

The Use of Terminology and Ontology Theory in the Development of a Language Assistant System for Backpackers

Hsing-I Wang, Department of Information Management,
Overseas Chinese University, Taichung, Taiwan

ABSTRACT

AS the living standards being elevated and the concept of demanding for better living qualities being planted in most of the people, leisure and entertainment have been considered more important in daily lives. Among the leisure activities, the independent travel is embraced the most. However, traveling abroad does not synchronize in the same growth trend as travelling domestically in Taiwan. It is believed that language capabilities hinder most travelers from travelling abroad. Although travelers may resolve the difficulties with recourse to tools, such as electronic translators and electronic dictionary, the tools that are popular and adopted by most people provide, in fact, very limited help to backpackers.

This research aims at building the ontology of the knowledge base of a language assistant system by applying terminology theory. This research first analyzes the contexts that most backpackers may encounter. The terminology theory is then adopted to categorize the contexts. Furthermore, this research constructs the ontology tree for the contexts. The conversations that may occur in all contexts are generated and placed in the ontology tree according to the rules of the ontology tree structure. A language assistant system in Chinese-English is developed to demonstrate how the backpackers would benefit from the system and how they are able to locate the proper dialogue from the knowledge database for a certain circumstance or a specific context.

Keywords: Terminology theory, Ontology, Backpackers, Language assistant system, Knowledge base

INTRODUCTION

AS the living standards being elevated and the concept of demanding for better living qualities being planted in most of the people, leisure and entertainment have been considered more important in daily lives. According to the 2009 investigation report of the traveling status of Taiwan, about 93.4% people traveled at least once on the island, and the number was about 88% in 2006 [Taiwan marketing research, 2009]. Among them, about 89% people planned the trip on their own. In the report, it also indicated that only 17.5% people traveled abroad at least once and Mainland China, Japan and Hong Kong were the most visited countries. Compare to traveling domestically, the ratio is quite low. The reasons may include the economic factors, time factors, and other personal reasons. The amount of time and costs that travelers spend on the collection of travel information may deter travelers from trying again [Chiou et. al, 2001]. In fact, previous researches did not give a specific and overall conclusions or suggestions of the reasons why people seldom adopt independent travels.

In Taiwan, previous researches regarding independent travels focused on information behaviors [Su, 2001, Lee, 2003], information collections [Li, 2004], the willingness of being an independent travelers [Hsu, 2008, Li, 2005], the characteristics of backpackers [Tseng, 1999]. The findings of Yang's [2004]

research did indicate that language capabilities were one of the major considerations that people were willing to go abroad on their own. The lack of confidence on language capabilities may explain why the ratio of traveling abroad was lower than traveling domestically in Taiwan.

Travelers would, therefore, ask for any tools in hand to resolve language difficulties. In fact, there are all kinds of products on the market or on the internet that provide instant translations in multiple languages. However, shortcomings are often found in such tools. For example, the system may function solely as a dictionary, or the grammar of the translation may be right but the meaning is totally difficult to understand, or the translation does not fit the context, etc. Moreover, most of the systems provide only the translations for the sentences that users enter, the users may not understand the response of the other side afterwards and the conversations would not go on either. In other words, current language tools that we find on the market or on the internet provide insufficient help to the backpacker on their journeys.

This research aims to develop a language assistant system for backpackers. The system will not be considered as a translation tool. Rather, a knowledge base which collects all possible dialogues that backpackers often encounter. The knowledge base is built based on ontology and terminology theory. This research first analyzes the contexts that most backpackers may encounter. The terminology theory is then adopted to categorize the contexts. Furthermore, this research constructs the ontology tree for the contexts. The conversations that may occur in all contexts are generated and placed in the ontology tree according to the rules of the ontology tree structure. A language assistant system in Chinese-English is developed to demonstrate how the backpackers would benefit from the system and how they are able to locate the proper dialogue from the knowledge database for a certain circumstance or a specific context.

This paper is organized as follows. In the literatures survey, this paper first compares current language assistant systems that most people familiar on the internet. Then the terminology and ontology theories are introduced. The third section describes the methodologies of this research. Section four demonstrates the system developed by this research and section five gives the conclusion.

