

Terminology in Korea: KORTERM

Key-Sun Choi and Young-Soog Chae

Korea Terminology Research Center for Language and Knowledge Engineering
Department of Computer Science
Korea Advanced Institute of Science and Technology
373-1 Kusong-dong Yusong-gu Taejon 305-701 Korea
{kschoi, yschae}@korterm.kaist.ac.kr

Abstract

Korterm (Korea Terminology Research Center for Language and Knowledge Engineering) had been set up in the late August of 1998 under the auspices of Ministry of Culture and Tourism in Korea. Its major mission is to construct terminology resources and their unification, harmonization and standardization. This mission is naturally linked to the general language engineering and knowledge engineering tasks including specific-domain corpus, ontology, wordnet, electronic dictionary construction as well as language engineering products like information extraction and machine translation. This organization is located under the KAIST (Korea Advanced Institute of Science and Technology) that is one of national university specifically under the Ministry of Science and Technology. KORTERM is only one representative for terminology standardization and research with relation to Infoterm..

1. Introduction

There are many aspects and disciplines in terminology research (ISO TC37; Sager, 1990). Korterm has responsibility to collect and distribute Korean terminology in all area under the auspices of Ministry of Culture and Tourism in the context of Korean language policy. Institute of Technology Standardization under the Ministry of Industry and Energy also designates Korterm for industrial terminology standardization in the context of industrial Korean Standards. On the other hand, Ministry of Science and Technology has supported by two kinds of reason to support Korterm: one is to pursue the Korean own scientific terms for the right communication and the

other to promote the information processing technologies.

Here, the current state of terminology standardization and research methodology will be introduced.

2. Organization

Korterm includes 35 resident researchers, 7 members of advisory committee, 40 KAIST professors for covering each area, and 15 members of non-resident R&D participants with cooperation with KAIST digital library and KOSEF (Korea Science and Technology Foundation). The resident researchers include quite interdisciplinary PhD researchers who cover the lexicography, formal semantics, thesaurus, Korean morphology and language

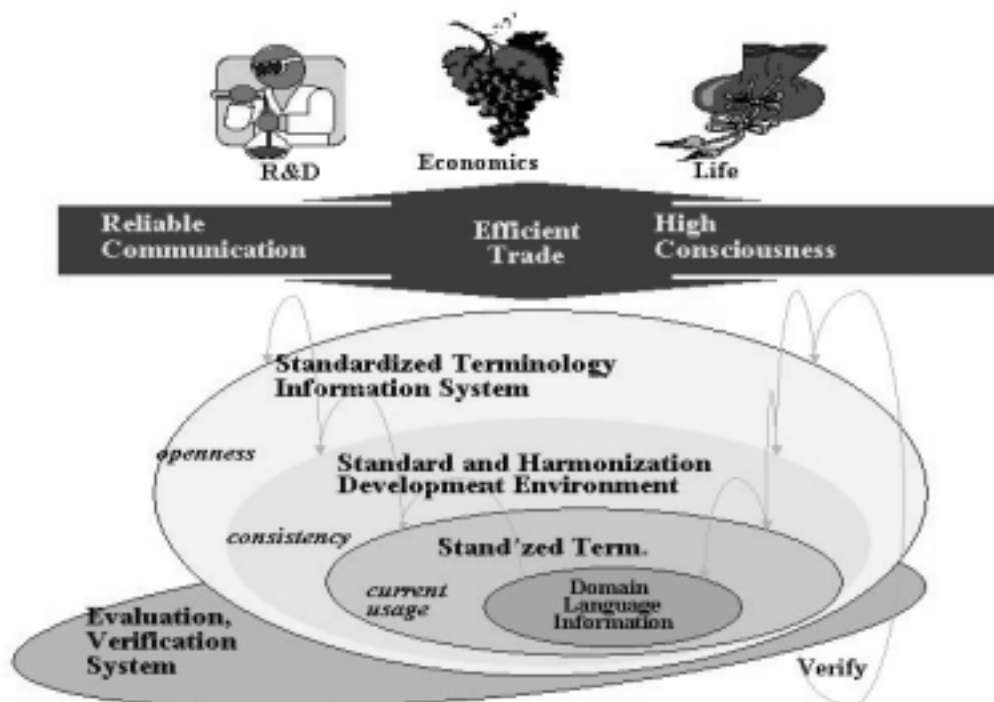


Figure 1. Strategies for Terminology Standardization and Harmonization.

engineering.

