Legends as RDF data. Poorman's Linked Data Toolkit [36] was used in order to convert the RDF data into Turtle format from an excel document. Turtle was also manually written when appropriate, such as when creating the creature type hierarchies.

An example of the Folk Legends data for one Tale Resource in Turtle is as follows:

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://example.org/> .
@prefix schema: <http://schema.org/> .
@prefix npg: <http://ns.nature.com/terms/hasSubject> .

<https://example.org/ColPBVolN1bd1> a ex:Tale;
schema:identifier "ColPBVolN1bd1";
schema:name "The Bargest of the Troller's Gill";
schema:pagination "4";
ex:typeofTale "Folk Legend";
ex:motifIndex "Thompson, G303.3.3.1.1. Devil in form of dog";
ex:motifInfo <https://sites.ualberta.ca/~urban/Projects/English/Content/g.htm>;
schema:location "Troller's Gill Ravine";
schema:location "Appletreewick";
schema:location "Yorkshire";
schema:location "England";
schema:locationAltName "The Gordale of Appletreewick" ;
ex:locationLinks <https://en.wikipedia.org/wiki/Trollers_Gill>;
ex:locationInfo <https://www.geonames.org/maps/wikipedia_54.053_-1.8976.html>;
ex:locationInfo <https://www.geonames.org/2657225/appletreewick.html>;
schema:keywords "yew tree" ;
schema:keywords "ravine" ;
schema:keywords "haunting" ;
schema:keywords "drawing on ground" ;
schema:keywords "charms" ;
schema:keywords "hound" ;
schema:keywords "whirlwind" ;
schema:keywords "fire" ;
ex:hasPlotPoints [
        a rdf:Seq;
        rdf:_1 "subject haunts location" ;
        rdf:_2 "search for subject" ;
        rdf:_3 "warning given" ;
        rdf:_4 "warning goes unheeded" ;
        rdf:_5 "subject summoned" ;
```

```
        rdf:_6 "subject disturbed" ;
        rdf:_7 "subject appears" ;
        rdf:_8 "storm appears" ;
        rdf:_9 "subject controls weather" ;
        rdf:_10 "death occurs" ;
        rdf:_11 "killed by subject" ;
] ;
ex:previouslyPrintedIn: <https://example.org/
YorkshireLegendsAndTraditions1stSeries> ;
npg:hasSubject  <https://example.org/ColPBVolN1bd1Subject> .
```

The dataset was then input into Apache Jena Fuseki [37], to store the RDF data. This data could then be tested to determine whether it would be suitable for the user needs by creating queries using SPARQL query language [38]. The number of triples when converted into RDF was 7801

5. Evaluation

5.1 User Needs

It was decided that the effectiveness and functionality of the data model and dataset would be tested by examining whether it was suitable for user needs. Information that a user might want to obtain or search for in a collection of folk legends was considered when designing the data model and building the dataset. These considerations also formed the basis of what the user needs would be. For example, users who want to find tales that meet their specific needs, but find it difficult to narrow down which tales to browse when searching a large collection.

A list of speculative user needs and use cases were defined in order to discover whether the dataset and data model would be able to fulfil the user needs.
The functional requirements for the user needs were:

1. Location - Searching by location, so users can narrow down their search results to a particular area of the country. This may be useful for discovering tales in users' local areas, and researching local history.

2. Creature subject- Searching for tales that contain a particular creature as the subject matter. This may be useful for users searching for all examples of tales with a particular creature, enabling them to find differences and similarities between subjects across different areas.

3. Keywords - Searching for tales that have certain keywords listed in them. This may be useful for users who want to search for tales about specific details, that are not contained in other sections, such as the plot points or characteristics sections, but are nevertheless important details. For example, searching for the keyword "cross"

or “monk” to investigate how religious customs were related to the supernatural.

4. Plot Point - Searching for tales that all contain a certain plot point or narrative structure. This may be useful for users that want to find all tales where a specific narrative event occurs, such as “subject changes shape/size” to find which creatures use the power of transformation in the tales. Also, searching for the plot point “supposed explanation for landscape feature” in order to find examples of the subjects creating landscape features, as this may be of interest for users investigating local history or geography. Furthermore, “cross protects against subject” in order to examine how customs developed related to Christianity and ways of guarding against the supernatural.

5. Motif Index and Tale Type - Searching for tales that are assigned specific types and motifs from the well established existing indices. This may be helpful if some users are already familiar with existing classification methods, and will enable them to search the dataset with these methods in conjunction with the newly proposed structured categories.

6. Creature subclass - Searching for a specific subclass of a creature. This may be helpful for users only wanting to examine a very particular type of creature. For example, searching only for Guardian Black Dogs, rather than the more common frightening types of Black Dogs, or searching only for Brownies, rather than other common types of Fairies.

7. Creature Characteristics - Searching for creatures with particular characteristics. For example, if users wanted to search for Fairies that “does housework”, or Dragons that “eats humans”. Creatures with particular characteristics can often appear in particular parts of the country, so this may be useful for users examining regional differences of the creatures.

8. Combining searches - It was also decided that it would be useful if users were able to combine searches in order to make the results more specific. For example, combining the Creature subject and Keywords search. Users could then search the Keywords for tales which mention “treasure”, then narrow the search further by limiting the results to only contain tales with the Keyword “treasure” when the tale also features only a “Dragon” as the Creature subject.

It was decided that a search for temporal information would not be included in the speculative user needs, as dates are not provided particularly often in the tales used for the dataset. When dates are provided, they can often be quite vague, such as ‘two or three centuries ago’.

It was decided that searching for Volume Number or Collection Part would also not be included in the speculative user needs, as for this research only Part B Volume I was used in the dataset. However, if the research was expanded in the future, this might also

be a useful search function for users wanting to find tales from a specific volume of the collection.

5.2 SPARQL Results

After having decided on some user needs, these queries were written using SPARQL query language, so that the dataset in Apache Jena Fuseki could be queried.

The prefix used were:
PREFIX schema: <http://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>

The results of the aforementioned speculative user needs are as follows:

1. Firstly, a query was made searching for a specific location. Tales with schema:location as “Devon” were searched for.

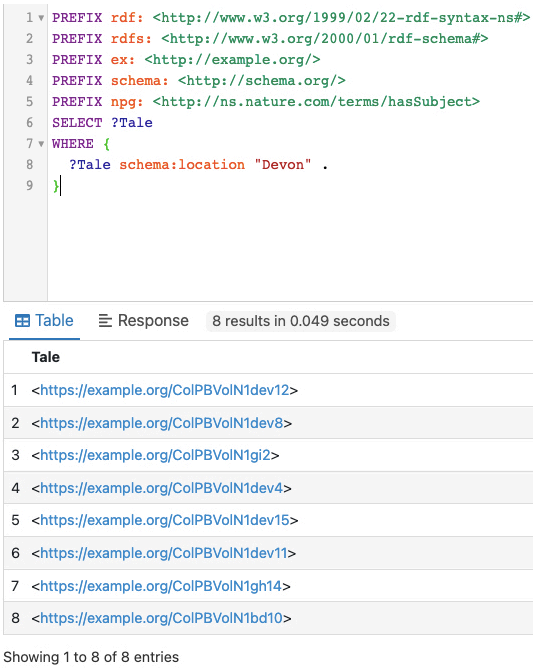


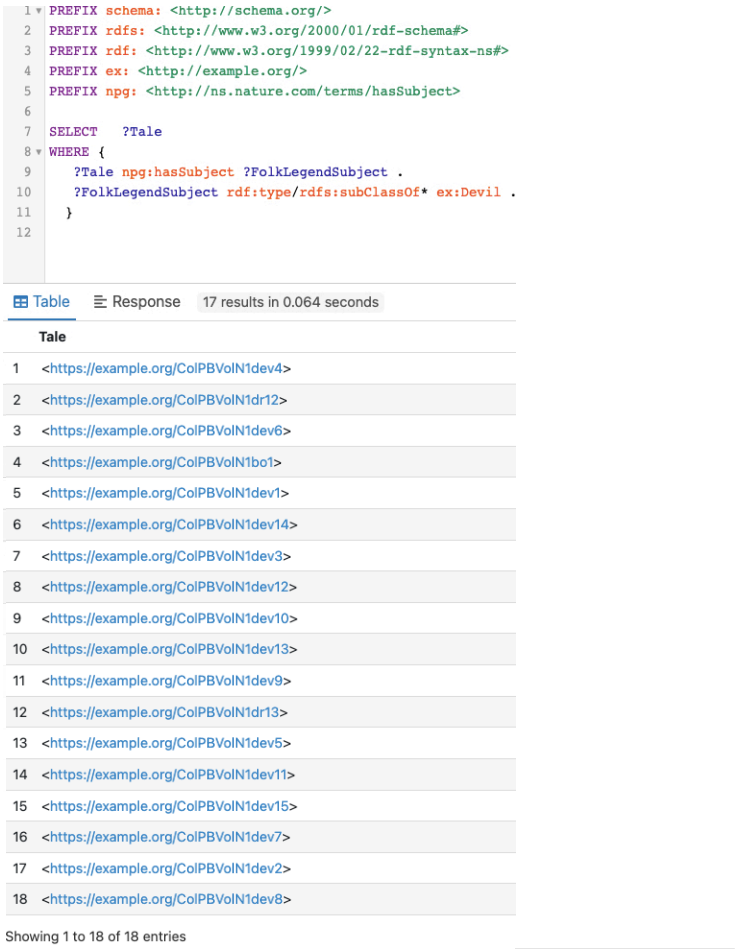
Figure 15. A screenshot of the query for location data “Devon”.

SELECT ?Tale
WHERE {

?Tale schema:location "Devon" .
}

The following tales appeared in the results: ColPBVolN1bd10; ColPBVolN1dev4; ColPBVolN1dev8; ColPBVolN1dev11; ColPBVolN1dev12; ColPBVolN1dev15; ColPBVolN1gi2; and ColPBVolN1gh14.

2. Next, a query was made searching for a specific Creature subject. Tales with npg:hasSubject ex:Devil were searched for.