LITERATURE SURVEY

This section first reviews language assistant systems that are currently adopted by most travelers. The terminology and ontology theories are then introduced to explain why and how the theories are applied to this research.

Language Tools for Backpackers

Figure 1 shows an example of using WordNet. WordNet is a large lexical database of English. In addition to take WordNet as an ordinary dictionary, users are able to navigate with the browser and search for meaningfully related words and concepts. Figure 2 is the example of using the Bilingual Ontological Wordnet (BOW) developed by the Academia Sinica of Taiwan. BOW was developed based on ontology theory and also owns a large database of Chinese-English. The system provides similar functions as of WordNet.

For backpackers who do not speak fluent English or knows just little English, they have to first figure a proper word to look at in the tool and then try to compose a complete sentence by themselves so that the travelers can express themselves clearly. The kinds of systems are therefore considered not suitable for backpackers to resolve their language difficulties.

Previous research indicated that the accurate rates of most of the translation systems were about 57% [Lu et. al, 2002]. Google translation is one of the most welcome online systems in Taiwan. Figure 3

shows that when enter “I need a taxi to the station” in Chinese, the system returns “For me, please call a car to the station”, which is obviously not correct. Moreover, a system that will ask a person to enter a appropriate sentence in order to get correct translation while he or she is already nervous about speaking in foreign language does not provide much help. On the contrary, a system may resolve not only the language difficulties and at the same time elevate the motivation of learning a new language if the system functions not as a translation system but a collection of terms for different events. The events are categorized according to the contexts so that backpackers can choose proper statement.

Terminology Theory

Wüster (1898-1977) is the pioneer of terminological theory. He devoted his life to promote theory of terminology to eliminate ambiguity for better communication, to help users benefiting from standardization, and to establish terminology as a discipline [Cabre, 2003]. Although the terminological theory is still in its infancy, its importance has never been neglected.

