# 의미분석 지식베이스를 위한 SENKOV 시스템의 구현

## Implementation of SENKOV System: A Knowledge Base for Semantic Analysis

문 유 진 (Yoo-Jin Moon) 한남대학교 경영정보학과

목 차



***Keywords:*** *SENKOV, User Interface, Semantic Networks, Isa-Hierarchies Co-occurrence Constraint, Selectional Restriction*

# Ⅰ. Introduction

World Wide Web(WWW) has recently changed information technologies dramatically. But as most information on the web is written in English, it is difficult for east-Asians using non-English to communicate with others through WWW. And as language processing technologies have been developed mainly for English, they have to be studied for east-Asian languages, too. Since language processing technologies and internet information technologies are tightly coupled as very essential issues, they are important for people including east-Asians who live in internet governance times. Under this recognition, this paper tries to implement a knowledge base for the semantic analysis of natural language processing(NLP) for Korean.

The semantic analysis is indispensable for the understanding of Korean speech and Korean text, for the lexical information technique in information communication and for integration of the different semantic information (이정민 등, 1997; Hernert, 1994; Montemagni 등, 1992). For efficient semantic analysis, the verb in the sentence should be recognized not as the string itself, but as the verb concept of the semantic network for the mutual relationships among the verbs.

Semantic networks for Korean verbs(SENKOV) is a necessary component for recognizing the verb concept. SENKOV is able to provide the verb concept with selectional restrictions in the role of a sentence and to efficiently perform NLP. Needless to say, it is necessary to construct SENKOV for efficient processing of the applications of Korean NLP and IR. Because constructing SENKOV needs a long time and needs many experts, it has not yet been completed. Many countries (U.S.A., Germany, France, Italy and Spain etc.) have been working for building semantic

networks of their own language.

The paper presents methodology and techniques for design and implementation of the SENKOV System based on the validation of set membership and dictionaries. And it performs verb concept classification available for establishing the co-occurrence constraint relationships among adverbs and verbs. The paper is important in that it has made the first attempt at classifying Korean verb concepts for the semantic analysis of NLP.

# Ⅱ. Literature Review

## 2.1 English WordNet

WordNet is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory (Miller 등, 1993). English nouns, verbs, adverbs and adjectives are organized into synonym sets, each representing one underlying lexical concept. WordNet presently contains approximately 120,000 word forms. WordNet may be viewed as semantic networks which represent hypernyms of English word senses in the form of isa-hierarchies.

WordNet, however, does not systematically classify top nodes of verbs, overclassifies verbs into similar concepts and does not distinguish intransitive verbs from transitive verbs.

## 2.2 Levin Verb Classes

Levin verb classes (Levin, 1997; Levin 등, 1996) contain various syntactically relevant and semantically coherent verb classes of English. It takes a semantic classification structure. Also it incorporates syntactic relationship in the semantic relationship for verbs, includes selectional restrictions of verbs and classifies approximately 3,000 verbs into 49 verb classes.

It is organized in order to group together verb classes by meaning. However, there is little hierarchical organization in comparison to the number of classes identified. And it does not fit top nodes for all of the Korean verbs.

## 2.3 Semantic Networks for the European Languages

Germany, Spain, France and Italy are working on constructing Euro-WordNet and their own semantic networks independently. They have been working on constructing their own semantic networks for nouns for many years, but not for verbs.

# Ⅲ. Design and Implementation of the SENKOV System

## 3.1 Design of the SENKOV System

In the design of the SENKOV System, the most important task is to come up with the way of representing the nodes of word concept and to establish the relation of their hierarchies. There are two ways of representing the word concept, that is, by the constructive theory and by the differential theory. The SENKOV System adopts the differential theory and represents the word concept with symbols differentiated for each concept. The symbols are represented by synonym sets (synsets) of English WordNet and those of Korean nouns, and their hierarchies are represented by hypernymous relations among these symbols.

With the strategies described in section 3.1.1., the SENKOV System consists of the Creation System and the Database Construction System.

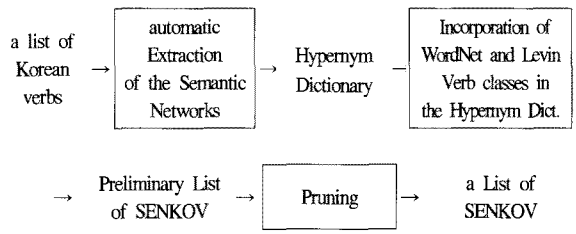
### 3.1.1 Design Strategies

The design strategies of the SENKOV System are as follow (문유진 등, 1994; 문유진, 1996; Montemagni 등, 1992; Sumita 등, 1993).

