

# 의미커널과 한글 워드넷에 기반한 지능형 채점 시스템

조 우 진<sup>†</sup> · 오 정 석<sup>†</sup> · 이 재 영<sup>\*\*</sup> · 김 유 섭<sup>\*\*\*</sup>

## 요 약

최근 인터넷 사용자가 급증하면서 원격교육의 발전과 함께 평가에서도 원격을 이용한 방법이 많이 사용되고 있다. 하지만 현재까지는 자연 언어처리의 어려움으로 객관식이나 단답식 평가가 주류를 이루고 있다. 본 논문에서는 서술형 주관식 문제의 빠르고 공정한 지능형 채점을 위하여, 다양한 언어 지식을 활용하였다. 이를 위하여, 가공되지 않은 말뭉치에서 의미커널을 구축하고, 수험자가 작성한 답안과 이미 구축된 정답을 벡터로 구성하여 이 답안간의 유사도를 의미커널을 통해 계산하여 정답여부를 자동으로 판단하도록 하였다. 의미커널을 구축하기 위하여 벡터 공간 모델에 기반한 은닉 의미 분석을 이용하였으며, 또한 한글 워드넷을 이용하여 답안의 정보부족 문제를 줄여보고자 하였다. 실험을 위하여 3000문항의 주관식 문제를 구축하였으며, 의미커널의 구축을 위하여 38,727개의 신문기사를 모아 말뭉치로 구성하고 75,175개의 색인어를 추출하였다. 의미커널에 기반한 자동 채점 시스템으로 실제 수험자에 의하여 작성된 답안을 채점한 결과, 출제자가 실제로 채점한 결과를 기준으로 하여 최고 0.894의 상관관계를 얻을 수 있었다.

키워드 : 원격교육, 지능형 채점 시스템, 의미커널, 워드넷, 벡터 공간 모델, 은닉 의미 분석

## An Intelligent Marking System based on Semantic Kernel and Korean WordNet

WooJin Cho<sup>†</sup> · JungSeok Oh<sup>†</sup> · JaeYoung Lee<sup>\*\*</sup> · Yu-Seop Kim<sup>\*\*\*</sup>

## ABSTRACT

Recently, as the number of Internet users are growing explosively, e-learning has been wide spread, as well as remote evaluation of intellectual capacity. However, only the multiple choice and/or the objective tests have been applied to the e-learning, because of difficulty of natural language processing. For the intelligent marking of short-essay typed answer papers with rapidness and fairness, this work utilize heterogenous linguistic knowledges. Firstly, we construct the semantic kernel from untagged corpus. Then the answer papers of students and instructors are transformed into the vector form. Finally, we evaluate the similarity between the papers by using the semantic kernel and decide whether the answer paper is correct or not, based on the similarity values. For the construction of the semantic kernel, we used latent semantic analysis based on the vector space model. Further we try to reduce the problem of information shortage, by integrating Korean WordNet. For the construction of the semantic kernel, we collected 38,727 newspaper articles and extracted 75,175 indexed terms. In the experiment, about 0.894 correlation coefficient value, between the marking results from this system and the human instructors, was acquired.

Key Words : Remote Education, Intelligent Marking System, Semantic Kernel, WordNet, Vector Space Model(VSM), Latent Semantic Analysis(LSA)

## 1. 서 론

최근 들어 인터넷 사용자가 급증하면서 교육 분야에서도 교내 망 또는 인터넷망을 통하여 다양한 학습정보를 제공하고 있으며 학생들은 시간과 공간의 제약 없이 학습할 수 있

는 원격 교육이 보편화되고 있다. 이제 원격 교육은 학습 자료의 검색뿐만 아니라 동영상 강의, 나아가 인터넷을 통해 실시간으로 시험을 치르는 영역으로까지 확대되고 있다 [1]. 하지만 대부분의 수험자 수학 능력 평가는 작성된 답안을 자연 언어 처리를 통하여 분석하는데 많은 어려움이 있기 때문에, 주관식 문제보다는 단답형 또는 선택형 문제를 통한 평가만이 가능하였다[2, 3]. 그러나 올바른 수학 능력의 평가를 위해서는 다양한 문제 및 답안 유형에 대하여 균등한 평가가 필요한데, 이를 위하여 주관식 문제 채점에 대한 필요성이 대두되었다.

\* 이 논문은 2004년도 한국학술진흥재단 지역대학우수교과자 지원사업 (R05-2004-000-10376-0)에 의하여 지원받았음.

