PanayHub: Digital Ontology with ChatBot on Folk Tales, Myths, and Legends from Panay Island

A Special Problem Proposal

Presented to

the Faculty of the Division of Physical Sciences and Mathematics

College of Arts and Sciences

University of the Philippines Visayas

Miag-ao, Iloilo

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Science in Computer Science by

CORDERO, Palmsdale Kevin

PEÑAFIEL, Jeremy Jobert

TRIFALGAR, Michael Rainiken

Francis DIMZON

Adviser

April 28, 2025

Abstract

Currently, little work is being done on the development of digital ontologies, particularly that of the folklore of Western Visayas. However, there exists a digital ontology developed by Dimzon and Dimzon (2015a) which stores various Western Visayan oral traditions, including folk narratives. To fill this digital preservation gap, the researchers enhanced and expanded the original ontology to accompany more depth of information and store more folk narratives from Panay Island, specifically myths, legends, and folk tales. In addition, the researchers developed a chatbot capable of providing insights and details on the stored Panayanon folk narratives. Specifically, the researchers to created a knowledge base of Panayanon folk narratives and subsequently developed and trained a chatbot to understand and answer inquiries about the Panayanon folk narratives.

Keywords: Philippine folk literature, Digital preservation, Ontology-based system, Chat bot

Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 Overview

Philippine folk literature is the body of oral literature of the Filipino people. Folk literature typically undergoes classification into three categories: folk narratives, folk speech, and folk songs. Myths, legends, and folktales are included in the category of folk narratives, a form of literature that provides a narrative through prose or verse, and will serve as the focus of this project. Myths and legends are both regarded as truthful accounts of the past that provide explanations for the origins of entities in the environment. However, myths are often sacred and linked with religion, whereas legends tend to be secular in nature. On the other hand, folktales are fictitious prose narratives typically employed for entertainment purposes (Eugenio, 2007). In addition to their roles in explaining origins or providing entertainment, these three forms of folk literature often function as mediums for the communication of morals, traditions, and beliefs of the Filipino people. Eslit (2023) explored 10 popular folklores in the Philippines, examining their portrayal of Filipino culture and identity. Common themes in the analyzed folklore include environmental importance, respect for elders, and justice. These forms of folk literature have played significant roles in the conveyance and instillment of key values, traditions, and identity within particular ethnolinguistic groups. However, as Eugenio (2007) notes, there is a significant lack of collections of Philippine folk literature. Consequently, research on Philippine folk literature presents difficulties due to its wide dispersion across the country, the necessity for translations, and the rapid decline of this literary form, which limits available research. While there has been some work addressing these challenges, access has been limited due to cost and dated nature.

According to (Dimzon & Dimzon, 2015a), there exists no digital ontology of Western Visayas folklore as digital ontology development was a new area of research. Their pioneering work serves as the start of the digitization of the Western Visayas folklore and is the basis of the researchers' work. With this, researchers propose the development of an ontology-based chatbot capable of answering questions and providing information about folk narratives, particularly those from Panay.

Jepsen (2009) offers a practical definition of ontology. Specifically, ontology as "a method of representing items of knowledge (ideas, facts, things—whatever) in a way that defines the relationships and classifications of concepts within a specified domain of knowledge." A chatbot is a software agent with the capability for engaging in human-like conversation. The researchers aim to provide the chatbot with knowledge and understanding of the relationships between concepts found in Panayanon folk narratives, which enables it to answer queries about them. Through the proposed system, the creation of a central hub of knowledge on Panayanon folk narratives facilitates the streamlining and accessibility of research and education on Panayanon folk narratives. Furthermore, the proposed system contributes to the preservation and promotion of cultural diversity and heritage, as globalization heightens the threat of the deterioration and disappearance of cultural heritage (UNESCO, 2001).

1.2 Problem Statement

The body of knowledge regarding Philippine cultural heritage, specifically Philippine folk literature, remains limited. Despite efforts to collect and analyze this literature, the accessibility of such research is constrained by the cost of resources and the outdated nature of existing works. Eugenio (2007) affirms the lack of comprehensive collections and accessible resources on Philippine folk literature, resulting in significant challenges in the study, documentation, and promotion of this literary form.

Damiana Eugenio, recognized as ”Ina ng Folklor ng Pilipinas” by the U.P. Folklorists, Inc. and the U.P. Folklore Studies Program, has made significant contributions to the preservation of Philippine cultural heritage. Her book Philippine Folk Literature: An Anthology—the first volume in a seven-volume series—compiled over 150 texts and selections of proverbs and riddles from across the Philippines. However, due to the rapid digitization of global information and the fact that her works are now over 15 years old, their accessibility continues to diminish.

Recent efforts have sought to address this issue, with projects like the Aswang Project, created in 2006 by Jordan Clark. This project serves as an online resource for Philippine folklore, featuring articles about various myths, creatures, and spirits found throughout the country. Furthermore, in the terminal report of Dimzon and Dimzon (2015a), they have collected and digitized Panayanon myths and legends by creating ontologies using Web Ontology Language (OWL). However, their work is not made publicly available and has not included folk tales from Panay; gaps remain in the collection of Panayanon folk narratives, which the researchers aim to explore further.

In the field of chatbots, Shawar and Atwell (2007) note that chatbots are designed to accommodate users' natural tendency to express their wishes through speaking, typing, or pointing (Zadrozny et al., 2000). Consequently, chatbots present potential as educational tools, particularly as information retrieval systems. By offering quick and convenient responses similar to human interaction, chatbots hold promise for facilitating research and education. This potential is evidenced by the rapid growth of OpenAI's ChatGPT, an artificial intelligence chatbot that gained one million users within days of its launch (Mortensen, 2024).

1.3 Research Objectives

1.3.1 General Objective

The researchers aim to further expand the original digital ontology by Dimzon and Dimzon (2015a), and develop a chatbot equipped with the ontology-based framework to answer questions about Panayanon folk narratives. Ultimately, the project output should be able to contribute to the preservation, accessibility and study of Panayanon folk literature.

1.3.2 Specific Objectives

Specifically, the researchers aim to:

1. Enhance the existing ontology by adding story elements as new classes, such as events and settings. Through this, additional details of the new folk narratives can be captured and queried.

2. Expand the scope of the existing ontology by adding new entities, attributes, and relationships from Panayanon myths, legends, and folk tales.

