

1 PanayHub: Digital Ontology with ChatBot on Folk Tales, Myths, and Legends

2 from Panay Island

3 A Special Problem Proposal

4 Presented to

5 the Faculty of the Division of Physical Sciences and Mathematics

6 College of Arts and Sciences

7 University of the Philippines Visayas

8 Miag-ao, Iloilo

9 In Partial Fulfillment

10 of the Requirements for the Degree of

11 Bachelor of Science in Computer Science by

12 CORDERO, Palmsdale Kevin

13 PEÑAFIEL, Jeremy Jobert

14 TRIFALGAR, Michael Rainiken

15 Francis DIMZON

16 Adviser

17 April 28, 2025

18

Abstract

19

20

21

22

23

24

25

26

27

28

29

30

31

32

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 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

1	Introduction	1
1.1	Overview	1
1.2	Problem Statement	2
1.3	Research Objectives	4
1.3.1	General Objective	4
1.3.2	Specific Objectives	4
1.4	Scope and Limitations of the Research	4
1.5	Significance of the Research	6
2	Review of Related Literature	7
2.1	Ontologies in Computer Science	7
2.1.1	Applications of Ontologies	7
2.2	Ontology Development	8
2.2.1	Ontology Construction	8
2.2.2	SPARQL for Ontology Querying	10
2.2.3	ApacheJena for Ontology Storage	11
2.3	Natural Language Question to SPARQL Translation	12
2.3.1	Natural Language Question (NLQ) Preprocessing	12
2.3.2	Entity and Relationship Extraction with Semantic Parsing	12
2.3.3	Semantic Parsing with SBERT	12
2.3.4	Query Construction/Generation	13
2.4	Chatbot Development	13
2.4.1	RASA Framework	13
3	Research Methodology	15
3.1	Research Activities	15
3.2	Ontology Development	16

59	3.2.1	Data Collection	16
60	3.2.2	Ontology Enhancement	17
61	3.2.3	Ontology Expansion	17
62	3.3	Chatbot Development	18
63	3.3.1	Chatbot Development Tools	18
64	3.3.2	User Experience	18
65	3.3.3	Rasa Framework	19
66	3.3.4	Chatbot Testing	21
67	3.4	System Deployment	22
68	3.4.1	Rasa Chatbot	22
69	3.4.2	PanayHub Website	22
70	4	Results and Discussions	25
71	4.1	Data Gathered	25
72	4.2	New Classes	26
73	5	Conclusion	31
74	5.1	Limitations	31
75	5.2	Recommendations	32
76		References	33
77	A	Resource Persons	36

78 List of Figures

79	1	Process Diagram of Special Project	16
80	2	Diagram of Rasa Framework	20
81	3	Diagram of NLU Pipeline	21
82	4	Homepage of PanayHub	23
83	5	Contact Us page of PanayHub	24
84	6	Chatbot page of PanayHub	24
85	7	Class Hierarchy of Original Ontology	27
86	8	Class Hierarchy of Enhanced Ontology	28
87	9	Object Property Hierarchy of Original Ontology	29
88	10	Object Property Hierarchy of Original Ontology	30
89	11	Data Property Hierarchy of Original Ontology	30

⁹⁰ **List of Tables**

⁹¹ 1 Hierarchy of Classes, Subclasses, and Subsubclasses 27

92 Chapter 1

93 1 Introduction

94 1.1 Overview

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

118 and dated nature.

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

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

139 **1.2 Problem Statement**

140 The body of knowledge regarding Philippine cultural heritage, specifically Philip-
141 pine folk literature, remains limited. Despite efforts to collect and analyze this
142 literature, the accessibility of such research is constrained by the cost of re-
143 sources and the outdated nature of existing works. Eugenio (2007) affirms the

144 lack of comprehensive collections and accessible resources on Philippine folk
145 literature, resulting in significant challenges in the study, documentation, and
146 promotion of this literary form.