† 준 회 원 : 한림대학교 대학원 컴퓨터공학과

\*\* 종신회원 : 한림대학교 정보전자공과대학 공대학장

\*\*\* 종신회원 : 한림대학교 정보통신공학부 조교수

논문접수 : 2005년 9월 15일, 심사완료 : 2005년 10월 7일

본 논문에서는 원격 교육에서의 주관식 평가에 있어 대두된 문제점들을 해결하기 위하여 주관식 답안의 자동 채점이 가능한 시스템을 제안한다. 자동 채점을 위하여 문제를 벡터 공간으로 사상하고, 말뭉치에서 추출하여 구성한 의미커널을 이용하여 해당 답안과 모범 답안 간의 유사도(similarity)를 계산하는 정보 검색 방법론을 사용하였다[2].에서는 벡터 공간 모델(Vector Space Model: VSM)을 적용한 정보 검색 방법론을 사용한 주관식 문제 채점 시스템을 제안하였다. 이것은 자동 채점의 문제를 벡터 공간으로 사상하여, 이들 벡터간의 유사도를 벡터의 내적으로 계산하여 추정함으로써 자동 채점을 한 것이다. 그리고 [3]에서는 [2]의 한계인 유사한 단어 간의 매칭 문제를 해결하기 위하여 자체적으로 제작한 소규모 유의어 사전을 사용하여 단순 매칭 방식이 아닌 의미 매칭 방식을 시도하였다. 또한 답안간의 유사도를 추정하기 위하여 휴리스틱에 기반한 유사도 계산 방법을 제안하였다. 그러나 이 방법은 유의어 사전의 규모가 작아 다양한 적용이 어렵고, 유의어를 통한 답안 확장으로 어휘의 의미를 부분적으로 반영할 수는 있지만 결과적으로 확장된 답안을 단순 매칭하여 유사도를 추정함으로써 어휘의 의미를 반영하는데 한계를 가지고 있었다.

벡터 공간 모델은 벡터 내부의 다양한 자질들 사이의 상관관계를 고려할 수 없다는 점에서 그 문제점이 지적되었다[4]. 이 문제에 대한 대안으로 은닉 의미 분석(Latent Semantic Analysis: LSA)이 있는데, 이 방법은 문서의 내용이 서술된 색인어 보다는 그 안에 표현된 개념에 기반한다는 점에 착안하여 제안된 모델이다. 이 방법을 사용하면, 문서들이 동일한 색인어를 가지고 있지 않더라도 문서간의 연관성을 계산할 수 있다. 즉 어떤 문서가 다른 문서와 개념이나 주제를 공유하는지 여부를 계산함으로써 문서의 유사도를 계산할 수 있다[5].

본 논문에서 제안된 시스템은 LSA를 기반으로 하여 의미커널을 구축하고 이를 통하여 주관식 답안의 채점을 가능케 하도록 하였다. 또한 한글 워드넷[6]을 이용하여 답안을 확장함으로써 수험자의 유의어 사용을 고려한 채점이 가능하도록 하였다. 한글 워드넷을 통한 답안의 확장은 대부분의 주관식 채점에서 문장의 문맥보다는 사용된 어휘의 의미에 가중치를 두어 채점한다는 것에 착안한 것이다.

실험을 위하여, 100명의 학생들로부터 얻은 3000문항의 주관식 답안을 출제자가 직접 채점한 결과와 본 논문에서 제시한 시스템으로 자동 채점한 결과를 비교하여 정확성과 신뢰성을 평가하였다. 그리고 기존에 제시되었던 벡터 공간 모델과의 성능 비교를 통하여 향상된 성능을 확인하였다.

2장에서는 자동 채점을 위하여 본 논문에서 활용한 의미커널 방법론에 대한 설명을 하고 3장에서는 한글 워드넷을 이용한 답안 확장 모델에 관하여 설명한다. 4장에서는 본 논문에서 제안한 방법을 적용시킨 주관식 채점 시스템의 구성을 상세하게 설명하고, 5장에서는 실험 결과를, 그리고 6장에서는 결론 및 연구 결과를 바탕으로 향후 연구 방향을 제시한다.

## 2. 의미커널 구축

### 2.1 의미커널(Semantic Kernel)

본 논문에서는 출제자와 수험자의 답안을 벡터 공간으로 사상하고, 변환행렬을 도입한 커널 함수를 사용하여 자동 답안 채점을 시도함으로써 출제자의 채점에 대한 부담을 줄이고자 하였다.

정보검색을 위한 벡터 공간 모델에서 두 문서  $d_i$ 와  $d_j$ 의 유사도는 기본적으로 다음과 같이 표현할 수 있다[4].

$$\text{sim}(d_i, d_j) = \cos(P^T d_i, P^T d_j) = \frac{d_i^T P P^T d_j}{\|P^T d_i\| \|P^T d_j\|}$$

$$d_i = (t_i(w_1), t_i(w_2), \dots, t_i(w_m))^T \quad (1)$$

$$d_j = (t_j(w_1), t_j(w_2), \dots, t_j(w_m))^T$$

$t_i(w_m)$ 은 단어  $w_m$  ( $1 \leq m \leq M$ )의 문서  $d_i$  내에서의 정수를 나타내는 것으로 보통  $d_i$  내에서의  $w_m$ 의 존재 유무나 빈도수를 나타낸다. 행렬  $P$ 는 기본 벡터 공간에서의 문서를 다른 자질공간으로 사상하기 위한 변환행렬이다. Cristianini[7]는 이를 SVM(Support Vector Machine)과 같은 커널(kernel) 방법론에서의 커널 관점에서 정의하였다. 커널을 이용한 방법은 특정 공간에서 상이한 입력들 사이의 내적을 구하는데 사용되며, 매우 복잡하며 무한한 차원 벡터들 사이의 내적을 계산하는데 효율적이다. 즉, 문서  $d_i$ 와  $d_j$ 가 주어졌을 때 특정 자질공간으로의 사상(mapping)  $\Phi$ 에 대해 커널 함수는  $k(d_i, d_j) = \langle \Phi(d_i), \Phi(d_j) \rangle$ 로 표현된다. 식 (1)에서 변환행렬  $P$ 는 결국  $\Phi$ 를 정의하는 것으로 볼 수 있으며, 이 때  $\Phi(d) = P^T d$ 로 주어진다.