3. Develop a prototype chatbot capable of understanding English questions and responding with accurate and appropriate information from the enhanced and expanded digital ontology.

1.4 Scope and Limitations of the Research

The primary focus of this project is on the expansion and enhancement of the original digital ontology, which was first developed by Dimzon and Dimzon (2015a). The scope of the folk literature analyzed for the digital ontology will be limited to folk narratives originating from the island of Panay, specifically myths, legends, and folk tales only. Further, these stories will be limited to those available during the project timeline, relying on existing research, expert consultations, and accessible resources. By building upon and expanding the original ontology, the researchers will ensure comprehensive coverage of the key entities and relationships within Panayanon folk narratives. Due to possible legal issues with ownership in data, the ontology will not contain the whole story but rather specific data that can help in research.

The native languages used in Panayanon folk narratives are Panayanon languages, namely Hiligaynon, Aklanon, and Karay-a. However, the language used in the development of the ontology and the chatbot will be in English. This is to ensure ease of use in academic and global contexts, thereby improving the accessibility of the ontology to a broader audience. Character names and other proper nouns will be kept in the original language to preserve authenticity. To enhance the scope of the ontology, new classes will be created, such as events and settings, which were not present in the original ontology. The researchers will be consulting with literature experts to ensure that the new classes are relevant. As such, more classes may be introduced based on the suggestions of experts.

The chat bot will primarily be used as the tool for information retrieval from the ontology. It will be developed as a prototype, focusing on demonstrating feasibility rather than full-scale deployment. Multilingual capabilities, advanced natural language processing for more complex queries, and deployment-level optimizations are beyond the scope of this project. Future projects may address these limitations.

1.5 Significance of the Research

The study holds significant value for the field of Panayanon cultural heritage and preservation for the following reasons:

The proposed system addresses the problem identified by Eugenio (2007) regarding the lack of published collections of Philippine folk literature. By serving as a central repository of knowledge for Panayanon folk narratives, the system is expected to facilitate easier access to Panayanon folk literature for researchers, students, educators, and the general public.

Additionally, the system seeks to address the issue of the decline of Panayanon oral literature by systematically collecting and digitizing these oral traditions, thereby contributing to their preservation for future generations.

Chapter 2

2 Review of Related Literature

This chapter discusses the features, capabilities, and limitations of existing research, algorithms, or software that are related/similar to the Special Problem.

2.1 Ontologies in Computer Science

This chapter contains a review of research papers that: One of the ultimate goals of ontology as a philosophy is to provide a definitive, exhaustive classification of entities across all spheres of being. However, in the context of computer and information science, this goal has transformed into the pursuit of creating a single unified system that resolves the differences of terminologies and concepts used across diverse data and knowledge-based systems (Smith, 2012). In fact, in their study on ontologies and knowledge-base systems, Kharbat and El-Ghalayini (2008) claimed that ontology has been an emerging computer science discipline for decades. They also concluded that ontologies formalize the semantics of a domain of knowledge by explicitly describing the elements that comprise the domain. This meant that ontologies consisted of concepts that describe the internal features or attributes of an entity, as well as properties that describe the relationships between these entities.

2.1.1 Applications of Ontologies

The aforementioned properties of ontologies in Kharbat and El-Ghalayini's study meant that ontologies are capable of performing a broad range of tasks across diverse research areas. The tasks that are relevant to the study include: the integration of heterogeneous data sources to overcome semantic heterogeneities (Lacroix & Critchlow, 2003); the creation of knowledge bases (Noy, McGuinness, et al., 2001); deriving aspects of information systems at run time (Guarino, 1998), and the construction of an ontology-based retrieval system that can assist end users in browsing and understanding domain concepts (Baker et al., 1999). Furthermore, Munir and Anjum (2018) stated that, with the recent dramatic increase in the use of knowledge discovery applications, there is a growing complexity in terms of the database search requests that the end users are supposed to write to retrieve the information that they wanted. Munir and Anjum (2018) stipulated that these difficulties are attributed to the need for the end users to have a good understanding of the complex structure of databases, and the semantic relationships that exist between different data within the database. It is through the use of ontologies for knowledge representation and interactive query generation that researchers were able to improve the interface between data and search requests, increasing the accuracy of the result sets to the user search requirements. Building upon these applications of ontologies, the study adopts a similar approach, creating an ontological knowledge base that consolidates, organizes, and classifies Panayanon myths, legends, and folk tales that also depicts the settings, character relationships, and themes that are embedded in these Panayanon stories.

2.2 Ontology Development

2.2.1 Ontology Construction

Yadav, Narula, Duhan, Jain, and Murthy (2016) further expounds on the core components that form an ontology. These components of ontologies include: a set of concepts that can serve as nodes in the representation of an ontology; an optional set of properties related to the concepts, these properties can also be summarized as the values of the concepts; a set of relational properties that implies relationship between two or more concepts, often generating a hierarchical path from one concept to another; a hierarchy of concepts and a hierarchy of properties as a result of the relational properties linking one concept to another; a transitive property relation that expands and allows for logical inference on relationships between properties; i.e., if Property A is related to Property B, and Property B is related to Property C, then Property A will be necessarily related to property C; symmetry and inverse symmetry relations among properties; domain values related to properties that define the level of properties within classes, indicating that concepts that share the same property values have the same domains; range values related to the properties which can either be an interval, a list of elements, or a character; and minimum and maximum cardinality for each concept-property pair that define how many properties are associated with a particular concept. These core components of ontologies will be applied in developing the ontology for this study.

Yadav et al. (2016) also listed the basic steps in constructing ontologies. According to their study, the first step in constructing ontologies is determining its scope. These include defining the structure of the ontology as well as the values that are associated with the ontology. Next, is the consideration of reusing ontologies. Yadav et al. (2016) stated that it's possible to re-use recent ontologies in defining the schema of the new ontology that is to be constructed. Third, is the enumeration of terms, where all terms must be clearly specified, together with the domain and range of the ontology. Fourth, is the definition of the taxonomy, where all terms are organized in a hierarchy. For example, if A is a subclass of B, then every instance of class A must be an instance of B. Fifth, is the definition of properties, which includes specifying the properties that link the classes while organizing them in a hierarchy. Next, is the definition of facets which is defined as the hierarchy of homogeneous terms that describe an aspect of the domain where each term in the hierarchy refers to a different concept (Giunchiglia, Dutta, Maltese, & Farazi, 2012). For example, if a domain is space, then facets might refer to bodies of water, land formations, and administrative divisions. Finally, the last step of ontology construction is the definition of instances within the ontology. The steps outlined by Yadav et al. (2016) will be applied in constructing the ontology for this study. This includes the reuse of an existing ontology, building upon it by incorporating additional concepts, classes, and all of the other aforementioned core components to expand the ontology's scope and application.