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

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

164 In the field of chatbots, Shawar and Atwell (2007) note that chatbots are de-
165 signed to accommodate users' natural tendency to express their wishes through
166 speaking, typing, or pointing (Zadrozny et al., 2000). Consequently, chatbots
167 present potential as educational tools, particularly as information retrieval sys-
168 tems. By offering quick and convenient responses similar to human interaction,
169 chatbots hold promise for facilitating research and education. This potential is
170 evidenced by the rapid growth of OpenAI's ChatGPT, an artificial intelligence

171 chatbot that gained one million users within days of its launch (Mortensen,
172 2024).

173 **1.3 Research Objectives**

174 **1.3.1 General Objective**

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

180 **1.3.2 Specific Objectives**

181 Specifically, the researchers aim to:

- 182 1. Enhance the existing ontology by adding story elements as new classes,
183 such as events and settings. Through this, additional details of the new
184 folk narratives can be captured and queried.
- 185 2. Expand the scope of the existing ontology by adding new entities, at-
186 tributes, and relationships from Panayanon myths, legends, and folk tales.
- 187 3. Develop a prototype chatbot capable of understanding English questions
188 and responding with accurate and appropriate information from the en-
189 hanced and expanded digital ontology.

190 **1.4 Scope and Limitations of the Research**

191 The primary focus of this project is on the expansion and enhancement of the
192 original digital ontology, which was first developed by Dimzon and Dimzon
193 (2015a). The scope of the folk literature analyzed for the digital ontology will

194 be limited to folk narratives originating from the island of Panay, specifically
195 myths, legends, and folk tales only. Further, these stories will be limited to
196 those available during the project timeline, relying on existing research, expert
197 consultations, and accessible resources. By building upon and expanding the
198 original ontology, the researchers will ensure comprehensive coverage of the key
199 entities and relationships within Panayanon folk narratives. Due to possible
200 legal issues with ownership in data, the ontology will not contain the whole
201 story but rather specific data that can help in research.

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

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

219 1.5 Significance of the Research

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

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

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

230 Chapter 2

231 2 Review of Related Literature

232 This chapter discusses the features, capabilities, and limitations of existing re-
233 search, algorithms, or software that are related/similar to the Special Problem.

234 2.1 Ontologies in Computer Science

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

248 2.1.1 Applications of Ontologies

249 The aforementioned properties of ontologies in Kharbat and El-Ghalayini's
250 study meant that ontologies are capable of performing a broad range of tasks
251 across diverse research areas. The tasks that are relevant to the study in-
252 clude: the integration of heterogeneous data sources to overcome semantic het-
253 erogeneities (Lacroix & Critchlow, 2003); the creation of knowledge bases (Noy,

254 McGuinness, et al., 2001); deriving aspects of information systems at run time
 255 (Guarino, 1998), and the construction of an ontology-based retrieval system that
 256 can assist end users in browsing and understanding domain concepts (Baker et
 257 al., 1999). Furthermore, Munir and Anjum (2018) stated that, with the recent
 258 dramatic increase in the use of knowledge discovery applications, there is a grow-
 259 ing complexity in terms of the database search requests that the end users are
 260 supposed to write to retrieve the information that they wanted. Munir and An-
 261 jum (2018) stipulated that these difficulties are attributed to the need for the end
 262 users to have a good understanding of the complex structure of databases, and
 263 the semantic relationships that exist between different data within the database.
 264 It is through the use of ontologies for knowledge representation and interactive
 265 query generation that researchers were able to improve the interface between
 266 data and search requests, increasing the accuracy of the result sets to the user
 267 search requirements. Building upon these applications of ontologies, the study
 268 adopts a similar approach, creating an ontological knowledge base that consol-
 269 idates, organizes, and classifies Panayanon myths, legends, and folk tales that
 270 also depicts the settings, character relationships, and themes that are embedded
 271 in these Panayanon stories.

