KUI: Self-organizing Multi-lingual WordNet Construction Tool

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

[1] Thai Computational Linguistics Lab.,
NICT Asia Research Center, Pathumthani, Thailand
[2] National Institute of Information and Communications Technology
3-5 Hikaridai, Seika-cho, Soraku-gaun, Kyoto, Japan 619-0289
{virach, thatsanee, krergrit}@tcllab.org, isahara@nict.go.jp

Abstract. This paper describes a multi-lingual WordNet construction tool, called KUI (Knowledge Unifying Initiator), which is a knowledge user interface for online collaborative knowledge construction. KUI facilitates online community in developing and discussing multi-lingual WordNet. KUI is a sort of social networking system that unifies the various discussions following the process of thinking model, i.e. initiating the topic of interest, collecting the opinions to the selected topics, localizing the opinions through the translation or customization and finally posting for public hearing to conceptualize the knowledge. The process of thinking is done under the selectional preference simulated by voting mechanism in the case that there are many alternatives. By measuring the history of participation of each member, KUI adaptively manages the reliability of each member's opinion and vote according to the estimated *ExpertScore*. As a result, the multi-lingual WordNet can be created online and produce a reliable result.

Keywords: Multi-lingual WordNet, KUI, *ExpertScore*, social networking system, information reliability.

1 Introduction

The constructions of the WordNet [1] for languages can be varied according to the availability of the language resources. Some were developed from scratch, and some were developed from the combination of various existing lexical resources. Spanish and Catalan Wordnets[1], for instance, are automatically constructed using hyponym relation, mono-lingual dictionary, bi-lingual dictionary and taxonomy [2]. Italian WordNet [3] is semi-automatically constructed from definition in mono-lingual dictionary, bi-lingual dictionary, and WordNet glosses. Hungarian WordNet uses bi-lingual dictionary, mono-lingual explanatory dictionary, and Hungarian thesaurus in the construction [4], etc.

---

[1] http://www.lsi.upc.edu/~nlp/

420   Virach Sornlertlamvanich et al.

A tool to facilitate the construction is one of the important issues related to the WordNet construction. Some of the previous efforts were spent for developing the tools such as Polaris [5], the editing and browsing for EuroWordNet, and VisDic [6], the XML based Multi-lingual WordNet browsing and editing tool developed by Czech WordNet team. To facilitate an online collaborative development and annotate a reliability score to the proposed word entries, we, therefore, proposed KUI (Knowledge Unifying Initiator) to be a Knowledge User Interface (KUI) for online collaborative construction of multi-lingual WordNet. KUI facilitates online community in developing and discussing multi-lingual WordNet. KUI is a sort of social networking system that unifies the various discussions following the process of thinking model, i.e. initiating the topic of interest, collecting the opinions to the selected topics, localizing the opinions through the translation or customization and finally posting for public hearing to conceptualize the knowledge. The process of thinking is done under the selectional preference simulated by voting mechanism in the case that there are many alternatives.

This paper illustrates an online tool to facilitate the multi-lingual WordNet construction by using the existing resources having only English equivalents and the lexical synonyms. Since the system is opened for online contribution, we need a mechanism to inform the reliability of the result. We introduce *ExpertScore* which can be estimated from the history of the participation of each member. The weight of each vote and opinion will be determined by the *ExpertScore*. The result will then be ranked according to this score to show the reliability of the opinion.

The rest of this paper is organized as follows: Section 2 describes the process of managing the knowledge. Section 3 explains the design of KUI for Collaborative resource development. Section 4 provides some examples of KUI for WordNet construction. And, Section 5 concludes our work.

2   Process of Knowledge Development

A thought is dynamically formed up by a trigger which can be an interest from inside or a proposed topic from outside. However, knowledge can be formed up from the thought only when managed in an appropriate way. Since we are considering the knowledge of a community, we can consider the knowledge that is formed by a community in the following manner.

- Knowledge is managed by the knowledge users.
- Knowledge is dynamically changed.
- Knowledge is developed in an individual manner or a community manner.
- Knowledge is both explicit and tacit.