The construction of the ontology will be done through Protege, an open-source knowledge requisition system written in Java (Yadav et al., 2016; Jain & Singh, 2013). More specifically, it's an ontology development editor that is capable of defining ontological concepts or classes, properties, taxonomies, and class instances. Protege supports ontology representation languages like OWL. Aside from constructing ontologies, Zhao, Zhang, and Zhao (2012) states that Protege is also capable of parsing an Ontology model using a Protege-based OWL API. Protege is able to: load an ontology model from the OWL file; collect the classes, subclasses, object properties, data properties; and find the domain and range relevant to a particular object property. The study will be using Protege Desktop v.5.6.4 in developing the ontological database for the Panayanon stories.

2.2.2 SPARQL for Ontology Querying

SPARQL 1.1 is a set of specifications that provide languages and protocols to query and manipulate RDF graph content on the Web or in an RDF store. The standard SPARQL Query Results are written in an XML Format, and in three other alternative formats: JSON, CSV, and TSV (Picalausa & Vansummeren, 2011). SPARQL 1.1 is the query language the Protege uses to retrieve, and manipulate ontological data.

2.2.3 ApacheJena for Ontology Storage

According to the Apache Community Development Project (n.d.), ApacheJena is able to provide a complete framework for building Semantic Web and Linked Data applications in Java. ApacheJena is also equipped with the following capabilities: parsers for Turtle, N-triples, and Resource Description Framework (RDF), and Extensible Markup Language(XML); an API for programming with Java; a complete implementation of the SPARQL query language for ontological querying; a rule-based inference engine for RDF Schema (RDFS) amd OWL entailments; a Triple Database (TDB) which is a non-SQL persistent triples store; a Semantic Database (SDB) which is a persistent triples store built upon a relational store, and Fuseki, an RDF server that uses web protocols. The Apache Software Foundation claims that ApacheJena complies with the relevant recommendations for RDF and related technologies from the World Wide web Consortium (W3C).

In a study conducted by Chokshi and Panchal (2022), they were able to construct a Job Search Ontology on Protégé, integrated the ApacheJena Fuseki Server with the ontology, and executed SPARQL queries on the ApacheJena Fuseki Server without using the Protégé tool. This study demonstrated that it is possible to construct a SPARQL endpoint with Apache Jena. ApacheJena will be mainly used for storing data about the study's ontology. An Apache Fuseki Server will publish the study's ontology as a SPARQL endpoint, making it available for querying and data sharing over the internet.

2.3 Natural Language Question to SPARQL Translation

2.3.1 Natural Language Question (NLQ) Preprocessing

spaCy is an open-source library for advanced natural language processing (NLP) in Python. spaCy is designed to handle preprocessing tasks with high efficiency and speed. spaCy's features and functionalities include: tokenization, lemmatization, part-of-speech (POS) tagging, and named entity recognition (Nawaz, 2023; SpaCy Documentation, n.d.). In the study, spaCy will be used to preprocess the NLQ through tokenization, and lemmatization.

2.3.2 Entity and Relationship Extraction with Semantic Parsing

According to Nawaz (2023), spaCy is capable of named entity recognition (NER) and dependency parsing. In the study, spaCy's NER and dependency parsing will be used to extract entities like folk tale titles, names of researchers, character names, and even the relationships between entities. These will be passed to the SPARQL query constructor to create a SPARQL query and retrieve information from the study's ontology.

2.3.3 Semantic Parsing with SBERT

Sentence Transformers or SBERT, is a Python module used for accessing, using, and training text and image embedding models. It can be used to compute embeddings using Sentence Transformer models or to calculate similarity scores using Cross-Encoder models. SBERT's features and functionalities include: semantic search, semantic textual similarity, and paraphrase mining. The Semantic Textual Similarity (STS) application aims to produce embeddings for all texts involved and calculate the similarities between them. The text pairs with the highest similarity score are considered to be the most semantically similar (SentenceTransformers Documentation, n.d.). In the study, STS will be used to embed phrases in the NLQ and compare them with the embeddings of the ontology's object and data property labels. STS will also be used to help resolve ambiguous queries where multiple relationships can potentially be extracted from the query.

2.3.4 Query Construction/Generation

RDFLib is a pure Python package made for working with RDF. RDFLib's features and functions include: parsers and serializers for RDF/XML, N3, NTriples, N-Quads, Turtle, TriX, JSON-LD, HexTuples, RDFa and Microdata; Store implementations like memory stores, and remote SPARQL endpoints; Graph interface either to a single graph or to multiple named graphs; and SPARQL 1.1 implementation (RDFLib Team, n.d.). In the study, RDFLib will be used to dynamically generate SPARQL queries together with the extracted entities, and relationships of the NLQ.

2.4 Chatbot Development

2.4.1 RASA Framework

Rasa Open Source is a Python framework that enables teams to build chatbots, voice assistants, and other automated conversation systems by connecting to messaging channels and third party systems through a set of APIs (Rasa Technologies, 2024).

In a study conducted by Mishra, Agarwal, Swathi, and Akshay (2022), they created a closed domain ontology for a hostel system usingProtégé, which was then referenced by an AI-powered chatbot through RASA that was able to formalize natural language queries into SPARQL to query knowledge bases. More specifically, in the study they were able to design a natural language query formalization pipeline that had intent recognition to determine the type of the user's natural language query, entity extraction, and query generation to translate the query's intent and extracted entities into a SPARQL query. The study by Mishra et al. (2022) has shown that it's possible to incorporate a NLQ to SPARQL pipeline within the chatbot. In the study, RASA open source will be used to construct the chat-bot.

Chapter 3

3 Research Methodology

This chapter lists and discusses the specific steps and activities carried out to complete the project.

