DOI: 10.3745/KIPSTA.2008.15-A.###

한국어 워드넷에서의 개념 유사도를 활용한 선택형 문항 생성 시스템

김 용 범[†] · 김 유 섭[††]

요 약

본 논문에서는 난이도를 고려하여 선택형 문항을 자동으로 생성하는 방법을 고안하였으며, 학습자 수준에 적합하도록 동적인 형태로 다양한 문항 제시를 할 수 있는 시스템을 구현하였다. 선택형 문제를 통한 평가에서는 적절한 규모의 문제 은행이 필요하다. 이와 같은 요구를 만족시키기 위해서는 보다 쉽고 빠른 방식으로 다양하고 많은 문제 및 문항을 생성할 수 있는 시스템이 필요한데, 본 논문에서는 문제 및 문항의 생성을 위하여 워드넷이라는 언어 자원을 이용한 자동 생성 방법을 고안하였다. 자동 생성을 위해서는 주어진 문장에서 형태소 분석을 통해 키워드를 추출하고, 각 키워드마다 워드넷의 계층적 특성에 따라 유사한 의미를 가진 후보 단어를 제시한다. 의미 유사 후보 단어를 제시할 때, 기존의 한국어 워드넷의 스키마를 개념간 의미 유사도 행렬을 구할 수 있는 형태의 스키마로 변경한다. 단어의 의미 유사도는 동의어를 의미하는 수준 0에서 거의 유사도가 없다고 볼 수 있는 수준 9까지 다양하게 제시될 수 있으며, 생성될 문항에 어느 정도의 유사도를 가진 어휘를 포함시키느냐에 따라서 출제자의 의도에 따른 난이도의 조정이 가능하다. 후보 어휘들의 의미 유사도 측정을 위해서, 본 논문에서는 두 가지 방법을 사용하여 구현하였다. 첫째는 단순히 두 어휘의 워드넷 상에서의 거리만을 고려한 것이고 둘째는 두 어휘가 포함되어 있는 트리 구조의 크기까지 추가적으로 고려한 것이다. 이러한 방법을 통하여 실제 출제자가 기존에 출제된 문제를 토대로 더 다양한 내용과 난이도를 가진 문제 또는 문항을 더 쉽게 출제할 수 있는 시스템을 개발할 수 있었다.

키워드 : 문항 생성 시스템, 한국어 워드넷, 개념 유사도, 문제 난이도 조절

A Question Example Generation System for Multiple Choice Tests by utilizing Concept Similarity in Korean WordNet

Kim, Young-Bum[†] · Kim, Yu-Seop[†]

ABSTRACT

We implemented a system being able to suggest example sentences for multiple choice tests, considering the level of students. To build the system, we designed an automatic method for sentence generation, which made it possible to control the difficulty degree of questions. For the proper evaluation in the multiple choice tests, proper size of question pools is required. To satisfy this requirement, a system which can generate various and numerous questions and their example sentences in a fast way should be used. In this paper, we designed an automatic generation method using a linguistic resource called WordNet. For the automatic generation, firstly, we extracted keywords from the existing sentences with the morphological analysis and candidate terms with similar meaning to the keywords in Korean WordNet space are suggested. When suggesting candidate terms, we transformed the existing Korean WordNet scheme into a new scheme to construct the concept similarity matrix. The similarity degree between concepts can be ranged from 0, representing synonyms relationships, to 9, representing non-connected relationships. By using the degree, we can control the difficulty degree of newly generated questions. We used two methods for evaluating semantic similarity between two concepts. The first one is considering only the distance between two concepts and the second one additionally considers positions of two concepts in the Korean Wordnet space. With these methods, we can build a system which can help the instructors generate new questions and their example sentences with various contents and difficulty degree from existing sentences more easily.

Keyword : Question Example Generation System, Korean WordNet, Concept Similarity, Difficulty Degree of Questions

1. 서 론

인터넷의 발전으로 오프라인에서 주로 시행되던 교육 및 평가 시스템이 온라인으로 확장되어 각종 교육기관에 의해 시행되고 있다. 과거에는 오프라인의 지필고사를 위주로 치

※ 이 논문은 2006년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2006-331-D00534)
† 준 회 원 : 한림대학교 정보통신공학부 학부과정
†† 종신회원 : 한림대학교 컴퓨터공학과 부교수 (교신저자)
논문접수 : 2007년 7월 9일, 심사완료 : 2008년 3월 20일

루어졌던 TOEFL과 같은 평가가 현재는 Internet-Based Test (IBT) 방식으로 바뀌고 있는 것이 예다 [1]. 이와 같은 국제시험 뿐만 아니라 많은 평가가 오프라인 영역에서 온라인 영역으로 이동하고 있으며, 이러한 시험을 대비하는 교육 방식 역시 온라인으로 급속히 이동하고 있다. 온라인 교육에서는 필히 온라인 평가가 요구되는데, 온라인 평가에서는 주로 문제 은행이라는 방대한 자료가 구축되어야 하지만 [2-4], 이를 위해서는 문제를 출제하는데 필요한 많은 자원이 소모된다. 만일 이러한 자원이 뒷받침되지 않으면 수험자는 반복되어 출제되는 문제의 유형 또는 내용을 암기하여 문제를 풂으로써, 정상적인 평가가 불가능하게 된다.

이와 같은 문제를 최소화하기 위해 문항 생성을 자동으로 하여 더 적은 비용으로 매우 다양한 문항을 반자동으로 생성하는 연구가 제시되고 있다[5]. 그러나 이 연구는 매우 많은 문항을 짧은 시간에 생성할 수는 있으나, 이 시스템은 어휘 온톨로지 상에서 주변에 있는 어휘들을 대체어로 제시하는 수준이기 때문에 주어진 문장과 유사한 문항을 생성할 수는 있지만, 유사한 정도를 조절할 수 없어 새로이 생성된 문제의 난이도를 제어할 수 없다는 문제를 가지고 있다.

따라서 본 연구에서는 자동으로 생성되는 문항의 유사도를 조절함으로써 문제의 난이도를 조절할 수 있도록 하였다. 기존의 [5] 연구에서도 워드넷(WordNet)[6]을 한국어로 번역한 한국어 워드넷[7]을 사용하여 유사한 문항을 동적으로 생성할 수 있었으나, 본 연구에서는 이러한 한국어 워드넷의 구조를 보다 적극적으로 이용하여 유사도를 조절함으로써 보다 다양한 난이도의 문제를 출제할 수 있도록 한다. 선택형 문제에서 오답은 문제의 난이도 및 학습자의 대상 어휘에 대한 이해도 평가에 중요한 역할을 한다[8]. 따라서 본 연구에서는 정답 문항의 어휘와 새로이 만들어지는 오답 문항의 어휘의 유사도를 워드넷에서 제공하는 Synset, Hypernym, Hyponym 과 같은 계층적 특성을 이용하여 조절함으로써 새로이 생성되는 문제의 난이도를 조절할 수 있도록 하였다. 어휘간 개념 유사도의 측정을 위해서, 본 논문에서는 [9], [10]에서 제시된 측정방법을 구현하여 사용자가 이들 두 가지 방법 중에서 본인의 의도에 더 적합한 방법을 선택할 수 있도록 하였다.