The environment of online community can successfully serve the requirement of knowledge management. Under the environment, the knowledge should be grouped up and narrowed down into a specific domain for each group. The domain specific group can then be managed to generate a concrete knowledge after receiving the consensus from the participants at any moment.

Open Source software development is a model for open collaboration in the domain of software development. The openness of the development process has

KUI: Self-organizing Multi-lingual WordNet Construction Tool   421

successfully established a largest software community that shares their development and using experience. The activities are dedicated to the domain of software knowledge development. SourceForge.net[2] is a platform for project based Open Source software development. Open Source software developers deploy SourceForge.net to announce their initiation, to call for participation, to distribute their works and to receive feedbacks concerning their proposed software. Developers and users are actively using SourceForge.net to communicate with each other.

Adopting the concept of Open Source software development, we will possibly be able to develop a framework for domain specific knowledge development under the open 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, we deliberate the process into 4 steps.

1) Topic of interest
The topic will be posted to draw the intention from the participants. The selected topics will then be further discussed in the appropriate step.

2) Opinion
The selected topic 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.

3) Localization
Translation is the straightforward implementation of the localization. Collaborative translation helps producing the knowledge in multiple languages in the most efficient way.

4) Public-Hearing
The result of discussion will be revised and confirmed by gathering the opinions to the final draft of proposal.

Fig. 1 shows the process of how knowledge is developed within a community. Starting from posting 'Topic of Interest', participants express their supports by casting a vote. Upon a threshold the 'Topic of Interest' 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

---

[2] http://www.sourceforge.net/

422   Virach Sornlertlamvanich et al.

community knowledge under the 'Selectional Preference' background. On the other hand, the under-threshold solutions become obsolete by nature of the 'Selectional Preference'.
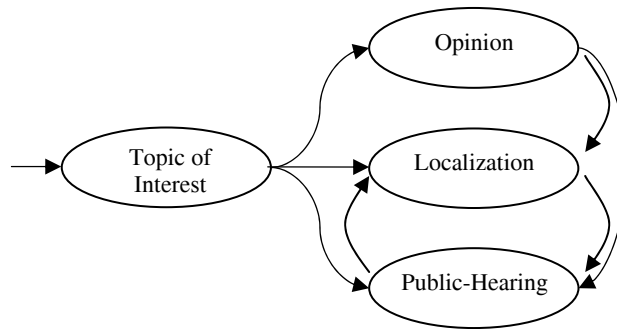

Fig. 1. Process of knowledge development

3   Knowledge User Interface for Knowledge Unifying Initiative

3.1   What is KUI?

KUI is a GUI for knowledge engineering, in other words Knowledge User Interface (KUI). It provides a web interface accessible for pre-registered members. An online registration is offered to manage an account by profiling the login participant in making contribution. A contributor can comfortably move around in the virtual space from desk to desk to participate in a particular task. A working desk can be a meeting place for collaborative work that needs discussion through the 'Chat', or allow a contributor to work individually by using the message slot to record each own comment. The working space can be expanded by closing the unnecessary frames so the contributor can concentrate on the task. All working topics can 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 tasks. A knowledge community can be formed and can efficiently create the domain knowledge through the features provided by KUI. These KUI features fulfill the process of human thought to record the knowledge.

   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 'Chatlog' 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 'Chatlog' to keep the trace of discussion or the comments to show the intention of knowledge composers.

3.2 Feature of 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 works**

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 and 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 the moment.

3.3 *ExpertScore*

KUI heavily depends on members' voting score to produce a reliable result. Therefore, we introduce an adjustable voting score to realize a self-organizing system. Each member is initially provided a default value of voting score equals to one. The voting score is increased according to *ExpertScore* which is estimated by the value of *Expertise*, *Contribution*, and *Continuity* of the participation history of each member. *Expertise* is a composite score of the accuracy of opinion and vote, as shown in Equation 1. *Contribution* is a composite score of the ratio of opinion and vote posting

424   Virach Sornlertlamvanich et al.

comparing to the total, as shown in Equation 2. *Continuity* is a regressive function based on the assumption that the absence of participation of a member will gradually decrease its *ExpertScore* to one after a year (365 days) of the absence, as shown in Equation 3.