```
PREFIX schema: <http://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>

SELECT   ?Tale
WHERE {
   ?Tale npg:hasSubject ?FolkLegendSubject .
   ?FolkLegendSubject rdf:type/rdfs:subClassOf* ex:Devil .
  }
```

Table | Response | 17 results in 0.064 seconds

	Tale
1	<https://example.org/ColPBVolN1dev4>
2	<https://example.org/ColPBVolN1dr12>
3	<https://example.org/ColPBVolN1dev6>
4	<https://example.org/ColPBVolN1bo1>
5	<https://example.org/ColPBVolN1dev1>
6	<https://example.org/ColPBVolN1dev14>
7	<https://example.org/ColPBVolN1dev3>
8	<https://example.org/ColPBVolN1dev12>
9	<https://example.org/ColPBVolN1dev10>
10	<https://example.org/ColPBVolN1dev13>
11	<https://example.org/ColPBVolN1dev9>
12	<https://example.org/ColPBVolN1dr13>
13	<https://example.org/ColPBVolN1dev5>
14	<https://example.org/ColPBVolN1dev11>
15	<https://example.org/ColPBVolN1dev15>
16	<https://example.org/ColPBVolN1dev7>
17	<https://example.org/ColPBVolN1dev2>
18	<https://example.org/ColPBVolN1dev8>

Showing 1 to 18 of 18 entries

Figure 16. A screenshot of the query for the subject “Devil”.

```
SELECT   ?Tale
WHERE {
  ?Tale npg:hasSubject ?FolkLegendSubject .
  ?FolkLegendSubject rdf:type/rdfs:subClassOf* ex:Devil .
 }
```

All 15 tales from the chapter about devils appeared in the search results, ColPBVolN1dev1~ColPBVolN1dev15. Tales ColPBVolN1dr12, ColPBVolN1dr13, and ColPBVolN1bo1 also appeared.

3. Queries searching for keywords were also conducted. Tales with schema:keywords “treasure” were searched for.

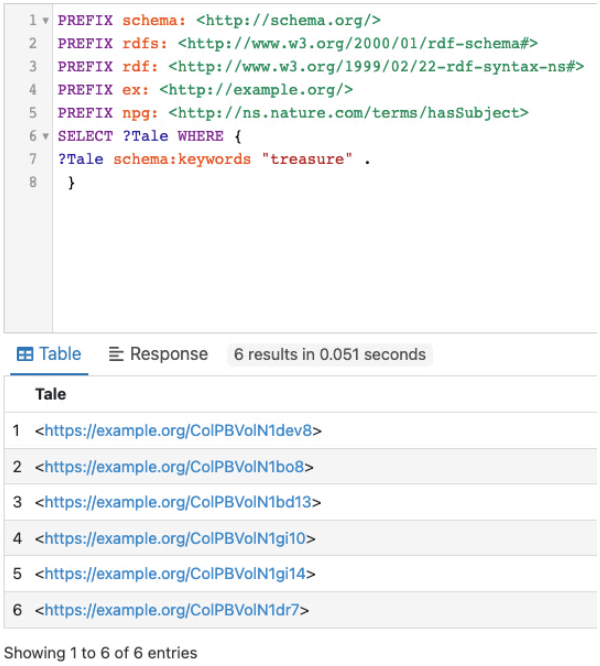


Figure 17. A screenshot of the query for the keyword “treasure”.

```
SELECT   ?Tale
WHERE {
?Tale schema:keywords "treasure"   .
}
```

The following tales appeared in the results: ColPBVolN1gi14; ColPBVolN1gi10; ColPBVolN1dr7; ColPBVolN1dev8; ColPBVolN1bo8; and ColPBVolN1bd13

Tales with schema:keywords “monk” were also searched for.

```
SELECT   ?Tale
WHERE {
?Tale schema:keywords "monk"   .
}
```

The following tales appeared in the results: ColPBVolN1bo4; ColPBVolN1dev8; ColPBVolN1dr12; and ColPBVolN1dr13

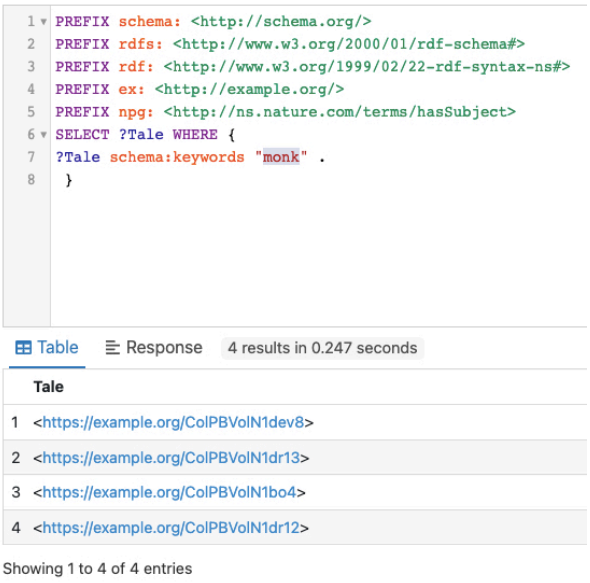


Figure 18. A screenshot of the query for the keyword “monk”.

4. Queries searching for plot points were also conducted. Tales with ex:hasPlotPoint “gift for subject” were searched for.

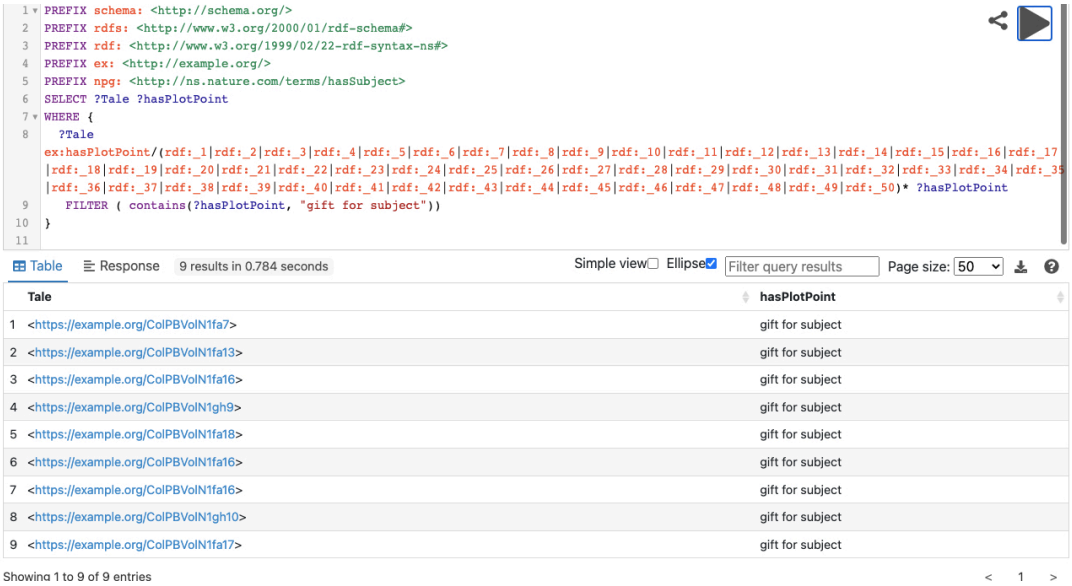


Figure 19. A screenshot of the query for the plot point “gift for subject”.

