Collaborative Platform for Multilingual Resource Development and Intercultural Communication

Virach Sornlertlamvanich[1], Thatsanee Charoenporn[1], Kergrit Robkop[1], and Hitoshi Isahara [2]

[1] Thai Computational Linguistics Laboratory, NICT Asia Research Center, Thailand
[2] National Institute for Communications Technology, Japan
{virach, thatsanee, kergrit}@tcllab.org, isahara@nict.go.jp

Abstract. The Internet has already connected people and established a mature online community. The use of WWW is also in the main stream of disseminating one's information and services. In the present information borderless society, we still need a lot of fundamental tools as well as the standard reference resources to facilitate our daily communications across the languages and cultures for better understanding or smoothing the communications. To develop such fundamental linguistic tools, we need to put efforts in cross language study and multilingual resource development. Online collaborative works are efficiently conducted among expert groups via many existing services such as Sourceforge, Wiki or Weblog. Each has already gained its user potential in several online communities. However, in the process of multilingual resource development and intercultural communication we still need to fulfill the requirements in well-structured design of the database, communication tools that provide necessary linkages between records of intention to particular assertions, and functions to realize selectional preference in case that there are more than one assertion. In this paper, we propose a new platform, called Knowledge Unifying Initiator (KUI), for building an online community for a particular purpose corresponding to the activities in human's thinking process: starting from issuing a topic of interest which can be branched into localization, polling and public-hearing corresponding to the maturity of the proposal. We conducted a study on multilingual medical text collaborative translation and the initiative in Asian WordNet development to evaluate our proposed platform.

Keywords: multilingual resource, intercultural communication, collaborative translation, WordNet

1 Introduction

Nowadays, the Internet plays a important role in our daily life communication. It can substitute almost all of our typical communication method i.e. letter, fax, and sometimes telephone in the case that we want to avoid direct speech which may interrupt our partner's inconvenient time. This handy communication environment can also be extended to create a web community where participants can efficiently make use of their available time to share their experiences. In the web community, we are aware of

its global scale of communication though the English language occupies the majority of the online population. In this paper, we proposed a web based communication tool for building a web community. The community members should be able to work collaboratively to create community knowledge disregarding the difference of the languages and be able to get support from others in order to make a community decision. Knowledge Unifying Initiator (KUI) [2], [7] was developed to maintain the process of knowledge creation by providing four possible stages of participation, namely (1) new task, to allow a participant to initiate a task, (2) opinion, to allow a participant to post his own opinion, (3) localization, to allow a participant to bring in a new knowledge into the community by translation, and (4) public-hearing, to allow a participant to post a draft of concept for conceptualizing the knowledge. A community decision on an opinion in any stages can be made by poll taking from the participants. Moreover, an online chat was also provided to keep record of discussion for each topic. These features of KUI were preliminarily integrated and evaluated in a task of Intercultural Collaboration Experiment (ICE) 2003 [1], [6] before extensive development in the current version of KUI. In the experiment, we selected a task of translation of a medical text, i.e. English emergency diagnosis phrases from MedSLT [4], for forming a web community to collaboratively translate the original English text into their native languages. The participants discussed to share their understanding of the original text to conduct a better translation by using the provided chat function. In the same time, they could refer to other translation which could help in improving their own translation. As a result, the multi-lingual translated medical text was produced and the behavior of participants in the web community was studied by analyzing the chat log. We are also providing a new task for Asian WordNet development by preparing records of English WordNet [5], [9] entries together with the word information to help discriminating the meaning of the head word.

The paper is organized in the following way. Section 2 explains the design of KUI for collaborative multilingual resource development. Section 3 describes the experiment of using KUI for medical text translation task. Section 4 discusses the nature of communication by analyzing the chat log. Section 5 describes the potential in adopting KUI for Asian WordNet development.

2 Collaborative Platform for Multilingual Resource Development

We developed KUI (Knowledge Unifying Initiator) for a knowledge development supporting tool of a web community. In this paper, we implemented KUI to be a Knowledge User Interface for a collaborative translation task of a medical text, MedSLT. The evaluation of KUI was also conducted by analyzing the result of communication in the chat log. Actually, KUI is a platform to unify the various thoughts created by following the process of thinking, i.e. (1) new task, to allow a participant to initiate a task, (2) opinion, to allow a participant to post his own opinion, (3) localization, to allow a participant to bring in a new knowledge into the community by translation, and (4) public-hearing, to allow a participant to post a draft of concept for conceptualizing the knowledge. The process of thinking is done under the selectional preference simulated by voting mechanism in the case that many alternatives occur.