변환행렬이  $P = I_M (I_M$ 은  $M \times M$  단위행렬)일 때는 단어 벡터 공간에서의 두 문서간의 내적에 의한 유사도이며, 역문서빈도를 해당 단어에 대한 가중치로 부여할 경우 변환행렬  $P$ 는 대각 행렬  $P_{IDF} = \text{diag}(idf(w_1), idf(w_2), idf(w_3), \dots, idf(w_m))$ 로 표현된다. 이와 같이  $P$ 가 대각 행렬인 경우 ( $P = I_M$ 인 경우도 포함하여)에는 각 단어들에 의해 표현되는 축이 벡터공간상에서 직교(orthogonal)를 이루므로 단어들의 상관관계를 고려할 수 없다는 점에서 그 문제점이 지적되곤 한다[7, 8, 9, 10].

기본 벡터 공간 모델의 이러한 문제점을 어느 정도 해소하기 위하여, 단어 간의 상관관계 또는 의미 관계를 고려할 수 있는 방향으로  $P$ 를 정의할 수 있는데, 이의 도입에 의한 문서 간 유사도 또는 커널 함수  $k(d_i, d_j)$ 을 “의미커널(semantic kernel)”이라 한다. 본 논문에서 사용하고 있는 LSA기법에서는 기존의 색인어-문서 행렬  $D$ 에 대하여 SVD(Singular Value Decomposition)[11]를 수행하고 ( $D = USV^T$ ) 상위  $k$  ( $k \leq \text{rank}(D)$ )개의 left singular vector 들을 이용하여 문서 간 유사도를 정의한다. 즉, LSA 적용에서의 변환행렬

은  $P_{LSA} = U_k$ 로 정의할 수 있다. 이에 대한 보다 자세한 설명은 다음 장에 나타나 있다.

2.2 은닉 의미 분석(Latent Semantic Analysis)

은닉 의미 분석은 다변량 통계분석 방법으로 고차원의 데이터 공간에 대해 축을 변경하여 그 데이터를 가장 잘 설명하는 새로운 축을 찾아내고 설명력이 낮은 축들은 걸러냄으로서 데이터에 은닉해 있는 의미 구조를 밝히는 기법이다. 새로운 축을 찾아내기 위해서는 SVD라는 대수학적 기법을 사용하는데, SVD를 거치면 설명력이 높은 순서로 원하는 수만큼의 고유벡터(eigen vector)를 얻을 수 있다. 이 때 설명력이 낮은 축은 잡음(noise)으로 간주하여 제거함으로써 데이터의 차원을 줄이는 효과도 얻게 된다. 이 고유벡터들은 원래의 데이터 행렬에는 드러나 있지 않았던 의미구조를 표상하게 되므로 보다 심층적인 수준에서의 의미 분석을 가능하게 된다[5].

말뭉치 안의 임의의 문서는 통사 구조를 가지고 있는 문장들의 집합으로 되어 있다. 그러나 은닉 의미 분석에서는 통사 구조를 무시하고 색인어들만을 고려하기 때문에 문서  $d_i$ 에는 어떤 색인어들이 있는가의 정보만 사용하여 벡터로 표현할 수 있다[10].

$$d_i = \langle t_1, t_2, \dots, t_n \rangle \quad (2)$$

$t_j$ 는 문서에 그 단어가 포함되어 있다면 1, 그렇지 않으면 0으로 나타낼 수도 있으며, 문서 안에서 몇 번의 빈도도 나타났는가에 따라서 정수로 표현할 수도 있는데, 현재 정보 검색 분야에서는 대부분  $tf \cdot idf$ 로 계산된 가중치 값을 사용한다.

각 문서들은 색인어의 집합인 벡터로 표현되기 때문에 문서들의 집합인 전체 말뭉치는 색인어-문서의 행렬로 나타낼 수 있다. 그런 색인어의 수와 문서의 수는 매우 큰 숫자이므로 이 행렬은 보통 대규모 희소행렬(sparse matrix)이 된다. 그러므로 이 행렬에 SVD를 적용한다.

예를 들어 색인어의 수가  $i$ 이고 문서의 수가  $d$ 인  $i \times d$  행렬  $D$ 는 SVD계산을 통해 다음과 같이  $U, \Sigma, V$ 의 세 개의 행렬로 나누어지게 되고, 이때  $U$ 는 단어 벡터들로 이루어진 행렬을,  $V$ 는 문서 벡터들로 이루어진 행렬을 의미한다.

$$D = U \Sigma V^T \quad (3)$$

여기서  $\Sigma$ 의 값을 특이치(singular value)라 하며,  $DD^T$  혹은  $D^T D$ 가 0이 아닌 고유값(eigen value)으로 구성된 대각 행렬(diagonal matrix)이며,  $U$ 와  $V$ 는  $DD^T$  혹은  $D^T D$ 의 0이 아닌  $r$ 과 연관된 직교 고유벡터(orthogonal eigen vec-

tor)이다. 분해된 행렬  $D$ 에 대한 근사 값을 구하기 위해 크기로 정렬되어 있는 특이치 값에서 상위  $k(k \leq rank(D))$ 개의 left singular vector들을 이용하면 식(3)은 다음의 식(4)과 같이 재 정의할 수 있다.

$$D_k = U_k \Sigma_k V_k^T \quad (4)$$

이와 같이 축소된 SVD를 수행하는 이유는 첫째 어휘간의 의미적인 연관성을 행렬로 표현하고 둘째 행렬의 희소성을 해소하여 분석의 효율성을 높이며, 마지막으로 잡음을 유발하는 축들을 걸러냄으로써 보다 정확도를 높이기 위함이다.