```
SELECT ?Tale ?hasPlotPoint
WHERE {
  ?Tale ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|
rdf:_10|rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17|rdf:_18|rdf:_19|rdf:_20|
```

```
rdf:_21|rdf:_22|rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|
rdf:_32|rdf:_33|rdf:_34|rdf:_35|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|
rdf:_43|rdf:_44|rdf:_45|rdf:_46|rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
   FILTER ( contains(?hasPlotPoint, "gift for subject"))
}
```

The following tales appeared in the results: ColPBVolN1fa7; ColPBVolN1fa13; ColPBVolN1fa16; ColPBVolN1fa17; ColPBVolN1fa18; ColPBVolN1gh9; and ColPBVolN1gh10.

Tales with ex:hasPlotPoint “subject leaves due to gift” were also searched for.

```
SELECT ?Tale ?hasPlotPoint
WHERE {
  ?Tale ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|
rdf:_10|rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17|rdf:_18|rdf:_19|rdf:_20|
rdf:_21|rdf:_22|rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|
rdf:_32|rdf:_33|rdf:_34|rdf:_35|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|
rdf:_43|rdf:_44|rdf:_45|rdf:_46|rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
   FILTER ( contains(?hasPlotPoint, "subject leaves due to gift"))
}
```

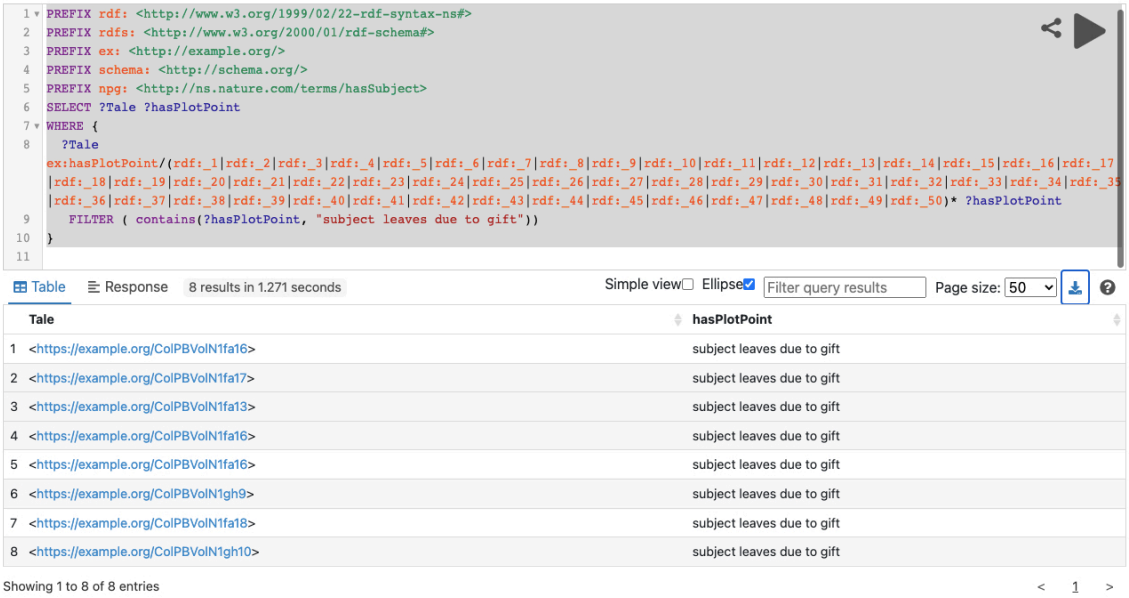
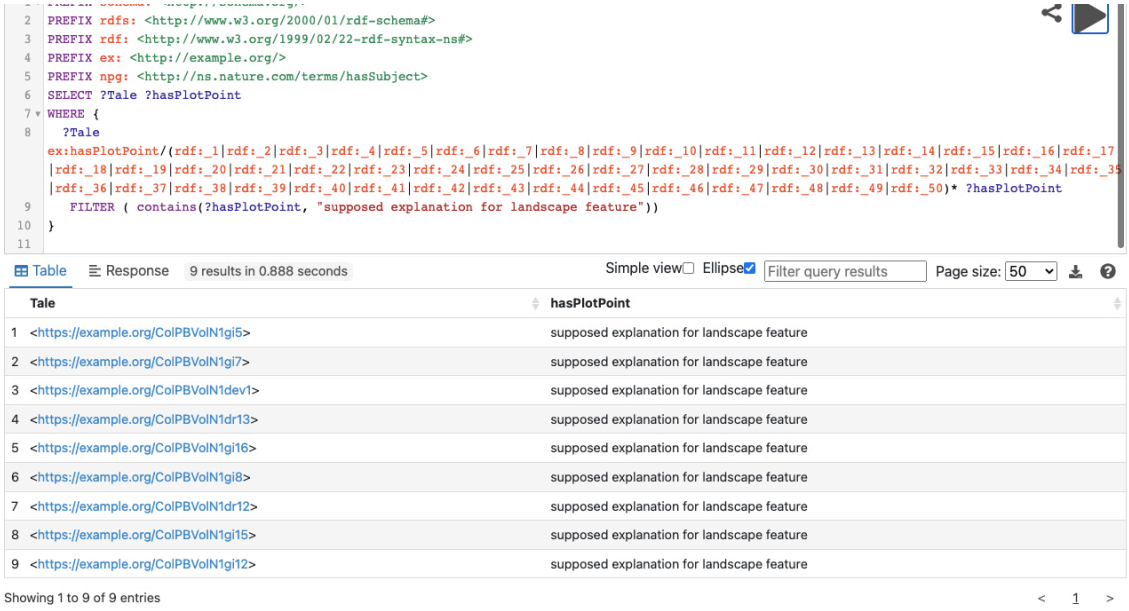


Figure 20. A screenshot of the query for the plot point “subject leaves due to gift”.

The following tales appeared in the results: ColPBVolN1fa13; ColPBVolN1fa16; ColPBVolN1fa17; ColPBVolN1fa18; ColPBVolN1gh9; and ColPBVolN1gh10

Tales with ex:hasPlotPoint “supposed explanation for landscape feature” were searched for.

```
SELECT ?Tale ?hasPlotPoint
WHERE {
  ?Tale ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|
rdf:_10|rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17|rdf:_18|rdf:_19|rdf:_20|
rdf:_21|rdf:_22|rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|
rdf:_32|rdf:_33|rdf:_34|rdf:_35|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|
rdf:_43|rdf:_44|rdf:_45|rdf:_46|rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
   FILTER ( contains(?hasPlotPoint, "supposed explanation for landscape feature"))
}
```



```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
SELECT ?Tale ?hasPlotPoint
WHERE {
  ?Tale
ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|rdf:_10|rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17
|rdf:_18|rdf:_19|rdf:_20|rdf:_21|rdf:_22|rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|rdf:_32|rdf:_33|rdf:_34|rdf:_35
|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|rdf:_43|rdf:_44|rdf:_45|rdf:_46|rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
   FILTER ( contains(?hasPlotPoint, "supposed explanation for landscape feature"))
}
```

Table | Response | 9 results in 0.888 seconds | Simple view | Ellipse | Filter query results | Page size: 50

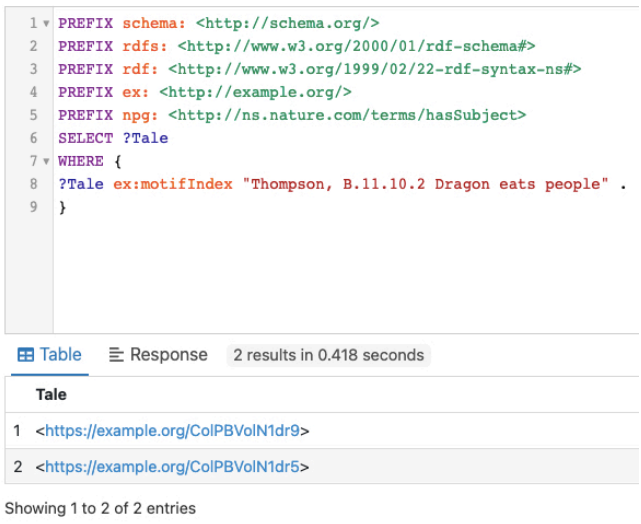
	Tale	hasPlotPoint
1	<https://example.org/ColPBVolN1gi5>	supposed explanation for landscape feature
2	<https://example.org/ColPBVolN1gi7>	supposed explanation for landscape feature
3	<https://example.org/ColPBVolN1dev1>	supposed explanation for landscape feature
4	<https://example.org/ColPBVolN1dr13>	supposed explanation for landscape feature
5	<https://example.org/ColPBVolN1gi16>	supposed explanation for landscape feature
6	<https://example.org/ColPBVolN1gi8>	supposed explanation for landscape feature
7	<https://example.org/ColPBVolN1dr12>	supposed explanation for landscape feature
8	<https://example.org/ColPBVolN1gi15>	supposed explanation for landscape feature
9	<https://example.org/ColPBVolN1gi12>	supposed explanation for landscape feature

Showing 1 to 9 of 9 entries < 1 >

Figure 21. A screenshot of the query for the plot point “supposed explanation for landscape feature”.

The following tales appeared in the results: ColPBVolN1dr12; ColPBVolN1dr13; ColPBVolN1gi5; ColPBVolN1gi7; ColPBVolN1gi8; ColPBVolN1gi12; ColPBVolN1gi15; ColPBVolN1gi16; and ColPBVolN1dev1

5. A query searching for a specific Motif Index was also conducted. Tales with ex:motifIndex “Thompson, B.11.10.2 Dragon eats people” were searched for.