In this section, we describe the concept behind KUI, the knowledge development process, various features in KUI, and the implementation of KUI for the collaborative medical text translation.

2.1 What is KUI?

Knowledge Unifying Initiator or KUI is a GUI for knowledge engineering, in other words Knowledge User Interface (KUI) [2], [7]. It provides a web interface accessible for pre-registered members only for the accountability reason. An online registration is offered to manage the account by profiling the login participant in making contribution to the community. A contributor can comfortably move around in the virtual space from desk to desk to participate in a particular task. A login member will be assigned to a desk when a participation task is defined. Members can then participate in the chat group of the same desk. A desk can be a meeting place for collaborative work that needs discussion through the chat function, or allow a contributor to work individually by using the message slot to record each own opinion. The working space can be expanded by closing the unnecessary frames so that the contributor can concentrate on a particular task. All working topics can also be statistically viewed through the provided tabs. These tabs help contributors to understand KUI in the aspects of the current status of contribution and the available tasks. A web community can be formed to create a domain specific knowledge efficiently through the features provided by KUI. These KUI features fulfill the process of human thought to record the knowledge.

In addition, KUI also provides a KUI look up function for viewing the composed knowledge. It is equipped with a powerful search and statistical browse in many aspects. Moreover, the chat log is provided to learn about the intention of the knowledge composers. We frequently want to know about the background of the solution for better understanding or to remind us about the decision, but we cannot find one. To avoid the repetition of a mistake, we systematically provide the chat log to keep the trace of discussion or the comments to show the intention of knowledge composers.

2.2 Knowledge Development in KUI

Adopting the concept of Open Source software development, we will be possibly able to develop a framework for domain specific knowledge development under the web community environment. Sharing and collaboration are the considerable features of the framework. The knowledge will be finally shared among the communities by receiving the consensus from the participants in each step. To facilitate the knowledge development, the process is deliberated into four steps (Sornlertlamvanich, 2006).

- *New Task*

A new task can be posted to draw intention from participants. The only selected tasks by a major vote will then be proceeded for further discussion in the requested type of task i.e. Opinion Poll or Localization or Public-Hearing.

- *Opinion Poll*

The selected task is posted to call for opinions from the participants in this step. Opinion poll is conducted to get the population of each opinion. The result of the opinion poll provides the variety of opinions that reflects the current thought of the communities together with the consensus to the opinions.

- *Localization*

Translation is a straightforward implementation of the localization. Collaborative translation helps producing the knowledge in multiple languages in the most efficient way. Multi-lingual texts are generated in this type of task.

- *Public-Hearing*

The result of discussion will be revised and confirmed by gathering the opinions to develop the final draft of the proposal. Suggestions for revision are ranked according to the vote. The author may consider the weight of suggestion to make decision on the final revision.

The developed knowledge is started from posting 'New Task', participants express their supports by casting a vote. Upon a threshold the ' New Task ' is selected for conducting a poll on 'Opinion', or introducing to the community by 'Localization', or posting a draft for 'Public-Hearing' to gather feedbacks from the community. The transition from 'Opinion' to either 'Localization' or 'Public-Hearing' occurs when the 'Opinion' has a concrete view for implementation. The discussion in 'Localization' and 'Public-Hearing' is however interchangeable due to purpose of implementation whether to adopt the knowledge to the local community or to get feedbacks from the community.

The knowledge creating is managed in 4 different categories corresponding to the stage of knowledge. Each individual in the community casts a vote to rank the appropriateness of solutions at each category. The community can then form the community knowledge under the 'Selectional Preference' background.

2.3 Features in KUI

Poll-based Opinion or Public-Hearing

A contributor may choose to work individually by posting an opinion e.g. localization, suggestion etc., or join a discussion desk to conduct 'Public-Hearing' with others on the selected topic. The discussion can be conducted via the provided 'Chat' frame before concluding an opinion. Any opinions or suggestions are committed to voting. Opinions can be different but majority votes will cast the belief of the community. These features naturally realize the online collaborative works to create the knowledge.

Individual or Group Work

Thought may be formed individually or though a concentrated discussion. KUI facilitates a window for submitting an opinion and another window for submitting a chat

message. Each suggestion can be cast through the 'Opinion' window marked with a degree of its confidence. By working individually, comments to a suggestion can be posted to mark its background to make it better understanding. On the other hand, when working as a group, discussions among the group participants will be recorded. The discussion can be resumed at any points to avoid the iterating words.

Record of Intention
The intention of each opinion can be reminded by the recorded comments or the trace of discussions. Frequently, we have to discuss again and again on the result that we have already agreed. Misinterpretation of the previous decision is also frequently faced when we do not record the background of decision. Record of intention is therefore necessary in the process of knowledge creation. The knowledge interpretation also refers to the record of intention to obtain a better understanding.