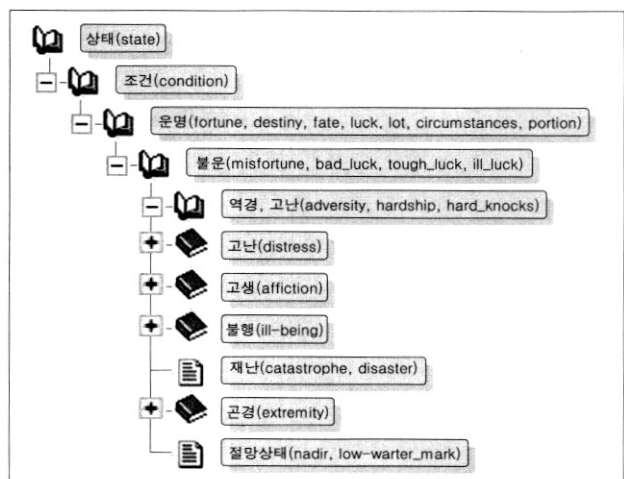
이렇게 새롭게 생성된 축소된 차원의 단어 벡터 행렬  $U_k$ 가 의미커널의 변환행렬( $P_{LSA} = U_k$ )로 사용되며 절의와 문서의 유사도는 식 (1)을 통하여 계산되어진다.

3. 워드넷을 이용한 답안 확장 모델

3.1 한글 워드넷

한글 워드넷에서의 어휘개념은 먼저 영어 워드넷의 동의어 집합으로 표현되고 영어 명사에 없는 어휘개념은 한글 명사에 알맞은 어휘개념을 만들어 표현하였다. 한글 명사 워드넷의 설계에 있어서 중요한 관건은 기본 골격인 어휘개념을 표현하는 형식과 계층구조 관계 설정이다. 한글 명사 워드넷의 어휘개념은 차별화 이론을 채택하여 개념별로 구별될 수 있는 상징에 의하여 구성되었고 한글 명사 워드넷의 계층구조는 이 노드들 간의 상하위개념 관계로 표현되었다[12].

이러한 의미론적 개념 분석에 기반하여 도메인 선정 작업을 하고 한글 명사 워드넷의 데이터베이스 구축 시스템을 설계하였다. 한글 워드넷 데이터베이스는 워드넷에 필요한 엔트리와 synset에 관한 정보를 데이터베이스로 구축한 것

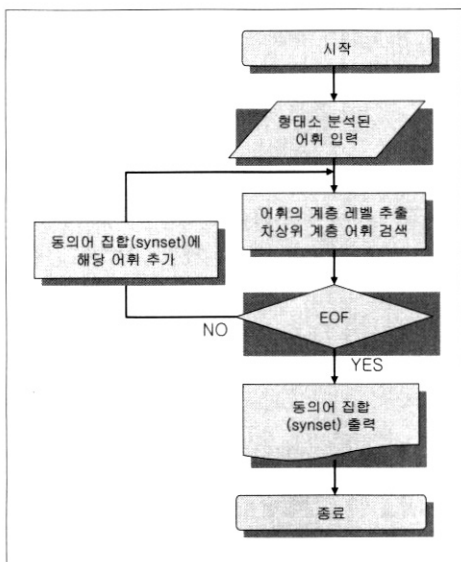


(그림 1) 한글 워드넷의 계층형 명사 구조

이다. (그림 1)은 ‘고난’이라는 명사 어휘에 대한 한글 워드넷의 계층적 구조와 위치를 나타낸 것이다.

### 3.2 워드넷을 이용한 답안 확장 모델

본 논문에서는 한글 워드넷 2.0[6]을 이용하여 모범답안과 수험자의 답안을 확장하였다. 시스템은 답안을 확장하기 위하여 입력된 답안에서 형태소 분석 프로그램[14]을 사용하여 명사 어휘를 추출한다. 이 추출된 명사 어휘를 가지고 데이터베이스 서버에 저장된 한글 워드넷을 검색하여 각 명사 어휘에 대한 복수개의 동의어 집합(Synset)을 추출한다. 동의어 집합을 이용하여 동일 의미 또는 유사 의미의 어휘를 추출하기 위하여, 먼저 추출된 명사 어휘를 데이터베이스에서 검색한 후 해당 어휘의 계층 레벨을 추출한다. 그런 후 추출된 계층상에 해당하는 단어와 직접적으로 하위(hyponym) 관계를 맺고 있는 명사 어휘를 모두 추출한다. 이렇게 하여 동일 의미 또는 유사 의미의 명사 어휘를 검색하게 된다. 이 때 형태소 분석기에서 추출된 명사 어휘와 관련된 동의어 집합이 더 이상 존재하지 않을 경우에는 시스템이 종료하게 된다. 하지만 다른 계층상에 동의어 집합이 존재할 경우에는 시스템은 답안 문장 백터에 동의어와 유사어를 추가함으로써 답안을 계속 확장하게 된다. (그림 2)는 한글 워드넷을 이용한 답안확장 순서도를, <표 1>은 한글 워드넷을 이용하여 명사 어휘를 확장하여 나타낸 예를 보여준다.