```
PREFIX schema: <http://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
SELECT ?Tale
WHERE {
?Tale ex:motifIndex "Thompson, B.11.10.2 Dragon eats people" .
}
```

Table | Response | 2 results in 0.418 seconds

	Tale
1	<https://example.org/ColPBVolN1dr9>
2	<https://example.org/ColPBVolN1dr5>

Showing 1 to 2 of 2 entries

Figure 22. A screenshot of the query for the Motif Index “Thompson, B.11.10.2 Dragon eats

```
SELECT   ?Tale
WHERE {
?Tale ex:motifIndex "Thompson, B.11.10.2 Dragon eats people"   .
}
```

The following tales appeared in the results: ColPBVolN1dr5, ColPBVolN1dr9.

6. A query searching for a specific Creature subclass was also conducted. Tales with npg:hasSubject ex:Brownie were searched for.

```
SELECT   ?Tale
WHERE {
  ?Tale npg:hasSubject ?FolkLegendSubject .
  ?FolkLegendSubject rdf:type/rdfs:subClassOf* ex:Brownie .
 }
```

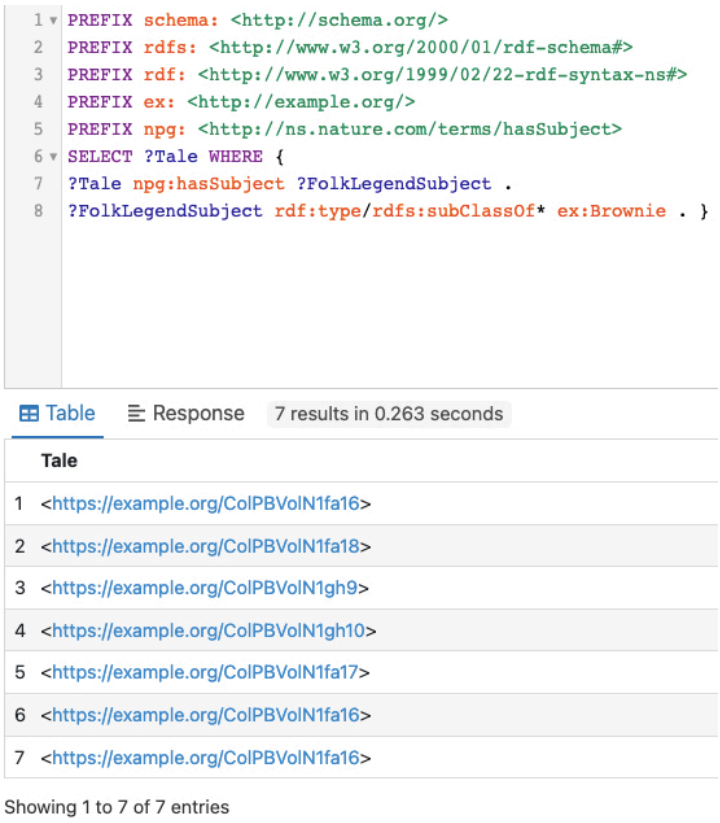


Figure 23. A screenshot of the query for the subclass ex:Brownie.

The following tales appeared in the results: ColPBVolN1fa16; ColPBVolN1fa17; ColPBVolN1fa18; ColPBVolN1gh9; and ColPBVolN1gh10.

7. Queries searching for specific creature Characteristics were also conducted. Creatures with ex:hasCharacteristics “shape-shifts” were searched for.

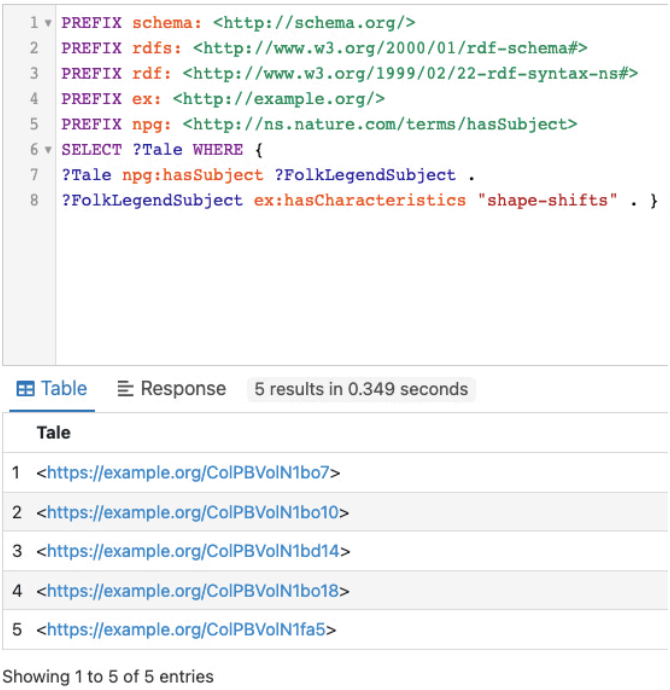


Figure 24. A screenshot of the query for the characteristic “shape-shifts”.

```
SELECT   ?Tale
WHERE {
  ?Tale npg:hasSubject ?FolkLegendSubject .
  ?FolkLegendSubject ex:hasCharacteristics "shape-shifts" .
 }
```

The following tales appeared in the results: ColPBVolN1bo7; ColPBVolN1bo10; ColPBVolN1bo18; ColPBVolN1fa5; ColPBVolN1bd14

8. A query combining searches with multiple elements was also conducted. Tales with npg:hasSubject ex:Giant and tales with ex:hasPlotPoint “supposed explanation for landscape feature” were searched for at the same time.

```
SELECT   ?Tale
WHERE {
  ?Tale npg:hasSubject ?FolkLegendSubject .
  ?FolkLegendSubject rdf:type/rdfs:subClassOf* ex:Giant .
  ?Tale ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|
rdf:_10|rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17|rdf:_18|rdf:_19|rdf:_20|
rdf:_21|rdf:_22|rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|
```

```
rdf:_32|rdf:_33|rdf:_34|rdf:_35|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|
rdf:_43|rdf:_44|rdf:_45|rdf:_46|rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
  FILTER ( contains(?hasPlotPoint, "supposed explanation for landscape feature"))
}
```

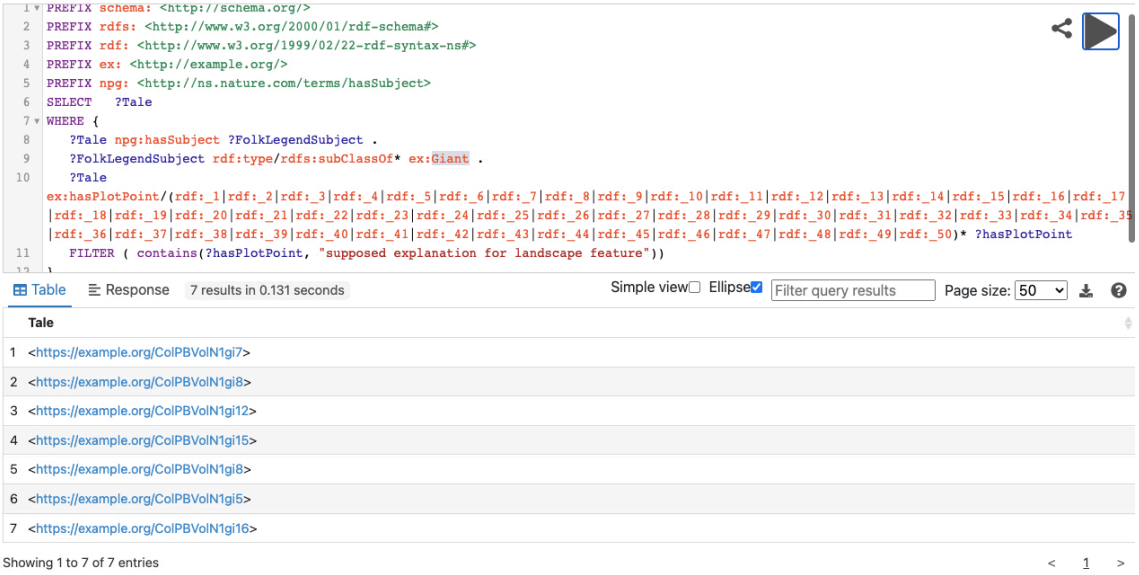


Figure 25. A screenshot of the query for the subject ex:Giant and plot point "supposed explanation for landscape feature"

The following tales appeared in the results: ColPBVolN1gi7; ColPBVolN1gi8; ColPBVolN1gi12; ColPBVolN1gi15; ColPBVolN1gi5; and ColPBVolN1gi16.

Tales with npg:hasSubject ex:Dragon and tales with ex:hasPlotPoint "supposed explanation for landscape feature" were also searched for at the same time.

Figure 26. A screenshot of the query for the subject ex:Dragon and plot point "supposed explanation for landscape feature"

