

Building Myanmar, Myanmar Sign and Japanese Languages Dictionary Based on Ontology

Nu Nan Dar Lin, Ye Kyaw Thu, Mitsuji Matsumoto and Yoshiyori Urano
Graduate School of Global Information and Telecommunication Studies,
Waseda University, Tokyo, Japan
nandarlin@akane.waseda.jp, yekyawthu@aoni.waseda.jp,
mmatsumoto@waseda.jp, urano@waseda.jp

Abstract

The World Wide Web is an information resource with virtually unlimited potential. However, this potential is relatively untapped because locating relevant information can often be time consuming and fruitless. If machines could understand the content of the web pages, searches with high precision and recall would be possible. So, our approach system is to find the meaning of word from Ontology database. Nowadays, Ontology is becoming increasingly widespread in research field and application area of computer science. Ontology is used not only in semantic web but also in database and applications that need to share domain information for specific area of knowledge. Ontology comprises as the concept, properties and restrictions on properties. So, we can build Ontology for every object and thing in our environment that has relationship and properties. This system was developed by categories grouping methods and then all of that groups are relation with ontology properties. So, users can see all of the words relation and study usage of the words from this dictionary. This dictionary prototype was developed with Protégé ontology Editor 4.2 Alpha version that supports OWL 2.0 languages.

1. Introduction

Ontology is a formal definition of a body of knowledge. An explicit formal description of concepts (classes) in a domain of discourses, properties of each concept describing various features and attributes of the concept, and restrictions on properties. Ontology together with a set of individual instances of classes constitutes

a knowledge base. Ontologies are used by people, databases and applications that need to share domain information (a domain is merely a specific area of knowledge, such as medicine, petroleum, financial management, etc.). Classes are the focus of most ontology. Classes describe concepts in the domain [9]. For example, a class of animal represents all animals. “Cat” is instances of this class. A class can have subclasses that represent concepts that are more specific than the super class. For example, the class of all animals into mammals, four feet animals and two feet animals can be divided. This system is mainly intended to develop dictionaries with words relationship. So, all of the words is collected with their concerning groups.

In Myanmar sign language, we didn't show the word's definition by using the words (or) sentences. Because some Myanmar sign languages have the action on the hand. If we showed that it with the words or sentences, it is so difficult to understand for users. Therefore, we presented it with the movie. We shot all of the fingerspelling movies by helping from Mandalay Deaf School in Myanmar.

2. Related Works

Ontologies are used in artificial intelligence, semantic web software engineering, biomedical information, library science and information architecture as a form of knowledge representation about the world or some part of it. We can develop many useful applications (such as dictionary, hospital system, etc...) by using Ontology. The most recent development in standard ontology languages is the web ontology

language (OWL) from the World Wide Web Consortium (W3C).

3. Ontology

Ontologies are metadata schema, providing a controlled vocabulary of concepts, each with an explicitly defined and machine processable semantics. By defining shared and common domain theories, ontologies help both people and machines to communicate concisely supporting the exchange of semantics and syntax. Moreover, ontology is a formal model of the kinds of concepts and objects that appear in the real world together with the relationships among them. There are roughly four kinds of ontologies: document ontologies, metadata ontologies, domain ontologies and service ontologies [7]. This system is intended to use with domain ontology.

3.1 Web Ontology Language (OWL)

Ontology describes the concepts in the domain and also the relationships that hold among those concepts. Different ontology languages provide different facilities. The most recent development in standard ontology languages is OWL from the World Wide Web Consortium (W3C) [9]. OWL which is short for Web Ontology Language is an ontology language designed to be compatible with the World Wide Web and the Semantic Web. OWL comprises as follow:

- ❖ Individuals: represent objects in the domain.
- ❖ Properties: Properties are binary relations on individual.
- ❖ Classes: interpreted as sets that contain individuals.