272 **2.2 Ontology Development**

273 **2.2.1 Ontology Construction**

274 Yadav, Narula, Duhan, Jain, and Murthy (2016) further expounds on the core
 275 components that form an ontology. These components of ontologies include: a
 276 set of concepts that can serve as nodes in the representation of an ontology; an
 277 optional set of properties related to the concepts, these properties can also be
 278 summarized as the values of the concepts; a set of relational properties that im-
 279 plies relationship between two or more concepts, often generating a hierarchical

280 path from one concept to another; a hierarchy of concepts and a hierarchy of
281 properties as a result of the relational properties linking one concept to another;
282 a transitive property relation that expands and allows for logical inference on
283 relationships between properties; i.e., if Property A is related to Property B,
284 and Property B is related to Property C, then Property A will be necessarily
285 related to property C; symmetry and inverse symmetry relations among prop-
286 erties; domain values related to properties that define the level of properties
287 within classes, indicating that concepts that share the same property values
288 have the same domains; range values related to the properties which can either
289 be an interval, a list of elements, or a character; and minimum and maximum
290 cardinality for each concept-property pair that define how many properties are
291 associated with a particular concept. These core components of ontologies will
292 be applied in developing the ontology for this study.

293

294 Yadav et al. (2016) also listed the basic steps in constructing ontologies.
295 According to their study, the first step in constructing ontologies is determin-
296 ing its scope. These include defining the structure of the ontology as well as
297 the values that are associated with the ontology. Next, is the consideration of
298 reusing ontologies. Yadav et al. (2016) stated that it's possible to re-use recent
299 ontologies in defining the schema of the new ontology that is to be constructed.
300 Third, is the enumeration of terms, where all terms must be clearly specified,
301 together with the domain and range of the ontology. Fourth, is the definition
302 of the taxonomy, where all terms are organized in a hierarchy. For example,
303 if A is a subclass of B, then every instance of class A must be an instance of
304 B. Fifth, is the definition of properties, which includes specifying the properties
305 that link the classes while organizing them in a hierarchy. Next, is the definition
306 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) and 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.

358 **2.3 Natural Language Question to SPARQL Translation**

359 **2.3.1 Natural Language Question (NLQ) Preprocessing**

360 spaCy is an open-source library for advanced natural language processing (NLP)
361 in Python. spaCy is designed to handle preprocessing tasks with high efficiency
362 and speed. spaCy’s features and functionalities include: tokenization, lemma-
363 tization, part-of-speech (POS) tagging, and named entity recognition (Nawaz,
364 2023; SpaCy Documentation, n.d.). In the study, spaCy will be used to prepro-
365 cess the NLQ through tokenization, and lemmatization.

366 **2.3.2 Entity and Relationship Extraction with Semantic Parsing**

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

373 **2.3.3 Semantic Parsing with SBERT**

374 Sentence Transformers or SBERT, is a Python module used for accessing, us-
375 ing, and training text and image embedding models. It can be used to compute
376 embeddings using Sentence Transformer models or to calculate similarity scores
377 using Cross-Encoder models. SBERT’s features and functionalities include: se-
378 mantic search, semantic textual similarity, and paraphrase mining. The Se-
379 mantic Textual Similarity (STS) application aims to produce embeddings for
380 all texts involved and calculate the similarities between them. The text pairs
381 with the highest similarity score are considered to be the most semantically
382 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 using Protégé, 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

408 of the user's natural language query, entity extraction, and query generation to
409 translate the query's intent and extracted entities into a SPARQL query. The
410 study by Mishra et al. (2022) has shown that it's possible to incorporate a NLQ
411 to SPARQL pipeline within the chatbot. In the study, RASA open source will
412 be used to construct the chat-bot.