```
SELECT   ?Tale
WHERE {
  ?Tale npg:hasSubject ?FolkLegendSubject .
  ?FolkLegendSubject rdf:type/rdfs:subClassOf* ex:Dragon .
  ?Tale ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|rdf:_10|
rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17|rdf:_18|rdf:_19|rdf:_20|rdf:_21|rdf:_22|
rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|rdf:_32|rdf:_33|rdf:_34|
rdf:_35|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|rdf:_43|rdf:_44|rdf:_45|rdf:_46|
rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
  FILTER ( contains(?hasPlotPoint, "supposed explanation for landscape feature"))
}
```

The following tales appeared in the results: ColPBVolN1dr12 and ColPBVolN1dr13.

6. Discussion

6.1 Results Interpretation

1. The results for the first query of location seemed to satisfy the user needs, as tales that took place in “Devon” were returned. It was interesting that many tales that took place in Devon had Devils appearing in them. This seems like an illustration of how location data can be searched to discover trends in folktales across specific location areas.

On the other hand, a limitation was discovered from the results of query 1. Although the results did seem to meet the user needs when searching for the location “Devon”, tale ColPBVolN1fa5 took place at the “Somerset-Devon border”, and this tale was not returned in the results list. This implies that some locations may be missed from the results if the names are less straightforward, such as these locations joined with a hyphen. This suggests that the way in which the location data is recorded needs to be examined further, so that border locations may also be included in the results.

2. The results for the second query about the type of creature subject seemed to satisfy the user needs, as all the tales that had Devils appear in them were returned in the results. All the tales in the dataset from the chapter about Devils in the collection were returned. In addition, three other tales were returned. These were marked as ex:Devil along with ex:Dragon and ex:Bogie, as they could potentially belong to more than one class.

Broadwell et al. suggested in their research that there are “borderlands between categories” and that with their approach to classification, agency is returned to the researchers, as they are able to read the tales in multiple contexts, that of a current classification, and where it potentially could have been classified [8]. In this way, it seems that by assigning a creature to more than one class, if there is uncertainty about its classification, the user is given the option to read the tales and decide which ones are relevant to their needs themselves. This seems preferable rather than only assigning a

creature to one class, and therefore potentially preventing users from finding a possibly relevant tale. These results seem to imply that the data model could be useful in this, and for searching for tales that might previously have been difficult for users to find, due to the different chapters in which they appear in the collection.

3. The results for the third query of keywords seemed to satisfy the user needs, as the relevant tales appeared in the results when searched for. This implies that the data model may be useful for users searching for tales that feature a certain topic, but a topic that is not the main subject of a tale, such as the type of creature featured. In this way, users could search for a wider range of concepts, such as the example given as "monk". For example, users wanting to examine how religion was portrayed in the folktales could search for keywords such as "monk", "cross" or "Christmas", and discover tales in which these words appear. In this way, the number of tales that the users would need to search through to find folktales that are relevant to their needs would be reduced.

On the other hand, there were limitations discovered when testing this query for keywords. Keywords that were very similar, or about the same topic but recorded differently, did not appear in the results unless the exact wording in which the keyword was recorded under was used. For example, the keywords "gambling" and "cards" are very similar and about the same topic. However, because some tales had "gambling" recorded in the keywords, and some had "cards", unless a user searched the keywords for both of these terms, they would not be able to find all the tales that feature this topic. This could leave potentially relevant tales from the results list. This necessitated that the recorded keywords were examined again, so that similar keywords could be recorded using the same terms, in order for them to appear in the query results together.

4. The results for the fourth query of plot points seemed to satisfy the user needs, as all the tales in which a subject is given a gift were returned. These were 5 tales about Brownies, or creatures classified as both Ghost and Brownie, and 2 other Fairies. Tale ColPBVolN1fa16 appears 3 times in Figures 19 and 20, as this tale actually contains three different stories about 3 different Brownies.

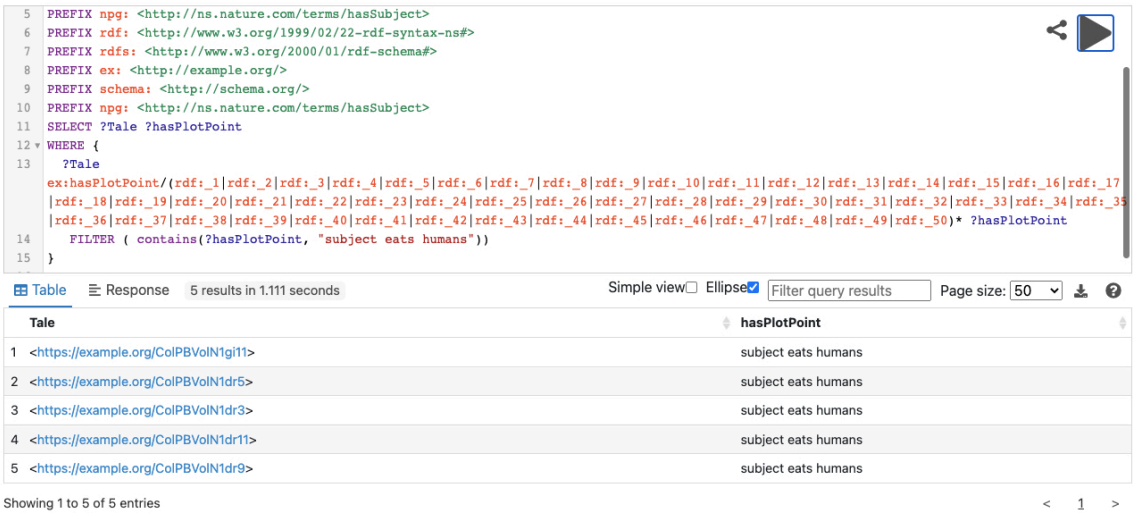
All the tales in which a subject leaves due to a gift were also returned. The results from this plot points query seem to imply that the data model is effective in enabling users to search the tales by examining the 'emic' units of their narrative structure, in conjunction with the changeable 'etic' units [8]. For instance, searching for ex:hasPlotPoint "subject leaves due to gift" returned 6 results: 3 classified as ex:Brownie, a subclass of ex:Fairy; 1 classified as ex:Trow, a subclass of ex:Fairy; and 2 classified both as ex:Brownie and ex:Ghost. In the collection, tale ColPBVolN1fa13 has been assigned the motif "Thompson, F.405.11 House spirit leaves when gift of clothing is left for it" [7]. Tale ColPBVolN1fa16 has been assigned "Thompson, F.381.3 Fairy leaves when he is given clothes" [7]. Tales ColPBVolN1fa17 and ColPBVolN1fa18 were not assigned a motif, as the Brownies left because they were given gifts of food rather than clothing. Tale ColPBVolN1gh9 is also assigned "Thompson, F.405.11 House spirit leaves when gift of clothing is left for it" [7] in addition to "Baughman, F.482.9 Getting rid of house-spirit"