(그림 2) 워드넷을 이용한 답안확장 순서도

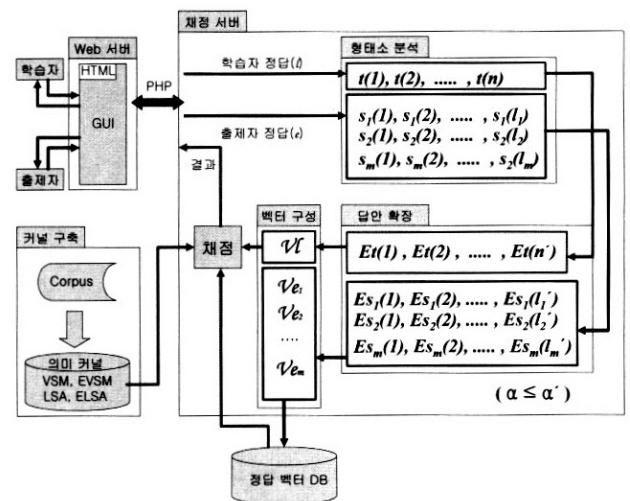
<표 1> 명사 어휘 확장의 예

명사 어휘	확장 어휘
문제	불운, 불행, 고민거리, 낭패, 곤경, 스캔들...
고난	고생, 재난, 곤경, 불행
요리	굽기, 튀기기, 끓이기
경영	지휘, 관리, 사법, 조직, 조직화, 정치

## 4. 의미커널과 한글 워드넷을 이용한 주관식 채점 시스템 구성

### 4.1 전체 시스템 구성

본 논문에서 제안하는 의미커널과 워드넷을 이용한 주관식 채점 시스템은 크게 세 가지 모듈로 구성되어 있다. 첫 번째는 웹서버 모듈로 출제자가 문제와 모범답안을 입력하고 수험자가 작성한 답안을 입력할 수 있는 모듈이다. 이 모듈에서 수험자는 제출한 답안의 획득 점수를 실시간으로 확인할 수 있다. 두 번째는 의미커널을 구축하는 모듈로 여기서 구축된 의미커널은 채점 모듈에서 유사도를 계산하는데 사용된다. 세 번째는 채점 서버 모듈로 수험자의 답안과 출제자의 답안을 형태소 분석한 후, 워드넷 확장과 커널 구축 모듈에서 이미 구축된 의미커널을 이용하여 유사도를 계산하는 본 시스템에서 가장 중심적인 모듈이다. (그림 3)은 본 논문에서 제안하는 시스템의 전체 구성을 나타낸다.



(그림 3) 의미커널과 한글 워드넷을 이용한 주관식 채점 시스템의 전체 구성

### 4.2 전체 시스템 작업 흐름

출제자는 수험자가 문제를 풀 수 있도록 웹서버에 접속하여 문제와 모범답안을 입력한다. 입력된 문제는 수험자가 웹서버에 접속하여 풀 수 있도록 웹서버에 저장되고 모범답안은 수험자가 입력할 답안과의 유사도 계산을 위해 채점서버로 전송되어 벡터형태로 저장된다. 벡터로 저장하기 위해서는 모범답안 문서를 우선 형태소 분석을 거쳐 형태소단위로 단위로 나누고 워드넷 검색을 통해 동의어 그룹으로 확장하여 확장답안을 구성한다. 구성된 답안은 기저벡터를 기반으로 벡터 형태로 DB에 저장되어 수험자가 채점을 요구할 때 사용된다.

수험자는 출제자와 마찬가지로 웹서버에 접속하여 문제를 풀고, 답안을 웹서버로 전송한다. 수험자가 웹서버에 접속하여 제시된 문제에 대한 답안을 작성하는 것으로 지능형 채점 시스템의 작업이 시작된다. 웹서버는 학습자 답안을 채



### 5. 실험 및 성능 평가

#### 5.1 실험 환경 및 실험 데이터

시스템의 성능 평가를 위하여 ‘고사성어’ 주관식 문제 30 문항을 가지고 100명의 대학생으로부터 수험자 답안을 수집, 총 3,000문항의 주관식 문제를 실험에 사용하였다. 유사도 계산을 위해 38,727개의 신문기사로 말뭉치(corpus)를 만들고 이를 형태소 분석 프로그램으로 색인어를 추출하여 75,175개의 어휘를 추출하였다. 이 어휘들은 유사도 계산을 위해 사용되는 벡터들의 기저벡터로 사용하였다.