본 연구에서는 말뭉치 데이터에서 추출되는 어휘의 관련도 (relatedness) 대신 워드넷의 계층구조에서 추출되는 어휘의 유사도 (similarity)를 활용하였다. 이는 새로운 문장을 생성하는데 있어, 그 전체적인 의미가 얼마나 변하는가에 따른 난이도 조절을 목표로 하기 때문이다. 만일 유사도가 아닌 관련도를 사용한다면, 새로이 생성된 문장은 문맥상 기존 문장과 비교하기 어려운 전혀 별도의 문장이 된다.

새로운 문장을 생성하기 위해서, 출제자는 먼저 자신이 참고할 만한 기존의 문장을 입력하거나 저장된 파일에서 불러오고, 원하는 유사도를 입력한다. 시스템은 한국어 형태소 분석기[11]를 사용하여 문장에서 키워드를 추출하고, 동시에 각 키워드와 주어진 유사도를 가지는 후보 어휘들을 한국어 워드넷에서 찾고 그 결과를 출제자에게 제시한다. 마지막으로 출제자는 후보 어휘 중에서 적절한 어휘를 선택하여 새로운 문항을 생성한다.

본 연구에서 구현한 난이도를 고려한 문항의 자동 생성 시스템을 이용하여 출제자는 보다 빠르고 간편하게 보다 다양한 문항을 생성할 수 있었으며, 테스트 결과 학생들에게 보다 높은 난이도의 문제를 간편하게 제시할 수 있었다.

2. 워드넷에서의 유사도 측정

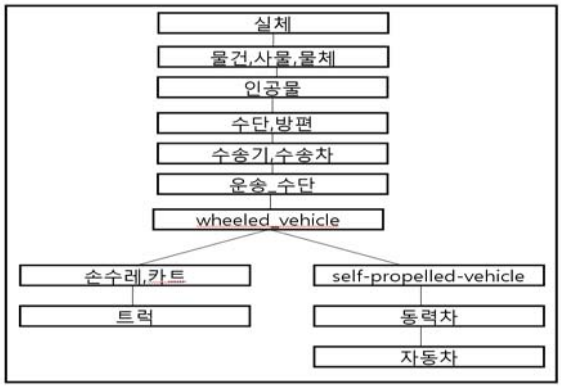
워드넷[6]은 1985년 프린스톤(Princeton)대학 인지과학연구실에서 연구되어 발전된 것으로써, 최초에는 어휘들의 연관성을 나타내기 위하여 고안되었으며 최근에는 언어처리, 정보검색 등의 분야에서 매우 활발하게 활용, 연구되고 있다[12].

(그림 1)은 한국어 워드넷 [7]의 기본 체계를 보여준다. 워드넷의 계층적 특성을 한국어 워드넷을 통하여 살펴보면, Synset이란 동의어(synonym)관계를 의미하며 (그림 1)에서의 {수단, 방편}과 같이 의미가 비슷하여 하나의 집합을 이루는 노드를 말한다. 이 Synset은 워드넷 상에서 하나의 개념을 표현한다. Hypernym은 Synset을 보다 추상적으로 표현한 상위단어를 말하며 (그림 1)>에서는 {인공물}이 {수단, 방편}의 Hypernym이 된다. Hyponym이란 상위단어에 대한 좀 더 구체적인 단어를 말하는데 (그림 1)에서는 {수송기, 수송차}가 {수단, 방편}의 Hyponym이 된다.

본 연구에서는 워드넷에서 Synset간의 Hyponym과 Hypernym 관계를 활용하여 어휘간 또는 개념간 유사도를 계산하였는데 [13]에서 제시된 바와 같이 성능면에서 큰 차이가 없고 대량의 데이터를 구축할 필요가 없는 [9]과[10]에서 제시한 방식을 사용하였다. [9]와 [10]에서 제시한 방식으로 어휘간 또는 개념간 유사도를 계산하였는데, 이 알고리즘에 대한 간략한 설명은 다음과 같다.

[9]에서 제시된 알고리즘은 두 개념 사이의 관계를 두 개념 사이의 경로(path)의 길이에 기반을 두어 다음과 같이 추정하였다.

$$rel_{HS}(c_1,c_2) = C - len(c_1, c_2) - k * d \qquad (1)$$



(그림 1) 한국어 워드넷의 기본 구조

여기서 C와 k는 상수이고, $len(a, \beta)$는 개념 a와 개념 β 사이의 가장 가까운 경로를 이루고 있는 간선의 개수이다. 그리고 d는 경로의 방향 전환 회수를 의미한다. 이 방법은 두 개념 사이의 간선의 개수가 적고, 경로의 방향 전환이 없는 경우에 더욱 유사한 개념으로 정의한다. 예를 들어 c_1이 {트럭}, c_2가 {자동차} 일 때 두 개념 사이의 유사도를 구하기 위해 (1)식을 따르면 $len(c_1, c_2)$는 5가 되고 {wheeled_vehicle}에서 경로의 방향이 전환되므로 C를 20, k를 0.5로 하였을 때 두 개념간의 유사도는 14.5가 된다. 반면에 '수송기'와 '수송차'는 동일 개념이므로 $len(c_1, c_2)$'가 0, 경로 방향 전환 횟수가 0이 되므로 유사도가 20이 되고, {트럭}과 {실체}는 '$len(c_1, c_2)$가 8이 되므로 유사도는 12가 된다.

한편, [10]에서는 개념 c_1과 c_2간의 유사도를 다음과 같이 추정한다.

