Analysis and Visualization of Narrative in \textit{Shanhaijing} Using Linked Data

Qian Wang*, Terhi Nurmikko-Fuller* and Ben Swift**

*College of Arts and Social Science, Australian National University
**Research School of Computer Science, Australian National University

I. INTRODUCTION

Existing Linked Data (LD) literature contains several examples (such as those based on Propp's \textit{Morphology of the Folktales} [16] or on ancient Sumerian mythologies [14]) where the fabula\(^1\) and syuzhet\(^2\) of Western folktales have been represented, and information regarding the stories themselves have been published in machine-readable formats such as RDF\(^3\). However, similar (linked) datasets and analyses to be largely non-existent for equivalent stories from Chinese mythology. This paper seeks to bridge that gap by creating, analyzing and publishing a case study example—the classic \textit{Shanhaijing (the Classic of Mountains and Seas 山海经)}.

We recount the complexities of representing ancient Chinese literary narratives, captured in unstructured data, using tools developed from Western perspectives and for complete and largely homogeneous, highly-structured data.

\textit{Shanhaijing} is an ancient encyclopedia. Its origins can be traced back to the pre-Qin period of China (4th century BC), its development continuing through to the early Han Dynasty (1st century AD). It covers broad areas such as ancient mythology, geography, witchcraft, religion, medicine, and other aspects [8]. \textit{Shanhaijing} occupies a significant position in the literary and mythological corpora of the East, and is representative of a wider spectrum of Eastern mythologies. Over thousands of years, numerous Chinese novels, literary fictions and dramas have been derived from the book, such as \textit{Zhuangzi} ([庄子]) [22] and \textit{Strange Tales of Liao Zhai} ([聊斋志异]) [18]. Mythologies from other Asian countries were influenced by it, e.g. \textit{Kaiishi choju zukan} [23] and \textit{Hyakki Yagyō} [19], both examples of Japanese folklore.

In this paper, we report on the state of existing work combining LD methodologies and approaches with literary compositions (Section II), and summarize the narrative of the \textit{Shanhaijing} (Section III). We outline our chosen methodology in Section IV. The custom-built user-interface (UI) is demonstrated in Section V, and we conclude the paper with a discussion of the complexities of the process in Section VI.

\(1\)The fabula is the chronological order in which events unfold in the story.
\(2\)The syuzhet refers to the narrative that is deployed to tell the story, and may include the recounting of events in non-chronological order, e.g., through flashbacks, or the use of frame stories.
\(3\)Resource Description Framework, a standard data model for exchanging information on the Web.

II. RELATED WORK

In the domain of literature, several publicly available datasets have been published as LD. For example, the Book-Sampo project [12], which provides information on fiction literature published in Finland going back to the 15th century, alongside rich descriptions of both content and context [11], or the Perseids Project [2], which provides a platform for creating, publishing, and sharing research data, in the form of textual transcriptions, annotations and analysis [1]. Other essential work in this space include the Brothers Grimm project [7], an ontology for Greek mythology [3], and the Aarne-Thompson’s Motif-Index project [6].

III. CLASSIC OF MOUNTAINS AND SEAS, 山海经

The present version of \textit{Shanhaijing} contains 18 chapters, approximately 31,000 words in total. It records ancient Chinese mythologies, where numerous monsters with fanciful descriptions are portrayed as possessing magical powers or as related to ancestor (totem) worship [10], such as the monster Lushu (鹿蜀, Stag-silkworm)\(^4\), which looks like a horse with a white head, a scarlet tail and tiger’s markings, and lives on Mount Cherrysunny ( Chesson之山). Whoever wears its fur will have a greater number of descendants.

We focused exclusively on the capture of the data for the monsters in \textit{Shanhaijing}. The reason for this is that the fascinating and detailed accounts of these creatures overwhelm the other aspects of the story; and these descriptions account for a notable proportion of instances of literary borrowings and inspirations in other cultures, increasing the likelihood of reuse and inter-linking with other ontologies.

IV. METHODOLOGY

The first stage of the project focused exclusively on the cornucopia of monsters (a total of 277). Through a close reading in both English and classical Chinese, we extracted structured data from the unstructured narrative.

In the second stage, we designed an ontological structure to model the domain [15]. After considering several pre-existing ontology software libraries, we concluded that the suitability of these vocabularies and resources for the representation of \textit{Shanhaijing} was limited. This necessitated the building of a

\(4\)This monster appears in the Chapter: \textit{Classic of the Mountains: Southern}
new ontology, which captures data types and represents the relationships between them as .TTL file. The ontology represents the characteristics of the monsters and the complex relationships between them. It contains a taxonomy of body parts, the characteristics, and the habitats of all monsters. We used a combination of top-down and bottom-up ontologies and schemas to identify Class and Property hierarchies, captured through rdfs:subClassOf and rdfs:subPropertyOf, and reused existing vocabularies such as DBPedia Ontology [5], BioTopOntology [21], Mahabarata Ontology [17], RDFS, and XML Schema. This approach enabled us to capture the specifics of the data (bottom-up approach), but also maximize the benefit from other well-developed and rich ontologies.