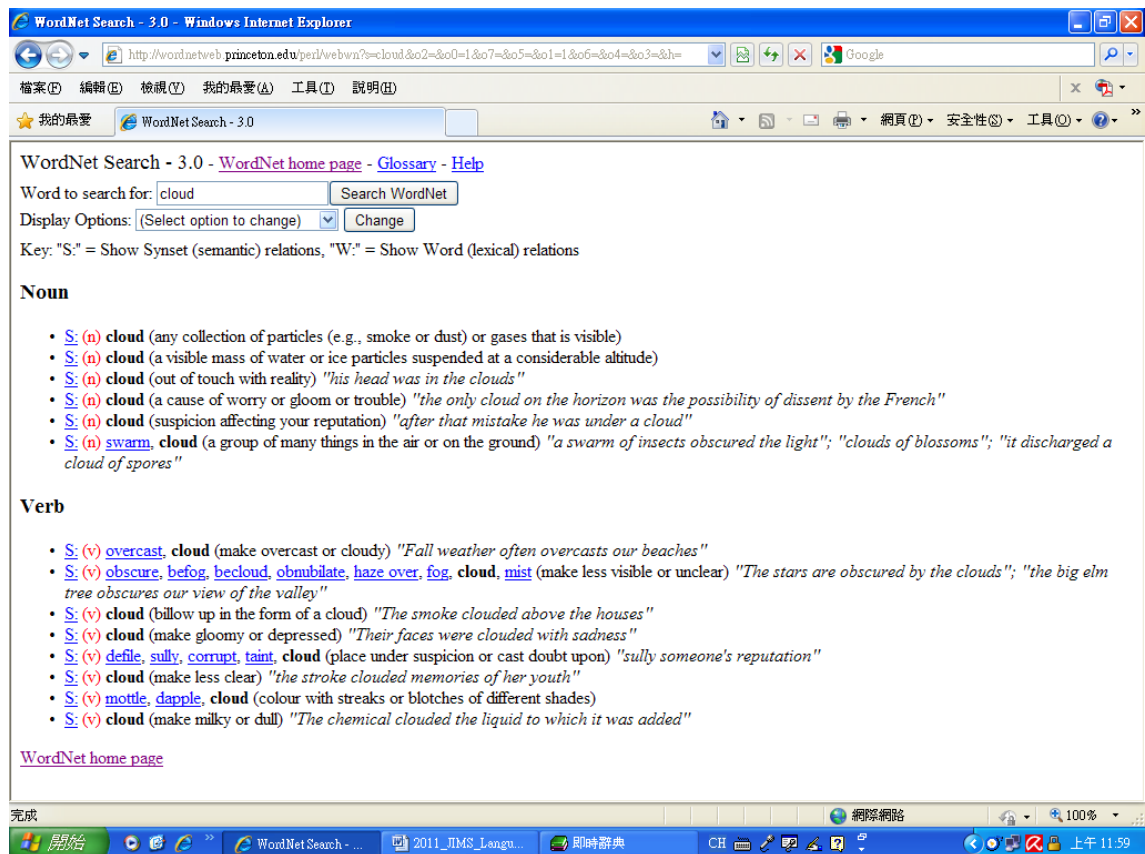


Figure 1: A sample page of WordNet



Figure 2: A sample page of the Academia Sinica Bilingual Ontological Wordnet [BOW, 2009]



Figure 3: A sample page of Google Translation

Terminology has constantly been mistaken as terms in the past. Bessé et al. [1997] defined terms as “a lexical unit consisting of one or more than one word which represent a concept inside a domain” and terminology as “the vocabulary of a domain or a subject field, i.e. a set of terms in a domain or a subject field”. Clearly, terminology is by no means equivalent to terms or words. In fact, in average, about 80 percent of the terminologies in scientific domains are made up of complex terms [Kageura, 2008]. Nevertheless, terminologies are connected to each other through shared terms [Kageura, 2008].

Myking [2001] proposed that to integrate terminology with linguistics is quite important especially the bound between onomasiology and conceptology. In the literatures, most researchers seem to share the same opinions. As pointed out in Kageura's [1995] study, he noted that almost all terminology studies started with concepts and took an onomasiological approach. The findings suggest that besides domain, the concept delivered within terminology should be taken into consideration while building the terminology database.

Terminological unit is another issue that has been discussed in terminological theory. From the above definitions, we notice that terminology is related to knowledge, language and communication. Therefore, a terminological unit actually covers a cognitive component, a linguistic component and a socio-communicative component [Cabre, 2003]. However, this definition does not make terminology distinguished from language unit, Cabre [2003] then specified terminological unit in more detail with respect to the three perspectives:

- They are context oriented; occupy a precise place in a conceptual structure and the meaning is determined by their place in this structure.
- They can have lexical and syntactic structure and belong to one of the broad semantic categories which are characterized by entities, events, properties or relations.
- They are acquired through a learning process and hence are handled by specialists in their field.

Finally, terminology is considered a set of needs, a set of practices to resolve these needs, and a unified field of knowledge [Cabre, 2003].

Ontology

Traditionally, "ontology" is about philosophical study of the nature of being or existence [Wikipeda]. It was then borrowed to AI to illustrate the relationships of the set of concepts and terms. Guarino [1995] defined ontology as an explicit specification of a conceptualization. It allows us to describe resources formally.

Ontology can be classified into terminological ontology, information ontology and knowledge ontology. In recent years, ontology has been widely adopted in the areas such education [Abel et. al, 2004], e-learning [Sampson et. al, 2002], knowledge management [Maedche et. al, 2003, Davies et. al, 2003] or even managerial fields [Storey, 1992].

Ontology can be expressed in the form of taxonomic tree; however, the kernel is much complicate. The representational primitives of ontology often include the information about the meaning and constraints on the logically consistent application [Grubr, 2008]. Swartout et. al [1997] and Handler [2001] considered ontology as a hierarchical structure of glossaries. It is useful in the description of terms of a specific domain including the words, the semantics, simple logics of the terms, and the most important, the relationships among terms. Therefore, building a learning or knowledge system with ontology will enhance sharing and reuse knowledge; the advance and the improvement of the knowledge of a certain domain will be verified as well.

