PanayHub: Ontology-Based ChatBot on Folk Literature of 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

November 8, 2024

Abstract

There is currently little work being done in developing of digital ontologies, particularly that of the folklore of Western Visayas. Thus, the researchers aimed to create a chatbot that is able to provide insights and details regarding Panayanon folk narratives which include, myths, legends, and folk tales. Specifically, the researchers aimed for the creation of a knowledge base of Panayanon folk narratives, and subsequently, development and training of a chatbot to understand and answer inquiries about Panayanon folk narratives.

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 (Eugenio, 2007). Myths, legends, and folktales are included under the category of folk narratives, a form of literature that provides a narrative through prose or verse. These three forms of folk literature will serve as the focus of this study. 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. These problems were the focus of Eugenio (2007)'s work. However, due to the cost of the book and its dated nature, access to the resource has been limited.

According to (Dimzon & Dimzon, 2015), 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 (2015), 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 add upon the ontology found in the pioneering work of Dimzon and Dimzon (2015). Subsequently, the researchers will develop a chatbot that is able to answer questions about Panayanon folk narratives.

1.3.2 Specific Objectives

Specifically, the researchers aim to:

1. Add Panayanon folk tales to the ontology created by Dimzon and Dimzon (2015)

2. Add story details to the myths, legends, and folk tales in the expanded ontology

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

1.4 Scope and Limitations of the Research

In this study, the development of a chatbot capable of understanding and responding to inquiries about Panayanon folk literature will be the primary focus. The scope of the folk literature analyzed will be limited to works originating from the island of Panay, specifically focusing on folk narratives, which include myths, legends, and folk tales. Additionally, the ontology used in the study will be based on the foundational work presented in the terminal report of Dimzon and Dimzon (2015), building upon and expanding these existing ontological frameworks to ensure comprehensive coverage of the key entities and relationships within Panayanon folk narratives.

The native languages used in Panayanon folk narratives are the Panayanon languages: Hiligaynon, Akeanon, and Karay-a. However, the language used in the creation of the ontology and the chatbot will be English. This is to ensure that the chatbot can reach a wider audience, especially those who are more proficient in English, while still preserving and representing the cultural richness of the Panayanon folk narratives.

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 opensource 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 PAROT for Natural Language Conversion

PAROT is a dependency based framework used for converting natural language (NL) queries of the user into SPARQL, the query language for databases making use of ontologies (Ochieng, 2020). PAROT was created to address the shortcomings of existing NL to SPARQL tools like AquaLog, CASIA@12, Querix, AutoSPARQL, K-Extractor, and SPARK in handling queries that contained the following conditions: opposing scalar adjectives, negation, numbered lists, and compound sentences. According to Ochieng (2020), these NL to SPARQL tools drastically decrease in their precision and recall values when encountering any of the four conditions stipulated. PAROT is able to handle compound, negation, numbered list, and scalar adjective based questions. According to Ochieng (2020), it is highly successful in converting NL to SPARQL, outper-

forming the gAnswer system when it comes to answering complex questions with the QALD9 dataset. The study will be using PAROT to convert natural language statements of the end user to SPARQL queries for retrieving data from the study's ontology.

Chapter 3

3 Research Methodology

This chapter lists and discusses the specific steps and activities that will be performed to accomplish the project.

3.1 Research Activities

3.1.1 Data Collection and Ontology Expansion

In collaboration with literature experts from the Center for West Visayan Studies, and the UPV Division of Humanities, the researchers will collect Panayanon myths, legends, and folktales from reliable resource persons and other sources, including written records, research papers, and digital archives. Then, the researchers will closely read and examine the story to look for relevant story elements and their relationships. From their findings, they will expand Dimzon and Dimzon (2015)'s ontology with entities such as settings, objects, and events, and their relationships with other entities to accommodate relevant story elements. The researchers will then populate the ontology with story elements from their analysis on the stories. Protégé will be utilized for ontology expansion for its extensive support in OWL files and SPARQL querying, reasoning and consistency checking features, as well as collaboration features. Throughout this whole process, the researchers will present and consult with literature experts from the UPV Division of Humanities on the created ontology to validate the findings and ensure consistency with conventions and practices in the field of literature.

This step is crucial in enhancing the current ontology by adding detailed elements of Panayanon myths, legends, and folktales. Thus, there will be a strong knowledge base for the chatbot to query from, ensuring support for accurate, contextual answers for the semantic queries from the users. The expected outcome is a detailed ontology that includes structured information on Panayanon myths, legends, and folktales. This step is scheduled to start in December 2024 and must be accomplished by the end of March 2025, with a total duration of four (4) months.

3.1.2 Chatbot Design and Development

In this step, the researchers will develop a chatbot 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 will use Python to employ natural language processing (NLP) techniques to convert English queries into SPARQL. PAROT (Ochieng, 2021) will also be utilized to support queries containing compound sentences, negation, scalar adjectives and numbered lists. The SPARQL queries will then be sent to an Apache Jena server where the expanded OWL file will be hosted. Finally, the query results will be formatted into English. 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.

With this chatbot, users will be able to interact with the ontology in natural language. This is in pursuit of data querying, which is Manansala, Bruskiewich, and Naval (2007)'s third and final pillar of ontology frameworks. The expected output is a chatbot that can semantically understand complex user questions, and answer them with accurate information from the ontology in a natural language format. This step is scheduled to start in January 2024 and must be accomplished by the end of May 2025, with a total duration of five (5) months.