All concepts related to monsters were collected, then split into terms. A term is considered as a Class if it has attributes pointing to other classes or literals, or it is a Superclass of other Classes. Otherwise, it is defined as a property. For example, the term Monster is defined as a Class because it has attributes linking it to other Classes, such as Mountain, through the property livesIn. However, term Noise is not a Class because it is not the domain of any attribute. Hence, it is considered as a property hasSameNoiseAs with Monster as its domain, and a literal as its range. The class Monster is defined as a subclass of Character in FRBRoo, allowing to use FRBRoo to represent the relations of works. Fig. 1 demonstrates a graphic version of the Shanhaijing ontology.

Instance-level data was normalized by mapping it to the ontology, and the Silk-Link Discovery Framework [20] was used to automatically link appropriately matched resources to external datasets (DBpedia, Wikidata and Schema.org) using owl:sameAs, owl:equivalentClass and owl:equivalentProperty.

An interactive data explorer software tool (iSHJ) was built for visualizing, querying (through a SPARQL interface) and analyzing the data. The dataset, ontology, and source code are available via GitHub. 

V. USER INTERFACE

iSHJ was built as a domain specific application for Shanhaijing. This tool has “Browse”, “Search” and “Visualization” interaction modes. Users are provided with quick search functions to explore the data by clicking buttons and inputting keywords rather than writing SPARQL queries directly, although they can be when more flexible and variable searches are needed (see Fig. 2). The results are displayed both in plain text and charts for visualization (see Fig. 3). We also provide a graphical version of Shanhaijing, where mountains are placed to represent the locations described in the book, the monsters correspondingly placed on the mountains where they live. A video of the UI is available.

VI. DISCUSSION

Despite the work on the narrative of separate regions of some prominent Western myths, projects focused on Chinese literature within this interdisciplinary field are rare. Existing LD methods have been developed almost exclusively in the context of Western culture, and predominantly for highly structured data. When facing ancient Chinese mythologies, there are two main unsolved challenges: non-existent structured datasets and the unavailability of reusable ontologies.

Before LD methods can be applied to the narrative of Chinese myths, a structured dataset capturing in-depth knowledge of Chinese mythologies must be constructed. However, the full potential of this pioneering project can only be tapped into once a greater number of external, disparate, but complementary datasets are published using the LD paradigm. That is to say, until other projects focusing on the analysis of Chinese literature engage in LD, there are limited opportunities for outward linkage.

The protagonists of Eastern and Western mythologies are not entirely similar. For example, in many ancient Chinese mythologies (such as Shanhaijing), numerous gods and creatures are described as a monstrous combination of different animals, falling somewhere between, for example, the human-like (both physically and emotionally) gods and heroes of Greek myths and the anthropomorphized animals of Aesop’s tales. Although there are some complementary aspects – e.g. in the Aarne-Thompson’s Motif-Index, the Nine-tailed Fox (九尾狐) [13] [9] [4], is recorded as B15.7.7.1; other motifs are the four-eyed tiger (B15.4.1.2.) and serpent with a jewel in its mouth (B103.4.2.) – these ontologies neither contain the narrative of Chinese myths nor are they created for Chinese folktales. Ultimately, the existing overlaps are insufficient.

Based on the differences in the narratives, most ontologies created for Western folktales are not completely suitable for the representation of ancient Chinese mythic classics and could not adequately demonstrate the characteristics of these gods and monsters in the Shanhaijing.

VII. CONCLUSION

We used LD methods for textual analyses and visualization of a book of Chinese mythology, Shanhaijing. We created and published a structured dataset, relevant LD, and an interactive explorer to represent the monsters within the text. An extensive review of existing ontologies for literary motifs and mythical creatures revealed that the was insufficient overlap between them and the needs of the dataset, necessitating the development of a new ontology. Future work will see us expand this analysis to all the contents of Shanhaijing. New ontologies will be generated.
from the one in this paper, and structures will be redetermined and improved to adapt to other mythologies. Other ontologies could be reused or interlinked to, increasing the number of linked elements.

We will also test the suitability of our ontology on other mythologies, ranging from Chinese mythologies appearing before and after Shanhaijing to other Asian mythologies such as Japanese tales. We will also apply our ontology to Western mythologies to assess the similarities and differences between Eastern and Western folk tales.

REFERENCES
Figure 2: Results of Quick Search Example “monster” in “Browse” Section in iSHJ

Figure 3: SPARQL Query Sample of Monster’s Tail Number with Visualization Results in iSHJ