[16]. Finally, tale ColPBVolN1gh10 is also assigned “Thompson, F.381.3 Fairy leaves when he is given clothes” [7]. These tales all deal with the story of a creature leaving a dwelling because it has been given gifts, and have many similarities. However, searching for the tales by using the Motif Index would not result in these 6 tales being found together. Although the tales are all very similar and share the same narrative structure, they are classified under different motifs due to the focus on the changeable ‘etic’ [8] elements of ‘fairy’ and ‘house spirit’. Also, there is no motif were a Fairy leaves because it has been given food, rather than clothing, so two of the tales would not be found when searching for tales based on the Motif Index, as they are not assigned any motif. By searching instead for “subject leaves due to gift”, focusing on the action of the plot point, rather than the details of what item was given and the subject it was given to, it becomes possible to search across the whole dataset for the tales that contain the same plot point, regardless of the changeable units surrounding them, such as the different creatures or subjects.

On the other hand, a limitation to be examined was also raised in the results of query 4, plot points. In addition to the plot point that was searched for, “gift for subject”, there is also a related plot point of “subject gives gift” which appears in ColPBVolN1fa7, ColPBVolN1fa11 and ColPBVolN1fa12. As these two plot points are quite similar, it might be interesting if there was a hierarchy for the actions performed in the plot points. For example, a user could search for tales about ‘giving gifts’, irrespective of who is performing the action of giving the gift. Users could then also search specifically for tales where the creature subject *receives* the gift, or tales where the creature subject *gives* the gift. This might enable the user to search for specific plot points dependant on the performer of the action, as well as specific plot points independent of the performer of the action.

5. The results of the fifth query of Motif Index seemed to satisfy the user needs, as the two relevant tales appeared in the results. This will enable users to use the existing classification systems with the proposed data model.

This result also demonstrates how the two systems can be used together, as “Thompson, B.11.10.2 Dragon eats people” returned two results. However, searching for ex:hasPlotPoint “subject eats humans”, which would be the equivalent in the proposed data model, returns 5 results, ColPBVolN1dr3, ColPBVolN1dr5, ColPBVolN1dr9, ColPBVolN1dr11, and ColPBVolN1gi11 (shown in Figure 27). These results are listed under three different Motif Index numbers, “Thompson, B.11.10.2 Dragon eats people” [7] (shown in Figure 22), “Thompson, F.531.3.11 Giant swallows men” [7] (shown in Figure 28), and “Thompson, B.11.10.3 Dragon devours children” [7] (shown in Figure 29). Tale ColPBVolN1dr11 is not listed as having any of these motifs, but the tale does state that the dragon eats young maidens. On the other hand, “subject eats humans” covers tales where the subjects eat people, irrespective of the type of subject, or who they are eating. This implies that using the two systems together can help in searching for both ‘etic’ and ‘emic’ [8] elements of the tales together.



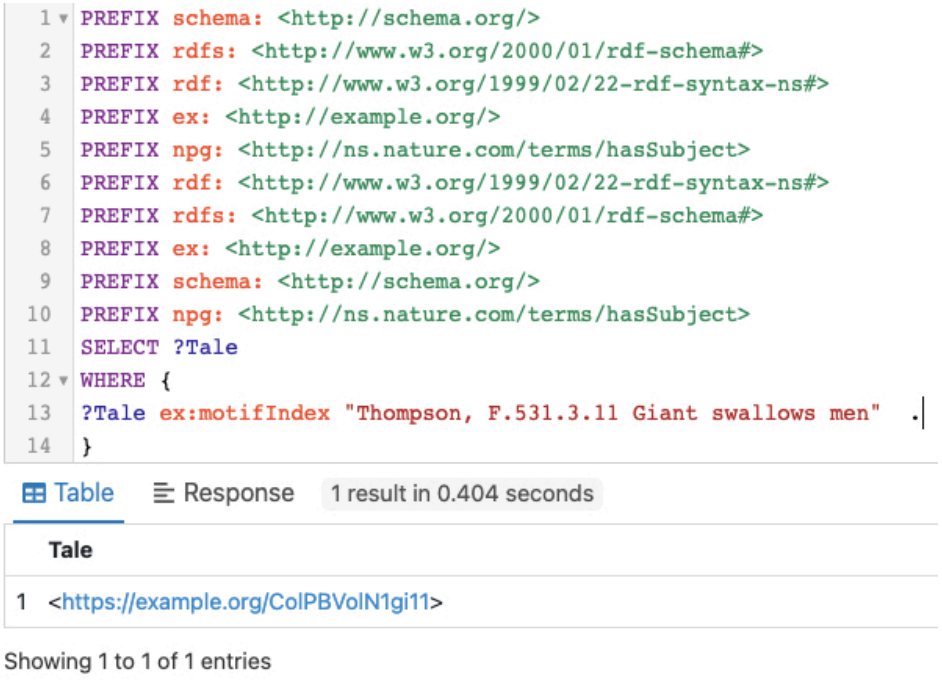
```
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex: <http://example.org/>
PREFIX schema: <http://schema.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
SELECT ?Tale ?hasPlotPoint
WHERE {
  ?Tale
ex:hasPlotPoint/(rdf:_1|rdf:_2|rdf:_3|rdf:_4|rdf:_5|rdf:_6|rdf:_7|rdf:_8|rdf:_9|rdf:_10|rdf:_11|rdf:_12|rdf:_13|rdf:_14|rdf:_15|rdf:_16|rdf:_17
|rdf:_18|rdf:_19|rdf:_20|rdf:_21|rdf:_22|rdf:_23|rdf:_24|rdf:_25|rdf:_26|rdf:_27|rdf:_28|rdf:_29|rdf:_30|rdf:_31|rdf:_32|rdf:_33|rdf:_34|rdf:_35
|rdf:_36|rdf:_37|rdf:_38|rdf:_39|rdf:_40|rdf:_41|rdf:_42|rdf:_43|rdf:_44|rdf:_45|rdf:_46|rdf:_47|rdf:_48|rdf:_49|rdf:_50)* ?hasPlotPoint
    FILTER ( contains(?hasPlotPoint, "subject eats humans"))
}
```

Table | Response | 5 results in 1.111 seconds | Simple view | Ellipse | Filter query results | Page size: 50

	Tale	hasPlotPoint
1	<https://example.org/ColPBVolN1gi11>	subject eats humans
2	<https://example.org/ColPBVolN1dr5>	subject eats humans
3	<https://example.org/ColPBVolN1dr3>	subject eats humans
4	<https://example.org/ColPBVolN1dr11>	subject eats humans
5	<https://example.org/ColPBVolN1dr9>	subject eats humans

Showing 1 to 5 of 5 entries < 1 >

Figure 27. A screenshot of a query for the plot point “subject eats humans”



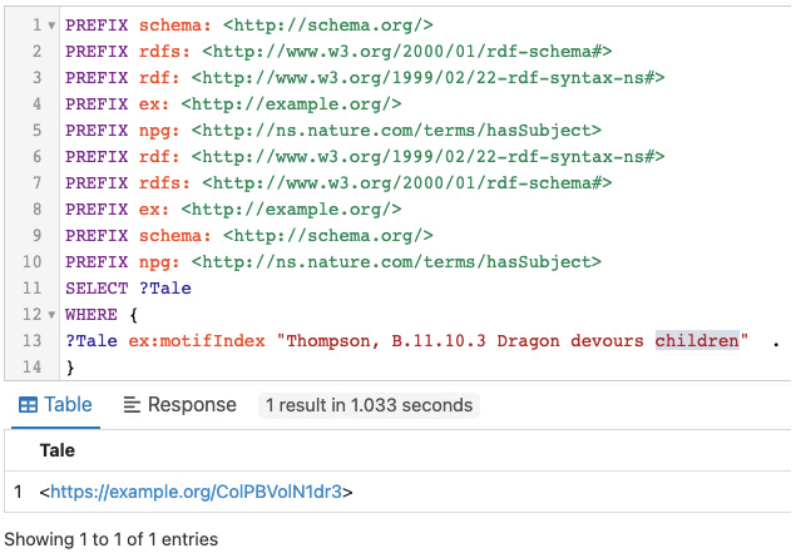
```
PREFIX schema: <http://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex: <http://example.org/>
PREFIX schema: <http://schema.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
SELECT ?Tale
WHERE {
?Tale ex:motifIndex "Thompson, F.531.3.11 Giant swallows men"  .
}
```

Table | Response | 1 result in 0.404 seconds

	Tale
1	<https://example.org/ColPBVolN1gi11>

Showing 1 to 1 of 1 entries

Figure 28. A screenshot of a query for the Motif Index “Thompson, F.531.3.11 Giant swallows men”.