3.1 Research Activities

As illustrated in Figure 1, the researchers conducted a series of research activities. They have first consulted domain experts to gather data, clarify its interpretation, and discuss enhancements on the ontology's structure. The gathered data was then encoded into the digital ontology, incorporating the suggested enhancements. Parallel to this, chatbot development began, using only some of the initially encoded data to expedite progress rather than waiting for the completion of encoding all gathered folk narratives. The chatbot was trained and tested on the basis of specified metrics, then deployed on a website.

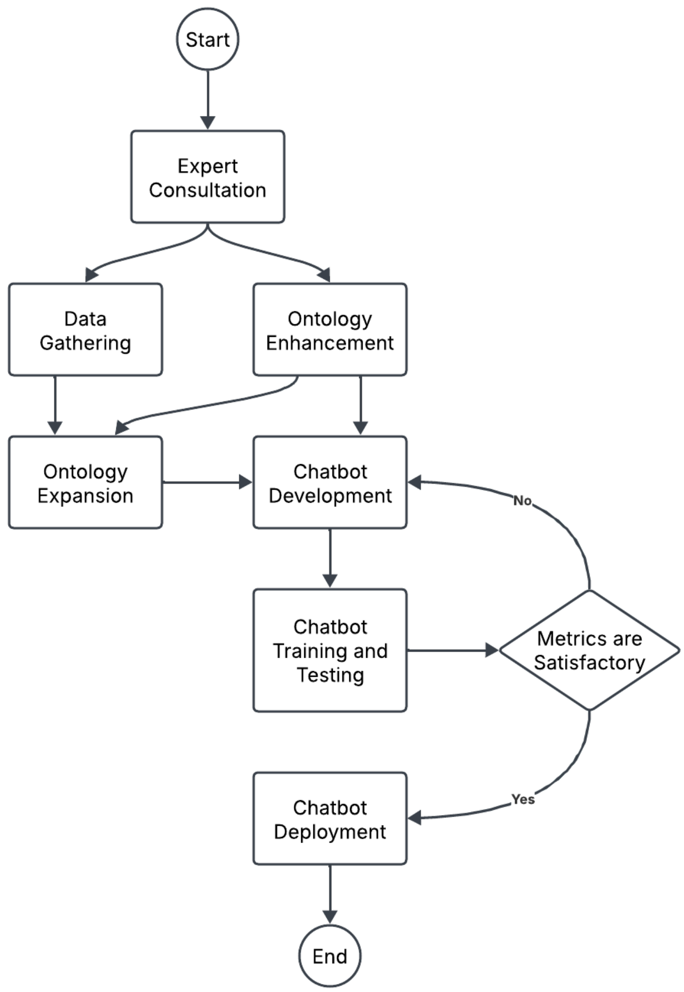

Figure 1: Process Diagram of Special Project

3.2 Ontology Development

3.2.1 Data Collection

The researchers gathered Panayanon myths and legends from the terminal report by Dimzon and Dimzon (2015b), which compiles various aspects of West Visayan culture and was presented to the U.P. Visayas Center for West Visayan

Studies (CWVS), a research and extension arm of UPV. In addition to this report, the researchers were provided with the accompanying digital ontology, which serves as the foundational groundwork for this special project.

Given the scope of the project, only stories from Panay Island were included and others were ignored. While the researchers considered incorporating additional stories from other sources, Mrs. Dimzon explained that research and anthologies from different sources might introduce classifications that differ from those used in her work. The researchers thought this may potentially overcomplicate the structure of the digital ontology; hence, they opted to work exclusively with her work.

3.2.2 Ontology Enhancement

Based on consultations with the domain expert Mrs. Dimzon, a few classes were created in the digital ontology. The new classes were created using Protégé, an open-source ontology editor that supports OWL (Web Ontology Language) for formalizing domain knowledge. Protégé has features such as logical constraints and reasoning, which were utilized in ensuring consistency and inferring class hierarchy. Each class was connected through relationships with other entities to create a structured and interconnected narrative representation. These additions enable the ontology to store more information, allowing users to apply specific filters when making more complex queries.

3.2.3 Ontology Expansion

The researchers have consulted with Mrs. Dimzon in how to properly identify key story elements, and the contextual relationships between entities in a story. With this, the researchers closely read and examined each story from the collected data, looking for relevant story elements and relationships. From their findings, they expanded the digital ontology by populating it with new stories, entities, and relationships based on the enhanced ontological structure.

Protégé was utilized for ontology expansion for its extensive support in OWL files and SPARQL querying, reasoning, and consistency checking features.

3.3 Chatbot Development

3.3.1 Chatbot Development Tools

The researchers developed a chatbot prototype that can handle English queries from users, query the ontology to search for relevant data, and present the information to the user in comprehensible English sentences. Specifically, the researchers utilized Python as the primary programming language to develop the chatbot, spaCy as the natural language processing (NLP) library to analyze and process user queries, Rasa as the machine learning framework to extract the entities and intents of user inputs, GraphDB as the knowledge base to host the ontology, and a natural language generation (NLG) server to generate conversational responses to the user.

3.3.2 User Experience

To enhance user experience, the chatbot was designed to exhibit a conversational tone while maintaining the accuracy and professionalism required for ontology-based information retrieval. By incorporating NLG techniques, the chatbot aims to engage users with dynamic and contextually appropriate responses that emulate human-like interaction. This conversational approach is expected to improve user engagement and satisfaction, especially when addressing more complex queries that may require clarification or follow-up interactions.

With this chatbot, users will be able to interact with the ontology in natural language in a conversational and friendly manner. This is in pursuit of data querying, which is Manansala, Bruskiewich, and Naval (2007)) third and final pillar of ontology frameworks.

3.3.3 Rasa Framework

Figure 2 shows the flow of data from user input to chatbot response generation. The ontology is converted into a graph database that can be queried for relevant information. The Rasa Agent works as a controller to easily orchestrate the dialogue flow of the chatbot, and manage the interaction of the different components.

When the user inputs a message, the NLU pipeline will first process it to extract its entities and intents. The extracted information is then passed through the Rasa Agent to the Dialogue Policies, which determine the appropriate action. If an external query is required, the Action Server requests information from the Knowledge Base. The data retrieved from the query are passed through an NLG server that contains Rasa's contextual response rephraser to generate more natural and conversational responses. The Rasa Agent receives this to then output the response to the user. The Tracker Store also receives the extracted entities, intents, and executed actions in order to maintain conversation history, which will help the Dialogue Policies make more context-aware decisions over time.