1) The SENKOV System adopts the differential theory and represents the word concept with symbols differentiated for each concept. The symbols are represented by synonym sets (synsets) of English WordNet and those of Korean nouns.
2) Hierarchies of the SENKOV System are represented by hypernymous relations among synsets.
3) The SENKOV System is designed on the basis of the definition in the Korean Dictionary.
4) The SENKOV System attempts to incorporate syntactic relationship in the semantic relationship for Korean verbs. The syntactic relationship deals with the predicate-calculus forms of verbs, and the semantic relationship deals with the sub-categorization information as well as concepts for verbs. This is because the syntactic and semantic relations simultaneously affect the semantic analysis of sentences.
5) The SENKOV System refers to the top nodes of Levin verb classes, which do not fit those of all Korean verbs and lack in the hierarchical organization. Also, it adjusts (deletes, updates and inserts) top nodes by referring to the Korean dictionary according to the characteristics of Korean verbs.
6) The SENKOV System deals with the verb classes of the root forms of Korean verbs. Later we intend to extend it to verb classes of the derived forms, passive forms and the causative forms of the verbs.
7) The SENKOV System distinguishes between intransitive verbs and transitive verbs of Korean. WordNet does not make this distinction.
8) The SENKOV System refers to the hierarchies of the Korean hypernym dictionary and WordNet. WordNet does not systematically classify top nodes of verbs and overclassifies verbs into similar concepts.

### 3.1.2 The Creation System

The general flow of the Creation System is illustrated in <Figure 1>. For the input Korean verbs, automatic extraction of semantic networks is performed using the definition of the Korean Dictionary. From the networks, a hypernym dictionary is built for the Korean verbs.




〈Figure 1〉 Flow of the Creation System

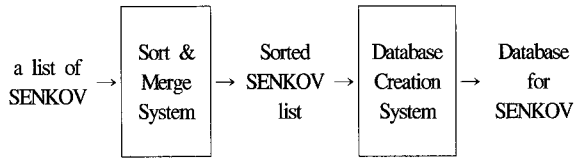
The incorporation of WordNet and Levin verb classes in the hypernym dictionary is performed automatically, and the output is the preliminary list of SENKOV. And then validation of set membership for SENKOV is performed. Finally, the list of SENKOV is produced.

In <Figure 1>, pruning is performed for the validation of set membership based on the design strategies as described in section 3.1.1.

### 3.1.3 The Database Construction System

The general flow of the Database Construction System for SENKOV is illustrated in <Figure 2>. The

list of SENKOV produced by the Creation System in section 3.1.2 is sorted and merged, and a sorted SENKOV list is produced. And with the SENKOV list, the database creation job is performed and the database for SENKOV is constructed. The database consists of a synset file and an index file.




〈Figure 2〉 Flow of the Database Construction System

### 3.2 Implementation of the SENKOV System

We selected 658 Korean verbs which are frequently used in the general domain, and implemented the prototype of the SENKOV System utilizing C++ compiler in the SUN workstation.

〈Table 1〉 Comparison of the Semantic Networks for Korean Verbs and Korean Nouns

| Part of Speech | # of the top node | average depth of the hierarchy |
|---|---|---|
| Korean Noun | 11 | 5.40 |
| Korean Verb | 44 | 2.35 |

According to the results of the experiment, the SENKOV System has 44 top nodes and average 2.35 depths in the hierarchy. <Table 1> compares Korean verbs with Korean nouns in the number of top node and in the average depth of the hierarchy. This means that Korean nouns have quite deep hierarchies and Korean verbs have groups of similar concepts with very shallow hierarchies.

"내쫓다"
2 senses of extradite
Sense 1
extradite
=> banish, evacuate, recall, extradite
=> remove, take, take away, take off
Sense 2
extradite, deliver, deport, surrender
=> expel, turf out, kick out, throw out, cast out
=> move, displace, make move

〈Figure 3〉 An Example of the Implemented SENKOV System

<Figure 3> illustrates an example of the implemented SENKOV System. It is the SENKOV list for the Korean verb "내쫓다 (extradite)", which has 2 word senses. The SENKOV System can be applied to the real system --- machine translation system, machine understanding system, information retrieval system, on-line shopping system and human-computer interface. <Index 1> shows a part of the SENKOV verb classes.

