

# 유의어 사전을 이용한 주관식 문제

## 채점 시스템 설계 및 구현

박희정<sup>†</sup> · 강원석<sup>††</sup>

### 요약

일선학교에서 평가는 주로 객관식 평가를 하여왔다. 주관식 평가는 객관식보다 높은 수준의 인지 능력을 측정할 수 있으나 채점의 객관성과 신뢰성의 문제로 기피하여 왔다. 본 논문은 주관식 평가를 위해 주관식 문제 채점 모형을 제안하고 유의어 사전을 이용한 채점 시스템을 설계, 구현하였다. 제안된 시스템은 주관식 문제의 다양한 유형을 처리할 수 있도록 하고 시스템의 사용법을 알지 못하는 교사라도 쉽게 이를 이용할 수 있는 편리성을 제공하였다. 이 시스템을 통해 채점의 시간과 노력을 덜고 채점의 객관성을 획득할 수 있다. 시스템의 실험 결과 시스템 채점이 교사 채점의 최고와 최저 사이에 오는 경우가 73%에 달하는 결과를 얻을 수 있었다. 앞으로 본 연구가 주관식 문제 채점 방법 연구의 기반이 될 것으로 기대된다.

## Design and Implementation of a Subjective-type Evaluation System Using Natural Language Processing Technique

HeeJung Park<sup>†</sup> · WonSeog Kang<sup>††</sup>

### Abstract

An instructor in education generally takes the objective-type evaluation for grading. The subjective-type evaluation has the merit that it can estimate the high-recognition ability, but the problem of the objectivity and reliability of the evaluation. This paper proposes the model which grades for the subjective-type evaluation, and designs and implements the evaluation system using the synonym thesaurus. This system can process the diverse and wide subjective-type questions and provide the easy usage for a beginner. It also can reduce the time and endeavor for evaluation and provide the objectivity of the evaluation. The system results the 73% success rate. We expect that this system will become a basis of the research on the subjective-type evaluation.

### 1. 서론<sup>1)</sup>

#### 1.1 논문의 목적 및 필요성

인터넷이 상용화되면서 교육 분야에서도 가상 교육 등 웹 기반 학습이 늘어나고 있다. 웹기반

학습은 언제 어디서든지 학습이 가능하다는 점과 함께 동영상 등의 멀티미디어를 활용할 수 있는 점에서 급속하게 발전하고 있다. 그러나 학습의 결과를 평가하는 연구는 아직 미흡한 실정이다.

평가는 교사와 학생들이 학습의 성취가 어떠한지를 알게 하고 학생들의 잘못된 이해를 교정

<sup>†</sup> 준회원 : 안동대학교 컴퓨터교육과 대학원생  
<sup>††</sup> 정회원 : 안동대학교 컴퓨터교육과 교수  
 논문접수: 2003년 6월 3일, 심사완료: 2003년 7월 15일

하며 질적 개선의 기회를 제공할 수 있기 때문에 어떤 방법을 사용해야 하는지는 매우 중요하다. 일반적으로 평가는 객관식과 주관식으로 할 수 있다. 객관식 평가는 채점의 객관성과 신뢰성이 높지만 평가의 질이 떨어진다. 대안인 주관식 평가는 자신의 생각과 의견을 서술할 수 있어 높은 수준의 인지 능력을 측정할 수 있는 교육적 가치가 높은 평가 방법이라고 생각된다[1]. 이와 같은 이유로 [2]는 채점에 시간과 노력을 줄이는 자동 주관식 채점 방법에 대한 연구를 하였으나 문제 출제와 채점 방법의 기준이 제한적이고 다양하지 못하여 일선학교에서의 이용가치가 낮다.

본 논문은 이와 같은 필요성에 따라 다양한 주관식 문제를 평가할 수 있는 주관식 문제 채점 모델을 설계하고 유의어 사전을 사용하여 채점하는 채점 시스템을 설계, 구현한다. 본 논문에서는 먼저 기준이 모호한 주관식 문제 채점 방법을 연구하여 객관성 있는 채점 모델을 제시한다. 이 모델을 통해 교사는 문제 채점의 기준을 입력하고 시스템은 그 기준을 토대로 학습자에게 신뢰성 있고 일관성 있는 채점 결과를 제시한다. 본 논문에서 제안한 주관식 채점 모델은 주관식 평가 방법의 기초가 되고 채점 시스템에 사용된 유의어 사전은 자연어처리 분야와 자연어 인터페이스 분야에 활용될 것으로 기대한다.

### 1.2. 연구의 내용 및 방법

본 논문에서 연구한 내용은 다음과 같다. 첫째 주관식 평가를 위해 문장 분석과 답안 채점에 필요한 기법에 대해 조사하고 기본 기술을 알아본다.

둘째, 주관식 문항에 관하여 조사하고 문항의 틀을 제안한다. 주관식 문항의 종류와 특징, 채점 방법 등을 분석하기 위해 일선 학교의 주관식 문제와 채점 방법을 조사하고 연구한다. 이를 기준으로 주관식 문제 채점을 위한 인터페이스를 제시한다.

셋째, 분석된 채점 방법을 토대로 채점 시스템을 제안한다. 조사된 8가지 주관식 문항에 따라 유의어 사전을 이용한 유의어 검사, 채점 유사도 계산 등을 통해 채점한다. 주관식 문항의 유형은 단답형, 논문형·서술형으로 나누고, 부분

점수의 유무, 정확히 일치하는지 여부, 핵심어의 적용 유무 등에 따라 8가지 유형으로 나뉜다.

마지막으로는 제안된 시스템의 결과를 검사, 분석한다.