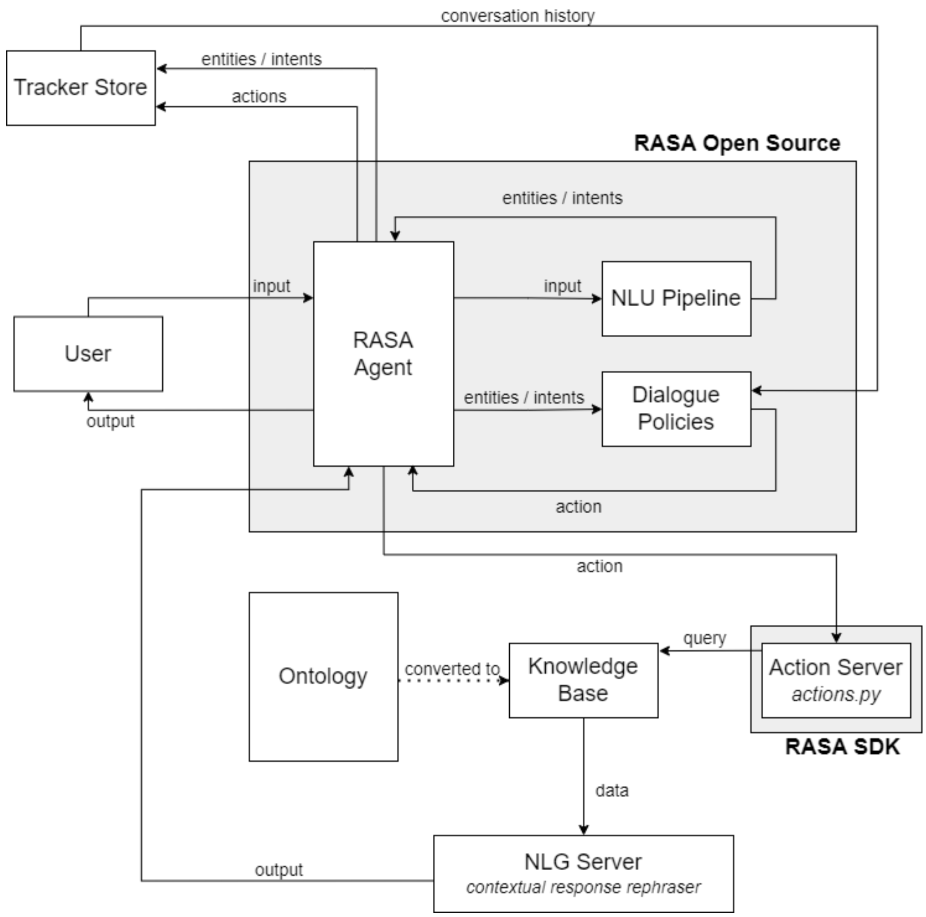

Figure 2: Diagram of Rasa Framework

Figure 3 illustrates Rasa's natural language understanding pipeline. The first component is the spaCy tokeniser, which splits the user's input into tokens. The second component is the spaCy featurizer, which is a dense featurizer that extracts features used for entity extraction, intent identification of the user's message, and response classification. The regex featurizer is a sparse featurizer that creates a vector representation of the user's message using regular expressions for the purpose of entity extraction and intent identification. Next is the lexical syntactic featurizer, which is a sparse featurizer that creates lexical and syntactic features for a user's message for the purpose of entity extraction. The fifth component is the count vectors featurizer, which generates a bag-of-words representation of the bot user's message, intent, and response for the purpose of intent identification and response selection. The DIET Classifier is a multi-task transformer architecture that is responsible for intent classification and entity extraction. The final component is the Entity Synonym Mapper that maps entities to their synonyms if they appeared in the training data. The extracted entities and identified intents in the NLU pipeline will then be passed to the chatbot dialogue policies to determine the appropriate actions that the bot will perform.

The components within the pipeline are used to process the user's input and extracts entities and intents from the user's input. This will be then queried through the knowledge base. Finally, the query results will be formatted in English using NLP techniques.

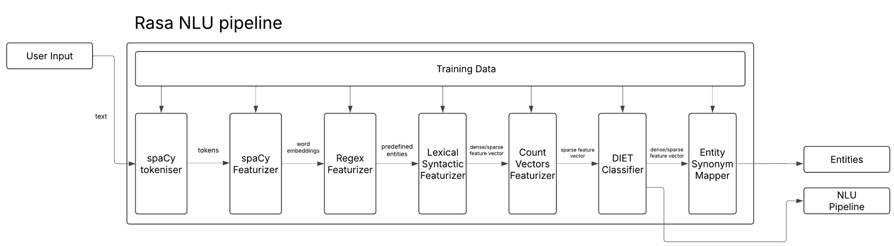

Figure 3: Diagram of NLU Pipeline

3.3.4 Chatbot Testing

The evaluation of the PAROT model by Ochieng (2020) involved the use of QALD-9 challenge metrics, including accuracy, recall, and F-measure. Similarly, the utilization of these metrics will be applied in the evaluation of the model under development of this special project.

With each iteration of the chatbot, the researchers will perform tests to verify chatbot query accuracy and response relevance, assess user interaction with the chatbot, and measure response times and optimize as needed.

3.4 System Deployment

3.4.1 Rasa Chatbot

The chatbot, developed using the Rasa framework, was deployed via the Google Kubernetes Engine (GKE), a managed Kubernetes service provided by the Google Cloud Platform (GCP). The use of GKE facilitated the creation of Kubernetes clusters without the manual setup of the Kubernetes infrastructure (Sullivan, 2023). The deployment process followed the guidelines outlined in the Rasa Open Source documentation.

3.4.2 PanayHub Website

Website hosting was facilitated through the Render.com platform, selected for its usability and streamlined deployment capabilities. Render.com provides infrastructure for hosting websites and servers, with integrated support for popular Git repository hosting services such as GitHub and GitLab. This integration enables automated deployment triggered by each new commit, enhancing development workflow efficiency.

Figures 4, 5, and 6 present the user interface of the website. Figure 4 illustrates the homepage, which functions as the primary landing page and navigation hub for users. Figure 5 depicts the *Contact Us* page, which provides users with a channel for communication with researchers and affiliated institutions. Lastly, Figure 6 displays the chatbot interface, allowing user interaction with the chatbot for inquiries related to the folk narratives of Panay.