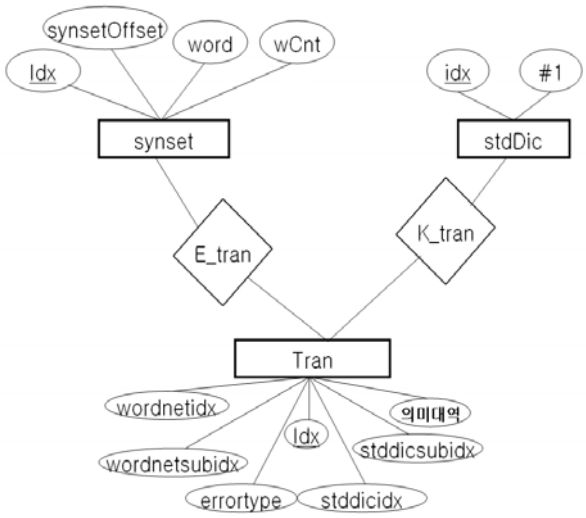
$$sim_{LC}(c_1, c_2) = -log\frac{len(c_1, c_2)}{2D} \quad (2)$$

여기서 D는 두 개념을 동시에 포함하는 계층 체계의 전체적인 깊이(depth)를 의미한다. 따라서 이 방법은 두 개념간의 경로를 이루는 간선의 개수가 작고 두 개념을 포함하는 전체 계층 구조의 깊이가 깊을수록 유사도가 높아지는 특징을 가진다. 그러나 워드넷에서의 유사도 계산은 동일 트리에 위치하는 개념들간에서만 이루어지기 때문에 사실상 비교에 있어서는 의미가 없다고 볼 수 있다. 따라서 단순히 최단 거리가 짧은 두 개념간의 유사도가 더 높은 결과를 보인다.

3. 한국어 워드넷 재구성

본 논문에서 사용한 한국어 워드넷[7]은 (그림2)의 구조로 구성되어 있어 이해가 쉽고 다양한 정보를 얻을 수 있다는 장점이 있다. 본 논문에서는 이 스키마를 이용하여 어휘 개념 유사도 행렬을 구성한다. 또한 주어진 키워드의 동의어 및 하위어는 실제 문장을 생성하는 단계에서 추출하기도 한다. 그러나 (그림 2)의 한국어 워드넷의 기본 구조는 많은 정보를 표현하기 위하여 어휘 개념 유사도 행렬 구성에는 필요하지 않은 데이터 필드를 가지고 있으며, 또한 데이터가 지나치게 세분화되어 유사도 행렬을 구성할 때 오버헤드를 유발시켜 성능을 저하시키는 문제를 가지고 있다. 따라서 본 논문에서는 이러한 문제를 해결하기 위하여 주어진 한국어 워드넷의 기본 구조를 유사도 행렬 구성에 맞추어 재구성하였다.

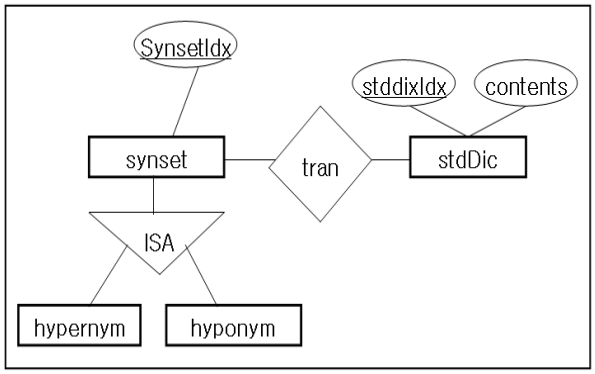
어휘 개념간의 유사도 측정에서는 한/영 키워드 변환 및 관련 정보 추출에서 관계 사이의 조인 연산이 빈번히 이루어지는데, 질의 처리에 있어서 가장 많은 시간을 사용하는 조인 연산을 줄이기 위해서 (그림 2)의 속성 중에서 불필요한 속성들을 제거하여 기본 구조를 재구성하였다. 또한 하나의 속성에 복수개의 값을 허용하면 속성의 크기의 최대값



(그림 2) 한국어 워드넷의 기본 Entity Relationship Diagram)

을 예상하기 어려워 구현시 비효율적인 공간 사용이 초래되고, 또한 질의를 통해 추출한 속성 값 전체를 실제 사용을 위하여 재가공하여야 되는 추가적인 과정이 필요하다.

본 논문에서는 재구성을 진행할 때, 실제로 전체 한국어 워드넷 공간에서 유일한 값을 가지는 synsetOffset 속성을 SynsetIdx라 명칭을 바꾸어 모든 관계의 주키(primary key)로 결정하고, Synset, Tran 관계에서 중복적으로 주키로 되어 있는 idx 속성을 제거하여 공간 사용을 줄이고 동시에 접근을 더욱 간단하게 하였다. 그리고 여러 번의 조인 연산을 통하여 입력된 한국어 어휘에 대한 영어 어휘를 찾아 이를 변환하던 것을 stdDic 및 synset 관계의 주키들의 비교를 통하여 보다 간소화시켰다. 또한 synset, hypernym, hyponym 관계는 모두 idx, offset과 같이 동일한 필드로 구성되기 때문에, ISA 관계로 이들을 하나로 결합시킴으로써 한국어 워드넷 구조를 더욱 단순화할 수 있었다. 재구성 결과를 함수종속성의 폐포와 후보키를 고려하여 BCNF (Boyce-Codd Normal Form)[14] 정규화의 조건을 맞추어 테이블을 재구성하면 (그림 3)과 같은 새로운 구조가 된다.

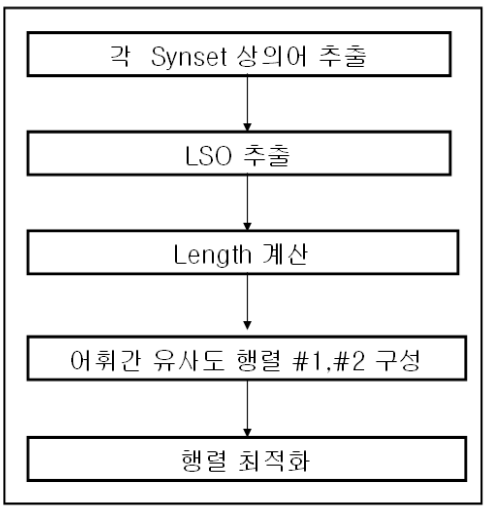


(그림 3) 재구성한 한국어 워드넷 ERD

4. 개념 유사도 행렬 구성

(그림 4)는 개념 유사도 행렬 #1, #2를 구성하는 전체 과정을 보여준다. 개념 유사도 #1은 [9]의 방법론을 그리고 개념 유사도 #2는 [10]의 방법론을 사용하여 구성하였다. 먼저 각 Synset들의 Hypernym들을 더 이상 Hypernym이 존재하지 않는 최상위 Synset 까지 순차적으로 순회하면서 상위어 리스트를 각 Synset마다 구성한다. 임의의 두 synset에 대하여, 위에서 구축된 최상위 Synset부터 내림차순으로 정렬이 되어 있는 리스트를 순회하면서 LSO(lowest super-ordianate)를 찾고 이를 통하여 식 (1)의 d값과 식 (2)의 D 값을 계산한다. 그리고 두 synset 사이의 경로를 구성하는 간선의 개수를 세어 *len()*을 계산한다. 계산 결과를 토대로 <표 1>과 같은 테이블을 구성하고 이를 최적화한다. 식(1)과 식(2)의 결과값은 분포를 고려한 10개의 구간 중의 하나의 구간에 할당되고, 유사도가 높을수록 0에 가까운 구간에 할당된다. 이 값들이 개념 유사도 행렬에 입력된다. 한국어 워드넷은 79,689개의 synset을 포함하고 있으므로, 79689 × 79689 행렬이 만들어진다.