## 2. 연구 배경

### 2.1 시소러스

시소러스(Thesaurus)는 단어에 대한 의미를 정의하기 위한 일종의 사전이다[7,12]. 이 사전을 통해 단어의 구체적인 의미와 동의어, 반의어, 상위어, 하위어 등을 알아낼 수 있다. 영어에 대한 시소러스는 이미 오랜 연구를 거쳐 실용가치가 있는 것이 사용중이다. WordNet[7]이 대표적인 사례이다. 그러나 한국어의 경우 필요성에 따라 많은 연구가 있지만 아직 가치가 인정된 보급화된 시소러스가 없는 실정이다. 이에 따라 본 논문에서는 유의어 사전을 정의하고 구성하여 주관식 문제 채점 시스템에 활용하고자 한다.

### 2.2. 주관식 문제의 유형과 특징

지필검사에서 사용되는 문항형식은 일반적으로 응답자의 반응형식을 기준으로 하여 선택형 문항(選擇刑 問項 : select-type item)과 서답형 문항(書答形 : supply-type item)으로 분류된다. 또 채점의 객관성을 기준으로 하여 객관식 문항(客觀式 問項 : objective-type item)과 주관식 문항(主觀式 問項 : subjective-type item)으로 분류되기도 한다[8].

단답형(short-answer item)
서답형(supply 완성형(completion item)-type)
서술형(description item)
논문형(essay item)

(그림 1) 주관식 문항의 분류

주관식 문항은 분석력과 종합력의 높은 수준의 인지능력을 측정할 수 있고, 또 학생들은 다른 어떤 형태의 시험보다 더 많은 노력을 해야 하기 때문에 많은 교사들이 주관식 문제들을 교육적인 가치가 가장 뛰어난 이상적인 시험 방법으로 생각하고 있다. 그러나 주관식의 가장 큰 단점으로는 채점이 주관적이어서 신뢰도가 낮고 채점에 시간과 노력이 많이 소모된다. 본 연구에서는 이와 같은 단점을 해소하고 사용자가 쉽게

사용할 수 있는 편리성을 제공하는 주관식 출제 및 채점 시스템을 구현하고자 한다. 먼저 주관식 문제의 유형과 특징을 살펴본다.

### 2.2.1 단답형·완성형

#### (1) 단답형

단답형 문항은 문제를 의문문이나 명령문으로 제시한 다음 한 두 개의 단어, 숫자, 기호, 혹은 짧은 구나 문장으로 답하도록 하는 문항 형식이다. 단답형을 단구적 단답형과 서술적 단답형으로 나눌 수 있다. 일반적으로 단답형 문항은 단구형 단답형을 의미한다. 서술적 단답형은 답의 길이가 길어지면 서술형이나 논문형으로 분류될 수 있기 때문이다.

#### (2) 완성형

완성형 문항은 문자의 일부분 또는 몇 부분을 비워 놓고 빈 자리에 단어, 구, 숫자, 기호 등을 써넣어 문장을 완성하게 하는 형식이다. 단답형 문항은 문장이 의문문으로 진술되는 데 비해 완성형 문항은 평서문으로 진술된다는 점이 다르지만 완성형과 단답형 문항은 본질적으로 같은 문항 형식이라고 말할 수 있다.

#### (3) 단답형과 완성형 문항의 장단점

단답형과 완성형의 장점은 간단한 지식, 사실, 공식, 계산결과, 용어 등을 손쉽게 출제할 수 있고, 추측의 요인을 최대한 배제할 수 있어 문항 제작에 별로 경험이 없는 사람도 쉽게 출제할 수 있다. 그리고 도표, 그림 지도 등 언어로 설명하기 어려운 자료에서 그 명칭, 기능, 구조 등을 묻는데 유용하다.

단점으로는 추리, 일반화, 해석, 비판, 비교, 분석 등 고등 정신 능력 등을 평가하기 어렵고 학생들의 학습 방향을 피상적이고 단순한 학습방향으로 몰아가게 할 가능성이 높다.

### 2.2.2 서술형·논문형

서술형 혹은 논술형 문항은 단답형과 완성형 문항의 단점을 개선하는 방식의 문항이다. 이 문항을 흔히 '주관식 평가'라고 한다. 이 평가 방식은 자신의 의견이나 생각을 직접 서술하도록 하는 평가 방식이다. 따라서 이 방식은 학생의 창의성, 문제 해결력, 비판력, 판단력, 통합력, 정

보 수집력 및 분석력 등 고등 사고 기능을 쉽게 평가할 수 있다[9].

서술형 평가와 논술형 평가가 자신의 생각을 글로 제시한다는 점에서는 차이가 없지만, 서술형 평가라고 할 경우에는 제시되는 글 중에 자기의 주장이 담기지 않으며 서술해야 하는 분량이 많지 않고 채점을 할 때 서술된 내용의 깊이와 넓이에만 중점을 둔다. 그러나 논술형 평가는 학생이 주장하는 바가 분명히 드러나야 하며 서술해야 할 분량이 상대적으로 많아 채점을 할 때에도 서술된 내용의 깊이와 넓이뿐만 아니라 학생의 생각이나 주장도 중요시한다.

다시 말하면, 논술형 평가도 일종의 서술형 평가이기는 하지만, 개인 나름의 생각이나 주장을 창의적이고 논리적이면서 설득력 있게 조직하여 작성해야 함을 강조한다는 점에서 일반적인 서술형 평가와 구별한다. 그러나 본 논문에서는 단답형이나 완성형과 비교했을 때, 요구되는 답의 길이가 현저히 길고, 학생의 생각과 주장에 대한 분석이 어렵기 때문에 서술형과 논술형을 구분하지 않고 같은 유형으로 분류한다. 그렇지만 이 서술형 문항의 유형은 서술할 내용의 제한하는 정도에 따라 다음의 두 가지로 분류할 수 있다.

