

한국어 서답형 자동채점을 위한 정답 템플릿 기술 방법

박일남[○], 노은희⁺, 심재호⁺, 김명화⁺, 강승식

국민대학교 컴퓨터공학부

pin0156@naver.com, noro@kice.re.kr, sim307@kice.re.kr, hwa@kice.re.kr, sskang@kookmin.ac.kr

Answer Template Description for Automatic Scoring of Korean Free-text or Constructed Answers^{*)}

IL-Nam Park[○], Eun-Hee Noh⁺, Jae-Ho Sim⁺, Myung-Hwa Kim⁺, Seung-Shik Kang
Kookmin University, Korea Institute for Curriculum and Evaluation⁺

요 약

한국어 서답형 문항의 자동채점 프로그램을 개발하기 위해서는 모범답안, 오답, 부분점수 부여를 위한 세부적인 내용을 채점 기준으로 기술해야 한다. 자동채점에 필요한 구체적인 사항들을 기술하기 위하여 XML 형식으로 정답 템플릿을 정의하였다. 채점에 필요한 내용을 단위 개념으로 정의하고 이를 컴퓨터가 액세스 가능한 형태의 정답 템플릿을 설계하였다. 정답 템플릿 형식에 맞게 편리하게 템플릿을 작성할 수 있는 작성 도구를 이용하여 학업 성취도평가 각 문항에 대한 채점 기준표를 정답 템플릿으로 작성하여 채점기준표를 작성하는 실험을 수행하였다.

주제어: 정답 템플릿, 자동채점 시스템, 채점 기준표, 서답형 문항, 문서 유사도

1. 서론

2011년에 60만 고등학생 2학년 학생들의 학업성취도 국어, 수학, 영어 과목을 평가함에 있어 7월 24일부터 8월 24일까지 한 달여 동안 채점자 3,988명과 122명의 관리자가 동원되었으며, 총 15억 7천만 원 가량의 비용을 소비하여 학생들의 답안을 채점하였다. 채점을 하기 위해 채점자들에 대하여 훈련을 시키고 실제 채점에 투입함에도 불구하고 인간채점자로서 사람마다 주관적인 요소가 반영이 되어 모든 학생의 답안에 대하여 공평하게 채점을 하지 못한다는 단점이 있다.[1] 이를 위해 컴퓨터를 이용한 한국어 서답형 자동채점에 대한 방법들이 고안되었으나 기존의 연구들은 대부분 정보검색에서 활용되고 있는 문서 유사도 기법을 적용하여 구 수준의 답안에 대해서는 채점 결과가 정확하지 못하다.

본 논문은 실제 사람이 채점 기준표를 보고 수작업으로 채점하는 환경을 컴퓨터를 이용한 자동채점 환경으로 이식시키기 위한 방법을 모색하고자 한 것이다. 권오용[2]은 모범 답안을 간단한 논리식 형태로 기술하여 학생 답안이 기술한 논리식을 따르는지 검사를 하는 방법을 제시하였는데, 이는 채점 기준표의 내용을 논리식으로 기술하여 사용하면 채점이 가능하나 정형화된 문장 형태의 문항에만 적용이 된다는 제약 사항이 있어 여러 유형의 서답형 문항에 대해서는 적용이 불가능하다. 영어 자동채점의 경우 개념(Concept)이라는 것을 도입하여 채점 기준을 반자동으로 구축하고, 개념A가 주어졌을 때, 학생 답 B에 대해서, B가 A인지 추론하여 자동채점을 하는

C-rater[3,4]가 있다. 본 논문은 한국어 서답형 자동채점을 위하여 컴퓨터가 채점 기준표를 이해할 수 있도록 정답 템플릿을 설계 및 정의하였다.

2. 한국어 서답형 답안의 유형 분석

서답형 답안 분석 자료는 채점 기준표에 따라 수작업으로 직접 채점한 2011년 국가수준 학업성취도 평가 데이터이다. 그림 1, 2는 각 문항별 1,000개의 학생 답안에 대한 답안 유형수를 나타내며 초등학교 6학년 국어/과학, 중학교 3학년 국어/과학, 고등학교 2학년 국어이다. 각 문항별 학생 답안 유형수를 살펴보면 유형수 500 이상의 문제들은 학생들에게 ‘~에 대한 이유를 쓰시오’와 같은 학생들의 개인적인 생각을 작성하는 문항이며, 유형수 100 이하의 답안들은 기호, 숫자, 특정 용어 등 1단어로 이루어진 단답형을 요구한다. 유형수 100~500 사이의 문항들은 대부분 1어절~3어절 정도를 요구한다.

성취도 평가 국어, 과학 문항들의 문항 유형수와 문항 채점 요구사항, 답안의 길이 등을 분석하여 패턴들을 정의하면 단순 스트링 매치를 요구하는 문항으로 형태소 분석이 필요 없는 단어 1개 수준의 답안 또는 모범 답안과 완벽히 일치해야 정답으로 간주하는 채점 유형이 있다. 다른 패턴으로는 형태소 분석을 통하여 얻은 어휘 토큰만 고려하는 채점 유형을 위한 것이 있다. 그 예로 ‘속담을 인용하고 있다’와 같은 답안이 모범 답안이라 할 때, 형태소 분석 결과 중 어휘 토큰인 ‘속담’과 ‘인용’ 인 어휘토큰 개념만 맞으면 정답으로 하는 채점 방식에 사용할 수 있다.