413 Chapter 3

414 3 Research Methodology

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

417 3.1 Research Activities

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

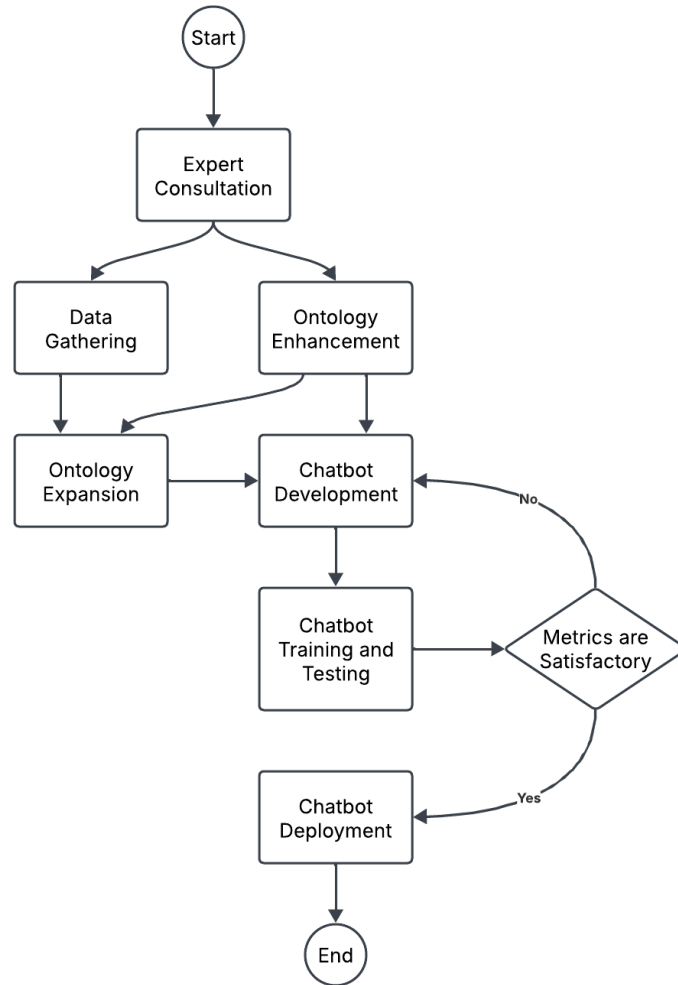


Figure 1: Process Diagram of Special Project

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

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

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

441 **3.2.2 Ontology Enhancement**

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

451 **3.2.3 Ontology Expansion**

452 The researchers have consulted with Mrs. Dimzon in how to properly identify
453 key story elements, and the contextual relationships between entities in a story.
454 With this, the researchers closely read and examined each story from the col-
455 lected data, looking for relevant story elements and relationships. From their
456 findings, they expanded the digital ontology by populating it with new stories,

457 entities, and relationships based on the enhanced ontological structure.

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

460 **3.3 Chatbot Development**

461 **3.3.1 Chatbot Development Tools**

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

471 **3.3.2 User Experience**

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

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

482 pillar of ontology frameworks.