#### (1) 응답 제한형

응답 제한형은 단답형이나 완성형보다는 대답의 폭이 다소 넓지만 한 두 개의 구 또는 문장으로 답을 쓸 수 있도록 응답의 범위와 형태를 엄격하게 제한한 형식이다. 학교에서 사용하는 대부분의 서술형 및 논술형 문항은 여러 가지 여건상 이러한 응답 제한형 문항을 활용할 가능성이 높다[10].

#### (2) 응답 자유형

응답 자유형은 범교과형, 특정 교과형, 단독 과제형, 자료 제시형으로 나뉘어진다. 범교과형은 응답해야 할 내용이 특정 교과에 국한되는 것이 아니라 여러 교과의 서로 관련이 있는 내용을 서로 관련시켜 논의하거나 통합적으로 구성하는 것이다. 특정 교과형은 특정 교과내용에 국한되는 것이고 단독 과제형은 답안 작성에 필요한 자료나 정보를 제시하고, 이를 바탕으로 응답하도록

하는 문항 형태이다. 자료 제시형은 답안 작성에 필요한 자료나 정보를 제시하고, 이를 바탕으로 응답하도록 하는 문항 형태이다.

서술형은 반응의 자유도가 보장되고 복잡한 정신 기능을 측정하는데 유용하며, 인성진단과 자료를 얻는데 유용하다. 그리고 바람직한 학습 방향을 제시해 주고 문항제작에 시간과 노력이 절약된다. 단점으로 채점에 신뢰도와 객관도가 부족하고 채점에 시간이 많이 들고 까다롭다.

### 2.3 주관식 문제 채점 관련 연구

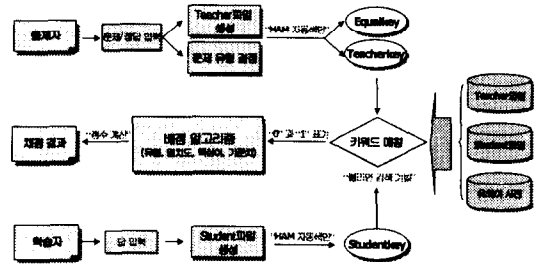
주관식 평가는 자신의 생각과 의견을 서술할 수 있어 높은 수준의 인지 능력을 측정할 수 있는 교육적 가치가 높은 평가 방법이다[1]. 이와 같은 이유로 연구의 필요성은 있으나 자연어 처리기술의 부담으로 연구가 많이 이루어지지 못하고 있는 실정이다. 관련된 논문이 제시되고 있으나 주관식 평가 방법이기보다는 평가 보조 시스템의 내용이 많다. [13]은 네트워크에서 학생 평가를 시도하였다. 이 논문은 문제 은행을 관리하면서 학생에게 문제를 제시하고 학생은 답안을 작성하는데 메뉴 선택을 통한 객관식 평가에 대한 연구이다. [14] 또한 학생의 답안 작성이 마킹을 통한 답안 작성으로 객관식 평가를 하였다. [15]는 주관식 문제를 채점하기 위한 보조 시스템으로 시작을 하였다. 이 논문에서는 채점자에게 객관적인 주관식 평가를 하기 위한 자료를 제시하여 평가의 일관성을 유지하도록 시도하였다. 그러나 평가는 채점자가 직접 평가를 한다. [16]도 학습 평가를 위해 자동화 시스템을 설계, 구축한 논문이다. 이 연구는 평가를 위해 필요한 자료와 문제 제시, 출제, 답안작성, 평가 등의 자동화를 시도하였다. 역시 이 논문도 평가를 위한 지원 환경의 자동화에 초점을 맞추었다. [2]는 주관식 채점의 방법을 제안한 연구이다. 이 연구는 시소러스와 벡터 유사도를 이용하여 채점을 시도하였다. 그러나 시소러스가 충분한 규모의 일반화되지 못한 것인 단점이 있다. 또한 이 채점 시스템의 상세한 내용을 모르는 일반 사용자가 이를 사용하려면 시스템의 사용법을 자세히 배워야 원하는 결과를 얻을 수 있다.

본 논문은 유의어 사전을 이용하여 어휘만의

비교가 아닌 의미의 비교를 할 수 있는 주관식 채점 방법을 제안하고 일반 사용자도 이를 쉽게 사용할 수 있도록 편리성을 제공하는 채점 시스템을 설계, 구현한다.

### 3. 주관식 문제 채점 시스템

주관식 문제 채점 시스템의 전체 구조도는 그림 2와 같다.



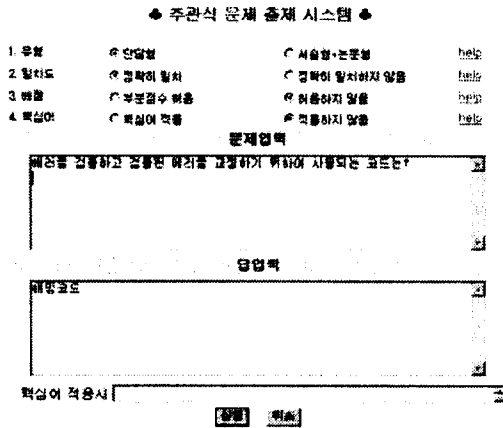
(그림 2) 시스템 전체 구조도

채점을 위해서 제일 먼저 해야 하는 작업이 출제자가 문제를 출제하고 정답을 입력하는 일이다. 정답을 입력하기 위해서는 시스템의 입력 절차를 따른다. 이때 문제의 유형이 결정된다. 정답의 내용은 형태소 해석하여 키워드 집합과 동의어 집합으로 저장된다.