반면에 ‘속담/을’과 ‘인용/하고’와 같이 ‘어휘/문법’ 토큰이 정확히 일치해야 정답으로 간주해야 하는 패턴과 1문장 이상의 답안에서 ‘속담을 인용하고 있

1) 이 논문은 한국교육과정평가원 2012년 기본과제 “대규모 평가에서의 서답형 자동채점 방안 연구”에서 프로그램 개발 의뢰를 받아 진행 중인 연구의 일환임.

다.’ 라는 개념이 등장하면 정답으로 인정하겠다는 경우를 위한 패턴이 있다. 본 논문은 서답형(단어, 구) 수준을 대상으로 하기 때문에 빈도수 500 이하의 문항들을 자동 채점하기 위한 정답 템플릿을 설계 및 정의하였다.

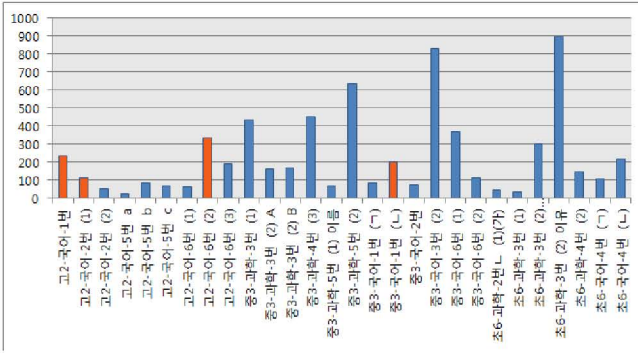


그림 1. 문항별 1000개 국어 학생 답안 유형수

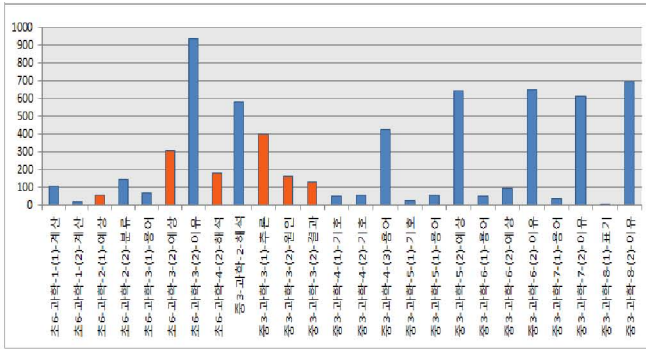


그림 2. 문항별 1000개 과학 학생 답안 유형수

3. 정답 템플릿 설계

기존의 한국어 서답형 자동채점 시스템들은 정보검색의 문서 유사도 기법을 사용하여 특정 단어들 이 출현 확률에 따라 점수를 부여하는 형식이었다. 따라서 표 1과 같이 채점 기준표에 대하여 원하는 점수 부여가 난해하다. 예를 들어, ‘속담’ 과 ‘인용’ 에 대하여 각각 한 개씩 단어들 이 출현하면 1점을, 동시에 두 개가 출현하면 2점을, 이렇게 점수를 부여하면 될 것이라고 생각될 것이다. 하지만 ‘속담, 인용’ 이라고 답한 학생의 점수는 1점으로 채점이 되어야 되는데 ‘속담’, ‘인용’ 이 있으므로 2점이 되게 된다. 또한, “인용을 속담하고 있다” 라는 학생 답안이 실제 존재했는데 이 학생 답안은 오답으로 채점해야 되나 2점으로 채점되는 큰 오류를 범하게 된다.

이러한 사항을 고려하여 채점 기준표를 컴퓨터가 이해할 수 있도록 정답 템플릿을 XML 기법으로 작성하였으며, 정답 템플릿에는 ‘개념(Concept)’ 을 정의하고 어떠한 개념이 출현하면 몇 점을 부여할 것인지 ‘점수 가이드(Scoring Guide)’ 를 기술할 수 있도록 설계하였다.

표 1. ‘고2-국어-1번’ 채점 기준표

모범 답안: 속담을 인용하고 있다.	
점수	정답 또는 정답 기준
2점	- 모범 답안과 같은 경우.
1점	- ‘속담’, ‘인용’중 한 가지만 제자리에 쓴 경우. - 문장으로 쓰지 않고, ‘속담’, ‘인용’만 순서대로 쓴 경우

3.1 정답 템플릿의 구조

개념 정의 부분은 하나의 문항에 대해 여러 개념들 (Concepts)을 정의할 수 있고 각 하나의 개념(Concept)은 단어, 구, 문장에 대하여 기술이 가능하도록 n개의 Word 라는 단위를 갖도록 하였다. Word는 1어절(또는 토큰)단위의 개념을 세부적으로 기술할 수 있도록 하였으며 Text태그를 사용하여 원문을 유지하도록 하였다.