0은 두 어휘가 동일 Synset에 위치하여 의미상으로 동일함을 보여주는데, 이는 두 어휘가 사실상 동일 개념이라는 의미로써, 유사도가 0에 가까울수록 두 어휘의 유사도는 증가한다고 할 수 있다. 만일 두 어휘의 LSO가 존재 하지 않을 경우, 즉 두 어휘가 완전히 다른 개념 트리에 속해 있는 경우에는 -1의 값을 갖는다. 개념 유사도 행렬은 대각 요소를 기준으로 전치행렬들은 동일한 값을 가지는 대칭적인 특징을 갖고 있기 때문에 사실상 전체의 절반 정도의 공간은 불필요하다. 또한 서로 다른 개념 트리간의 의미 유사도 계산은 사실상 불가능하기 때문에 이러한 값을 위한 공간 역시 불필요하다. 이러한 불필요한 공간을 줄여 보다 최적화된 행렬을 구현하였다.



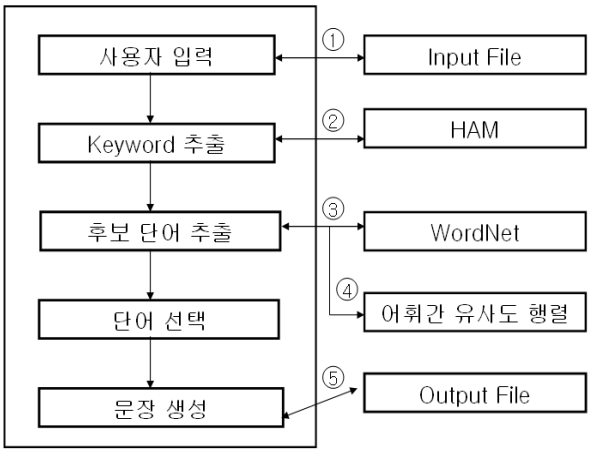
(그림 4) 개념 유사도 행렬 구성 과정

〈표 1〉 개념 유사도 행렬

Word / Word	W_1	W_2	W_3	…	W_n
W_1	0	1	2		-1
W_2	1	0	1		-1
W_3	2	1	0		
:				…	
W_n	-1	-1	-1		0

5. 전체 시스템 개요

본 논문에서 제안하는 시스템은 (그림 5)와 같이 총 4단계의 처리 과정과 5가지의 자원으로 이루어진다.



(그림 5) 전체 시스템 구성도

5.1 사용자 입력단계

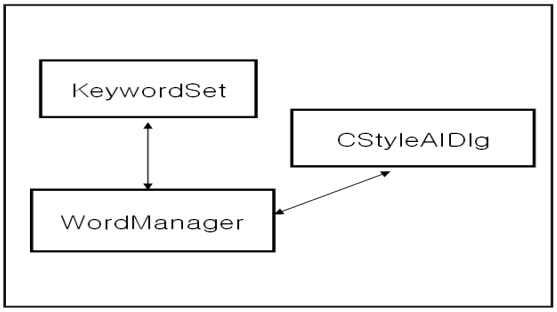
이 단계에서는 새로운 문장을 생성하기 위하여 참고할 문장을 입력하고, 생성될 문장과 기존의 문장의 유사 정도를 선택하여 최종적으로 문제의 난이도를 제어할 수 있도록 한다. 새로운 문장은 입력창을 통하여 직접 입력하거나 기존에 작성된 입력 파일을 불러들여 작성할 수 있다. 예를 들어 "파일을 압축하는 목적은 저장 공간의 절약과 통신 속도의 향상이다." 라는 문항과 관련한 새로운 문장을 생성한다고 가정한다면 문장의 입력 후 다음 중에서 하나를 선택하고 다음의 ② 또는 ③을 선택하였을 때 유사 정도를 직접 입력한다.

① 하위어 or 동의어를 추출
② type-1((1) 수식으로 계산된 유사도)
③ type-2((2) 수식으로 계산된 유사도)

사용자가 입력한 문장은 사용자가 원할 경우 다시 (그림 5)의 ① 입력 파일에 저장된다. 저장된 파일의 문장들은 사용자의 다음 사용 시에 불필요한 입력의 수고를 덜어준다.

5.2 Keyword 추출 단계

이 단계에서는 (1) 단계에서 입력된 문장에서 대체될 단어를 추출하는데, (그림 6)의 클래스들의 구성과 흐름에 따라 진행된다. 단계(1)에서 사용된 문장이 입력되면, 사용자 인터페이스를 처리하는 CStyleAIDlg 클래스에서 이를 받아 WordManager 클래스을 통하여 한국어 형태소 분석기인 HAM(Hangul Analysis Module)[11]을 사용하여 {파일, 압축, 목적, 저장, 공간, 절약, 통신, 속도 ,향상}으로 이루어진 Keyword Set을 추출하여 문장 내에서의 순서에 따라 인덱스를 부여 하여 인덱스, Keyword, 조사로 이루어진 KeywordSet클래스를 <표 2>와 같이 구성한다. 여기서 조사 속성은 향후 완전한 문장을 생성할 때 사용된다.



(그림 6) Keyword 추출에 따른 클래스 관계도

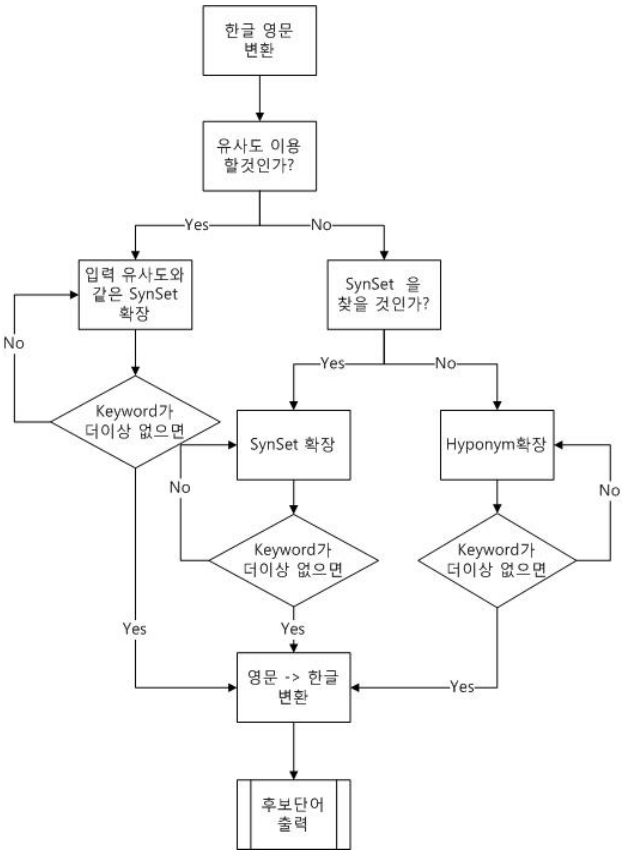
〈표 2〉 KeywordSet 클래스 구성

Index	Keyword	조사
0	파일	을
1	압축	하는
2	목적	은
3	저장	NULL
4	공간	의
5	절약	과
6	통신	NULL
7	속도	의
8	향상	이다.