$$Expertise = \alpha \frac{count(BestOpinion)}{count(Opinion)} + \beta \frac{count(BestVote)}{count(Vote)} \ . \tag{1}$$

$$Contribution = \gamma \frac{count(Opinion)}{count(TotalOpinion)} + \rho \frac{count(Vote)}{count(TotalVote)} \ . \tag{2}$$

$$Continuity = 1 - \left(\frac{D}{365}\right)^4 \ . \tag{3}$$

where,

$$\alpha + \beta + \gamma + \rho = 1$$

D is number of recent absent date ($0 \leq D < 365$)

As a result, the *ExpertScore* can be estimated by Equation 4.

$$ExpertScore = \left(1 - \left(\frac{D}{365}\right)^4\right)$$
$$\times \left\{ \begin{array}{l} \alpha \frac{count(BestOpinion)}{count(Opinion)} + \beta \frac{count(BestVote)}{count(Vote)} \\ + \gamma \frac{count(Opinion)}{count(TotalOpinion)} + \rho \frac{count(Vote)}{count(TotalVote)} \end{array} \right\} \ . \tag{4}$$

If *ExpertScore* is less than one, it is reset to one. The value of *ExpertScore* is ranged between 1 to 365 according to the accuracy and the rate of contribution of each member. This means that reliable members are rewarded better score for each vote. However, the expertise of the member is decreased according to the *continuity* of the participation. By means of the *ExpertScore*, we can rank the opinions precisely and yield reliable results, especially for the results produced by an online community.

4   KUI for Multi-lingual WordNet Construction

There are some previous efforts in developing WordNets for Asian languages, e.g. Chinese, Japanese, Korean [7], [8], [9], [10], [11] and Hindi [12]. The number of languages that have been successfully developed their WordNets is still limited to some active research in this area. However, the extensive development of WordNet in other languages is important, not only to help in implementing NLP applications in each language, but also in inter-linking WordNets of different languages to develop multi-lingual applications to overcome the language barrier.

KUI: Self-organizing Multi-lingual WordNet Construction Tool   425

We adopt the proposed criteria for automatic synset assignment for Asian languages which has limited language resources. Based on the result from the above synset assignment algorithm, we provide KUI (Knowledge Unifying Initiator) [13], [14] to establish an online collaborative work in refining the WorNets.

KUI allows registered members including language experts revise and vote for the synset assignment. The system manages the synset assignment according to the preferred score obtained from the revision process. The revision history reflects in the *ExpertScore* of each participant, and the reliability of the result is based on the summation of the *ExpertScore* of the contributions to each record. In case of multiple mapping, the record with the highest score is selected to report the mapping result.

As a result, the community WordNets will be accomplished and exported into the original form of WordNet database. Via the synset ID assigned in the WordNet, the system can generate a cross language WordNet result. Through this effort, an initial version of Asian WordNet can be established.

Table 1 shows a record of WordNet displayed for translation in KUI interface. English entry together with its part-of-speech, synset, and gloss are provided if exists. The members will examine the assigned lexical entry whether to vote for it or propose a new translation.

Table 1. 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;

Fig. 2. KUI Participation page.

426   Virach Sornlertlamvanich et al.

Fig. 2 illustrates the translation page of KUI[3]. In the working area, the login member can participate in proposing a new translation or vote for the preferred translation to revise the synset assignment. Statistics of the progress as well as many useful functions such as item search, record jump, chat, list of online participants are also provided. KUI is actively facilitating members in revising the Asian WordNet database.
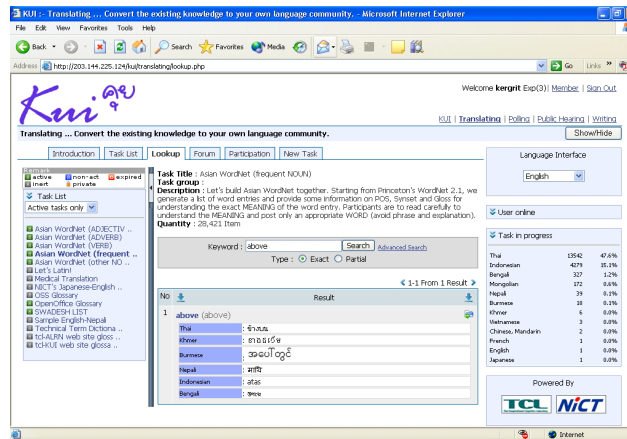

Fig. 3. KUI Lookup page.