In the review of terminological theory, literatures indicated that terminologies are connected to each other through shared terms. Therefore, an ontology-based terminological database would be more appropriate in the development of a terminology system.

METHODOLOGIES

In the first phase, this research reviewed secondary data including literatures, online resources, travel related books etc. to collect all possible contexts that backpackers may encounter during their journeys. In addition, the interviews to travelers who have had years of experiences on independent travels were conducted by this research. The experts were introduced by travel agencies, travel related magazines or were recommended by many internet users due to the outstanding travel blogs of the experts. The experts were invited to review as well as to modify the contexts this research summarized and to provide all possible “terminologies” (i.e. all possible conversations under each context).

This research then categorized the data collected according to terminology theory. The terminology theory suggests that each terminology contains five components:

- Class: refers to the first layer of the category.
- Domain: refers to the specific knowledge of a class.
- Context: refers to all the activities in the domain.
- Concept: refers to the idea that the terminology wants to express.
- Description: refers to the explanation of the terminology including the knowledge, the language and communication.

In the literatures survey, this research notices that most language assistant systems classify the words or the terms in four different ways:

- Dictionary style. The terms are organized alphabetically. The meanings of the terms are explained and some examples are given.
- According to the object. The class represents entities and is named with a noun such transportations, places, equipments and items.
- According to the event. The class represents events and is named with a verb such as booking, tour, ride, etc.
- According to the subject. The class represents a certain subject and is named with a noun such as business, entertainment, airport, dining, etc.

This research adopts the concept of event-driven as the basis to categorize the class, the domain as well as the contexts of the research. The reasons are because that the description of even-driven is easier and clearer. Furthermore, event-driven is easier for users to understand and the users are able to choose the right context within a second. The ontology is applied in this research to establish the relationships of class, domain, context, concept, description and responses.

The ontology tree was built according to the result of the categorization. Protégé was adopted in this research to accomplish the task. Protégé is a freeware developed by SMI-Stanford Medical Informatics. The graphical interface of Protégé simplified the establishment, the retrieval, as well as the maintenance of the ontology and knowledge base. Figure 4 shows the basic ontology of this research. In figure 4, ins1 and ins2 are the instances of context, q1 and q1 are the instances of concepts in a certain context. R1 and r2 are the responses for q1. In the real situation, an answer or a response may apply to more than one question or case. For example, the response “the seats have been sold out” may refer to booking a theater ticket or buying train tickets. Hence, the relationships between r and q are not father-son relations. Rather, this research considers them as association relationships.