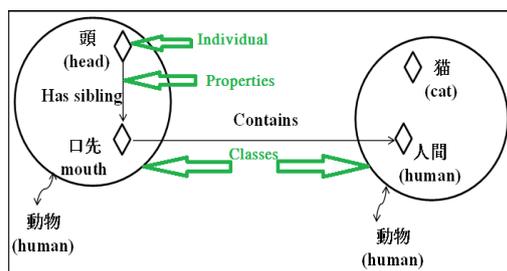


Figure.1 Representation of Web Ontology Language

Image adapted from [9]

4. Myanmar Language

Myanmar language is the official language of the Republic of the Union of Myanmar. There are a total of 111 languages spoken by the people who living in Myanmar. Because of Myanmar is comprised with 135 ethnic groups, also known as “Nationalities” [4]. Myanmar script is a writing

system constructed from consonants, consonant combination symbols (i.e. Medials), vowel symbols related to relevant consonants and diacritic marks indicating tone level (niggahita, visajjaniya). Overall writing direction is from left to right. Myanmar language alphabet is recognized as containing 33 consonants and 12 basic vowels, viz [6].

4.1. Myanmar Sign Language and Finger Spelling

Myanmar sign language is a language which uses visually transmitted sign patterns (such as manual communication, body language). Sign patterns are combining hand shapes, orientation and movement of the hands, arms or body and facial expressions to fluidly express a speaker’s thoughts. That is mainly used for communicating among deaf people [11]. But only sign language is not enough to communicate with people because it is needed to spell words (such as people’s names, city names, places and etc.). Therefore, finger spelling is invented as a part of sign language. Fingerspelling is the representation of the letters of a writing system and sometimes numeral systems, using only hands [5]. But these manual alphabets may differently depend on several countries in the world.

There are two different fingerspelling character sets for Myanmar language; one is used in northern Myanmar (e.g. used at “Mandalay School for the Deaf” in Mandalay city) and the other is used in southern Myanmar (e.g. used at “Mary Chapman School for the Deaf” in Yangon city). In 2007, Myanmar sign language dictionary book was published by Department of Social Welfare, Ministry of Social Welfare, Relief and Resettlement [10]. We also used that book for building the Sign language dictionary. Myanmar fingerspelling consonants are based on the American manual alphabet such as “a” for “အ” (အ), “b” for “ဘ” (ဘ), “d” for “ဒ” (ဒ), “z” for “ဇ” (ဇ) [8] [10]. This Fingerspelling order is the same as the Myanmar language hand writing order.

4.2. Japanese Language

Japanese language ranks as one of the world’s most important languages with over 130 million speakers [3]. Japanese language is written with a combination of three scripts: Chinese characters called Kanji, and two syllabic scripts. That syllabic scripts are made of modified Chinese characters called Hiragana and Katakana. And also have Latin alphabet, called Romaji, is used in modern Japanese language. It is especially used for company names and logos, advertising and when entering Japanese text into a computer. Two numerals of language are used

in Japanese language. There are Arabic numerals used for general and traditional Shino – Japanese numerals that are also used in common place [2].

5. Group Design for Dictionary

This part presents what the idea for dividing the categories groups is and how much relationship between these words of dictionary has. Firstly, we are thinking that we must have the general categories groups for words of all dictionaries. All of the categories groups are classified depending on usage of words. In this case, we found most of the simple dictionaries software and dictionaries book are dividing their categories group depending on consonants of their specialized languages. That is suitable if we don't have the relationship between these words and also they don't want to show the same meaning words (synonym). This consonant grouping method is general and also it cannot show the usage of environments.

Generally, all of the words have at least one relationship with other words (such as the same meaning or opposite meaning). This system has divided 23 general groups and many sub-groups. After that, we add the words to their concerning groups. Figure.2 shows the sample group structure of "Animal".