## Ⅳ. Verb Concept Classification Available for the Selectional Restrictions

The SENKOV System has been applied to the verb concept classification available for the selectional restrictions of adverbs.

Korean adverbs are divided into two kinds of adverb --- constituent adverbs and sentence adverbs (남기심 등, 1985). Constituent adverbs qualify the constituent of the sentence, and sentence adverbs qualify the sentence itself. Constituent adverbs are divided into characteristic adverbs, symbolic adverbs and demonstrative adverbs. Characteristic adverbs qualify the verb, the adjective or the noun in the sentence.

Symbolic adverbs represent the onomatopoeic word and the mimetic word. Characteristic adverbs and symbolic adverbs are related with their selectional restrictions for verbs.

<Figure 4> shows that the SENKOV System has been applied to the selectional restriction of characteristic adverbs for verbs. The Korean adverb "세게" means "strongly", and the Korean adverb "잘게" means "finely". The numbers in <Figure 4> are class numbers for Korean verbs. The SENKOV System has been proved to be valid for the selectional restriction of characteristic adverbs.

```
세게 + (18.1, 9.3 ~ 9.7, 10.4, 10.7, 11.2, 12, 15.1, 17, 18,
        21, 22, 38, 40.1 ~ 40.3, 45, 51.3, 57)

잘게 + (21, 10.7)
```

〈Figure 4〉 Examples of the Application of the SENKOV System for the Selectional Restrictions

```
procedure selectional_Restriction_of_Adverbs_for_Verbs(in : SENTENCE)
/* SENTENCE : an input sentence  */
/* ADVERB, VERB : a pair of adverb and verb for selectional restriction  */
/* SEL_RESTRICT : set of selectional restrictions of the adverb for verbs */
/* VERB_CLASS   : SENKOV verb class of the verb to be processed   */
{ ADVERB = VERB = null;
  WELL_FORMED = false;
  find pairs of ADVERB and VERB from the SENTENCE;
  while  (pair of ADVERB and VERB)
{ SEL_RESTRICT = set of selectional restriction of ADVERB;
  VERB_CLASS = SENKOV verb class of VERB;  //find verb concept of VERB.

/* compares the SENKOV verb class (verb concept) of the input verb with each verb class (verb concept) of
   the selectional restriction of the input adverb. */
for  I  in SEL_RESTRICT
{ if (VERB_CLASS == I)
    { WELL_FORMED = true;  break; }
 SUB_RESTRICT = set of SENKOV subclass of I;
 for  J in SUB_RESTRICT
   if (VERB_CLASS == J)
     { WELL_FORMED = true; break; }
   if WELL_FORMED break;
}  //for

// If they match, the algorithm outputs that the adverb and the verb are well formed.
if WELL_FORMED
  { print("Well_Formed", ADVERB, VERB);
    WELL_FORMED = false;
   }  //if
  }  //while
}  //procedure
```

〈Figure 5〉 Algorithm of the Selectional Restriction of Adverbs for Verbs

<Figure 5> displays the algorithm of the selectional restriction of adverbs for verbs. From the input sentence, the algorithm finds all pairs of the adverb and the verb. For each pair of the adverb and the verb, the algorithm checks validity of the selectional restriction of the adverb. It compares the SENKOV verb class (verb concept) of the input verb with each verb class (verb concept) of the selectional restriction of the input adverb. If they match, the algorithm outputs that the adverb and the verb are well formed.

Most of symbolic adverbs have the selectional restrictions on only one or two verbs. <Figure 6> illustrates an example of the application of the SENKOV System for the selectional restrictions on the symbolic adverb "땡땡 (with a clang)". The number in <Figure 6> is the class number for Korean verbs. It shows that the SENKOV System doesn't give a big aid to the selectional restrictions of symbolic adverbs.

| 땡땡 + ("치다 (hit)" in 18.1) |
|---|

〈Figure 6〉 An Example of the Application of the SENKOV System for the Symbolic Adverb

# V. Conclusions

For ease of use and user friendliness in the WWW sites, HCI should be designed with ease and flexibility for users. To accomplish these goals, many aspects of HCI may be considered, one of which is the semantic analysis of NLP.

The paper presented methodology and techniques for design and implementation of the SENKOV System for the semantic analysis of Korean NLP, based on the validation of set membership and dictionaries. And it performs verb concept classification available for establishing the co-occurrence constraint relationships between adverbs and verbs. The SENKOV System is designed on the basis of the definition in the Korean Dictionary. And it refers to Levin verb classes for the establishment of its top nodes and WordNet for the establishment of its hierarchies.