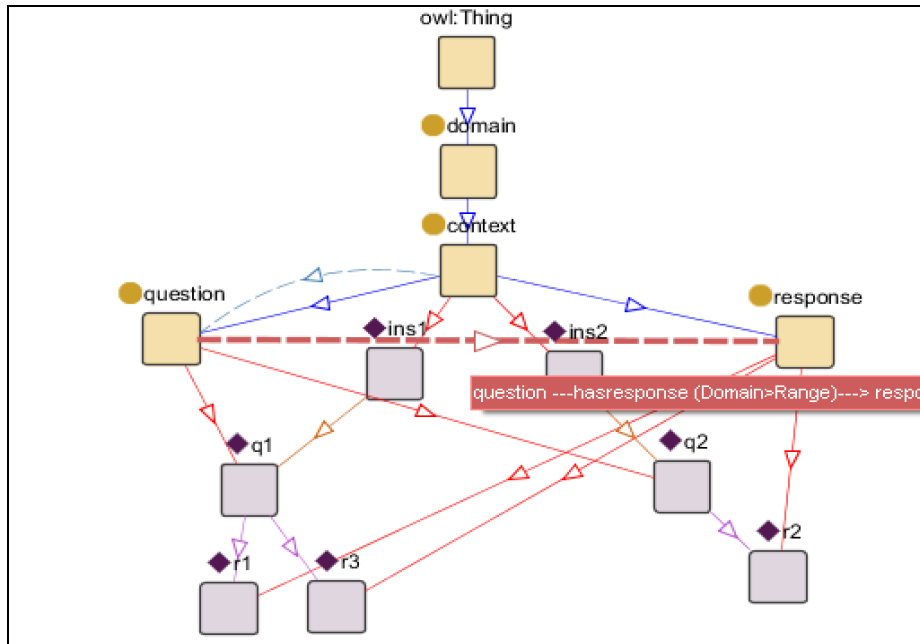


Figure 4: The ontology tree of this research

SYSTEM DEMONSTRATION

This research has compiled more than 200 terminologies that backpackers usually encounter during their journeys. According to the definition of the five components of terminology described in the previous section, i.e. class, domain, context, concept and description, respectively. The class of this research is independent travel. The terminologies are categorized into six domains including tourism, entry and leaving the country, social activities, shopping, accommodation and emergency context, respectively. The system provide English-Chinese version and can be connected thru any mobile device. The design of the system focuses on practical utilization since most of the mobile devices are light and with small screen. Once the users connect to the system, the six contexts will be presented on the first page in the tree-structure. Figure 5 shows the home page of the system.

User may expand any node of the tree to find relative contexts. Figure 6 gives the example of the expansion of tourism. There are eight contexts under tourism including the matters regarding map, route, attractive point, shopping, transportation, special performances, fares and photos, etc. Once the user chooses a certain context, all possible terminologies are listed on the right of the screen. User reads the Chinese and picks one that he/she wants to express (as shown in figure 7). The system will demonstrate the English accordingly. There may be more than one English statement for users to choose. In addition, the system will mark the key word of the statement so that the users may use only the key word if he cannot read the complete sentence. Notice that in figure 7, a speaker is given to the left of the statement. Once the users click on the icon, the system will read back the statement to the user.

Furthermore, figure 8 demonstrates that once the users click on the English statement, such as the example of “Can I have a city map” in figure 7, the system will show possible responses that the backpackers may hear from the other side. In the example of “Can I have a city map”, figure 8 shows two possible answers. Each answer has its Chinese version on the left cell as well.

Figure 5: The home page of the system

Figure 6: The contexts of tourism

Figure7: The terminologies for a context

Figure 8: The example of responses to “Can I have a city map”

CONCLUSIONS

This research applies terminology and ontology theory to establish the context-based language structure for independent travel. An English-Chinese system was developed to demonstrate the effect of the approach this research adopted. The test of the system proved that backpackers could use the system online thru their cell phone or PDA whenever they encountered language difficulties. The system gives proper and correct statements for backpackers to employ. The system also provides the possible answers the backpackers might hear and give them the opportunities to continue the conversations.

In summary, this research contributes the following:

- This research utilizes the concept of terminology theory in language areas. On one hand, it extends the usage of terminology theory. On the other hand, it expands the flexibility of the system of the research.
- Unlike other translation systems, this research builds a knowledge base for the system. The knowledge base collects all the terminologies that most backpackers will exercise during their journeys. The terminologies are categorized into context and concept accordingly. Moreover, the knowledge base collects corresponding responses for each terminology. The relationships between terminologies and the responses and the links among terminologies are established as well.

As for the benefits the system provides, it is believed that the system developed by this research assures the backpackers:

- An easy way to locate the correct statement to express themselves.
- A knowledge base that collects most of the possible statements that backpackers will need on their journeys.
- A knowledge base that also collects different feedbacks to each statement in the knowledge base.
- A system that can be connected with any mobile device.

For the purpose of continuous study of this research, the ontology tree established by this research has reserved the possibility to expand. Further works will establish the relationships among the terminologies from different angles. For example, backpackers will find proper statement not only from context to the concept; the system will analyze the simple sentence and bring all statements that may fit the context. Hopefully, the system that is designed based on the terminology and ontology theory will amend the shortcomings as well as to fill up the insufficiencies that current systems have been denounced.