5.3 후보 단어 추출

이 단계는 크게 두 가지 경우로 나누어진다. 첫째, 한국어 워드넷의 synset과 hyponym 관계만을 사용하여 후보 단어를 추출하는 방법이 있다. 이 방법에서는 KeywordSet에 있는 단어들마다 한국어 워드넷을 이용하여 후보단어를 추출 한다. 한국어 워드넷은 영문을 기준으로 계층적으로 구성되어 있기 때문에 후보 단어 추출과정에서 한글로 구성된 KeywordSet의 단어들마다 의미가 같은 영어 단어로 대치하고 그 단어의 하위어 또는 동의어 등 원하는 관계의 Synset을 찾아낸 다음 다시 한글 단어로 대치하여 후보 단어를 만들어 낸다. 둘째는 한국어 워드넷에서 정밀한 유사도 측정을 위해 필요한 정보를 추출하여 재구성한 의미 유사도 행렬을 한국어 워드넷과 함께 이용하여 사용자가 입력한 유사도에 따른 후보 단어를 추출하는 방법이다. 이 방법에서 유사도를 한국어 워드넷을 통하여 실시간으로 계산하면 엄청난 오버헤드가 발생하기 때문에, 한국어 워드넷의 단어들의 Index를 PrimaryKey로 하여 각 단어마다 상위어, 하위어, 다른 단어와의 Edge의 개수 등을 미리 계산하여 유사도 행렬을 미리 구축 하였다. 사용자가 유사도를 입력하면 KeyWord를 한국어 워드넷을 이용하여 영문으로 대치한 후, 그에 해당하는 Index를 통해 유사도 행렬을 검색하여 후보 단어들을 추출할 수 있게 하였다. (그림 7)은 이러한 모든 과정을 보여준다.

한국어 워드넷은 영문을 기준으로 의미의 계층이 이루어지기 때문에 키워드의 영문 변환 후에 유사도를 측정하여야 하며 유사도에 따른 후보단어 확장 이후 다시 한글 변환이 필요하다. 한국어 워드넷에서 동의어를 추출하기 위해서는 키워드의 한국어 워드넷 상에서의 offset을 계산하고 이와 같은 offset을 가지는 단어들을 후보단어로 추출하며, 하위어를 추출할 때에는 키워드에 연결된 하위어들의 인덱스를 이용하여 hypernym table에서 추출하여 후보단어를 추출한다. <표 3>은 유사도에 따른 추출된 후보 단어의 예를 보여준다.



(그림 7) 후보단어 추출 순서도

〈표 3〉 유사도에 따른 후보 단어 예시

Level	압축	통신
0	{압착,수축,....}	{전달, 커뮤니케이션...}
1	{푸싱, 집중.. }	{메세지, 의사소통...}
2	{누르기, 농축..}	{전화메세지,방송..}
3	{밀기, 방출...}	{소포우표, 국제전보...}
4	{분만, 출발..}	{서론,그림문자 ...}
5	{발판, 행동..}	{외형, 모습...}
6	{진행, 상호작용..}	{주체성, 개성..}
7	{교차,접촉...}	{스탠드,동일시...}
8	{교류,융합...}	{인프라,주름...}
9	{관계,전례...}	{활기,은둔...}

5.4 단어 선택

이 단계에서는 (3)의 과정을 통해서 얻은 각 Keyword의 후보 단어들을 동적으로 콤보박스에 나열하여 줌으로써 사용자가 원하는 단어를 선택하여 새로운 문항을 생성할 수 있게 한다. 예를 들어 사용자는 '목적'에 대하여 '의도'를, '공간'에 대하여 '공란'을 '통신'에 대하여 '전달'을 그리고 '향상'에 대하여 '증대'를 선택하여 향후 생성될 문장에 사용할 수 있다. 그리고 이 단계에서는 후보 단어를 동적으로 생성하면서 이에 맞는 어미 조합형 코드를 이용하여 받침의 유무를 판단하고 (2)번 과정에서 얻은 조사를 후보단어에 어울리는 조사로 변형한다.

5.5 문장 생성

이 단계에서는 (4)번 과정에서 선택한 후보단어와 조사를 이용하여 새로운 문항을 생성한다. (1)에서 입력된 '파일을 압축하는 목적은 저장 공간의 절약과 통신 속도의 향상이다'라는 문항에 대하여 결과적으로 '파일을 압축하는 의도는 저장 공란의 절약과 전달 속도의 증대이다'라는 문항이 생성된다. 또한 사용자는 생성된 문항 중에서 일부를 (그림 5)의 출력 파일에 보관하여 재사용 할 수 있다.