483 3.3.3 Rasa Framework

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

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

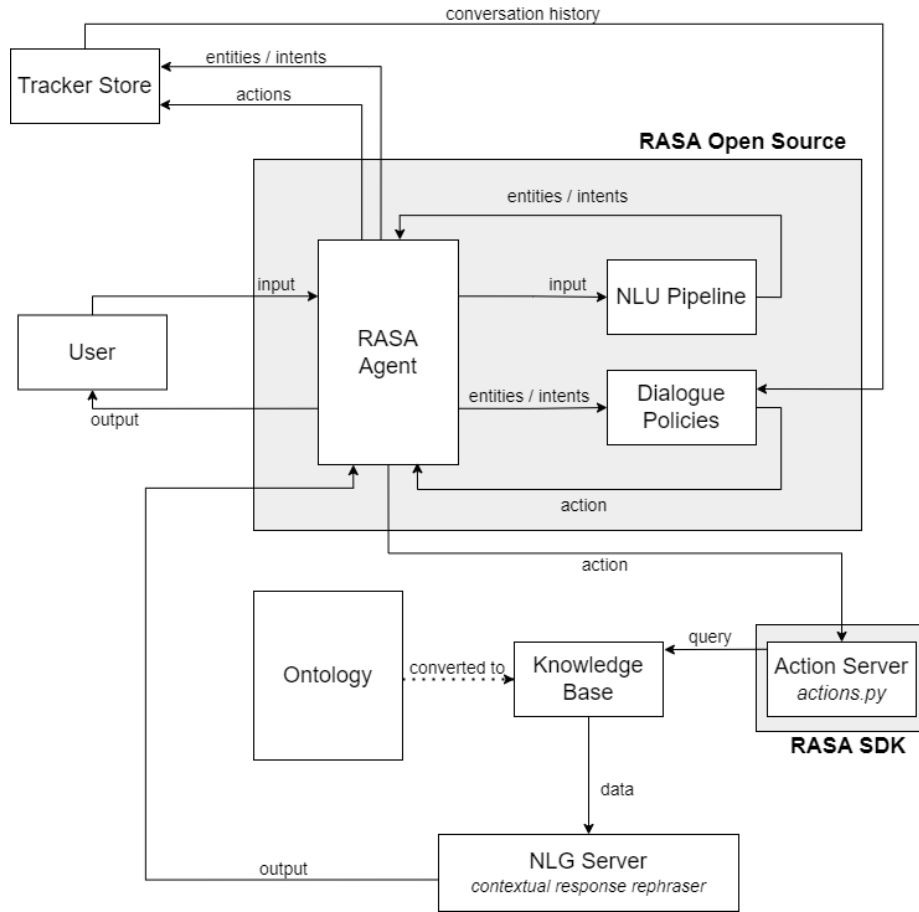


Figure 2: Diagram of Rasa Framework

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

508 syntactic features for a user’s message for the purpose of entity extraction. The
 509 fifth component is the count vectors featurizer, which generates a bag-of-words
 510 representation of the bot user’s message, intent, and response for the purpose of
 511 intent identification and response selection. The DIET Classifier is a multi-task
 512 transformer architecture that is responsible for intent classification and entity
 513 extraction. The final component is the Entity Synonym Mapper that maps en-
 514 tities to their synonyms if they appeared in the training data. The extracted
 515 entities and identified intents in the NLU pipeline will then be passed to the
 516 chatbot dialogue policies to determine the appropriate actions that the bot will
 517 perform.

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

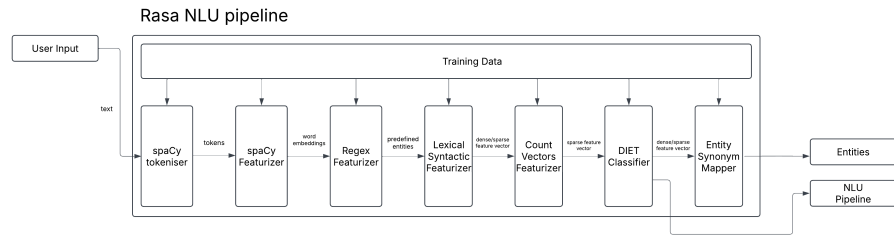


Figure 3: Diagram of NLU Pipeline

522 3.3.4 Chatbot Testing

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

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

530 **3.4 System Deployment**

531 **3.4.1 Rasa Chatbot**

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

538 **3.4.2 PanayHub Website**

539 Website hosting was facilitated through the Render.com platform, selected for
540 its usability and streamlined deployment capabilities. Render.com provides in-
541 frastructure for hosting websites and servers, with integrated support for popu-
542 lar Git repository hosting services such as GitHub and GitLab. This integration
543 enables automated deployment triggered by each new commit, enhancing devel-
544 opment workflow efficiency.

545 Figures 4, 5, and 6 present the user interface of the website. Figure 4 illus-
546 trates the homepage, which functions as the primary landing page and naviga-
547 tion hub for users. Figure 5 depicts the *Contact Us* page, which provides users
548 with a channel for communication with researchers and affiliated institutions.
549 Lastly, Figure 6 displays the chatbot interface, allowing user interaction with
550 the chatbot for inquiries related to the folk narratives of Panay.