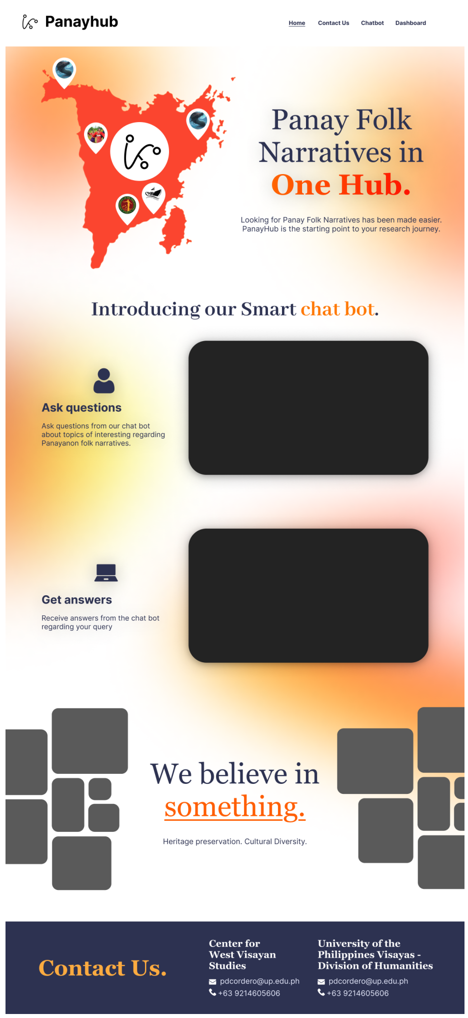

Figure 4: Homepage of PanayHub

Figure 5: Contact Us page of PanayHub

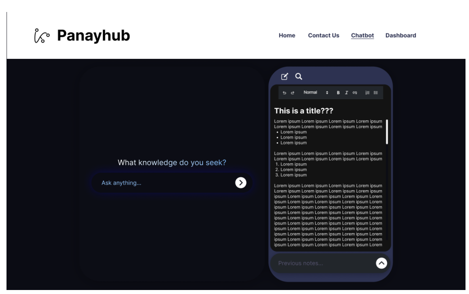

Figure 6: Chatbot page of PanayHub

Chapter 4

4 Results and Discussions

4.1 Data Gathered

The researchers have contacted their resource person Prof. Dimzon on her collection of folk narratives. She gave a Terminal Report Dimzon and Dimzon (2015b) on her completed project on collecting myths and legends from Western Visayas. It listed a total of 189 stories, 28 being myths, 161 being legends and 9 categorized as others. Each folk narrative has already been categorized into their respective types, including etiological legends, non-etiological legends, and others. Below is a list of the different types of folk narratives collected, their subtypes, and their count.

I. Myths: 28

II. Legends: 161

- A. Etiological Legends: 69
  - i. How Legends: 59
    - a. Origin of Animals: 14
    - b. Origin of plants and forms of plant life: 4
    - c. How places and things got their names: 41
- B. NonEtiological Legends: 83
  - i. Heroic Legends - great men, culture heroes: 18
  - ii. Religious/Saints Legends: 9
  - iii. Legends on Supernatural/Enchanted Beings: 56
- C. Others: 9

However, some of these folk narratives were out of scope for the special project as it focuses purely on stories from Panay Island only. Below is the number of stories after filtering for Panayanon specific narratives. In summary, there were 21 myths, 108 legends, and 7 categorized as others.

I. Myths: 21

II. Legends: 108

    A. Etiological Legends: 43

        i. How Legends: 36

            a. Origin of Animals: 9

            b. Origin of plants and forms of plant life: 3

            c. How places and things got their names: 24

    B. NonEtiological Legends: 64

        i. Heroic Legends - great men, culture heroes: 11

        ii. Religious/Saints Legends: 9

        iii. Legends on Supernatural/Enchanted Beings: 44

    C. Others: 7

4.2 New Classes

As per consultation with Mrs. Dimzon, the classes in Table 1 were added to expand on the original ontology.

Class	**Subclass**	**Subsubclass**
GeographicFeature		
	Landform	
		Mountain
		Hill
		Forest
		Forest
		Island
		Volcano
	BodyOfWater	
		River
		Sea
		Ocean
		Lake
Language		
	Hiligaynon	
	Kinaray-a	
	Akeanon	

Table 1: Hierarchy of Classes, Subclasses, and Subsubclasses

Figure 7 presents the class hierarchy of the objects in the original ontology as presented in Protege. Classes are categories or types of things in the ontology, representing a group of objects or individuals that share common characteristics. Instances of these classes are called individuals in Protege, representing a specific thing that belongs to the class.

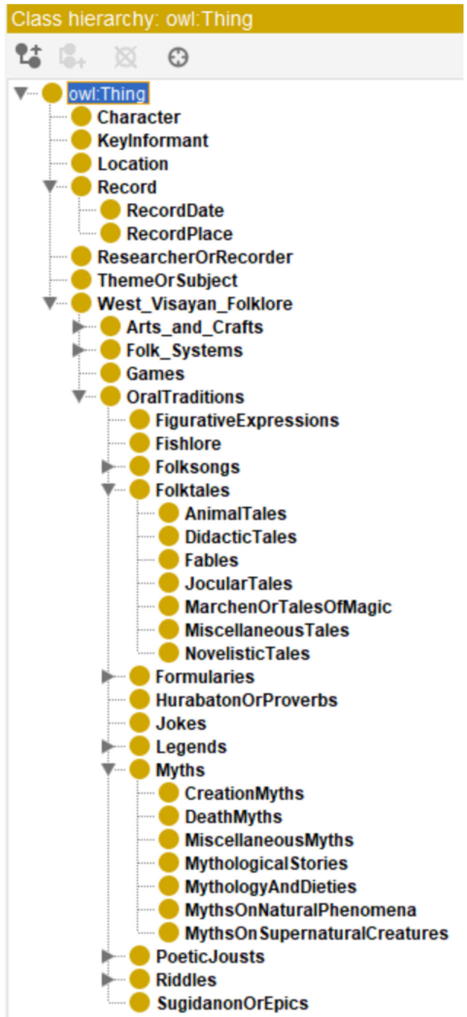

Figure 7: Class Hierarchy of Original Ontology

With the newly introduced classes, Figure 8 presents the class hierarchy of the objects in the enhanced ontology as presented in Protege. Specifically, the classes from Table 1 were added.