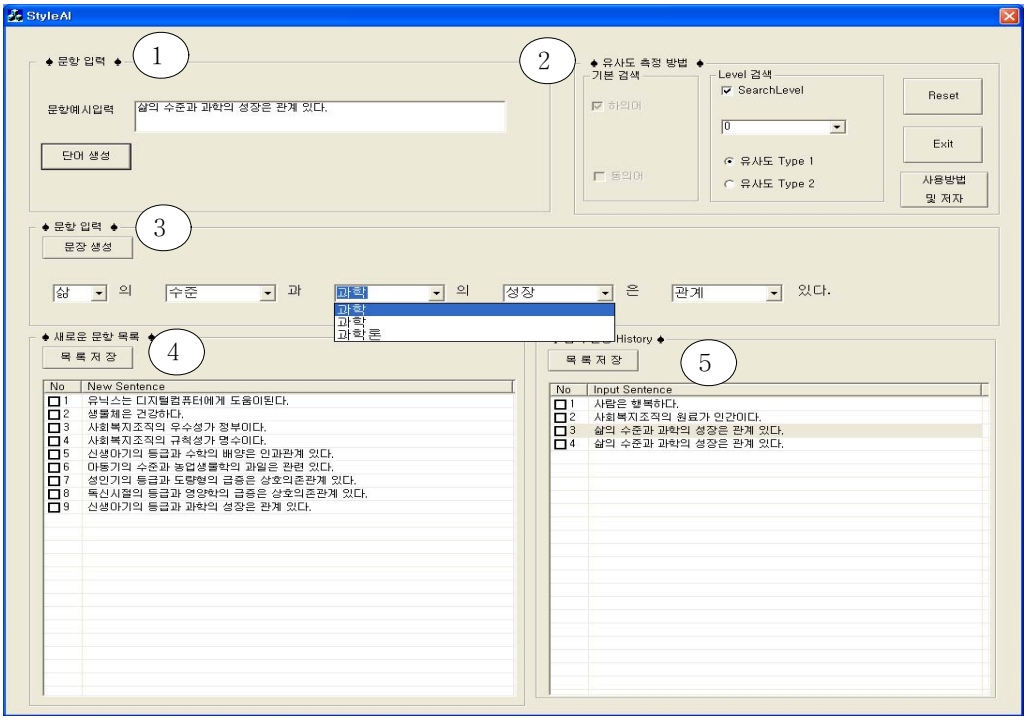
6. 구현 및 실험 결과

6.1 사용자 인터페이스

(그림 8)은 본 논문에서 제시한 시스템의 사용자 인터페이스를 보여주고 있으며, 다음과 같은 세부 부분으로 구성되어 있다.

① 문항 입력 부분
② 유사도 측정 방법 및 Level을 선택하는 부분
③ 후보단어 선택을 위한 콤보 박스 생성부분
④ 최종 문장 목록 부분
⑤ 입력 문장 목록 부분

①은 새로운 문장을 생성할 때 참고를 할 문장을 입력하는 창으로써, 사용자가 직접 타이핑하여 입력할 수도 있고 ⑤에서 체크박스에 표기함으로써 목록에서 불러올 수 있도록 하였다. ②는 생성된 문항을 통하여 문제 전체의 난이도



(그림 8) 사용자 인터페이스

를 조절하는 부분이다. 만일 높은 유사도를 가진 단어를 선택하면, 매우 작은 의미의 차이가 있는 문장이 생성되므로 기존의 정답 문항과의 차이점을 찾기 어려워진다. 반면에 낮은 유사도를 가진 단어를 선택하여 새로운 문항을 생성하면 정답 문항과의 의미의 차이를 찾기에 용이하기 때문에 그만큼 문제의 난이도 역시 낮아진다. 여기서는 두 가지 방식으로 대체어를 찾을 수 있도록 하였는데, 첫째는 동의어 및 하위어만을 제시하는 방법이고, 둘째는 워드넷 계층 공간에서의 어휘 의미 유사도를 이용하여 유사 수준을 사용자가 직접 결정하는 방법이다. 두 방법 중 하나를 반드시 선택하여야 한다. ③은 문장에서 추출된 키워드의 대체어를 보여주는 부분이다. 박스를 클릭하면 대체 후보 단어들이 나열되고, 이 중에서 선택을 하도록 한다. ④에서는 선택된 대체어들로 이루어진 새로이 생성된 문장을 보여주고, 이 문장을 파일에 저장할 수 있도록 목록을 보여주는 부분이다. 마지막으로 ⑤는 생성 시스템이 가동된 이후에 새로이 생성된 문장들의 리스트를 보여주고 이 중에서 원하는 문장을 저장할 수 있도록 해주는 부분이다.

6.2 실험 및 평가

6.2.1 문항의 난이도 조절 사례

문제의 난이도는 기존 문항의 어휘를 유사도를 토대로 대체함으로써 조절한다. 본 논문에서는 난이도를 상향 조절하는 것을 실험하였다. 난이도를 높이기 위해서 정답을 토대로 오답을 구성하거나, 역시 정답을 토대로 정답을 재구성하는 방법을 이용하였다. 즉 의미상 정답이라 볼 수 있는 문항에서 일부 어휘를 유사도가 높은 어휘로 대체하고, 그 결과 의미가 동일하다고 보이는 문항은 그대로 정답 문항으로 사용하고, 의미가 틀려졌다고 보이는 문항을 오답 문항으로 사용하였다. 이렇게 함으로써, 새로운 정답 문항은 정답으로 인정할 수 있는 의미를 가지지만, 기존 정답 문항과는 약간 다른 내용을 갖게 되고, 새로운 오답 문항은 정답으로 인정하기 어려우나 기존 정답 문항과 매우 유사한 내용을 갖게 되어 수험자의 혼란을 가져오게 된다.

예를 들어 '파일 압축에 대한 설명 중 옳은 것은?' 이라는 문제의 정답으로 '디스크 저장 공간을 효율적으로 활용하기 위해서 압축을 한다.' 라는 문항이 사용되었다고 가정하자. 이 문항에 있는 명사 어휘들을 하나씩만 유사도 1의 관계에 있는 어휘들로 바꾸었을 때, 문맥상 별 문제가 없는 문장은 다음 <표 4>에 나열되어 있다. <표 4>의 문항들은 생성된 모든 문항 중에서 출제자가 문맥상 큰 문제가 없다고 판단한 문항만을 선택한 것인데, 이 경우에는 전체 83개의 생성된 문항 중에서 18개의 문항이 선택되었다. 그러나 실제 실험에서는 2개 이상의 어휘를 대체하기도 한다.

이와 같이 생성된 문항 중에는 정답 문항을 그대로 대체할 수 있는 문항이 있고, 오답 문항으로 사용되는 것이 바람직한 문항도 있다. 예를 들어, '디스크 저장 공간을 경제적으로 활용하기 위해서는 압축을 한다.' 와 같은 문장은 정답 문항으로 사용하여도 의미상 문제가 되지 않지만, '디스크 저장 간격을 효율적으로 활용하기 위해서는 압축을 한다.'

〈표 4〉 유사도 1의 어휘로 기존 어휘를 대체한 결과

변경어휘	생 성 문 장
원 문	디스크 저장 공간을 효율적으로 활용하기 위해서는 압축을 한다.
'저 장'	디스크 비축 공간을 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 적재 공간을 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 축적 공간을 효율적으로 활용하기 위해서는 압축을 한다.
'공 간'	디스크 저장 면적을 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 저장 위치를 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 저장 장소를 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 저장 지역을 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 저장 지대를 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 저장 층을 효율적으로 활용하기 위해서는 압축을 한다.
	디스크 저장 간격을 효율적으로 활용하기 위해서는 압축을 한다.
'효 율'	디스크 저장 공간을 경제적으로 활용하기 위해서는 압축을 한다.
'활 용'	디스크 저장 공간을 효율적으로 이용하기 위해서는 압축을 한다.
	디스크 저장 공간을 효율적으로 사용하기 위해서는 압축을 한다.
'압 축'	디스크 저장 공간을 효율적으로 활용하기 위해서는 부호화를 한다.
	디스크 저장 공간을 효율적으로 활용하기 위해서는 암호화를 한다.
	디스크 저장 공간을 효율적으로 활용하기 위해서는 축소를 한다.
	디스크 저장 공간을 효율적으로 활용하기 위해서는 집중을 한다.