Korterm is only one representative for terminology research authority in Korea with relation to Infoterm. In East Asia region, one terminology forum has been operated under the name of Eafterm (East Asian Forum on Terminology) in liaison office of Beijing for joint terminology association of China, Japan, Korea, and Mongolia with cooperation of Infoterm.

3. Mission and Strategies

3.1. Mission and Plan

The major missions of KORTERM include the basic researches for multi-lingual terminology information base construction, the terminology history recording for consistent usage survey, the international cooperation with related organization, the international meeting hosting/participation for terminology research, and the terminology distribution.

Korterm has 10 year plan to be divided into four phases: Phase 1 (1998-2000) is to construct the R&D environment and basic data collection in the field of basic science and technology (that is, mathematics, physics, chemistry, biology, earth science, mechanics and economics), phase 2 (2001-2003) is to build up the high-quality and valued-added terminology under the quality

control process development, phase 3 (2004-2007) is planned to construct the multi-lingual large-scale terminology integration, and then the continuous extension and management in phase 4 (2008-).

3.2. Strategies

Because of the total deficiency of terminology work in Korea, we had to set up strategies to pursue the earliest accomplishment of terminology infrastructure. Here, we have questions to do that: (1) what subject fields have priority for terminology R&D according to national need, industrial need, necessity for people's everyday life? (2) What are the common methodology and component technology for terminology R&D? (3) What application systems need terminology information? (4) What are appropriate formats of terminology and dictionaries for interoperability and quality assurance/evaluation system? (5) What procedures and systems are necessary for standardization and harmonization? (6) What supports are necessary for continuous extension and maintenance of R&D products? (7) What supports are necessary for more effective distribution and utilization by more users?

Standardized terminologies must come from the current usage in the society, which are extracted and surveyed from the domain-specific language information as shown in figure 1. We need a development environment for standardized and harmonized terms that