```
PREFIX schema: <http://schema.org/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ex: <http://example.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX ex: <http://example.org/>
PREFIX schema: <http://schema.org/>
PREFIX npg: <http://ns.nature.com/terms/hasSubject>
SELECT ?Tale
WHERE {
?Tale ex:motifIndex "Thompson, B.11.10.3 Dragon devours children"  .
}
```

Table | Response | 1 result in 1.033 seconds

	Tale
1	<https://example.org/ColPBVolN1dr3>

Showing 1 to 1 of 1 entries

Figure 29. A screenshot of a query for the Motif Index “Thompson, B.11.10.3 Dragon devours children”.

6. The results of the sixth query of creature subject subclass also seemed to satisfy the user needs, as all tales featuring Brownies were returned in the results. 3 of the tales returned featured subjects that were classified as ex:Brownie, and 2 tales that featured subjects that were classified as ex:Brownie and ex:Ghost. Tale ColPBVolN1fa16 appears 3 times in Figure 23, as this tale actually contains three different stories about 3 different Brownies. The 3 Brownie tales appeared in the chapter about Fairies in the collection, and the other two tales appeared in the Ghosts chapter of the collection. This result implies that the data model could be useful for searching for tales that might previously have been difficult for users to find, due to the chapters in which they appear in the collection.

7. The results of the seventh query of creature characteristics also seemed to satisfy the user needs, as all tales featuring creatures that "shape-shifts" were returned in the results.

On the other hand, there were limitations discovered when testing this query for creature characteristics. Characteristics that were very similar, or about the same attribute but recorded differently, did not appear in the results unless the exact wording in which the characteristic was recorded was used.

For example, tales about Black Dogs often feature descriptions about their eyes, such as: "saucer sized eyes"; "eyes bright as fire"; "saucer eyes"; "glowering eyes"; "eyes like balls of fire"; "great glassy eyes"; and "eyes blazed like fire". These are all very similar and are about the same feature. However, unless these characteristic were all recorder using the same terms, they would not appear together in the search results. This could leave potentially relevant creature subjects from the results list. This necessitated that the recorded characteristics were examined again, so that similar characteristics could be recorded using the same terms, in order for them to appear in the query results together.

Another limitation was also discovered, as there were some characteristics recorded that resulted in many results being returned, and so they were not particularly relevant or helpful, such as searching for the characteristic "large". This also necessitated that the recorded characteristics were examined again, so that very common or irrelevant characteristics could be edited.

8. Finally, the results of the eighth query combining subject and plot points also seemed to satisfy the user needs, as all tales featuring Giants that have the plot point "supposed explanation for landscape feature" were returned in the results, and all the tales featuring Dragons that have the plot point "supposed explanation for landscape feature" were also returned. Tale ColPBVolN1gi8 appeared twice in Figure 25, as there were 2 Giants in this single tale. This demonstrates how users could utilise combined searches to narrow down results further, as Figure 21 shows that 9 tales appear in the results for the plot point query "supposed explanation for landscape feature", Figure 25 shows that

this can be narrowed down to 6 tales if searching for Giants with this plot point, and Figure 26 shows that this can be narrowed down to 2 tales if searching for Dragons with this plot point.

From these results, although issues have been raised that need examining, it seems that the data model is relatively effective, as the tales can be searched and the results narrowed down to find more specific information for the users.

6.2 Limitations

There are a number of limitations with the research. Firstly, the sample size is small. Only 109 tales were used in the dataset, spread across the seven subject areas of Black Dogs, Bogies, Devils, Dragons, Fairies, Ghosts and Giants. The vast amount of other tales that the collection contains were unable to be covered. In particular, there are a huge amount of tales in the Fairies and Ghosts chapters of the collection that were unable to be used. It would be interesting to examine whether the data model would be effective with a larger sample of tales featuring just one creature subject, such as the ex:Fairy class, to ascertain whether the data model would still be effective when just dealing with a large number of tales about one single creature subject. As the current data model only has subclasses for the subjects that appear in tales that are currently in the dataset, if the number of tales were to be expanded, then the data model would also need to be revised and it would be necessary to include more subclasses, to encompass the additional subject types that would appear with the addition of new tales in the dataset. Therefore, it would be beneficial to extend the dataset to include more tales from Part B Volume I in the future.

Furthermore, in addition to the tales that were not used in the dataset from Part B Volume I, there are also all the Folk Legends in Part B Volume II of the collection which were also unable to be used. Part B Volume II includes the subjects Historical Traditions, Local Legends, Origin Myths, Saints, The Supernatural, Witches and Miscellaneous Legends. The data model did not include and was not tested on any of these subjects, so there is a lack of understanding on whether the data model would have been effective if tales on these subject matters were included in the dataset. If these tales were to be included as part of future work, in order for the data model to be effective with these additional subjects, more subject classes and subclasses would need to be added to the data model.

Secondly, no user studies were carried out as part of this research, so the user needs and use cases for the data model and dataset were hypothesised based on previous experience and impressions gained from informal conversations with various individuals engaged in educational professions. In order to gain a better understanding of how the dataset would be used in reality, and if it would be relevant to user needs, user studies need to be performed. Data collection could be carried out by making use of interviews, questionnaires, or carefully designed lab style user studies, where a target user would be assigned a task to simulate the user needs, and their actions recorded.

Thirdly, there is a limitation with the sample selection. There are only 13 tales in the Dragon chapter. Although this small number of tales about Dragons might be acceptable in the current small size of the dataset, if the dataset was extended to include more tales from Part B Volume I of the collection, the lack of tales about this subject might be an issue, because the data model could only be tested for dragons on those 13 tales.

Also, further examination into how the ex:hasPlotpoint information is extracted from the collection of tales would be beneficial. In section 6.1, when discussing the results of the SPARQL queries, it seemed that searching for tales using the ex:hasPlotPoint appeared to be useful for finding tales that had similar elements but were not listed as having the same Tale Types or Motifs in other classification systems. Therefore, it may be beneficial to change the data model in order to make the plot points resources, rather than the literals that are currently used, in order for users to be able to easily understand what values would be suitable for querying. A list of the plot points extracted from the tales would need to be created and standardised, to ensure that they would be easy to search.

It would also be interesting to examine whether a hierarchy of plot points would be useful for users focusing on the narrative structure of the tales. For example, plot points focusing on similar actions or topics could be placed into different classes and subclasses, enabling users to search for tales with very specific actions, or tales with a broader category of actions.

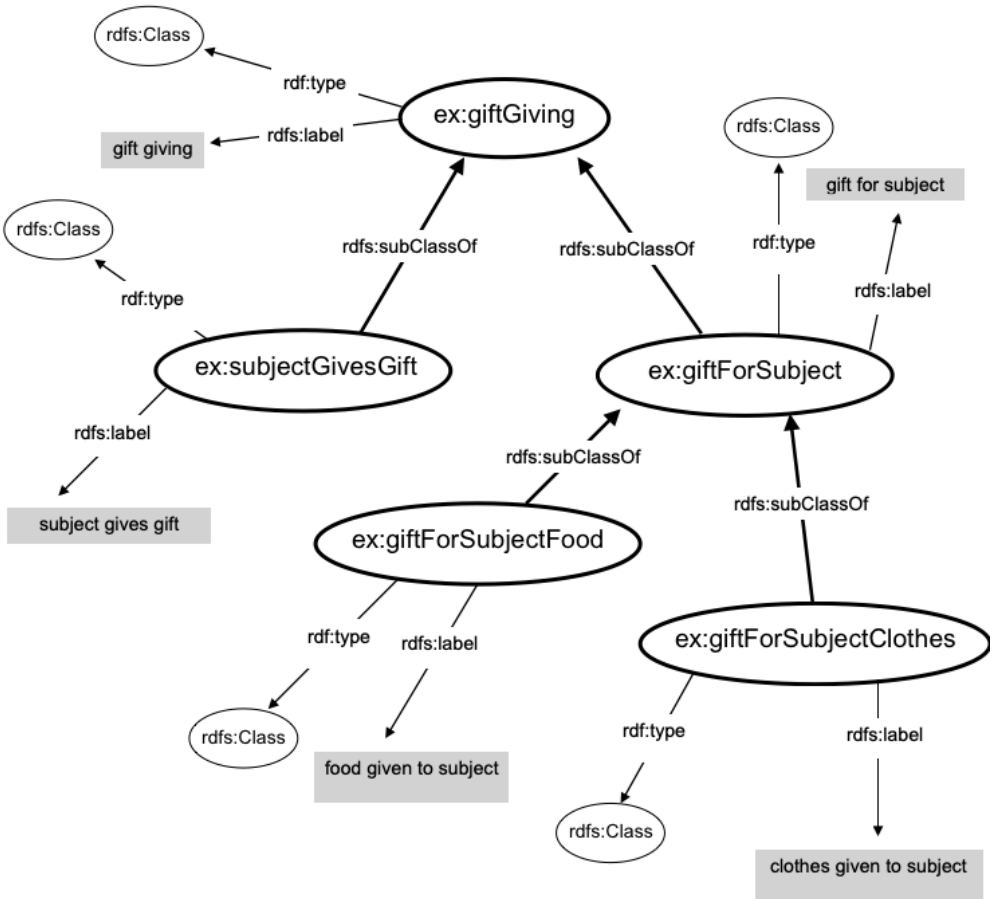


Figure 30. An example of how a hierarchy of classes for the plot point resources might be arranged.

For instance, as shown in Figure 30, a user could search for tales under the category of "gift giving", and tales that contain plot points in the various subclasses of this category (such as "subject gives gift", "gift for subject", "food given to subject", "clothes given to subject") would appear in the results. The user could also search for tales where only the plot point "food given to subject" appears, to get more specific results.

However, there would be limitations involved with this approach, as new tales being added to the dataset would likely have plot points that would not have been recorded in the data model or standardised list created. Therefore, maintaining the data model and vocabulary list would be difficult, as the list would need to be added to and edited frequently, and new classes and subclasses would need to be added to the data model.

7. Conclusion