3.1.3 Documentation

The researchers will document relevant results and information throughout the project. It shall cover data, methodology, results, and analysis. Applying software engineering principles, the researchers will also create diagrams such as data flow diagrams, use case diagrams, and sequence diagrams. Google Docs will be used for its simplicity and familiarity with the researchers, and Overleaf will be utilized for final formatting. For diagrams, computer assisted software engineering (CASE) tools will be utilized.

This step ensures that all information has been transparently communicated for future reference to be used by other researchers and interested parties. The expected output is complete project documents, including technical details, and a final project report. This step is scheduled to start in December 2024 and must be accomplished by the end of May 2025, with a total duration of five (6) months.

3.2 Calendar of Activities

Table 1 shows a Gantt chart of the activities. Each bullet represents approximately one week worth of activity.

Table 1: Timetable of Activities

Activities (2025)	Dec	Jan	Feb	Mar	Apr	May
Data Collection and Ontology Expansion	●●●●	●●●●	●●●●	●●●●		
Chatbot Design and Development		●●●●	●●●●	●●●●	●●●●	●●●
Documentation	●●●●	●●●●	●●●●	●●●●	●●●●	●●●●

Chapter 4

4   Preliminary Results/System Prototype

4.1   Original Ontology

As illustrated in Figure 1, the original ontology by Dimzon and Dimzon (2015) does not contain story details but rather the classification of the different oral traditions found in the cultures of Western Visayas. This presents the knowledge gap that the researchers propose on exploring. Specifically, the ontology will be expanded with story elements for the Myths, Legends, and Folktales entities present in the current iteration of the ontology.
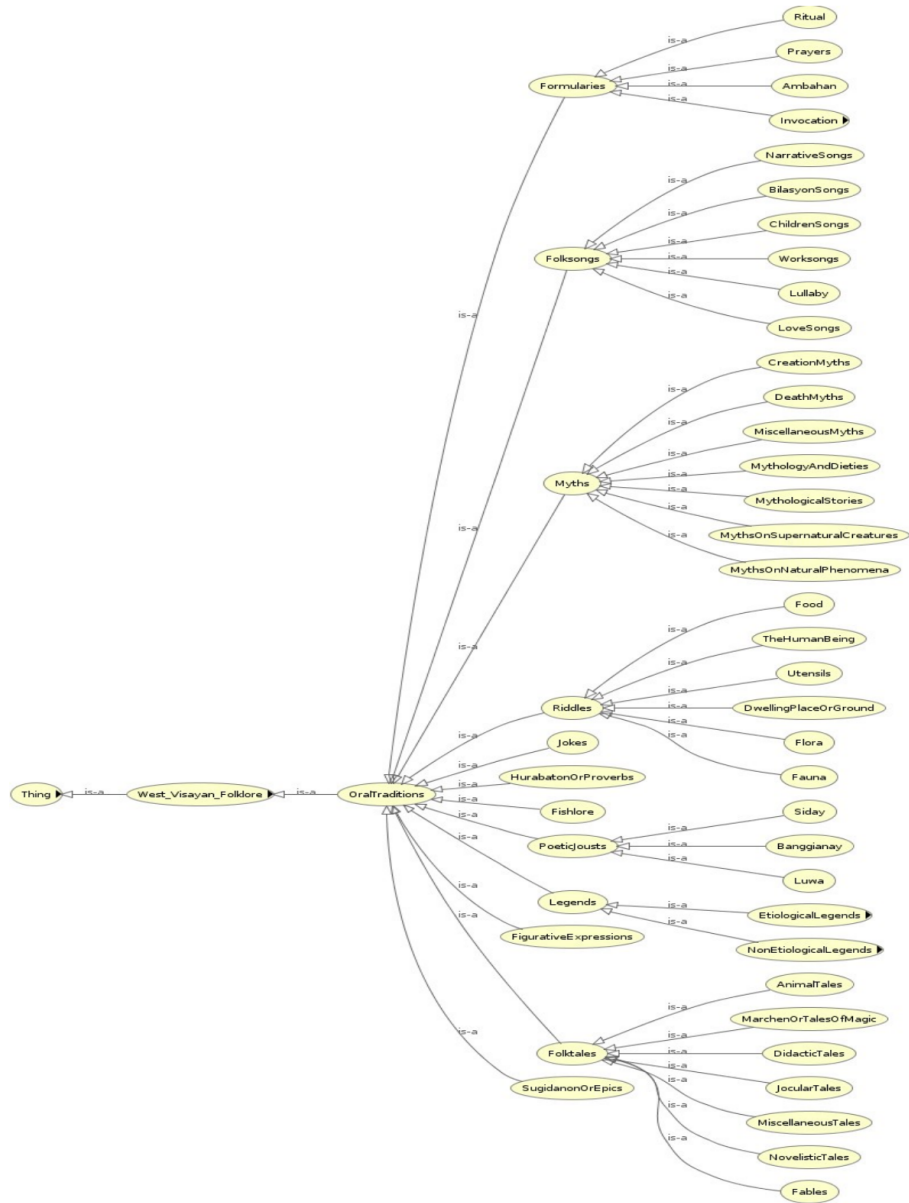
Figure 1: Diagram of The Original Ontology by Dimzon and Dimzon (2015)

References

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.

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

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 Applications*, *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.

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.

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 . . . .

Ochieng, P. (2020). Parot: Translating natural language to sparql. *Expert Systems with Applications: X*, *5*, 100024.

Ochieng, P. (2021). Parot: Translating natural language to sparql. *Expert Systems with Applications*, *176*, 114712. Retrieved from https://www.sciencedirect.com/science/article/pii/S0957417421001536 doi: https://doi.org/10.1016/j.eswa.2021.114712

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).

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.

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`

....