각각의 문항은 출제자가 미리 입력한 복수의 모범답안이 벡터형식으로 DB에 저장되어 있고 수험자가 입력한 답안은 기저벡터 형식으로 변환되어 유사도가 계산된다. 의미커널

〈표 5〉 수험자 답안지별 각 모델의 채점 데이터

	VSM	EVSM	EVSM_MAX	LSA	ELSA	ELSA_MAX	Human
평균	16.50	17.49	18.51	20.44	20.77	21.38	20.18
표준편차	5.16	5.32	5.41	4.73	4.70	4.74	5.81

〈표 6〉 문항별 각 모델의 채점 데이터

	VSM	EVSM	EVSM_MAX	LSA	ELSA	ELSA_MAX	Human
평균	0.55	0.58	0.62	0.68	0.69	0.71	0.67
표준편차	0.39	0.41	0.40	0.36	0.39	0.36	0.39

은 VSM과 LSA를 이용하여 구축하고 이 커널을 통하여 답안을 확장한 경우에 그렇지 않은 경우에 대하여 유사도를 계산하였다.

(그림 6)과 (그림 7)은 한개 문항에 대한 채점 시스템으로 30개의 문항에서 한 문제를 랜덤으로 출제하여 수험자가 답을 입력할 수 있도록 하고 있다. 수험자가 해당 문제에 대한 답을 입력하면 LSA와 ELSA(Extended LSA)로 채점된 결과를 화면에 출력하여 주고 그 최대값을 표시한다.

#### 5.2 성능 평가

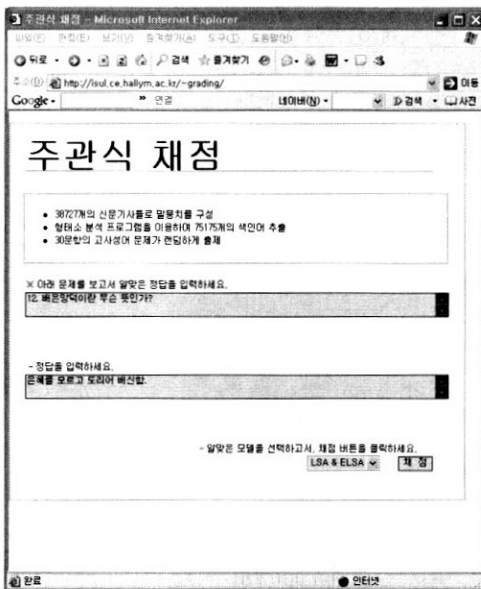
성능평가를 위해 수험자 답안지별(100항목), 문항별(3000항목)의 점수평균, 표준편차와 상관계수[13]를 구하여 각 모델간의 성능을 비교하였다.

각 항목은 계산된 유사도를 한 문항당 1점 만점으로 환산하여 채점하였고 출제자도 0~1점 사이의 점수로 채점하였다. <표 5>에서 VSM 항목은 VSM 의미커널을 이용한 수험자 답안의 유사도 계산 결과이고, EVSM 항목은 한글 워드넷을 이용하여 수험자 답안 및 모범답안이 확장된 VSM 의미커널을 이용한 유사도 계산 결과이다. EVSM\_MAX 항목은 VSM과 EVSM 두 가지 방법으로 계산한 것 중 최대 유사도를 산출한 항목이다. LSA 세가지 항목도 VSM과 같은 방법으로 표기하였다. 마지막 HUMAN 항목은 출제자가 0~1 사이의 부분점수로 채점한 결과이다. <표 5>는 수험자 100명의 답안지 채점결과들의 평균과 표준편차를 측정된 결과이다.

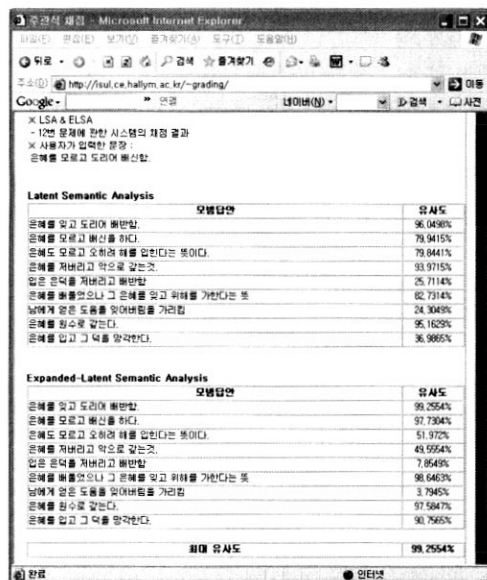
<표 6>은 100명의 수험자 1명당 30문제씩, 총 3000문항에 대한 답안지 채점결과들의 평균과 표준편차를 측정된 결과이다.

본 평가에서는 시스템의 채점결과와 출제자의 채점결과와의 유사도 측정을 위하여 출제자가 채점한 데이터와 각 모델들간의 상관계수(Corr(X,Y))를 [13]에서 설명하고 있는 Pearson의 상관계수를 이용하여 계산하였다.