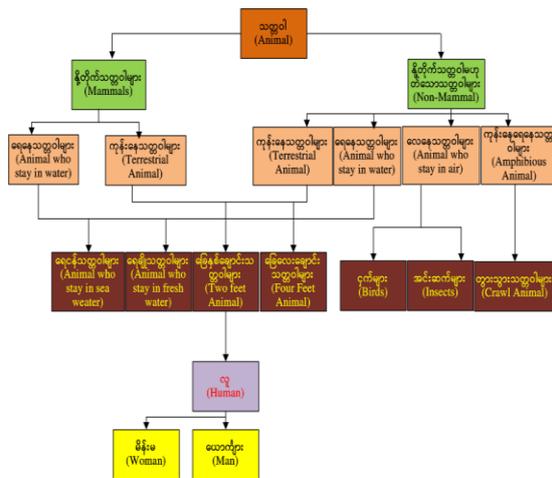


Figure.2 Sample Group Structure of Animal

5.1. Ontology Database Design for Dictionary

This system is developed with Ontology facilities and web ontology language (OWL). We need to use a software editor in building the ontology database. We can see many ontology software editors. But we chose Protégé editor for this system. Protégé tool is developed at Stanford University in collaboration with the University of Manchester [1]. That tool is a famous open source ontology editor and is also used in developing knowledge-based application with

ontology. It can strongly support to build knowledge acquisition system and also it has many plugs-in. One of these plugs-in is supporting Unicode font system.

And then, we would like to present the sample ontology database design of the dictionary. Figure.3 shows the relationship of “odori/おどり” (dance) word. It is included in “Artist Group”, grammar type is “Noun” that shows with “hasGrammartype” properties. And it continues to show the Myanmar definition of that word and synonym word. Some of the words exist in more than one categorizes groups. “odori/おどり” (dance) also has more than one categorizes group that is “Music Group”. Myanmar sign language dictionary’s database design is also the same like figure 3.

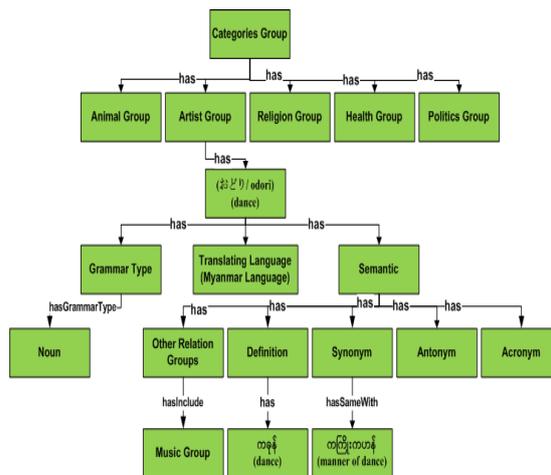


Figure.3 Sample Ontology Building for Japanese to Myanmar Dictionary

6. Implementation of the System

This system is an implementation of Japanese language, Myanmar language and Myanmar sign language terms dictionaries using web ontology language. This system is developed by using Java Script language and HTML 5 (Hyper Text Markup Language version 5).

All of the sign language data are taken from Myanmar sign language dictionary [10]. It has 1572 words in total. But we can add nearly 700 videos in this dictionary. Other words are continuing shooting for adding to the dictionary. And we already added nearly 5000 words to Japanese to Myanmar language dictionary.

Figure.4 shows the search result interface of Myanmar word “ရှင်ပြုသည်” (initiate a novice). The definition of “ရှင်ပြုသည်” (initiate a novice) is also showing not only Sign language movie but also finger spelling.

- [5] <http://en.wikipedia.org/wiki/Fingerspelling>
- [6] “Development of Efficient Input Methods for Myanmar Language Short Messaging Service (SMS)” -- Ye Kyaw Thu (March 2006)
- [7] “Capturing and Building Knowledge Repository: Ontology Approach.” (2009, University of Computer Studies Mandalay, Myanmar). -- Khin Myo Htet.
- [8] “Studies on Input Interface of Asian Characters Based on Common Syllabic Writing Systems” -- Ye Kyaw Thu (October 2011)
- [9] "A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools Edition 1.2" Copyright The University Of Manchester, March 13, 2009.-- Matthew, Horridge.
- [10] “Direct Keyboard Mapping (DKM) Layout for Myanmar Fingerspelling Text Input – Study with Developed Fingerspelling Font (mmFingerspelling.ttf)”—Ye Kyaw Thu, Sai Aung Win Maung and Yoshiyori Urano.
- [11] http://en.wikipedia.org/wiki/Sign_language