와 같은 문장은 거의 유사하다고 볼 수 있으나 엄밀히 본다면 정답 문항으로 인정할 수 없다. 이 같은 차이는 출제자가 생성된 문장을 보고 개별적으로 판단하여 결정한다.

유사도가 2의 어휘로 대체된 문항들은 기존 문항과의 유사도가 상대적으로 떨어진다. 물론 '디스크 저장 개체를 효율적으로 활용하기 위해서는 압축을 한다.'와 같이 의미가 거의 유사하여 기존 문항을 대체하여 사용할 수 있는 경우도 있으나, 주로 '디스크 저장 공간을 효율적으로 재활용하기 위해서는 압축을 한다.'와 같이 문맥상 문제는 없으나 기존 문항과는 전혀 의미가 다른 문항이 생성된다. 이처럼 보다 그 차이가 확실한 문항은 오답 문항으로 사용한다. 그리고 유사도가 3이 넘어가면 대부분 문맥상 이해하기 어려운 문항들이 생성되었으며, 또한 문맥이 틀리지 않더라도 기존 문장과는 그 의미가 완전히 다른 문항들이 생성되어 향후 오답 문항으로 사용될 경우 해당 문제의 난이도는 더욱 낮아진다.

본 연구에서 제시된 자동 생성 방법은 후보 어휘들을 생성 시스템이 자동으로 생성하여 출제자에게 제시하고, 이를 출제자가 문맥상 하자가 없는지를 검토하고, 이를 통과한 문항을 기존 문항을 대치할 수 있는 문항, 기존 문항으로의 선택을 방해함으로써 문제의 난이도를 높일 수 있는 문항, 그리고 기존 문항과 그 의미의 차이가 커서 쉬운 난이도의 문제를 구성할 수 있는 문항으로 구분하여 활용한다. 후자의 경우에는 정답 문항으로 여러 후보 문항을 생성할 때, 유사도 2 이상의 어휘로 생성하여 문맥을 검증하게 되면 기존 문항과 전혀 뜻이 다른 문항들이 생성된다. 따라서 정답 문항으로 의미의 차이가 큰 문항을 생성한 후, 이 오답 문항을 사용하여 문제를 구성하게 되면 학생들은 문항의 오답 여부를 보다 쉽게 판단하게 되어 문제의 난이도를 낮출 수 있게 된다.

6.2.2 난이도 조절 실험 결과

본 실험은 자동 생성 시스템을 통하여 실제 문제의 난이도를 얼마나 높일 수 있었는가에 대한 평가 실험이다. 먼저 본 실험을 위하여 컴퓨터 활용 능력 2급 필기 문제[15] 중에서 난이도를 고려하여 20 문제를 선택하였다. 이 문제를 임의로 10 문제씩 2개의 부분으로 구분하였는데, 한 부분은 원래 문제 그대로인 상태이고 다른 하나는 본 시스템을 활용하여 문항을 변경한 상태이다. 본 시스템을 가지고 문항을 변경할 때에는 유사도가 2 이하인 대체어를 선택하도록 하였다. 각 부분들은 모두 동일한 10명의 학생들로 하여금 풀게 하였다.

실험 결과, 10개의 기존 문제를 풀게 하였을 때에는 학생들이 평균 7.2점을 획득한 반면에 본 시스템을 활용하여 문항을 변경하여 풀게 하였을 경우에는 평균 6.4점을 획득하였다. 보다 자세한 실험 결과는 <표 5>에 나타나 있다. 이러한 결과는 출제자가 유사어를 가지고 기존의 문항과 의미상 유사해 보이는 문항을 생성함으로써 학생들의 판단에 혼란을 주었기 때문으로 보인다.

본 실험의 통계적인 검증을 위하여 paired t-검정 [16]을 사용하였다. t-검정은 기본적으로 2개의 집단의 평균값이 차이가 있는가를 검정하는 것인데, 본 연구에서 얻은 표본의 경우 한 학생이 원문 및 변경문에 모두 답하였으므로, 이 두 유형의 값이 서로 독립적이라고 보기가 어렵다. 따라서 두 유형의 결과 값의 차이가 0인지를 검정하는 paired t-검정을 적용하였다. 검정시 기본가설은 원문 유형과 변경문 유형간 점수 차이가 없다는 것이며, 대립가설은 두 유형간 점수 차이가 통계적으로 유의하게 차이난다는 것이다. ($H_0 : A - B = 0$ vs $H_1 : A - B \neq 0$) 다음 <표 6>에 검정 결과가 정리되어 있다.

<표 6>을 보면 t-값이 임계값보다 크기 때문에, 검정통계량이 기각역에 속하게 되므로 기본 가설을 기각하게 되므로, 본 연구의 결과로 두 유형간 점수 차이가 통계적으로 차이난다고 볼 수 있다. 즉, 원문의 평균이 변경문의 평균보다 유의하게 크다고 판정되어 본 실험의 난이도 조절 효과가 있었다고 판단된다.

또한 실험의 객관성을 확보하기 위하여, 위와 같은 실험을 동일한 방식으로 다른 2개의 부류를 대상으로 행하였다.

〈표 5〉 10명의 학생들의 시험 결과

	1	2	3	4	5	6	7	8	9	10	평균
원문	9	9	6	7	7	6	8	9	6	5	7.2
변경문	9	8	4	6	5	7	8	7	5	5	6.4

〈표 6〉 paired t-검정 결과 요약

평균	표준오차	t-값	임계값
0.8	0.327	2.45	2.26

주: 유의수준 0.05에서의 임계값임.

〈표 7〉 추가 실험 결과

부류	문제	1	2	3	4	5	6	7	8	9	10	평균
1	원문	9	6	7	6	6	8	8	7	9	10	7.60
	변경문	7	4	7	5	5	6	6	8	8	8	6.40
2	원문	7	9	10	6	7	5	9	7	6	8	7.40
	변경문	8	8	10	5	5	4	7	9	9	6	7.10

〈표 8〉 추가 실험의 t-검정 결과

개 수	평 균 값	표준 편차	표준 오차 (=표준편차/$\sqrt{N}$	t-값	p-값
30	0.67	1.37	0.25	2.66	0.0126

각 부류는 10명의 학생으로 이루어져 있으며 각각 10개의 원문 및 변경문에 대하여 답하도록 하였다. <표 6>은 두 부류 학생들의 시험 결과를 보여준다.