다음으로 학습자는 출제자가 제출한 문제를 확인하여 답안을 작성한다. 그 결과를 형태소 해석하여 키워드 집합으로 보관한다. 다음 불리언 검색기법을 사용하여 학생 답안의 키워드와 정답의 키워드 및 동의어 집합과 비교를 한다. 그리고 채점 알고리즘에 따라 문제유형, 일치도 여부, 핵심어 적용여부, 유사도의 기준치 초과 여부를 고려하여 최종 점수가 계산된다.

#### 3.1 시스템 문제 및 정답 입력 인터페이스

본 논문의 시스템은 사용자가 쉽게 사용할 수 있도록 인터페이스를 설계하였다. 즉, 출제자가 시스템에서 정의한 주관식 문제 유형과 선택사항을 알지 못하더라도 옵션을 선택할 수 있도록 인터페이스를 만들었다. 시스템의 입력 인터페이스의 화면은 그림 3와 같다.



(그림 3) 문제 및 정답 입력 인터페이스

### 3.2 키워드 추출

답안에 포함되어 있는 키워드를 추출하기 위하여 먼저 형태소 해석을 한다. 본 연구의 형태소 해석기는 공개되어 있는 [11]를 이용하였다. 학생 답안의 경우 키워드만 추출하고 출제자의 정답에 대해서는 유의어 사전을 이용하여 동의어도 추출하여 불린 매칭을 위한 데이터를 준비한다.

### 3.3 유의어 사전

한국어에서는 WordNet[7] 같은 대규모의 범용 시소러스 사전이 없다. 따라서 본 논문에서 사용하기에 적합한 유의어 사전을 구축한다. 유의어 사전의 성능은 시스템의 질을 좌우한다. 따라서 유의어 사전의 정의와 구축이 일관성이 있어야 하고 객관성이 있어야 한다. 소량의 데이터에 대한 유의어 사전을 구축하더라도 사전 정의와 구축 과정이 일관성과 객관성을 유지할 수 있다면 대량의 사전 데이터로의 확장에 대한 정당성을 뒷받침할 수 있다. 본 논문에서는 신뢰성을 얻을 수 있도록 유의어 사전 구축의 알고리즘을 정의하고 이 과정에 따라 사전의 데이터를 구성하였다.

유의어 사전 구축의 알고리즘은 Bible 컴퓨터 용어 대사전, Terms 온라인 컴퓨터 용어사전, 네이버 국어사전, 야후 영한/한영사전과 [12]를 기초로 다음 휴리스틱 규칙으로 구성되었다. 사전의 단어는 먼저 대상 시험지의 답안에 들어있는 단어로 정의한다. 차후 단어를 확장할 예정이다.

- (1) 각 사전에서 키워드를 설명하는 내용에 들어 있는 유의어(유의어 태그 표시가 된 것)를 추출한다.
- (2) 각 사전에서 키워드를 설명하는 내용에 들어 있는 “~ 하는 것” 유형에서의 ~에 해당하는 단어를 추출한다.(예: 통신 - 소식이나 정보를 교환하는 일 => 교환)
- (3) 각 사전에서 키워드를 설명하는 내용에 들어 있는 “~의 총칭” 유형에서의 ~에 해당하는 단어를 추출한다.(예 : 음란 - 범죄의 총칭 => 범죄)
- (4) 사전에서 키워드를 설명하는 내용에 들어 있는 마지막 단어를 추출한다.(예 : 신호 - 소리·색깔·빛 따위의 부호에 의해 의사를 전하는 일, 또는 그 부호 => 부호)
- (5) 사전에서 키워드를 설명하는 내용에 들어 있는 “~를 말함” 유형에서의 ~에 해당하는 단어를 추출한다.(예 : 버스 - 데이가 이동하는 통로를 말함 => 통로)
- (6) “~의 준말” 유형에서의 ~를 추출한다. (예 : 단말 : 단말기의 준말 => 단말기)(4) 영어의 표기의 경우 영한 사전의 엔트리를 추출한다.(예 : star - 스타형, 별형)
- (7) 영어 표기를 우리말로 쓴 것에 대해 한영 사전의 엔트리를 추출한다.(예 : 링형 - ring)

### 3.4 주관식 문제 채점방법 조사

먼저 채점 방법을 정하기 위하여 현장에서의 채점 과정을 살펴보았다. 「데이터통신」 과목을 담당하는 현직 교사 5명을 대상으로 중간고사 평가를 검토하여 주관식 14문항에 대해 고등학교 2학년 학생들에게 시험을 치르게 하였다. 그리고 5명의 교사의 채점 결과를 살펴보았다.

수집한 데이터통신 중간고사 문항을 살펴보면, 객관식에 대한 주관식 문항의 비율이 20~30%로 현저히 적었다. 주관식 문항 중에서도 단답형에 편향되어 있어 단답형·완성형문항이 90%를 차지하고 있다. 이는 앞에서 말한 주관식 평가의 어려움이 있기 때문이라고 볼 수 있다. 주관식 문항의 문제점을 검토하여 다음 사실을 알아내었다.

첫째, 문제에서 정답의 유형의 설명에 대한 언급이 부족하여 채점의 일관성을 유지하기 어렵다. 예를 들어 정답이 “VAN”일 경우 영어로 답할 것인지 “부가가치통신망”이라는 용어를 요구할 것인지에 대한 것이 불분명하여 교사마다 채점이 다르게 나타났다.

둘째, 배점이 있는 경우와 없는 경우의 구분이 모호하여 채점의 일관성을 유지하기 어려웠다. 배점을 정확히 명시하면 채점의 명확성을 향상시킬 수 있게 된다.

셋째, 문항의 유형이 제한되어 있다. 대부분이 출제하기 쉽고 채점하기 쉬운 단답형으로 국한되어 있기 때문에 학습자들에게 다양한 형태의 주관식 평가를 할 수 없었다.