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (5)$$



(그림 6) 답안 입력 화면



(그림 7) 점수 출력 화면

여기서  $X$ 는 각 모델의 채점 데이터,  $Y$ 는 출제자의 채점 데이터를 의미하고  $\overline{X}$ 는  $X$  값의 평균을 의미한다. 상관계수는  $-1 < Corr(X,Y) < 1$ 의 값을 가질 수 있으며 그 값이 1에 가까울수록 두 데이터간의 양의 상관관계가 높다고 할 수 있다[13]. <표 7>은 출제자의 채점 데이터와 각 모델의 간의 상관계수를 계산한 결과이다.

<표 7> 출제자의 채점 데이터와 각 모델간의 상관계수

	VSM	EVSM	EVSM_MAX	LSA	ELSA	ELSA_MAX
수험자별	0.8533	0.8603	0.8699	0.8859	0.8940	0.8914
문항별	0.5658	0.5584	0.5776	0.5875	0.5940	0.6129

위 결과에서 VSM에 비하여 LSA와 ELSA가 사람이 채점한 결과와 비슷한 평균과 표준편차, 상관관계를 보였다. ELSA\_MAX의 값은 LSA와 ELSA 두 모델의 최대값을 가져왔기 때문에 사람이 채점한 결과보다 약간 점수가 높게 채점되었다. VSM은 벡터 내부의 다양한 자질들 사이의 상관관계를 고려할 수 없는 등 방법론적인 한계로 LSA 모델에 비해 사람이 채점한 점수와의 차이가 많이 났다. 그리고 여기서 수험자별 상관관계가 문항별 상관관계에 비하여 매우 높은 값을 보이고 있는데, 이는 상관관계가 비율이나 평균값(상수)인 경우 더 높은 값을 보이는 성질 때문에 나타난 현상이라 볼 수 있다. 그러나 채점결과가 수험자별로 주어진다는 사실을 고려한다면, 이는 충분히 의미 있고 또한 매우 높은 상관관계 값이라 할 수 있다.

5.3 계산오류 원인 분석

5.2절의 성능평가를 보면 상관계수( $Corr(X,Y)$ ) 수치상 약 0.1정도의 채점오류가 발생하였다. 이와 같은 채점 오류가 발생한 원인을 분석해 본 결과, 형태소 분석시의 오류, 색인어 부족으로 인한 오류, 방법론적인 한계 등 크게 세 가지로 그 원인을 나타낼 수 있었다. <표 8>은 각 원인별 채점 오류 발생 비율을 나타낸 것이다.

형태소 분석오류의 경우 <표 2>에서 “한번의 노력으로 두가지 이익을 얻다”라는 문장이 있을 때 ‘한번’ 이라고 형태소 분석이 되지 않고 ‘한번의’ 라는 원치 않은 단어로 형태소 분석이 되어 채점 오류를 야기한 경우가 되겠다. 색인어 부족의 경우로는 사람이 봤을 때 의미상으로는 같지만 모범답안이 이를 미처 표현하지 못한 경우, 맞춤법이나 띄어쓰기가 정확하지 않아 시스템이 그 어휘를 제대로 채점에 반영할 수 없는 경우가 있겠다.

<표 8> 원인별 채점 오류 발생 비율

	발생 비율(%)
형태소분석 오류	70%
색인어 부족	22%
기타(방법론의 한계)	8%

실험에서 같은 문장을 가지고 유사도 계산을 했을 때 LSA와 비교하여 ELSA에서 현저히 낮은 점수가 나오는 경우가 있었다. <표 9>는 “자기 자신을 칭찬하는 말”이라는 수험자 답안이 있을 때 계산되는 값을 나타내고 있다. <표 9>에서 수험자 답안 매칭, 모범 답안 매칭 은 형태소 분석된 단어가 75,175개의 기저 벡터에 매칭된 횟수를 나타낸다.

<표 9>와 같이 모범답안은 확장이 되어 기저벡터에 많이 나타나지만 수험자답안은 문장이 짧음, 형태소 분석 실패 등의 이유로 답안 확장이 이루어 지지 않은 경우, 유사도 계산시의 분모에 해당하는 값만 커지게 되어 LSA와 비교하여 현저히 낮은 점수가 나오는 것을 볼 수 있었다.

<표 9> ELSA 계산 오류의 예

LSA	계산 모델	ELSA
0.8337	유사도 점수	0.0025
3	수험자 답안 매칭	3
6	모범 답안 매칭	63

5. 결론 및 향후 연구과제

인터넷의 발달과 함께 교육분야에서도 원격을 이용한 학습 및 평가가 활발하게 이루어지고 있다. 수험자의 수학 능력 평가에서는 자연언어처리의 어려움으로 객관식이나 단답식 평가가 주류고 이루고 있으나 본 논문에서는 VSM, LSA, 한글 워드넷 등 다양한 언어 지식을 이용하여 서술형 주관식 문제의 빠르고 공정한 채점을 가능하게 하는 시스템을 보였다. 여러 가지 모델로 의미커널을 구축하고 성능을 측정한 결과 LSA와 한글 워드넷을 이용하여 구축한 시스템에서 사람이 채점한 내용과 비교하여 최대 0.894의 상관관계를 얻을 수 있었다.