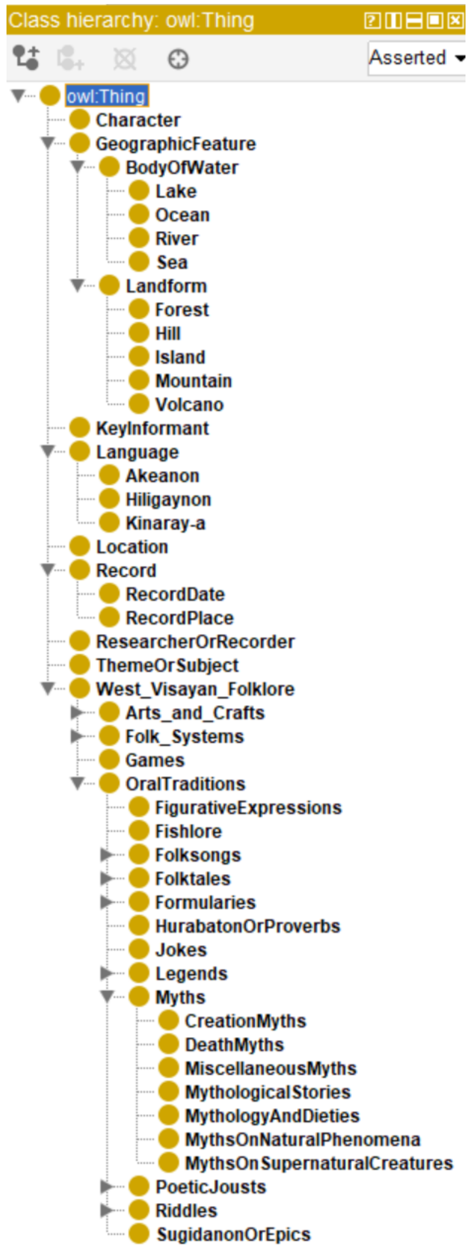

Figure 8: Class Hierarchy of Enhanced Ontology

Figure 9 presents the hierarchy of the object properties in the original ontology as presented in Protege. Object properties define relationships between two individuals in the ontology, and are used to link classes or instances. In the ontology enhancement phase, the researchers will introduce new object properties to accommodate the new classes.

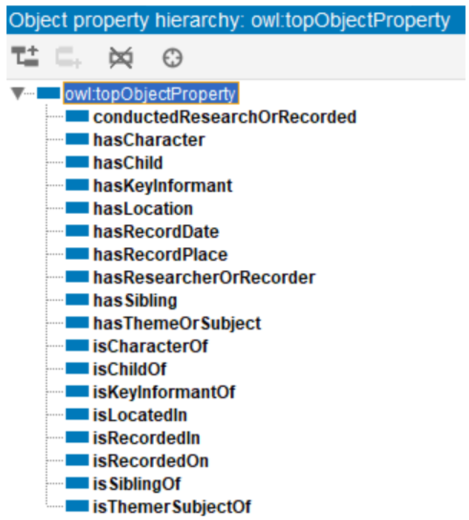

Figure 9: Object Property Hierarchy of Original Ontology

With the newly introduced classes, Figure 10 presents the the hierarchy of the object properties in the enhanced ontology as presented in Protege. Specifically, the object properties 'hasLanguage', 'isLanguageIn', 'hasGeographicFeature', and 'isGeographicFeatureIn' were added.

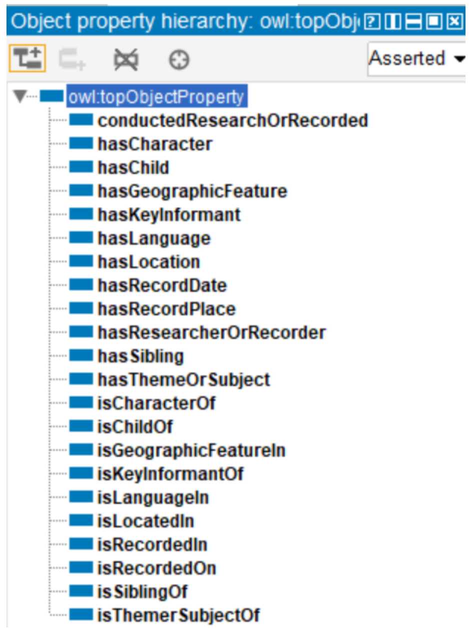

Figure 10: Object Property Hierarchy of Original Ontology

Figure 11 presents the hierarchy of the data properties in the original ontology as presented in Protege. Data properties define relationships between an individual and a literal value, such as a string, number, or date. There was no need to add new data properties with the introduction of the new classes.

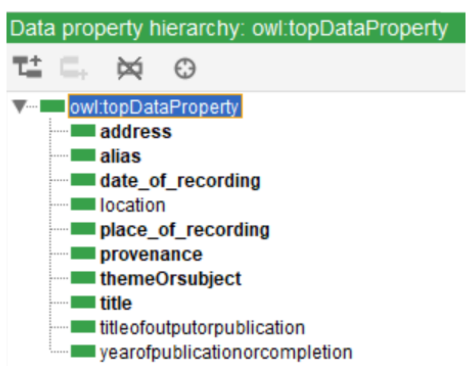

Figure 11: Data Property Hierarchy of Original Ontology

Chapter 5

5 Conclusion

In this special problem project, the researchers built upon the foundational work of (Dimzon & Dimzon, 2015b) in the digital preservation of folk narratives from Panay Island. They enhanced the original ontology by introducing new classes and relationships that enable a richer and more detailed modeling of the stories' elements, thereby increasing its utility for complex scholarly analysis.

The researchers successfully gathered a total of 137 Panayanon folk narratives, systematically encoding specific story elements into the expanded digital ontology based on the enhanced form. This effort contributed to the preservation of folk narratives by making them accessible in a structured, machine-readable, and queryable format. The resulting ontology serves as a foundational resource for future academic research in the humanities.

To accomplish Manansala et al. (2007)) third and final pillar of ontology frameworks, the researchers developed a chatbot as an intuitive querying tool. Designed to be user-friendly, the chatbot enables users to retrieve the gathered data without requiring technical knowledge of ontology structures or query languages. The chatbot and the digital ontology were deployed through a website, ensuring that the repository of Panayanon folk narratives can be accessed.