넷째, 주관식의 경우, 개념적인 질문의 경우 채점의 결과가 채점자에 따라 다르게 나타났다. 예를 들어 “데이터 통신에 대한 개념”을 묻는 문항에서 A학생의 답안을 보고 5명의 채점자가 각각 5, 3, 0, 0, 0점을 주었다. 똑같은 답안에 대해 채점자에 따라 채점의 차이를 보이게 되었다.

다섯째, 복수 개의 답을 요구할 경우 일치하는 개수에 따라 다른 배점이 나타났다. 예를 들어 “3개이상 나열하십시오”라는 문항에 대해 정답과 비교해서 2개가 일치하는 경우 정답으로 처리한 경우도 있고, 부분점수인 5점으로 채점한 경우, 0점으로 채점한 경우 등 채점자에 따라 다양하게 나타났고 부분 배점의 점수도 다양하게 나타났다. 예를 들어 배점이 5점인 경우 0점~5점까지 다양하게 나타났다.

여섯째, 점수 계산의 오류이다. 배점을 잘못 준 경우 뿐 아니라 각각 문항의 덧셈이 잘못되어 총점수가 높거나 낮게 나타남을 볼 수 있었다.

이외에도 채점시간이 많이 걸리고, 학생의 인상이나 글씨체 등이 점수에 영향을 미치는 위험성이 있었다.

이러한 문제점들은 정확한 답을 요구하는 단답형이나 완성형의 경우는 비교적 채점 결과가 동일하였으나 길게 답을 요구하는 서술형이나 논문형의 경우는 아주 다르게 나타났다. 이 사실은 일관성 있고 객관성 있는 주관식 채점 시스템이 필요함을 뜻한다.

### 3.5 문제 채점 알고리즘

본 논문의 문제 채점 알고리즘은 검토한 문제 점을 고려하여 다음과 같이 구성한다.

#### (1) 문제 유형 결정

첫 번째로 고려할 내용이 문제 유형 결정이다. 문제 유형은 단답형과 서술형 또는 논문형의

두 가지 가운데 하나를 선택한다. 앞에서 언급하였듯이 본 논문에서는 서술형과 논문형을 학습자의 의견을 기술한다는 측면에서 같다고 간주하였다. 이 단계는 단지 채점 유형을 선택하는 단계이다.

- ① 단답형 : 간단한 단어, 숫자등 제한된 형태로 대답하게 하는 형식
- ② 서술형+논문형 : 단답형보다 답의 길이가 길고, 의견을 직접 서술하도록 하는 형식

#### (2) 정답과의 일치 여부

채점자가 요구하는 답과 학습자가 기재한 답이 정확히 일치해야 하는 경우와 정확히 일치하지 않아도 되는 경우를 선택한다.

- ① 정확히 일치하는 경우 : 교사가 입력하는 답과 학생의 답이 정확히 일치해야 정답으로 처리한다.
- ② 정확히 일치하지 않아도 되는 경우 : 교사가 입력하는 정답과 학생에게 요구되는 답이 비록 정확히 일치하지 않아도 어느 정도 유사하거나 의미가 같다고 판단되면 정답으로 간주하는 것이다. 주로 서술형과 논문형에 해당된다. 이 경우 제한한 주관식 문제 채점 알고리즘을 따른다.

정답과의 일치 여부는 단순히 어휘간의 일대일 매칭여부를 검사하지 않고 유의어 사전을 이용하여 동의어와의 매칭여부를 검사한다. 따라서 어휘가 다를지라도 의미가 같은 경우를 찾아내어 채점의 정확성을 올린다. 정확히 일치하지 않는 경우는 유사도에 근거하여 채점을 하게 된다. 유사도에 대한 것은 다음 절에서 설명한다.

#### (3) 부분 점수의 적용여부

다음은 부분 점수를 적용하느냐 하지 않느냐를 선택한다.

- ① 부분 점수를 적용하는 경우 : 교사가 입력한 정답과 학생이 입력한 정답이 기준치 이상 일치하면 부분점수를 부여한다.
- ② 부분 점수를 적용하지 않는 경우 : 교사가 입력한 정답과 학생이 입력한 답이 기준치 이상 일치해도 배점을 적용하지 않기 때문에 오답으로 처리한다.

부분 점수의 적용여부는 배점에 있어서 이진 점수가 아닌 중간 점수 부여 여부를 말한다. 따라서 채점의 확실성을 탈피하는데 목적을 둔다.

#### (4) 핵심어 사용 여부

채점시 핵심어의 사용 유무에 따라 채점을 달

리하는 기준이다. 핵심어에 대한 것은 평가에서 우선권을 부여하는 방법이다. 지정된 핵심어가 포함될 경우 우선 배점한다.

- ① 핵심어를 적용하는 경우 : 답이 많이 길어지더라도 중요 핵심어가 포함되거나, 여러 개의 답 중에서도 특정 핵심어가 든 답을 원할 때 이 조건을 선택한다. 즉 핵심어만 포함되면 정답으로 처리하는 방법이다.
- ② 핵심어를 적용하지 않는 경우 : 핵심어를 적용하지 않을 때는 정답과 같아야만 배점을 하는 방식이다.

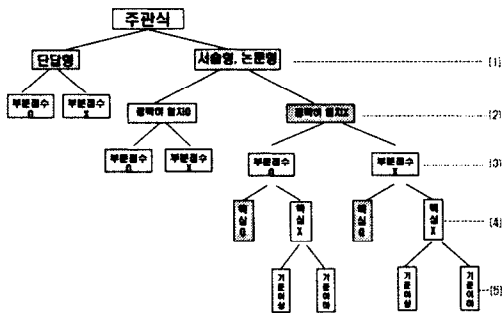
위의 배점 알고리즘을 트리 모양의 형태로 나타내면 그림 4와 같다. 이 트리는 정의한 기준에 의해 그림과 같이 나누어진다. 우측의 번호는 그 기준에 해당한다. 첫 번째의 분기는 단답형 여부에 따라 발생하고 단답형의 경우 부분 점수 부여

하고 그 일치 정도가 기준치를 넘으면 점수를 부여한다. 그 일치 정도는 유사도 계산식에 따른다.