We selected 658 Korean verbs which are frequently used in the general domain, and implemented the prototype of the SENKOV System utilizing C++ compiler in the SUN workstation. The experiments say that the SENKOV System has 44 top nodes and average 2.35 depths in the hierarchy. In addition, the SENKOV System has been applied to the selectional restriction of characteristic adverbs and symbolic adverbs for Korean verbs.

The paper is important in that it has made the first attempt at classifying Korean verb concepts for user friendliness in the WWW sites.

Future works to be done is as follows. Firstly, the SENKOV System should be expanded to more than 3,000 Korean verbs. Secondly, it should be constructed with a consideration for the selectional restrictions of the verb for the subject and the object in sentences.

## References

남기심, 고영근, "표준 국어 문법론," *탑출판사*, 1985, pp. 169-174.

문유진, 김영택, "한영기계번역에서 개념기반의 동사 번역," *한국정보과학회 논문지*, 제22권 제8호, 1995, pp. 1166-1173.

문유진, "의미론적 어휘개념에 기반한 한국어 명사 WordNet의 설계와 구축," *서울대학교 대학원 컴퓨터공학과 박사학위 논문*, 1996.

이정민, 강범모, 남승호, "한국어 술어의 의미구조연구," *제2회 소프트과학워크숍 학술회의*, 1997.

Hernert, P., "KASSYS : A Definition Acquisition System in Natural Language," *Proc. of COLING-94*, Aug. 1994, pp. 263-267.

Levin, B., "English Verb Classes and Alterations : A Preliminary Investigation," *The MIT Press*, 1997.

Levin, B. and Hovav, M., "Unaccusativity : At the Syntax-Lexical Semantics Interface," *The MIT Press*, 1996.

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K., "Introduction to WordNet : An On-line Lexical Database," in Five Papers on WordNet, CSL Report, Cognitive Science Laboratory, Princeton University, 1993.

Montemagni, S. and Vanderwende, L., "Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries," *Proc. of COLING-92*, Aug. 1992, pp. 546-552.

Sumita, E., Furuse, O., and Iida, H., "An Example-Based Disambiguation of Preposition Phrase Attachment," *Proceedings of TMI*, 1993, pp. 80-91.

# 〈Index 1〉 A Part of the SENKOV Verb Classes

11 Verbs of Sending and Carrying : move, displace, make move (목적어 : 물건)
- 11.1 Send Verbs : transmit, transfer, conduct, transport, channel
  보내다, 운송하다, 배달하다, 부치다, 돌려보내다, 수송하다, 밀수하다, 전달하다, 운반하다
- 11.2 Slide Verbs : slide, roll, float
  미끄러지게하다, 튕기다, 굴리다, 떠내려보내다, 부유하다
- 11.3 Bring and Take : bring, take
  가져오다, 가져가다
- 11.4 Carry Verbs : bring, carry, convey
  나르다, 끌다, 올리다, 차다, 끌어당기다, 예인하다, 밀다
- 11.5 Drive Verbs : transport, move in a conveyance
  운전하다, 날리다, 젓다

12 Verbs of Exerting Force: Push/Pull Verbs : push, pull, move with force
끌어올리다, 잡아당기다, 누르다, 밀다, 당기다

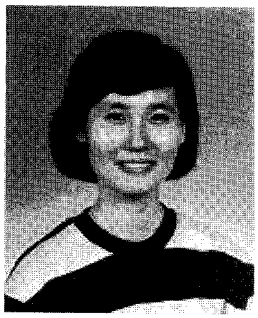
13 Verbs of Change of Possession : transfer (목적어 : 물건)
- 13.1 Give Verbs : give
  먹이다, 주다, 빌리다, 임차하다, 빌려주다, 지불하다, 갚다, 환불하다, 팔다, 교환하다
- 13.2 Contribute Verbs : give, present, give as a present, make a present a present of
  기여하다, 분배하다, 나누어주다, 기증하다, 기부하다, 제출하다
- 13.3 Verbs of Future Having : assign, allot, portion
  할당하다, 배당하다, 상주다, 양보하다, 유증하다, 설정하다
- 13.4 Verbs of Providing : provide, supply
  - 13.4.1 Verbs of Fulfilling : present, give formally
    제공하다, 선물하다, 공급하다, 맡기다, 지급하다
  - 13.4.2 Equip Verbs : equip, saddle
    설치하다, 설비하다, 얹다, 장치하다, 갖추다
- 13.5 Verbs of Obtaining : get, obtain, take into one's possession
  - 13.5.1 Get Verbs : get, buy, catch
    Obtain Verbs
    얻다, 사다, 잡다, 전세내다, 벌다, 가져오다, 찾다, 획득하다
- 13.6 Verbs of Exchange : exchange, change, interchange
  바꾸다, 대체하다, 교환하다, 교체하다, 갈다, 갈아넣다, 대신하다
- 13.7 Berry Verbs : fish, catch fish
  낚다, 잡다