Fig. 3 illustrates the lookup page of KUI. The returned result of a keyword lookup is sorted according to the best translated word of each language. The best translated word is determined by the highest vote score. As a result, the user can consult the WordNet to obtain a list of equivalent words of the same sense sorted by the languages. The *ExpertScore* provided in KUI will help selecting the best translation of each word.

5   Conclusion

KUI is a platform for composing knowledge in the Open Source style. A contributor can naturally follow the process of knowledge development that includes posting in 'Topic of interest', 'Opinion', 'Localization' and 'Public-Hearing'. The posted items are committed to voting to perform the selectional preference within the community. The results will be ranked according to the vote preference estimated by the *ExpertScore* for the purpose of managing the multiple results. 'Chatlog' is kept to indicate the record of intention of knowledge composers. A contributor may participate KUI individually or join a discussion group to compose the knowledge. We are expecting KUI to be a Knowledge User Interface for composing the knowledge in the Open Source style under the monitoring of the community. The statistical-base visualized

---

[3] http://www.tcllab.org/kui/

KUI: Self-organizing Multi-lingual WordNet Construction Tool   427

'KUI look up' is also provided for the efficient consultation of the knowledge. We introduce KUI for Asian WordNet development. The *ExpertScore* efficiently ranks the results especially in the case where there is more than one equivalent.

References

1. Fellbaum, C. (ed.): WordNet: An Electronic Lexical Database. MIT Press, Cambridge, Mass (1998)
2. Atserias, J., Clement, S., Farreres, X., Rigau, G., Rodríguez, H.: Combining Multiple Methods for the Automatic Construction of Multilingual Word-Nets. In: Proceedings of the International Conference on Recent Advances in Natural Language, Bulgaria (1997)
3. Magnini, B., Strapparava, C., Ciravegna, F., Pianta, E.: A Project for the Construction of an Italian Lexical Knowledge Base in the Framework of WordNet. IRST Technical Report # 9406-15 (1994)
4. Proszeky, G., Mihaltz, M.: Semi-Automatic Development of the Hungarian WordNet. In: Proceedings of LREC2002. Spain (2002)
5. Louw, M.: Polaris User's Guide. Technical report. Belgium (1998)
6. Horák, A., Smrž, P.: New Features of Wordnet Editor VisDic. J. Romanian Journal of Information Science and Technology. 7(1–2), 201–213 (2004)
7. Choi, K. S.: CoreNet: Chinese-Japanese-Korean wordnet with shared semantic hierarchy. In: Proceedings of Natural Language Processing and Knowledge Engineering. Beijing (2003)
8. Choi, K. S., Bae, H. S., Kang, W., Lee, J., Kim, E., Kim, H., Kim, D., Song1, Y., Shin, H.: Korean-Chinese-Japanese Multilingual Wordnet with Shared Semantic Hierarchy. In: Proceedings of LREC2004. Portugal (2004)
9. Kaji, H., Watanabe, M.: Automatic Construction of Japanese WordNet. In: Proceedings of LREC2006. Italy (2006)
10. Korlex: Korean WordNet. Korean Language Processing Lab, Pusan National University, 2007. Available at http://164.125.65.68/ (2006)
11. Huang, C. R.: Chinese Wordnet. Academica Sinica. Available at http://bow.sinica.edu.tw/wn/ (2007)
12. Hindi Wordnet: Available at http://www.cfilt.iitb.ac.in/wordnet/webhwn/ (2007)
13. Sornlertlamvanich, V.: KUI: The OSS-Styled Knowledge Development System. In: Proceedings of the 7th AOSS Symposium. Malaysia (2006)
14. Sornlertlamvanich, V., Charoenporn, T., Robkop, K., Isahara, H.: Collaborative Platform for Multilingual Resource Development and Intercultural Communication. In: Proceedings of the First International Workshop on Intercultural Collaboration (IWIC2007), LNCS4568, pp. 91–102 (2007)