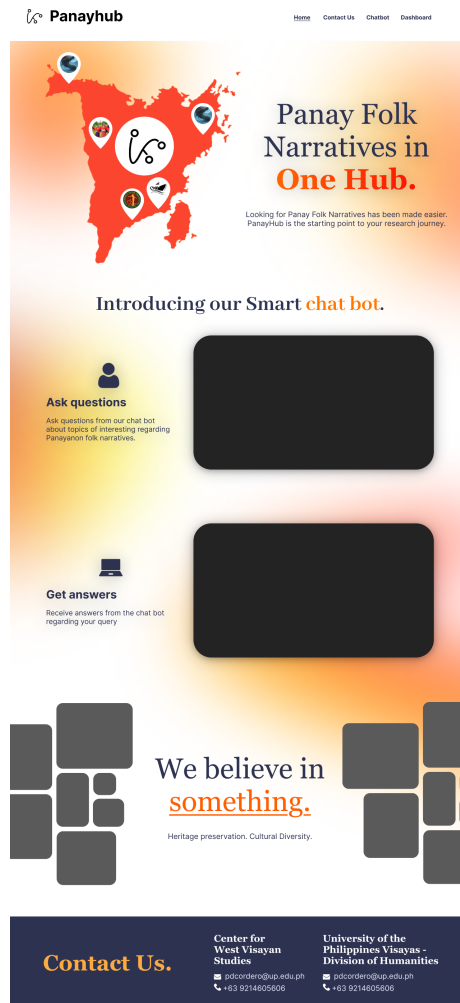


Figure 4: Homepage of PanayHub

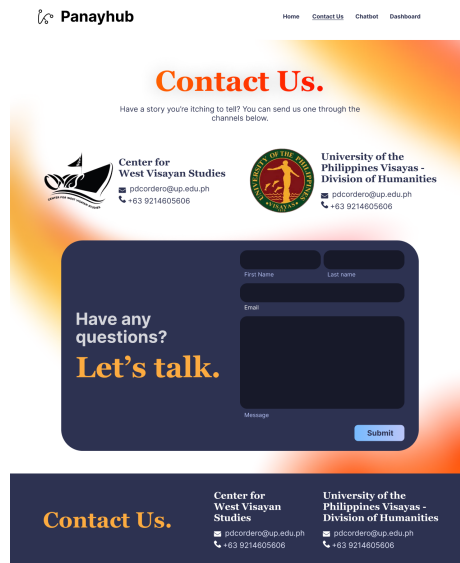


Figure 5: Contact Us page of PanayHub

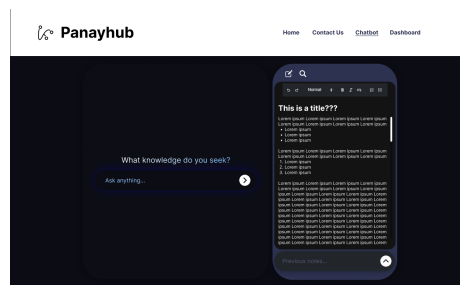


Figure 6: Chatbot page of PanayHub

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

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

578 I. Myths: 21

579 II. Legends: 108

580 A. Etiological Legends: 43

581 i. How Legends: 36

582 a. Origin of Animals: 9

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

584 c. How places and things got their names: 24

585 B. NonEtiological Legends: 64

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

587 ii. Religious/Saints Legends: 9

588 iii. Legends on Supernatural/Enchanted Beings: 44

589 C. Others: 7

590 4.2 New Classes

591 As per consultation with Mrs. Dimzon, the following classes were added to
592 expand on the original ontology.

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

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

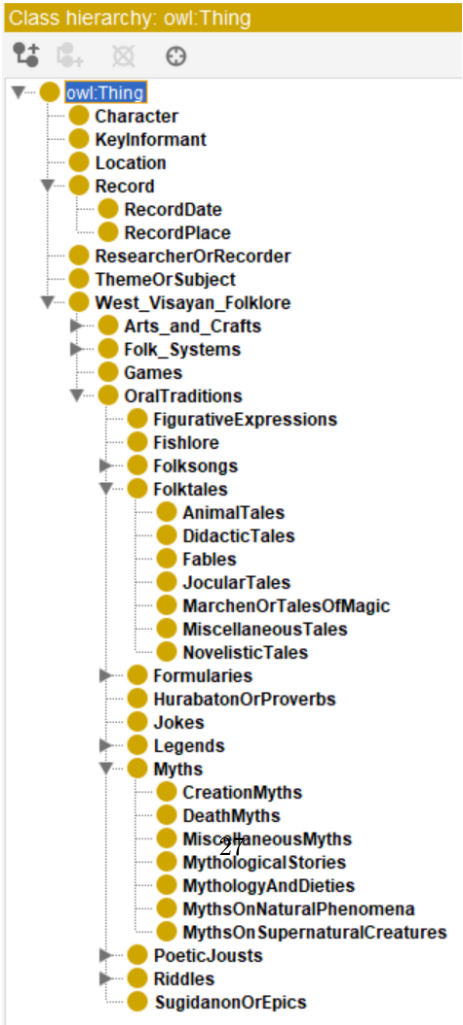


Figure 7: Class Hierarchy of Original Ontology

598 With the newly introduced classes, 8 presents the class hierarchy of the
 599 objects in the enhanced ontology as presented in Protege.