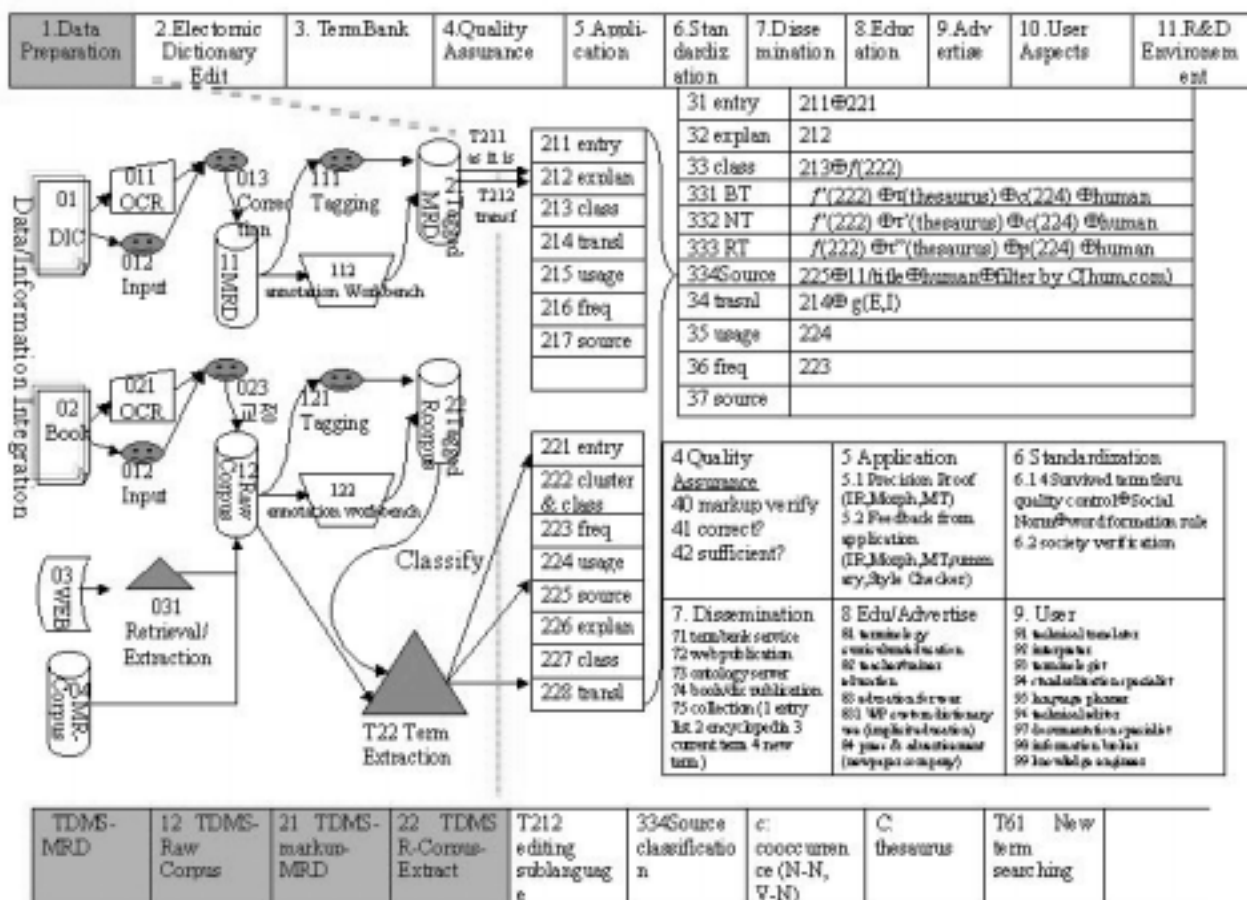


Figure 2. Terminology development process: 10 processes in the top line of boxes. In the bottom line of boxes, there are necessary tools and information, where TDMS means "Text and Data Management System" for dealing with dictionaries and corpora (Choi, *et al.*, 1997). "R-corpus" means "raw corpus", "MR-corpus" is "machine-readable corpus", "MRD" is "machine-readable dictionary". Functions for construction of BT, NT and RT are not explained here.

will be investigated in the next section.

4. Terminology Development

4.1. Terminology Development Process

A terminology development process (figure 2) is designed to consist of 11 components: (1) data preparation, (2) electronic dictionary edition, (3) termbank construction, (4) quality verification, (5) application testing for information technology (e.g., information retrieval and machine translation), (6) unification, standardization and harmonization of terminology and concepts, (7) distribution issues for termbank service, electronic publication, encyclopedic knowledge base and current new term discovery, (8) education methodology for terminology and customization for the real use as in word processor, (9) user adaptation, for example, for technical translator, technical editor, information broker, documentation specialist and so on, (10) R&D environment like terminology extraction, management method for terminology, knowledge and corpora, information extraction application, and so on.

Term history is crucial to maintain the real usage in time line consistently. With rational history of term change, no one wants to use the term. Terms in special domain are not naturally created but some person defines them in some article. Even if it is defined, the term may not be used if the expert group does not agree with it. This task starts from the corpus collection and construction.

This terminology study needs the workbench for terminology browser and corpus retrieval environment in

time line, multi-lingual classified lexicon, terminology identifier and extractor, terminology usage retrieval like KWIC, multi machine readable dictionary searcher for integrated search of each term and so on, as shown in the bottom line of figure 2.