In conclusion, a data model was proposed in order to help users find relevant tales that meet their specific requirements from the Folk Legend section of a collection of British folktales. Attention was paid to what information the users may find relevant and how to go about breaking down the tales into detailed structured elements that would enable users to find the specific information that they want.

A dataset was created from 109 tales from Briggs' 'A Dictionary of British Folktales, Part B Folk Legends Volume I', using a number of tales from each chapter of the volume, in order to cover the various subject matter that is included in this volume, such as Black Dogs, Dragons and Giants. The information from the selected tales was recorded and applied to the proposed data model. Attention was paid to the information in the tales that a user might find relevant or useful.

The information from the dataset was converted into RDF format using Turtle syntax, and inserted into Jena Apache Fuseki. A list of use cases and user needs were written as SPARQL queries and the dataset was queried. The results were analysed to determine whether the user needs were met, and what changes might be applied to the data model and dataset next time.

In conclusion, it appears that the data model satisfied the user needs, but there are issues and limitations that need to be addressed, such as:

1. How the location data, keywords, and creature characteristics are recorded, so as not to miss border locations and similarly recorded keywords and characteristics from the results, and to edit very common or irrelevant creature characteristics.

2. How the plot points are recorded, as a hierarchy of plot point classes and subclasses might be useful for users who want to search for specific events regardless of who is performing the action in the plot point. Plot points as resources may also be beneficial.

3. Determining if the data model will be effective with a larger number of tales. This includes a large number of tales featuring just one subject type, such as ex:Fairy, to

examine whether the data model will work with one particular subject in depth, rather than over a relatively small number of tales covering a range of subject types. Also, determining if the data model will be effective with tales featuring creatures and subject matter not covered in Part B Volume I of the collection.

4. Determining whether the hypothesised user needs are relevant in reality.

7.1 Future Work

In the future, the data model would need to be tested on a larger dataset, to examine whether it would be effective with a larger number of tales. In particular, examining whether the data model would be effective on a large number of tales about the same creature subject. Enlarging the dataset to encompass the whole of Briggs' Part B Volume I of the collection would be the next step in extending the dataset.

Also, further examination into how the ex:hasPlotpoint information is extracted from the collection of tales would be beneficial. For example, changing the data model in order to make the plot points resources, rather than the literals that are currently used, in order for users to be able to easily understand what values would be suitable for querying. Furthermore, examining whether a hierarchy of plot points would be useful for users focusing on the narrative structure of the tales.

Finally, conducting user need studies to ascertain actual user needs and use cases, rather than using speculative user needs, would be beneficial for discovering how the data model could be improved and how it might be used in real life settings.

8. Bibliography

[1] Uther, H. J. (2004). The types of international folktales : a classification and bibliography, based on the system of Antti Aarne and Stith Thompson. Helsinki: Suomalainen Tiedeakatemia, Academia Scientiarum Fennica.

[2] Crevasse, I. E. (2020). 'Once Upon A Scheme: Conceptualizing Digital Fairy Tale And Folklore Classification Systems for Extra-Academic Professionals'. Master's Thesis. University of North Carolina, Chapel Hill.

[3] UNESCO (2001). International Round Table: "Intangible Cultural Heritage – Working Definitions" Piedmont, Italy, 14 -17 March 2001 Annotated Agenda. Available at: https://ich.unesco.org/doc/src/00075-EN.pdf

[4] UNESCO (1997). 'Preservation and conservation of expressions of folklore: the experience of Africa'. UNESCO-WIPO World Forum on the Protection of Folklore. Phuket, Thailand, 1997. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000220167

[5] Oxford Reference. (n.d.). Folktale. Available at: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095826550

[6] McCallum, R. (2000). 'Approaches to the Literary Fairy Tale', in Zipes, J. (ed.) Oxford Companion to Fairy tales. Oxford: Oxford University Press, pp. 17-21.

[7] Thompson, S. (1955-1958). Motif-Index of Folk-Literature: A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-Books, and Local Legends. Bloomington: Indiana University Press.

[8] Broadwell, P. M., Mimno, D., Tangherlini, T. R. (2017). 'The Tell-Tale Hat: Surfacing the Uncertainty in Folklore Classification', *Journal of Cultural Analytics*, Available at: https://culturalanalytics.org/article/11069-the-tell-tale-hat-surfacing- the-uncertainty-in- folklore-classification (Accessed: 09 February 2021)

[9] Briggs, K. M. (1971). A Dictionary of British Folktales, Part B Folk Legends Volume I. London: Routledge & Kegan Paul.

[10] Cox, M. R. (1893). Cinderella; three hundred and forty-five variants of Cinderella, Catskin, and Cap o'Rushes, abstracted and tabulated, with a discussion of mediæval analogues. London: The Folk-lore Society.

[11] Teverson, A. (2013). Fairy Tale. New York: Routledge.

[12] Aarne, A. (1910). Verzeichnis der Märchentypen. Helsinki : Suomalaisen Tiedeakatemian Toimituksia.

[13] Aarne, A. & Thompson, S. (1961).The Types of the Folk-Tale: A Classification and Bibliography. Antti Aarne's Verzeichnis der Märchentypen, translated and enlarged by Stith Thompson. Helsinki: Suomalainen Tiedeakatemia, Academia Scientarum Fennica.

[14] Thompson, S. (1964). *The Folktale*. New York: The Dryden Press.

[15] Christiansen, R. (1958). The Migratory Legends: a proposed list of types with a systematic catalogue of Norwegian variants. Helsinki : Suomalainen Tiedeakatemia.

[16] Baughman, E. W. (1966). Type and Motif-Index of the Folktales of England and North America. The Hague: Mouton & Co.

[17] Propp, V. (1968). Morphology of the Folktale. 2nd ed. Austin: University of Texas Press.

[18] Ilyefalvi, E. (2018). 'The Theoretical, Methodological and Technical Issues of Digital Folklore Databases and Computational Folkloristics', *Acta Ethnographica Hungarica*, 63(1), pages 209-258.

[19] Meder, T., Karsdorp, F., Nguyen, D., Theune, M., Trieschnigg, D., Muiser, I. E. C. (2016). 'Automatic Enrichment and Classification of Folktales in the Dutch Folktale Database'. The Journal of American Folklore. Vol. 129, No. 511. pages 78-96.

[20] Meertens Institute (n.d.) The Dutch Folktale Database. Available at: https://www.verhalenbank.nl/

[21] Giannoulakis, S., Tsapatsoulis, N., Grammalidis, N. (2018). 'Metadata for Intangible Cultural Heritage, The Case of Folk Dances'. VISAPP 2018: Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications. Pages 634-645. Funchal, Madeira, Portugal. 27-29 January 2018.

[22] Berners-Lee, T. (2006). Linked Data. Available at: https://www.w3.org/DesignIssues/LinkedData.html

[23] Tom Heath and Christian Bizer (2011) *Linked Data: Evolving the Web into a Global Data Space* (1st edition). Synthesis Lectures on the Semantic Web: Theory and Technology, 1:1, 1-136. Morgan & Claypool. Available at:http://linkeddatabook.com/editions/1.0/

[24] Urban, S. (n.d.). Unpacking World Folk-literature: Thompson's Motif Index, ATU's Tale Type Index, Propp's Functions and Lévi-Strauss's Structural Analysis for folk tales found around the world. Available at: https://sites.ualberta.ca/~urban/Projects/English/Motif_Index.htm

[25] Brown, T. (1958). The Black Dog. *Folklore*. **69**(3), 175-192.

[26] GeoNames (n.d.). GeoNames. Available at: http://www.geonames.org/

[27] Ordnance Survey (n.d.). Get Outside. Available at: https://getoutside.ordnancesurvey.co.uk/

[28] Historic England (n.d.). The Most Important Historic Places in England Are Listed. Available at: https://historicengland.org.uk/

[29] Wikipedia (n.d.). John Strode (c.1561-1642). Available at: https://en.wikipedia.org/wiki/John_Strode_(c._1561_%E2%80%93_1642)

[30] Wikipedia (n.d.). John Strode (died 1679). Available at: https://en.wikipedia.org/wiki/John_Strode_(died_1679)

[31] Briggs, K. (1976). An Encyclopedia of Fairies - Hobgoblins, Brownies, Bogies, And Other Supernatural Creatures. New York: Pantheon Books.

[32] Tongue, R. L. & Briggs, K. M. (1970). Forgotten Folk Tales of the English Counties. London: Routledge & K. Paul.

[33] Bowker, J. (1878). Goblin Tales of Lancashire. London: W. Swan Sonnenschein & Co.

[34] Briggs, K. M. (1970). A Dictionary of British Folktales, Part A Folk Narratives Volume I. London: Routledge & Kegan Paul.

[35] W3C. (2014). RDF 1.1 Turtle. Available at: https://www.w3.org/TR/turtle/

[36] Poorman's Linked Data Toolkit. Available at: https://github.com/jp-textbook/jp-textbook.github.io/wiki/Toolkit.en

[37] The Apache Software Foundation. (n.d.) Apache Jena Fuseki. Available at: https://jena.apache.org/documentation/fuseki2/

[38] W3C (2013). SPARQL 1.1 Query Language. Available at: https://www.w3.org/TR/sparql11-query/