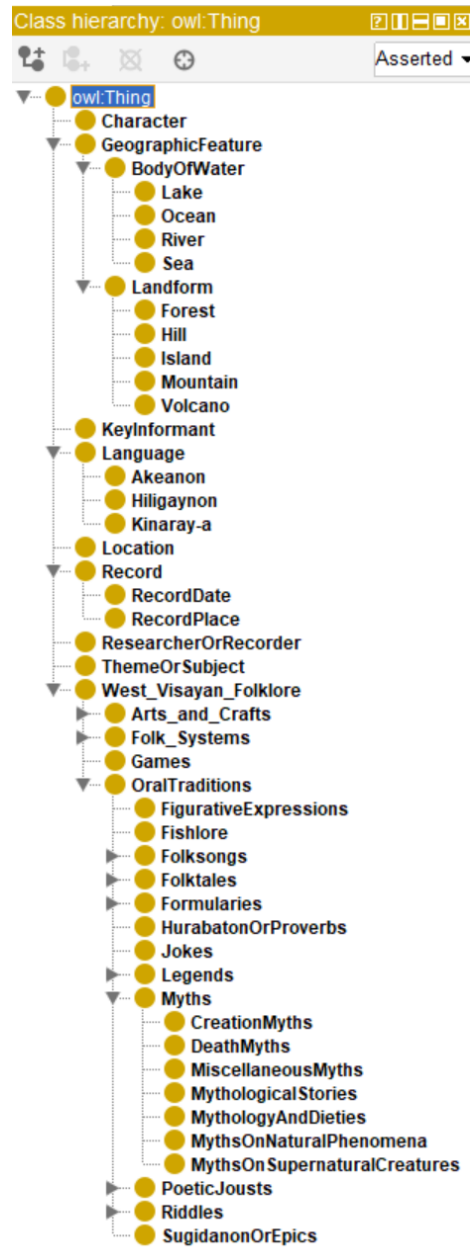


Figure 8: Class Hierarchy of Enhanced Ontology

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.

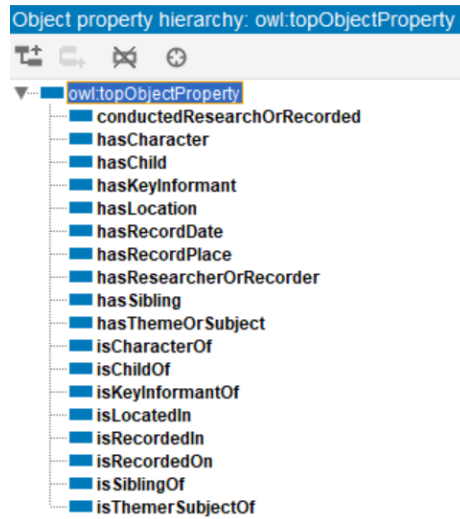


Figure 9: Object Property Hierarchy of Original Ontology

With the newly introduced classes, 10 presents the the hierarchy of the object properties in the enhanced ontology as presented in Protege.

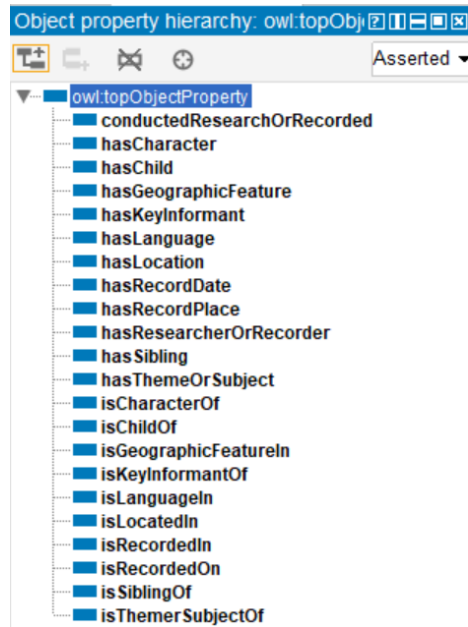


Figure 10: Object Property Hierarchy of Original Ontology

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.