- ① 기준치 이상 : 채점 유사도를 계산하여 기준치 이상이면 정답과 유사하다고 판단되어 정답으로 간주한다.
- ② 기준치 이하 : 유사도 계산했을 때 기준치 이하면 정답과 거리가 멀다고 판단하여 오답으로 간주한다.

본 논문에서는 유사도 계산식을 정보검색 분야에서 많이 사용되고 있는 벡터공간모델의 코사인 함수를 사용하지 않고 불린 모델을 사용한다. 벡터공간 모델의 가중치는 일반적으로 빈도수를 이용한다. 주관식 답안의 성격상 서술형의 답일지라도 그 길이가 그렇게 길지 않아 사용된 어휘가 그렇게 많지 않다. 따라서 어휘가 일반화될 정도의 충분한 양이 아닌 주관식 답안의 채점에 불린 모델을 사용하였다.

유사도 계산식은 데이터 자료를 근거로 하여 휴리스틱한 식으로 정의되었다. 학습 데이터의 양이 적어 일반화된 코사인 함수 등의 식보다 휴리스틱한 식을 선택하였다. 유사도 식을 얻기 위해 문제의 정답과 학생정답의 키워드를 추출하였다. 표 1에서 정답 색인수 8/3/11은 정답과 일치하는 수가 8개, 일치하지 않는 색인 수가 3개, 전체 색인어 수가 11개임을 말한다. 여기에서의 색인수는 유의어 사전을 통해 동의어 색인어까지 포함한 것이다. 일치도는 정답과 일치하는 비를 말한다. 나 학생과 다 학생을 보면 정답과 일치하는 수가 둘 다 7로 같다. 그러나 실제 채점에서는 각각 10점과 5점으로 채점되었다. 이 사실은 정답과의 일치 여부 이외에 일치하지 않는 비율을 고려해주어야 한다는 것을 의미한다. 표에 나타나듯이 1/3과 1/4이 실제 채점과 가장 유사한 것을 볼 수 있다. 1/3은 실제 시스템에서 사용해 보니, 일치하지 않는 수의 비율을 많이 차지해 1/4를 채택하였다. 그림 4.3에서 이를 사용하고 표 4.1에서 0.7이상이면 정답, 0.5이상이면 부분 점수, 그 이하는 오답으로 처리한다고 볼 수 있다.



(그림 4) 채점 기준 트리

여부에 따라 다시 분기된다. 서술형 및 논문형의 경우 정답과의 일치 여부에 따라 분기되고 다음으로 부분 점수 부여 여부에 따라 나누어진다. 네 번째 기준인 핵심어 적용 여부에서 핵심어 적용할 경우는 핵심어 이용하여 채점하고 그렇지 않을 경우 다시 기준치 적용을 한다. 기준치 적용은 다음 절에서 정의한 유사도 계산식에 의해 결정된다.

### 3.6 휴리스틱 유사도 계산식

본 논문의 문제 채점 알고리즘은 기본적으로 유의어 사전을 이용하여 이루어진다. 먼저 학생의 답안의 어휘를 추출하고 유의어 사전을 이용하여 어휘의 동의어를 구한다. 그리고 동의어 집합과의 매칭여부를 검사하여 판단을 하게 된다. 그렇지만 서술형의 주관식 답안은 정답과 정확히 일치하지는 않아도 그 뜻이 일치하면 되기 때문에 정답과의 동의어 집합과의 일치 여부를 검사

<표 1> 채점 유사도 점수 계산을 위한 식

계산 학생	정답수	일치도	1/2	1/3	1/4	1/5	채점 결과
가	8/3/11	1	0.86	0.91	0.93	0.95	○
나	7/7/14	0.88	0.63	0.71	0.75	0.78	○
다	7/20/27	0.88	0.50	0.63	0.69	0.73	△
라	2/14/16	0.25	-	-	0.03	0.08	×
마	4/11/15	0.50	0.13	0.26	0.32	0.35	×

이 표에서 채점 교사 채점 결과랑 가장 유사한 1/4을 적용하여 휴리스틱한 채점 유사도 식을 도출하였다. 이를 일반화하기 위해 대량의 학습 데이터에 대해 조정할 필요가 있다고 생각된다.

$$M = \frac{S}{T} - \frac{1}{4} \left( \frac{NS}{S+NS} \right)$$

T : Teacherkey 색인어 수

S : Studentkey의 정답과 일치하는 수

NS : Studentkey의 정답과 일치하지 않는 수

(그림 5) 휴리스틱 채점 유사도 식

#### 4. 주관식 문제 채점 시스템 실험 결과

시스템은 주로 C언어로 구현하였으며 자세한 사항은 표 2와 같다.

<표 2> 개발 환경

구분	사항
운영체제	리눅스
개발언어	C, PHP
편집 도구	vi편집기, Edit plus v2.11

본 논문의 주관식 시스템 채점을 실험하기 위해 주관식 문제 연구에 사용했던 14문항의 주관식 문항을 실제로 채점 시스템에 적용하였다. (그림 6)과 같이 5명의 교사가 직접 채점한 점수 중 최상과 최하 점수를 제외한 3명의 평균 점수와 시스템 채점을 비교하였다. 평균 오차는 3.9점이며, 최고 오차는 10점이며 정확히 일치하는 채점 결과도 볼 수 있었다. 특히 고무적인 것은 시스템의 채점 결과가 3명 교사의 평가 최저와 최고 사이에 있는 경우가 11/15(73%)가 되고 표에 나

와 있지 않은 최저 최고점수의 교사까지 포함하여 조사해보니 시스템의 채점이 5명 교사의 평가 최저와 최고 사이에 있는 경우가 15/15(100%)가 되었다. 이는 이 시스템이 교사의 채점만큼 신뢰성 있음을 설명한다.