Selectional Preference
Opinions can be differed from person to person depending on the aspects of the problem. It is not always necessary to say what is right or what is wrong. Each opinion should be treated as a result of intelligent activity. However, the majority accepted opinions are preferred at the moment. Experiences could tell the preference via vote casting. The dynamically vote ranking will tell the selectional preference of the community at each moment.

2.4 KUI for Collaborative Translation Task

In this collaborative text translation, participants of different mother language work online as a virtual group by using KUI. After registering the system, KUI automatically provides a group of discussion for each task. The group consists of participant from different languages. Multi groups operate in parallel. Before translating, they are encouraged to discuss by the provided chat function about the topic in question, system, personal information and so on.

3 Experiment

According to the purpose of developing a multi-lingual medical text, we set an experiment for online collaborative translation task in the ICE experiment [1], [6]. The source text used in the experiment is a collection of English emergency diagnosis phrases from MedSLT, an Open Source project for developing a medical speech translation system [4]. The translation task is done collaboratively online via KUI interface. The volunteer translators join the discussion group to translate the source text (English) into their native language sentence by sentence. Each group can be participated by translators of more than one language. During the translation task, they all are encouraged by the group communication to build their own community.

As a result, we obtain both the translated medical text and "chat log" which is considered to be the background intention of translation. This chat log will later be a resource for further analysis on cross language communication.

3.1 Medical Text

As described in the previous section, the source text for the translation is the English emergency diagnosis phrases from MedSLT which doctors suspect that patients may be suffering from. These phrases includes the range of utterances of standard examination question about chest pain and factors that increase or decrease such pain, which can be accomplishedly communicated by one or two word responses or gestures [8]. The patterns of the question utterances are grammatically enough to ask about most domain concept in a natural way.

In terms of sentence pattern, the questions are usually limited to be the basis form of the followings. Table 1 shows the type of utterance.

Do you … ?, Have you… ?
How long … ? How (usually, often, ever, …) … ? When … ?
or
Does (it) cause (a symptom)?

Table 1. Type of utterance.

Type of Utterance	Sample
Sentence	Have you had pain for weeks? How long do your headaches usually last? Do you ever have chest pain in the morning? Do your chest pains appear at night is the pain gradual?
Phrase	muscle aches after meals high blood pressure heat cheese

Currently, there are 915 utterances in the experiment including noun phrases, verb phrases and simple sentences.

3.2 Participation

The volunteer participants are to translate the English medical text into their native languages, by using KUI. They act as a virtual group and participate in the translation via this web interface. With different backgrounds and degrees of translation abilities, they, therefore, are encouraged to chat, discuss or exchange their opinion while translating each utterance. The communication is not only for getting to know each other, but also for better understanding of the utterance before translation. Figure 1 shows the participation work flow.

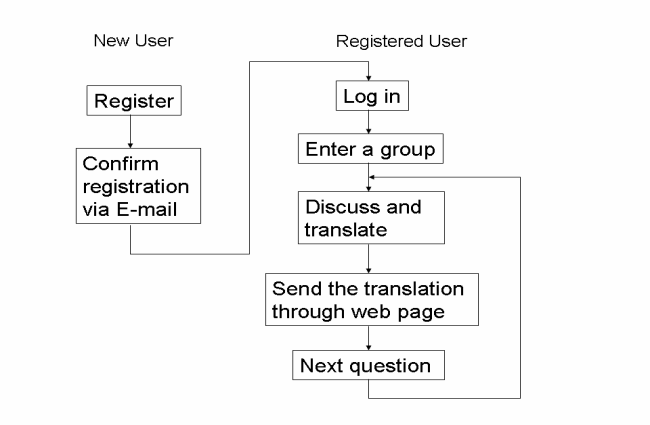

Fig. 1. Participation work flow of collaborative translation task.

There were 3 tracks in the experiment. Each consisted of participants from two different languages as shown in Table 2. Each group was assigned automatically to translate different questions into their languages.

Table 2. Participation in the ICE experiment.

Country pair	Language pair	No. of Group	No. of Participant
Thailand – China	Thai – Chinese	2	4
Thailand – Japan	Thai – Japanese	5	10
Thailand – Korea	Thai – Korean	3	6
Total		10	20

And after two days of experiment, 539 utterances were translated from these 20 participants.

4 Analysis of Chat Log

From the chat log, we can divide the topics on which participants talk together during the experiment into two groups. One is the group of topics related to the utterance in question and another is the group of topics related to personal data enquiry.