5.1 Limitations

This study focused on a single anthology of Panayanon folk narratives, building upon the foundational work established by Dimzon and Dimzon (2015a). While this approach ensured consistency with the existing ontology structure, it necessarily limited the scope of the data to a single source.

Furthermore, the developed chatbot was constructed as a prototype querying tool and remains constrained by the researchers' limited expertise in the domains of the humanities and literature. Particularly, they were challenged in understanding how to structure queries in a manner conventional to these fields.

The deployment of the application was constrained by certain limitations, primarily due to resource restrictions associated with the use of the GCP free tier. In an effort to minimize costs, the researchers opted for this tier, which provides limited allocations of CPU and RAM. However, the resource requirements of the Rasa chatbot approached the bounds of these allocations, thereby posing challenges to the stability and scalability of the deployment.

5.2 Recommendations

To enhance the breadth and richness of the digital ontology, future work should consider incorporating additional folk narratives documented by other researchers. Expanding the corpus in this way would not only diversify the representation of Panayanon traditions but also contribute more substantially to the preservation of West Visayan cultural heritage in digital form. Such efforts could serve as valuable references for future research in related fields.

Moreover, extensive user testing involving students, faculty, and researchers specializing in literature and folklore is recommended. Insights from these user groups would provide critical feedback for refining the chatbot's response accuracy, improving its handling of complex queries, and enhancing its overall utility as a querying tool.

Finally, future iterations of the application would benefit from deployment under a paid cloud hosting tier. Access to expanded computational resources would enable greater reliability, support higher user load, and allow for more effective scalability.

References

Apache Community Development Project. (n.d.). *Apache Jena.* Retrieved from `https://projects.apache.org/project.html?jena`

Baker, P. G., Goble, C. A., Bechhofer, S., Paton, N. W., Stevens, R., & Brass, A. (1999). An ontology for bioinformatics applications. *Bioinformatics (Oxford, England)*, *15*(6), 510–520.

Chokshi, H. J., & Panchal, R. (2022). Using apache jena fuseki server for execution of sparql queries in job search ontology using semantic technology. *International Journal of Innovative Research in Computer Science & Technology*, *10*(2), 497–504.

Dimzon, E., & Dimzon, F. (2015a). *Developing a digital ontology of the west visayan folklore heritage (phase 1: The oral traditions).* (Unpublished)

Dimzon, E., & Dimzon, F. (2015b). *Myths and legends terminal report.* (Unpublished)

Eslit, E. R. (2023). Resilience of philippine folklore: An enduring heritage and legacy for the 21st century. *IJELR: International Journal of Education, Language, and Religion*, *5*(1), 9–20.

Eugenio, D. L. (2007). *Philippine folk literature: An anthology* (Vol. 1). UP Press.

Giunchiglia, F., Dutta, B., Maltese, V., & Farazi, F. (2012). A facet-based methodology for the construction of a large-scale geospatial ontology. *Journal on data semantics*, *1*, 57–73.

Guarino, N. (1998). *Formal ontology in information systems: Proceedings of the first international conference (fois'98), june 6-8, trento, italy* (Vol. 46). IOS press.

Jain, V., & Singh, M. (2013). Ontology development and query retrieval using protégé tool. *International Journal of Intelligent Systems and Applica-*

tions, *9*(9), 67–75.

Jepsen, T. C. (2009). Just what is an ontology, anyway? *IT Prof.*, *11*(5), 22–27.

Kharbat, F., & El-Ghalayini, H. (2008). *Building ontology from knowledge base systems.* INTECH Open Access Publisher.

Lacroix, Z., & Critchlow, T. (2003). *Bioinformatics: managing scientific data* (Vol. 6) (No. 2). Academic Press.

Manansala, K., Bruskiewich, R., & Naval, P. (2007, 01). An ontology framework for a crop information system.

Mishra, D. S., Agarwal, A., Swathi, B., & Akshay, K. C. (2022). Natural language query formalization to sparql for querying knowledge bases using rasa. *Progress in Artificial Intelligence*, *11*(3), 193–206.

Mortensen, O. (2024, Apr). *How many users does chatgpt have? statistics facts (2024).* Retrieved from

Munir, K., & Anjum, M. S. (2018). The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, *14*(2), 116–126.

Nawaz, M. (2023, 4). *NLP Preprocessing using Spacy.* Retrieved from `https://soshace.com/2023/04/05/nlp-preprocessing-using-spacy/`

Noy, N. F., McGuinness, D. L., et al. (2001). *Ontology development 101: A guide to creating your first ontology.* Stanford knowledge systems laboratory technical report KSL-01-05 and ....

Picalausa, F., & Vansummeren, S. (2011). What are real sparql queries like? In *Proceedings of the international workshop on semantic web information management* (pp. 1–6).

Rasa Technologies. (2024, 12). *Introduction to Rasa Open Source Rasa Pro.*

Retrieved from https://rasa.com/docs/rasa/

RDFLib Team. (n.d.). *RDFLiB 7.1.1.* Retrieved from https://rdflib.readthedocs.io/en/stable/

Shawar, B. A., & Atwell, E. (2007). Chatbots: are they really useful? *Journal for Language Technology and Computational Linguistics*, *22*(1), 29–49.

Smith, B. (2012). Ontology. In *The furniture of the world* (pp. 47–68). Brill.

Sullivan, D. (2023). Computing with kubernetes. In *Google cloud certified associate cloud engineer study guide* (p. 157-177).

UNESCO. (2001). *Unesco - text of the convention for the safeguarding of the intangible cultural heritage.* Retrieved from https://ich.unesco.org/en/convention

Yadav, U., Narula, G. S., Duhan, N., Jain, V., & Murthy, B. (2016). Development and visualization of domain specific ontology using protege. *Indian Journal of Science and Technology*, *9*(16), 1–7.

Zadrozny, W., Budzikowska, M., Chai, J., Kambhatla, N., Levesque, S., & Nicolov, N. (2000). Natural language dialogue for personalized interaction. *Communications of the ACM*, *43*(8), 116–120.

Zhao, H., Zhang, S., & Zhao, J. (2012). Research of using protege to build ontology. In *2012 ieee/acis 11th international conference on computer and information science* (pp. 697–700).

A Resource Persons

Prof. Eliodora Dimzon

Faculty

University of the Philippines Division of Humanities

`eldimzon@up.edu.ph`

....