14 Learn Verbs : learn, acquire knowledge, gain knowledge, acquire skills
습득하다, 주입하다, 배우다, 암기하다, 읽다, 공부하다, 익히다
15 Hold and Keep Verbs : hold, keep
15.1 Hold Verbs : take, get hold of
잡다, 쥐다, 죄다, 붙잡다, 사로잡다, 부여잡다, 붙들다, 움켜쥐다, 맞잡다, 손잡다, 움켜잡다, 휘어잡다, 마주잡다, 옴키다, 조르다, 옥죄다
15.2 Keep Verbs : have, have got, hold
저장하다, 남기다, 비축하다, 보관하다
16 Verbs of Concealment : hide, conceal
16.1 Conceal Verbs (자동사)
숨다
16.2 Conceal Verbs (타동사)
감추다, 숨기다, 막다, 고립시키다, 은폐시키다, 대피시키다, 가로막다, 틀어막다, 묻다, 덮다
17 Verbs of Throwing : throw, pelt, propel
17.1 Throw Verbs(자동사) : project through the air, move forward
Pelt Verbs
튀다, 퍼붓다
17.2 Throw Verbs(타동사)
튀기다, 퍼붓다, 쏘다, 사격하다, 던지다, 망치질하다, 난타하다, 두들겨패다, 찌르다, 놓다, 꽂다, 갈기다, 내던지다, 풀러치다, 놓다, 쑤시다
18 Verbs of Contact by Impact
: touch, make contact with, contact, come in contact with
18.1 Hit Verbs : hit, deal a blow to, beat
Swat Verbs
Spank Verbs
Poke Verbs
때리다, 치다, 툭툭치다, 차다, 두드리다, 패다, 휘갈기다, 걷어차다, 까다, 손대다
18.4 Non-Agentive Verbs of Contact by Impact
: strike, deliver a blow to, deliver a stroke to
부딪치다, 충돌하다, 맞부딪치다, 치다, 부딪히다
18.5 Verbs of Contact: Touch Verbs : caress, touch affectionately
툭치다, 만져주다, 애무하다, 입맞추다
21 Verbs of Cutting : cut
21.1 Cut Verbs : separate, force apart, divide, part, take apart, put apart
Split Verbs
짜르다, 토막내다, 썰다, 톱질하다, 쪼개다, 긁어내다, 찢다, 까다, 타다, 패다, 파다
21.2 Carve Verbs : carve, crush, grind
21.2.1 Carve Verbs(자동사)
우그러지다
21.2.2 Carve Verbs(타동사)
조각하다, 우그리다, 분쇄하다, 깎다, 자르다, 갈다, 으깨다, 새기다, 마르다, 밀다, 제기다, 후리다

## 〈Abstract〉

The paper presents methodology and techniques for design and implementation of the SENKOV System based on the validation of set membership and dictionaries. And it performs verb concept classification available for establishing the selectional restriction relationships among adverbs and verbs. The paper is important in that it has made the first attempt at classifying Korean verb concepts for the semantic analysis. We select about 600 Korean verbs which are commonly used in the daily life, and implements the SENKOV System. According to results of the experiments, SENKOV has 44 top nodes and depth of average 2.35, and that it can be utilized to classify Korean verb concept for the selectional restrictions among adverbs and verbs.

# ◐ 저 자 소 개 ◑



**문 유 진 (yjmoon@eve.hannam.ac.kr)**

한국외국어대학교를 졸업하고, 성균관대학교 경영대학원을 수료하였으며 Pennsylvania State University의 전산학과에서 이학석사 학위를 취득하였다. 그리고 서울대학교 컴퓨터공학과에서 공학박사 학위를 취득하였다. 호남대학교 컴퓨터공학과에서 부교수로 재직하였으며 현재 한남대학교 경영정보학과의 부교수로 재직 중이다. 그리고 University of Pennsylvania에서 Post-Doc. 연수를 하였다. 관심분야는 인공지능, 자연언어처리, 지식베이스 및 전자상거래 등이다.