그리고 <표 7>의 실험과 <표 5>의 실험을 모두 합하여 이를 하나의 실험으로 설정하여 다시 paired t-검정을 실시함으로써 원문 및 변경문의 평균의 차이가 통계적으로 유의함이 있는지를 검정하였다. <표 8>은 t-검정의 결과를 보여준다.

<표 8>의 결과에서 t-값은 2.66이며, p-값은 0.0126으로 유의수준인 0.05보다 훨씬 작다. 따라서 귀무가설은 기각되며, 평균값이 0.67으로 양수의 값을 가지므로, 변경문에 대한 시험 결과가 원문에 대한 시험 결과보다 통계적으로 유의하게 낮다는 것을 알 수 있다. 즉 시험의 난이도가 향상되었다고 판단할 수 있다.

7. 결 론

본 연구에서는 한국어 워드넷의 계층 구조를 이용하여 주어진 문항을 가지고 원하는 정도의 유사도를 가진 새로운 문항을 생성하는 시스템을 구현하였다. 본 시스템은 대체어의 선택시 단순히 주어진 어휘의 동의어 또는 하위어만을 고려하는 부분과 워드넷의 계층 구조상에서의 유사도를 고려하여 대체어를 선택하는 부분으로 이루어진다. 이렇게 다양한 정도의 유사성을 가진 문항을 생성함으로써 원하는 정도의 문제 난이도를 구현할 수 있었다.

워드넷은 자체적으로 어의 중의성 문제를 가지고 있다. 즉 하나의 어휘가 워드넷 계층 구조 상에서 복수의 위치에 나타남으로써 실제 문장에 나타난 어휘의 위치를 판단해야 하는 문제가 있다. 또한 워드넷은 분야별로 새로이 등장하는 어휘를 즉각적으로 반영하지 못하는 문제도 가지고 있다. 따라서 향후 이러한 워드넷의 한계를 극복하기 위하여 첫째, 문제의 영역(domain)에 따라서 말뭉치를 구축하고 이러한 말뭉치에서 추출되는 정보를 통하여 어의 중의성을 해소하여야 하며, 둘째, 새로이 등장하는 전문 어휘들을 워드넷 상에 위치시킬 수 있는 방법에 대한 연구가 필요하다. 또한 다양한 문제 난이도를 구현하기 위해서는, 문항의 개별 키워드별로 별도의 유사도를 반영하여 대체어를 찾을 수 있는 방법도 구현하여야 할 것이다. 마지막으로 시험의 성격을 반영한 시스템의 구축이 진행되어야 할 것이다. 본 연구에서 활용한 시험은 그 성격상 전문 지식의 인지 여부를 물어보는 시험이다. 이러한 시험의 경우에는 많은 문항들이 특정 용어 및 수치로 이루어진다. 이러한 성격의 시험에서는 본 연구에서 제안된 방법이 제한적으로 활용 될 수 밖에 없다. 따라서 본 연구에서 제안된 방법을 [8]의 연구에서 활용한 성격의 시험에 적용하고 이에 맞추어 최적화하는 연구를 진행하여야 할 것이며, 동시에 전문 지식을 물어보는 시험에 최적화된 방법론을 새로이 개발하여야 할 것이다.

참 고 문 헌

[1] Educational Testing Service, *http://www.ets.org*.

[2] 황대준, "사이버 스페이스상의 상호참여형 실시간 원격교육 시스템에 관한 연구", *한국정보처리학회* 제4권 3호, 1997. 5.

[3] 조은순, "원격교수-학습을 위한 사고의 전환: 하드웨어에서 소프트웨어로", *한국정보처리학회* 제4권 3호, 1997. 5

[4] 원대희, 강태호, 김원진, 방훈, 이재영, "임의 추출 분할 방식을 이용한 동적 문제 출제 시스템," *한국정보과학회 추계학술대회*, 2001.

[5] 오정석, 추승우, 조우진, 김유섭, 이재영, "한글 워드넷을 이용한 동적 문제 출제 시스템 설계," *한국정보기술학회* 논문지 4권 5호 pp.37-44, 2006.

[6] G. A. Miller, "WordNet: An On-Line Lexical Database", International Journal of lexicography, 1990.

[7] 이은령 임성신, "WordNet2.0의 한국어 번역 작업과 결과물," 부산대학교 한국어정보처리연구실.

[8] 최수일, 임지희, 최호섭, 옥철영, "사용자 어휘지능망과 자동문제생성기술을 이용한 한국어 어휘학습시스템," *제 18회 한글 및 한국어 정보처리 학술대회* 논문집, pp. 15-21, 2006.

[9] Graeme Hirst and David St-Onge, "Lexical chains as representations of context for the detection and correction or malapropisms," In Fellbaum, pp.305-332, 1998.

[10] Claudia Leacock and Martin Chodorow, "Combining local context and WordNet similarity for word sense identification," In Fellbaum, pp. 265-283, 1998.

[11] 강승식, 범용 형태소 분석기 "HAM Ver 6.0.0", 국민대학교 자연언어 정보 검색 연구실, http://nlp.kookmin.ac.kr

[12] Wikipedia, http://en.wikipedia.org/wiki/WordNet.

[13] Budanitsky, A., and G. Hirst, "Semantic Distance in WordNet: An Experimental Application-oriented Evaluation of Five Measures", Workshop on WordNet and Other Lexical Resources, in the North American Chapter of the Association for Computational Linguistics (NAACL-2001), Pittsburgh, PA, June 2001.

[14] Date, C. J., "An Introduction to Database Systems: 7th edition," Addison Wesley Longman, 1999.

[15] 대한상공회의소 감정사업단, http://www.passon.co.kr.

[16] Goulden, C. H., "Methods of Statistical Analyssi 2nd ed.," New York: Wiley, pp.50-55, 1956.

김 용 범

e-mail : stylemove1@hallym.ac.kr
2008년 한림대학교 정보통신공학부 컴퓨터공학전공(학사)
관심분야 : 자연언어처리, 텍스트 마이닝, 가상현실, 컴퓨터 게임 등

김 유 섭

e-mail : yskim01@hallym.ac.kr
1992년 서강대학교 전자계산학과(학사)
1994년 서울대학교 대학원 컴퓨터공학과 (공학석사)
2000년 서울대학교 대학원 컴퓨터공학과 (공학박사)
현 재 한림대학교 정보통신공학부 부교수
관심분야 : 전산금융, 자연언어처리, 기계학습, e-learning 등