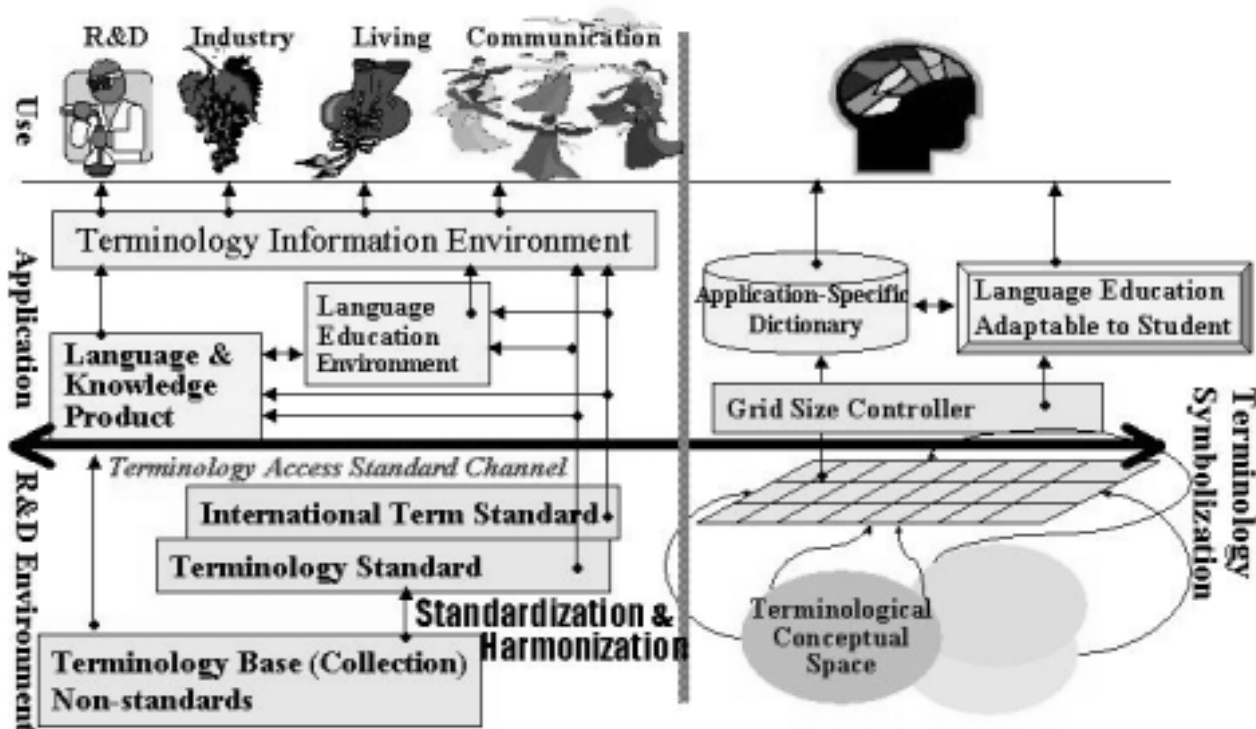
4.2. Terminology Formation

90% of terms in Korea are derived from foreign language like English, Japanese, and other European languages. Every term has to be translated or transliterated to Korean (Kang, *et al.*, 2000). When we build up the corpus, bi-lingual texts are preferably selected for the comparison study. Terms are tested through several processes. First, their occurrences are counted in the corpus. Second, their word formation adequacy is tested through linguistic evaluation (Song, 2000). Third, the current usage is investigated through that domain's dictionary and other domains' dictionary for the concept harmonization. Then, finally they are evaluated by that domain's experts as well as by the other domain's experts in the cross-domain checking manner.