REFERENCES

- Abel, M.-H., Benayache, A., Lenne, D., Moulin, C., Barry, C. and Chaput, B. (2004). Ontology-based organizational memory for e-learning, *Journal of Educational Technology & Society*, vol. 7, pp. 98-111.
- BOW, the Academia Sinica Bilingual Ontological Wordnet, <http://bow.sinica.edu.tw/>, visited in 2009.
- Cabre, M.T. (2003). Theories of terminology, *Terminology*, vol. 9, pp. 163-199.
- Chiou, Y-W, Chen, C-H, and Liang, C-S. (2009). Using ontology to construct an itinerary semantic web service system, *Journal of Information Management (Nan Hua University)*, issue 9, pp.67-93.
- Davies, J., Fensel, D. and Harmelen, F. V. (2003). *Towards the semantic web: ontology-driven knowledge management*, New York: Wiley, 2003.
- De Bessé, B., Nkwenti-Azeh, B. and Sager, J C. (1997). Glossary of terms used in terminology, *Linguistics*, vol. 35, pp. 861-877.
- Gruber, T. (2008). *Ontology*, *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag.
- Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation, *International Journal of Human-Computer Studies*, vol. 43, pp.625-640.

- Handler, J. (2001). Agents and the semantic web, *IEEE Intelligent Systems*, vol. 16, no. 2, pp 30-37.
- Hsu, S-H. (2008). Exploring the effects of technology readiness on perceived service quality, perceived risk, relationship quality and continuance intention for the self-travel service websites, unpublished master thesis, Department of Business Management, Tatung University.
- Kageura, K. (1995). Toward the theoretical study of terms, *Terminology*, vol. 2, pp. 239-258.
- Kageura, K. (2008). An Analysis of the motivatedness structure of Japanese terminologies, *Mathematical Linguistics*, vol. 26, pp. 241-263.
- Lee, K-C (2003). The research on consumer internet information search behavior—backpacker as an example, unpublished master thesis, Department of Business Administration, National Sun Yat-sen University.
- Li, W-C. (2004). The influence of lifestyle and environmental preference on accommodation site choice. —A case study of Lushan Hot Springs and Chingjing Area, unpublished master thesis, Department of Tourism, Providence University.
- Li, Z-L. (2005). A study of the influence on repurchase intention of service quality and customer satisfaction—take independent overseas travelers as an example, unpublished master thesis, Department of International Business Management, Da Ye University.
- Lu, W-H, (2003). Term Translation Extraction Using Web Mining Techniques, unpublished dissertation, Department of Information Engineering, National Chiao Tung University.
- Maedche, A., Motik, B., Stojanovic, L., Studer, R. and Volz, R. (2003). Ontologies for enterprise knowledge management, *IEEE Intelligent Systems*, vol. 18, pp. 26-33.
- Myking, J. (2001). Against precriptivism? The ‘Sociocritical’ challenge to terminology, *IITF Journal*, vol. 12, pp. 49–64.
- Sampson, D., Karagiannidis, C. and Cardinali, F. (2002). An architecture for web-based e-learning promoting re-usable adaptive educational e-content, *Journal of Educational Technology & Society*, vol. 5, pp. 27-37.
- Storey, V. C. (1992). Real world knowledge for databases, *Journal of Database Administration*, Winter, pp. 1-19.
- Su, H-C. (2001). The plan of the travel website based on the information behavior of travelers, unpublished master thesis, Department and Graduate Institute of Library and Information Science, National Taiwan University.
- Swartout, B., Partil, R., Knight, K. and Russ. T. (1997). Toward distributed Use of large-scale ontologies, *Spring Symposium Series on Ontological Engineering*, AAAI Press, Stanford, pp.30-40.
- Taiwan Marketing Research, (2009). The 2009 investigation report of traveling status of Taiwan, Tourism Bureau, Ministry of Transportation and Communication, Taiwan.
- Tseng, H-M. (1999). A study of traveling experience of women individual travelers unpublished master thesis, Graduate School of Forestry, National Taiwan University.
- Wikipedia, <http://en.wikipedia.org/wiki/Ontology>.
- Yang, K-H, (2004). Study of vacation lifestyle and purchase behavior of travel products by Taiwan's young-Adult backpackers engaging in outbound travel, unpublished master thesis, Department of Tourism, Shih Hsing University.

ACKNOWLEDGE

The author would like to give her thanks to S-Y Huang, C-W Ji, H-S Lai, W-J Lee, Y-W Ye, and K-X Yang who dedicated their time on the development of the system of this research.