학생번호 (일의)	A (연동)	B (역성)	E (영주)	평균	시스템 채점	오차 점수
1	50	55	58	54	53	1
2	45	50	47	47	40	7
3	25	35	41	34	28	6
4	65	60	77	67	60	7
5	50	45	67	54	48	6
6	55	60	64	60	65	5
7	45	40	53	46	36	10
8	45	45	55	48	43	5
9	50	65	64	60	58	2
10	45	60	63	56	58	2
11	50	65	69	61	61	0
12	30	20	28	26	30	4
13	30	25	28	28	28	0
14	40	50	51	47	48	1
15	45	45	40	43	40	3
평균점수	44.7	48.0	53.7	48.8	46.4	3.9

(그림 6) 채점 결과 비교

(그림 6)에서 우리는 결과를 더 분석하였다. 위 분석에서는 절대 점수 부여의 차원에서 검사를 한 것이다. 이와 함께 상대 점수 부여의 차원에서 검사를 하였다. 즉, 점수별로 등수를 매기고 등수의 결과가 얼마나 차이가 나는지를 분석하였다. 채점자 평균을 기준으로 하여 각 채점자의 결과와 시스템의 결과를 검토하였다.

학생번호	A	B	C	평균	시스템
1	3	6	7	6	6
2	7	7	11	9	10
3	15	13	12	13	14
4	1	3	1	1	3
5	3	9	3	7	7
6	2	3	4	3	1
7	7	12	9	11	12
8	7	9	8	8	9
9	3	1	4	3	4
10	7	3	6	5	4
11	3	1	2	2	2
12	13	15	14	15	13
13	13	14	14	14	14
14	12	7	10	10	7
15	7	9	13	12	10
오차1	0.4(6/15)	0.53(8/15)	0.8(12/15)		0.6(9/15)
오차2	0.73(11/15)	0.87(13/15)	0.93(14/15)		0.93(14/15)
오차3	0.8(12/15)	0.93(14/15)	0.93(14/15)		1.0(15/15)

(그림 7) 채점 결과의 상대평가표

(그림 7)은 (그림 6)의 결과를 상대평가를 한 것이다. 표의 각 항의 내용은 등수이다. 아래 오차1, 오차2, 오차3의 행은 기준인 평균점수의 등수와 오차가 1이내의 것, 2이내의 것, 3이내의 것



에 대한 성공률을 나타낸다. 표를 보면 1번 학생의 경우 A 교사의 경우 3등으로 부여되었고 B 교사의 경우 6등, C 교사의 경우 7등으로 부여받았다. 평균 점수로 6등을 부여받았고 시스템도 6등으로 부여하였다. 오차 2 이내의 경우 1번 학생은 B, C 교사와 시스템은 성공하였다고 볼 수 있다. 결과를 보면 시스템은 오차 1 이내의 경우 60%, 오차 2 이내의 경우 93%, 오차 3 이내의 경우는 100%의 성공률을 보인다. 이에 비해 교사의 채점은 시스템의 성공률보다 낮다. 이 사실은 교사의 채점보다 본 논문의 채점 시스템이 객관성이 있음을 보여주고 있다.

(그림 8)은 문항별 채점 결과 비교이다. 15명의 학생을 임의의 번호별로 오차를 비교한 것으로 9점, 8점등의 약간의 오차가 발생하였지만 오차 없이 정확히 일치하는 채점결과를 볼 수 있다. 특히 3번, 6번은 평균 오차가 0.1로 거의 유사하게 채점되었고, 1번, 8.2번, 10번은 1.5이상의 평균 오차로 비교적 높게 나타남을 볼 수 있다.

문제 번호	백점	교사채점과 시스템 채점의 오차 비교															오차 합계	평균 오차	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15			
1	5	0	0	0	2	3	1	0	1	0	3	0	5	4	1	0	3	23	1.5
2.1	5	0	0	2	0	1	4	2	2	0	0	0	0	0	0	0	11	0.7	
2.2	5	0	0	0	0	1	0	0	1	0	0	1	0	0	0	1	4	0.3	
3	10	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	0.1
4	10	1	0	4	1	1	4	0	0	0	0	8	0	0	0	0	19	1.3	
5	5	0	0	0	0	1	1	0	0	0	0	0	1	0	1	1	5	0.3	
6	5	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.1	
7	5	0	0	0	0	2	2	2	2	1	2	1	0	2	2	0	16	1.1	
8.1	10	0	0	0	0	0	0	0	0	1	0	0	5	0	0	0	7	0.5	
8.2	10	0	0	0	3	2	1	2	3	1	2	9	0	1	3	0	27	1.8	
8.3	10	2	0	4	0	0	0	0	0	2	4	3	0	0	0	2	17	1.1	
9.1	5	0	0	0	1	4	1	0	0	0	0	2	0	0	0	0	8	0.5	
9.2	5	0	0	1	0	4	1	0	0	1	0	2	0	0	1	0	10	0.7	
10	10	1	1	5	4	2	1	6	1	1	1	5	4	0	1	4	37	2.5	
총계	(10)																		

(그림 8) 문항별 채점 결과 비교

이 채점 시스템을 사용함으로써 문제 유형과 배점 등의 선택만으로 채점이 결정되므로 사용하기가 쉽다. 채점 면에서는 애매한 답이나 각기 다른 배점은 모든 학생에게 동일하게 적용된 채점 결과를 얻을 수 있었고, 점수 계산 오류나 주관성이 배제된 채점이 가능하게 되었다.