4.3. Terminology Processing

There are 400 terminology dictionaries published in Korea. Among them, we can buy around 200 books. They are used for the starting resources to make a standardized term set. In case of economic terms, there are at least seven subcategorizations, for example, trade, management, stock, banking, accounting, *etc.* The first process is to extract terms from corpus and to make confirm how frequent the terms occur in the current usage. The new terms are extracted by the word formation rule as well as

Figure 3. Terminology R&D and Use: In the R&D environment, the major task is to connect the conceptual space with the word formation, which is expressed by "Terminology Symbolization". "Grid size controller" is to control the usage of word depending on the domain and the application.



by the statistical way (Oh *et al.*, 1999).

The study on terminology is not limited to noun. Verbs are the first citizen of terminology standardization. For example, "run" and "pass" are translated into the same verb "dollida" in Korean respectively for two phrases: "run the program" and "pass the paper" in computer field and office field respectively.

All terms are collected and compared with the multi-lingual basis. Especially, Korean terms are compared and formulated with Japanese, Chinese and English terms.

4.4. Terminology Use and Application

Figure 3 shows the terminology use and application. One of application is found in Lee, *et al.* (1999). Terminology dictionaries can be utilized to the automatic document classification. The right term use helps the education adaptable to the students, that is expressed here by "grid size controller". "Terminology symbolization" is to select the proper terms that are matched with the concepts. "Terminology access standard channel" has been discussed in the context of ISO TC37.

5. Conclusion

In this paper, the overall introduction to Korterm and its activity was introduced.

Korterm is only one representative for terminology research authority in Korea with relation to Infoterm in Austria. In East Asia region, one terminology forum has been operated under the name of Eafterm (East Asian Forum on Terminology) in liaison office of Beijing for joint terminology association of China, Japan, Korea, and Mongolia with cooperation of Infoterm.

The overall work is shown in <http://korterm.or.kr/> (in Korean) and <http://korterm.org/> (in English). Other works can be shown in <http://csfive.kaist.ac.kr/kcp/> for the concordance service of Korean and Japanese corpora. The compound noun formation and word frequencies are searched. The Korean corpora and related works can be found in <http://kibs.kaist.ac.kr/> in Korean language. Other works are in <http://korterm.kaist.ac.kr/nlplab/>.

6. Acknowledgement

Korterm is sponsored by Ministry of Culture and Tourism under the program of King Sejong Project. Many fundamental researches are supported by the R&D fund "Large Scale Korean Language Processing and Quality Management" under the plan "Software Technology Enhancement 2000" by Ministry of Science and Technology. The Korea Institute of Technology Standardization supports ISO related tasks.

7. References

- Choi, Byung-Jin, Jae Sung Lee, Woon-Jae Lee, Key-Sun Choi (1997) The Processing of Dictionary Definitions and Maintenance in Text and Dictionary Management System (TDMS). In *Proceedings of the 17th International Conference on Computer Processing of Oriental Languages (ICCPOL'97)* (pp. 443-447). Hong Kong.
- Eafterm, <http://www.eafterm.org/>.

Infoterm, <http://www.infoterm.or.at/>.

ISO TC37, <http://www.iso.ch/meme/TC37.html>.

Kang, Byung-Ju and Key-Sun Choi (2000) *Automatic Transliteration and Back-Transliteration by Decision Tree Learning*. In LREC-2000 Proceedings, Athens.

Lee, Kyungsoon and K.S. Choi (1999) *Document Classification System based on Terminology and Information Extraction*. In the Proceedings of Hangul and Korean Language Processing (pp. 79-84). (in Korean).

Oh, Jonghoon and K.S. Choi (1999) *Foreign Word Extraction Model from Technical Documents Using HMM*. In the Proceedings of Hangul and Korean Language Processing (pp. 137-141). (in Korean).

Sager, J.C. (1990) *A Practical Course in Terminology Processing*. Amsterdam: John Benjamins Publishing Company.

Song, Youngbin (2000) *On History and State of Terminology in Korea*. In the Journal of Japan Terminology Association. (in Japanese)