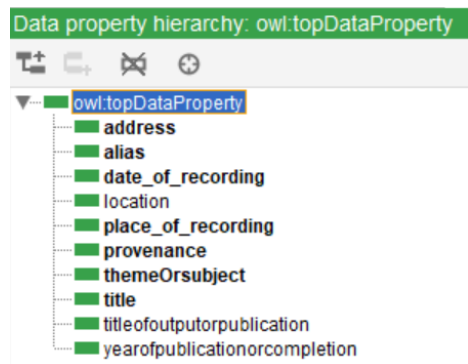


Figure 11: Data Property Hierarchy of Original Ontology

611 Chapter 5

612 5 Conclusion

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

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

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

630 5.1 Limitations

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

635 Furthermore, the developed chatbot was constructed as a prototype query-

ing 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.

661 References

- 662 Apache Community Development Project. (n.d.). *Apache Jena*. Retrieved from
663 <https://projects.apache.org/project.html?jena>
- 664 Baker, P. G., Goble, C. A., Bechhofer, S., Paton, N. W., Stevens, R., & Brass,
665 A. (1999). An ontology for bioinformatics applications. *Bioinformatics*
666 (*Oxford, England*), 15(6), 510–520.
- 667 Chokshi, H. J., & Panchal, R. (2022). Using apache jena fuseki server for
668 execution of sparql queries in job search ontology using semantic technol-
669 ogy. *International Journal of Innovative Research in Computer Science &*
670 *Technology*, 10(2), 497–504.
- 671 Dimzon, E., & Dimzon, F. (2015a). *Developing a digital ontology of the west*
672 *visayan folklore heritage (phase 1: The oral traditions)*. (Unpublished)
- 673 Dimzon, E., & Dimzon, F. (2015b). *Myths and legends terminal report*. (Un-
674 published)
- 675 Eslit, E. R. (2023). Resilience of philippine folklore: An enduring heritage and
676 legacy for the 21st century. *IJELR: International Journal of Education,*
677 *Language, and Religion*, 5(1), 9–20.
- 678 Eugenio, D. L. (2007). *Philippine folk literature: An anthology* (Vol. 1). UP
679 Press.
- 680 Giunchiglia, F., Dutta, B., Maltese, V., & Farazi, F. (2012). A facet-based
681 methodology for the construction of a large-scale geospatial ontology.
682 *Journal on data semantics*, 1, 57–73.
- 683 Guarino, N. (1998). *Formal ontology in information systems: Proceedings of*
684 *the first international conference (fois'98), june 6-8, trento, italy* (Vol. 46).
685 IOS press.
- 686 Jain, V., & Singh, M. (2013). Ontology development and query retrieval using
687 protégé tool. *International Journal of Intelligent Systems and Applica-*

688 *tions*, 9(9), 67–75.

689 Jepsen, T. C. (2009). Just what is an ontology, anyway? *IT Prof.*, 11(5),
690 22–27.

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

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

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

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

700 Mortensen, O. (2024, Apr). *How many users does chatgpt have? statistics facts*
701 (2024). Retrieved from

702 Munir, K., & Anjum, M. S. (2018). The use of ontologies for effective knowledge
703 modelling and information retrieval. *Applied Computing and Informatics*,
704 14(2), 116–126.

705 Nawaz, M. (2023, 4). *NLP Preprocessing using Spacy*. Retrieved from
706 <https://soshace.com/2023/04/05/nlp-preprocessing-using-spacy/>
707

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

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

714 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).

735 **A Resource Persons**

736 **Prof. Eliodora Dimzon**

737 Faculty

738 University of the Philippines Division of Humanities

739 eldimzon@up.edu.ph

740