5.3절에서 밝힌 계산 오류 원인에 대한 문제를 해결함으로써 본 시스템의 성능을 더욱 향상시킬 수 있을 것이다. 형태소 분석에서의 오류와 색인어 부족 등과 같은 문제는 정교한 형태소 분석을 통한 어휘의 누락을 방지하고 좀 더 강력한 시소러스의 구축으로 한글 워드넷에서 어휘 누락 및 확장 실패를 방지할 수 있을 것이다. 다음으로 다양하고 강력한 커널의 구축으로 최적의 커널 모델을 찾고 시스템의 성능을 향상시키는 연구를 계속 진행하는 것은 방법론의 한계를 극복하는 방법이 될 것이다.

또한 벡터 유사도 방식의 한계인 단어들의 위치에 따른 상관관계를 표현하지 못하는 문제를 해결하는 것, 그리고 한 어휘가 워드넷 상의 여러 계층에 존재할 경우 가장 적합한 계층을 찾아내어 답안 확장을 보다 정확히 하는 것은 지능형 채점 시스템의 성능 향상을 위한 좋은 연구 과제가 될 것이다.

참고 문헌

[1] 김경아, “웹 기반 자동 문제 출제 시스템”, 동국대학교 교육대

학원 석사학위논문, 2000.

- [2] 정동경, “벡터유사도와 시소러스를 이용한 주관식 답안의 채점방법”, 동국대학교 교육대학원 석사학위논문, 2001.
- [3] 박희정, 강원석, “유의어 사진을 이용한 주관식 문제 채점 시스템 설계 및 구현”, 컴퓨터교육학회논문지, 2003.
- [4] 장정호, 김유섭, 장병탁, “헬름홀츠머신 학습 기반의 의미커널을 이용한 문서 유사도 측정”, 한국정보과학회 춘계학술대회, 2003.
- [5] 신동호, “Latent Semantic Analysis를 이용한 내용 기반 정보 검색 시스템”, 서울대학교 대학원 협동 과정 인지과학전공 석사학위논문, 2000.
- [6] 이은녕, 임성신, WordNet 2.0의 한국어 번역 작업과 결과물, 부산대학교 한국어정보처리연구실.
- [7] Cristianini, N., Shawe-Taylor, J., and Lodhi, H., “Latent Semantic Kernels,” Journal of Intelligent Information System, Vol.18, No.2/3, pp.127-152, 2002.
- [8] Deerwester, S., S. T. Dumais, G. W. Furas, T. K. Landauer, and R. A. Harshman, “Indexing by latent semantic analysis,” Journal of the American Society for Information Science, ature Medicine, Vol.41, No.06, pp.391-407, 1990.
- [9] Frey, B. J., “Graphical Models for Machine Learning and Digital Communication,” The MIT Press, 1998.
- [10] Landauer, T. K., P. W. Foltz, and D. Laham, “An Introduction to Latent Semantic Analysis,” Discourse Process, Vol.25, pp.259-284, 1998.
- [11] Landauer, T. K. and Laham, D., “Learning Human-like Knowledge by Singular Value Decomposition,” A Progress Report, NIPS, Vol.10, pp.45~51, 1997.
- [12] 의미론적 어휘개념에 기반한 한국어 명사 WordNet의 설계와 구축, 서울대학교 박사학위논문, 1996.
- [13] L. Bain and M. Engelhardt, “Introduction to Probability and Mathematical Statistics,” Thomson Learning, pp.179-190, 1987.
- [14] 강승식, 범용 형태소 분석기 “HAM Ver 6.0.0”, 국민대학교 자연언어 정보 검색 연구실, <http://nlp.kookmin.ac.kr>



**조우진**

e-mail : woojini@hallym.ac.kr  
 2004년 한림대학교 정보통신공학부(학사)  
 2004년~현재 한림대학교 컴퓨터공학과 석사과정  
 관심분야: 원격교육, 컴퓨터통신, 무선인터넷 콘텐츠 및 솔루션 개발



**오정석**

e-mail : bil78@hallym.ac.kr  
 2003년 한림대학교 수학과(학사)  
 2005년 한림대학교 컴퓨터공학과(공학석사)  
 2005년~현재 한림대학교 컴퓨터공학과 박사과정  
 관심분야: 원격교육, 컴퓨터통신, 무선인터넷 콘텐츠 및 솔루션 개발



**이재영**

e-mail : jylee@hallym.ac.kr  
 1978년 한양대학교 전자공학과(학사)  
 1981년 서울대학교 전자공학과(공학석사)  
 1996년 서울대학교 대학원 전자공학과(공학박사)  
 1984년~현재 한림대학교 정보통신공학부 교수  
 2004년~현재 한림대학교 정보전자공과대학 공대학장  
 관심분야: 원격교육, 컴퓨터통신, 무선인터넷 콘텐츠 및 솔루션 개발



**김유섭**

e-mail : yskim01@hallym.ac.kr  
 1992년 서강대학교 전자계산학과(학사)  
 1994년 서울대학교 컴퓨터공학과(공학석사)  
 2000년 서울대학교 컴퓨터공학과(공학박사)  
 2000년~2001년 서울대학교 컴퓨터신기술 공동연구소 전문연구요원  
 2001년 (주)아이시티 연구소장  
 2001년~2002년 이화여자대학교 과학기술대학원 연구전임강사  
 2002년~현재 한림대학교 정보통신공학부 조교수  
 관심분야: 전산금융, 자연언어처리, 기계번역, 데이터마이닝, 기계학습