It is noted that the conversation texts are not so grammatical. One possible reason is that the conversations were done promptly and the addressers did not pay so much attention on the texts but rather on the quick responding time. This also shows that non-native English speakers are not quite familiar with the expression and they choose to communicate in short using a common text straight to the meaning.

Lacking of revision time in writing the conversation text, we can find the influence of the culture and the mother tongue of the addressers. Some particles are seen added to the sentence. For example, a particle “na” at the end of sentences such as “but i still in the medicine na” in case of Thai addressers. Copula such as verb to be

and articles such as “the” are frequently omitted because there is no such kind of word used in the expression. Noun-verb agreement for example in “that because it take time to take you messages from Server” is often neglected. These can be seen in Thai, Japanese and Chinese addressing texts. Except for the common usage mistakes in such as misspelling, capitalization, omission etc., Table 3 shows the analysis result of the influence of culture and language background of addressers.

Table 3. Type of grammatical errors influenced by the culture and language background.

Type of grammatical error	**Example**
Omission	
- Punctuation mark	what do you think about better how about your name
- Article	sometime we translate same sentence
Misuse	
- Copula be	system is delay i'm write in oppose meanging...yeap gogo what's your name mean..in thai?
- Subject-verb agreement	better mean... worse?? headache decrease It seem my computer don't reset
- Word class	Type so quick but typing quick doens't make time follow quick
- Tense	now it came back again.
Addition	
- Preposition	the ideal thing is, that we help to each other
- Particle	go ne? but I still in medicine na
- Mixture of mother tongue	hello p'som chai mai

Being interested in the culture and language background, we decided to show the original text without any modification for thorough understanding.

Topics related to the utterance in question

The topics related to the sentence in question can be classified into four sub-topics as followings.

- Requiring more information about the system, KUI

There are some questions asking about how to work with KUI, components and features of KUI. For example:

English text: Does it hurt in the left chest?
Thai1: what does desk 1, desk 4 mean? **Thai2**: It means you do different topics **Thai1**: but I still in medicine na **Thai2**: yeah, medicine has a lot of topics **Thai1**: but i still in the medicine na

Thai1: all right
Thai2: if you stay in the same group so you do in the same item in that topic
Thai2: sorry same 'Desk' not same group
Thai1: can i use Thai here?
Thai2: of course, yes
Thai1: why I don't see my message?
Thai2: every things in here show in Unicode, you can type every languages
Thai2: that because it take time to take you messages from Server

- Discussion on the question for detail understanding

The participants are encouraged to discuss to understand the question before translating. This is because the original utterances are defined for a given English or Western context. For example:

English text: Can exertion sometimes cause chest pains?
Japanese1: ok
Japanese2: if you have same words like 'some times' do you use the translation you did before?
Japanese1: i am trying to mention the possibility

English text: Can exertion sometimes give you chest pains?
Japanese1: because sometimes means there are occasion that symptom doesn't occur
Japanese2: that's right.. and also I am not sure the frequency of 'sometimes'

English text: Can your headache be caused by exertion?
Japanese2: for example two or three times par day... or something like that..but 2 or 3 par day is a lot ! :)

English text: Can exertion cause abdominal pains?
Japanese1: even in the Japanese language definition of 'sometimes' in the Japanese language definition of 'sometimes' = tokidoki difficult
Japanese1: excuse me. my last sentence was difficult to read
Japanese2: and also we should take care not to translate into particular symptoms..
Japanese2: abdominal pain is general I think
Japanese1: yes I think it's correct
Thai: maybe
Japanese1: it doesn't immediately mean 'cancer' or some particular diseases ..

- Problems found during translation relating to the system. For example:

English text: Does it hurt in the lower abdomen?
Thai: it's hard to use the tools. cursor jumps all the time :(
Thai: yes
Thai: when i write chat box, it often jumps to translation box.
Thai: i have to wait until start icon activates, right?
Thai: Anybody there? can u read me?
Chinese: yes, i ok now, go ahead
Chinese: yes, i ok now, go ahead

- The comment on the utterance

During the translation, some participants expressed their comment concerning to the utterance such as linguistic knowledge, ambiguity of the utterance, etc. For example:

English text: a stabbing pain
Thai1: quite difficult...
Thai1: Are U OK Ou?
English text: after meals
Thai2: Do you think it's difficult?
Chinese: has login

Topic related to the personal data enquiry

This kind of topic includes general conversations which are not related to the translation task, such as greeting, persuading to join the experiment next time, background of the participants (nationality, country, address, age and so on). For example:

English text: Is there family history of heart disease?
Thai1: hello guys!
Japanese2: hello!
Thai1: you are so on time!
Thai1: it's just 3.00 p.m. in Thai
Japanese1: hello!
Japanese2: yes, here in France, it's 10 in the morning
Thai1: oh i c
Thai2: how are u?
Japanese2: so we are all three or more?
Thai1: did your guys type anything?
Japanese2: type? where? here? or translation box?
Thai1: i see nothing in the translation tab
Japanese2: ah, translated already?
Thai1: yes, i just done it
Thai1: and yes for the upper question, we have 3 person in our group
Japanese2: hello, yes, I came here to know about the situation

English text: Is your headache caused by bright light?
Thai: have you ever been out of your country? **Japanese**: what do you like? **Thai**: i like movies, traveling, reading, swimming, hanging out with friends **Japanese**: I never go to out of japan **Thai**: i also like the beaches **Japanese**: it is nice **Thai**: yeap here has many nice beaches

5 Initiative in Asian WordNet Development

Adopting the same translation framework, we can make use of KUI for making a translation version of WordNet [5], [9] and also establish a link between the translated WordNet's via the synset ID. We convert the Princeton's WordNet into KUI's internal database. In the Princeton's WordNet 2.1 [9], there are 145,103 nouns, 24,890 verbs, 31,302 adjectives, and 5,720 adverbs differentiated by the synset ID. Since there are a large number of records in noun, we group them into a group of frequent noun and the rest. As a result, we have a group of frequent noun of 28,421 nouns according to the frequency count of noun in Penn Treebank WSJ corpus [3]. The group of frequent noun may have a priority in translation.

In each record, only necessary information for understanding the meaning of the head word will be depicted. Table 4 shows a record of WordNet displayed for translation.

Table 4. A record of WordNet.

Car
[Options] **POS**: NOUN **Synset**: auto, automobile, machine, motorcar **Gloss**: a motor vehicle with four wheels; usually propelled by an internal combustion engine

Each record is assigned a synset ID. Translated word will be attached to each synset ID. As a result, a multi-lingual WordNet will be aligned according to the synset ID, and each language WordNet can be generated by substitution of the words in the synset.

Translation of WordNet is an appropriate task for open collaborative work and can yield an effective result. Word by word translation is simple enough for open participation. The provided synset is also a good information source for meaning differentiation and measurement to find the relationship between a particular concept pair.

6 Conclusion

We proposed an efficient online collaborative framework in producing and maintaining the multi-lingual resources. KUI was designed to support an open web community by introducing a voting system and a mechanism to realize the function of selectional preference. It was efficiently introduced to encourage the communication among participants from different language background by providing a task of translating a list of medical text. We collected and analyzed their communication resulting in a set of common segment of the online communication. This will lead to an efficient retrieval system of the response to either the request of knowledge about the system or the topic in question. KUI was also proved to support the collaborative work in producing the multilingual medical text. The translated text will be voluntarily maintained by the online participants under the selectional preference based on the voting function. In future, we plan to provide a task for Asian WordNet development. Each word entry is given the information of part-of-speech, synset and gloss to determine the meaning. Translation of the word entry will establish a link between the synset ID and the translated word. This will result the aligned word entries between languages and each language WordNet can be generated by replacing the translated words in the synset.

References

1. ICE. http://character.kuis.kyoto-u.ac.jp/ICE/ (2006)
2. KUI. http://www.tcllab.org/kui/ (2006)
3. Marcus, M. P., Santorini, B. and Marcinkiewicz, M.A.: Building a Large Annotated Corpus of English: The Penn Treebank. Computational Linguistics, 19(2), 313-330 (1993)
4. MedSLT. http://sourceforge.net/projects/medslt/ (2006)
5. Miller, G. A.: WordNet: A Lexical Databases for English. Communications of the ACM, 39-41, November (1995)
6. Nomura, S., Ishida, T., Yasuoka, M.: Open Source Software Development with Your Mother Language: Intercultural Collaboration Experiment 2002. HCI International (2003)
7. Sornlertlamvanich, V.: KUI: The OSS-Styled Knowledge Development System. Handbook of The 7th AOSS Symposium, Kuala Lumpur, Malaysia (2006)
8. Starlander, M., Bouillon, P., Rayner, M., Chatzichrisafis, N., Hockey, B. A., Isahara, H., Kanzaki, K., Nakao, Y. and Santaholma, M.: Breaking the Language Barrier: Machine Assisted Diagnosis using the Medical Speech Translator. Proceedings of the XIX Internation Congress of the European Federation for Medical Informatics MIE, Geneva, Switzerland (2005)
9. WordNet. http://wordnet.princeton.edu/ (2006)