이 시스템의 결과를 검토하여 다음과 같은 개선점이 있음을 알 수 있었다.

- 문제 2번, 3번, 6번 문항과 같이 답이 명확한 유형은 시스템과 교사의 채점이 오류가 거의

- 없고 비교적 정확하게 일치하였다.
- 1번, 4번, 8.2번, 10번 문항은 서술형의 유형으로 약간의 오차 발생했다. 키워드와는 일치하지 않지만 문맥으로 보아 일치하는 경우가 있기 때문이다.
- 유의어 사전에 등록되지 않은 유의어로 정답을 할 경우에는 약간의 오차가 발생하였다. 대량의 자동화된 시소러스 사전이 필요하다.
- 학생의 경우 오타나 띄어쓰기를 문법에 맞게 입력해야 좋은 점수를 얻을 수 있다. 왜냐하면 키워드 추출을 형태소 해석기가 하기 때문에 문법에 맞지 않는 경우는 오타나 잘못된 단어를 키워드로 하기 때문이다.

문제점을 분석해 본 결과 자동화된 시소러스 개발이 필요하다고 생각된다. 또한 보다 정확한 채점을 위해서는 키워드 중심의 표현이 아닌 구조 또는 의미구조 중심의 표현법이 개발되어야 할 것으로 사료된다. 그리고 본 논문의 주관식 채점 모델을 보다 일반화된 주관식 문항을 채점할 수 있는 주관식 채점 모델로 개선해야 할 것이다.

### 5. 결론 및 향후 과제

주관식 평가는 교사들이 시간과 노력이 많이 들고 객관성을 잃는다는 이유로 기피하고 있다. 본 논문은 이와 같은 문제점을 해결하고자 주관식 문제 채점 시스템을 설계, 구현하였다. 본 논문의 시스템은 시스템 사용법을 모르는 초보자들에게도 편리하게 채점을 할 수 있도록 편리한 인터페이스를 구비하였다. 이 시스템을 사용을 통해 채점의 시간과 노력을 덜고 객관성을 얻어 채점의 공정성을 기할 수 있다. 이 시스템의 개발에 사용된 노하우는 향후 주관식 문제 채점 연구의 기반이 될 것으로 기대된다.

앞으로의 과제는 보다 정확한 채점을 위해 주관식 문제 채점의 선택 사양을 보다 세분화하고 다양화하여야 할 필요가 있다. 그리고 대규모의 시소러스 구축, 구문분석과 의미분석의 활용 등의 고급 자연어 처리 기술의 활용이 필요하다.

### 참고문헌

[1] 포항공대 대학교육 개발 센터, 학습 평가, 교수학습 자료.  
 [2] 정동경, 백터유사도와 시소러스를 이용한 주

관식 답안의 채점방법, 동국대학교 교육대학원 석사논문, 2001.

- [3] 황도삼, 최기선, 김태석 공역, 자연언어 처리, 홍릉과학출판사, 1998.
- [4] J. Allen, Natural Language Understanding, Benjamin/Cummings Publishing Co., 1987.
- [5] G. Salton, Automatic Text Processing, Addison-Wesley, 1989.
- [6] 김명철, 김덕봉, 김유성, 김재훈, 박혁로, 이하규 공역, 최신정보검색론, 홍릉출판사, 1999.
- [7] G. A. Miller, R. Beckwith, C. Fellbaum, D.Gross, and K. Miller, "Introduction to WordNet:An On-line Lexical Database," Report of WordNet, Princeton University, 1990.
- [8] 권대훈, 교육평가, 안동대학교 교수학습자료, 2001.
- [9] 한국교육과정평가원, 고등학교 사회과 수행평가의 이론과 실제, 한국교육과정평가원, 1999.
- [10] 김옥환, 김정규, 교육평가, 형설출판사, 1980
- [11] 강승식, 범용 형태소 분석기 "HAM Ver 6.0.0", 국민대학교 자연언어 정보검색 연구실, <http://nlp.kookmin.ac.kr/>
- [12] 조평옥, 한국어 명사의 의미계층 구조 구축, 울산대학교 교육대학원 석사논문, 1996.
- [13] 김은미, 추교흠, 심효지, 김창수, "클라이언트/서버환경에서 수준별 학습을 위한 학생평가 시스템 구현 연구," 1998년도 정보처리학회 춘계학술대회 논문집, 5권 1호, 1998.
- [14] 이종희, 김태석, 이근왕, 오해석, "학습자 모니터링을 이용한 학습 평가 시스템," 2002년도 정보처리학회 춘계학술대회 논문집, 9권 2호, 2002.
- [15] 방훈, 강태호, 김원진, 원대회, 이재영, "웹기반 주관식 답분류 채점시스템," 2001년도 한국정보과학회 춘계학술대회 논문집, 28권 2호, 2001.
- [16] 정용기, 최은만, "웹기반 학습평가 자동화시스템의 설계 및 구현," 한국정보처리학회 논문지 D, 9-D권 2호, 2002.



### 박희정

2001 안동대학교 컴퓨터  
공학과 (공학학사)  
2003 안동대학교 컴퓨터  
교육과(교육학석사)

관심분야: 컴퓨터교육, WBI  
E-Mail: happy@andong.ac.kr



### 강원석

1985 경북대학교 전자공학과  
(공학학사)  
1988 한국과학기술원  
전산학과 (공학석사)  
1995 한국과학기술원  
전산학과 (공학박사)

1995 - 현재 안동대학교 컴퓨터교육과 교수  
관심분야: 컴퓨터교육, 자연어처리, 정보검색  
E-Mail: wskang@andong.ac.kr