원문 단어에 대하여 형태소 분석을 하여 얻어진 어휘/문법 토큰(LToken/GToken)을 각각 기입할 수 있도록 하였으며, 각 어휘 토큰과 문법 토큰은 대체 가능한 토큰 (SimLToken/GimLToken)을 기입할 수 있도록 하였다. 대체 시에 감점 또는 가산의 점수를 기입할 수 있도록 하였는데, 이는 ‘속담을 인용하고 있다’ 에서 ‘속담’ 이라는 어휘 토큰 대신에 ‘재담’ 이 오면 감점을 주겠다는 기술을 가능케 한다. 유사 단어(SimWord)는 ‘속담을’ 대신에 ‘외국어를’ 과 같이 대체 가능한 단어를 Word 단위로 기입할 수 있도록 하였고 마찬가지로 대체 시에 부여할 가중치 값을 정의할 수 있도록 하였다.

SimLToken, SimGToken, SimWord는 대체 가능한 토큰, word 단위에 ‘+’ 와 ‘*’ 기호를 사용하여 와일드카드 기능을 제공할 수 있도록 하였다. ‘속담을’ 의 SimGToken으로 ‘+’를 기입하고 -1점을 주면 ‘을’ 대신 다른 문법 토큰이 오면 1점을 감점하겠다는 것을 뜻하며 ‘속담으로’, ‘속담’ 에 대해서는 1점을 감점한다는 표현을 나타낸다. 반면에 ‘*’ 를 SimGToken으로 기입하면 ‘+’ 와 기능이 유사하지만 문법 토큰이 없는 것도 대체 가능하다는 표현을 나타내므로 ‘속담’ 이라는 어휘 토큰만 존재하는 Word는 감점 대상에서 제외가 된다.

그림 3은 ‘속담을 인용하고 있다’ 라는 3음절 답안에 대하여 하나의 개념을 정의한 예제이다. <word ord=“1”> 부분을 보면 SimWord에 와일드카드 문자를 사용하여 ‘속담을’ 위치에 다른 것들이 오면 -1점을 감점하는 정보가 기술되어 있다. 즉, ‘속담을 인용하고 있다’ 가 기본 개념이고 ‘__ 인용하고 있다.’ 라는 속담 이외의 단어들 이 있는 경우에 대해서는 -1점을 감점하는 것을 의미한다. <word ord=“3”>은 ‘있다’ 에 대하여 기술하였는데, SimGToken을 사용하여 ‘습니다’ 가 오면 0점의 가중치를 주겠다고 하여 3번째 어절은 ‘있다’ 와 ‘있습니다’ 가 올 수 있다고 기술한 것을 볼 수 있다.

정의된 개념이 학생 답안에서 출현하면 해당 답안에 대하여 몇 점을 부여할 것인지는 ‘점수 가이드’ 를 통하여 기술하면 된다. 점수 가이드는 그림 4와 같이 설계

```

1 <Concepts count="1">
2   <Concept ord="1" word_count="3">
3     <Word ord="1">
4       <Text>속담을</Text>
5       <LToken>속</LToken>
6       <GToken>을</GToken>
7       <SimLTokens count="0" />
8       <SimGTokens count="0" />
9       <SimWords count="1">
10        <SimWord ord="1" weight="-1">
11          <Text>+</Text>
12          <LToken>+</LToken>
13          <GToken>+</GToken>
14        </SimWord>
15      </SimWords>
16    </Word>
17    <Word ord="2">
18      <Text>인용하고</Text>
19      <LToken>인</LToken>
20      <GToken>고</GToken>
21      <SimLTokens count="0" />
22      <SimGTokens count="0" />
23      <SimWords count="0" />
24    </Word>
25    <Word ord="3">
26      <Text>있다</Text>
27      <LToken>있</LToken>
28      <GToken>다</GToken>
29      <SimLTokens count="0" />
30      <SimGTokens count="1">
31        <SimGToken ord="1" weight="0">습</SimGToken>
32      </SimGTokens>
33      <SimWords count="0" />
34    </Word>
35  </Concept>
36 </Concepts>

```

그림 3. 개념 정의 예제

```

<ScoringGuides count="3">
  <ScoringGuide ord="1" score="2">C1</ScoringGuide>
  <ScoringGuide ord="3" score="1">C2</ScoringGuide>
  <ScoringGuide ord="3" score="1">C3C4</ScoringGuide>
</ScoringGuides>

```

그림 4. 점수 가이드 예제

매우 커질 수 있다. 이러한 정답 템플릿을 채점 전문가가 일일이 모범 답안, 단서어, 개념 정의, 점수 가이드 부분을 작성하기에는 많은 시간이 소요될 뿐 아니라 각 태그들에 대하여 보기 좋게 들여쓰기 하는 것은 매우 어렵다. 정답 템플릿 작성의 편의를 위하여 그림 5와 같이 정답 템플릿 작성 도구를 개발하여 반자동으로 쉽고 빠르게 정답 템플릿 작성이 가능하도록 하였다. 왼쪽 상단에는 모범 답안 점수와 모범 답안 점수를, 오른쪽 상단에는 단서어를 기입할 수 있도록 구성하였다. 하단부에는 학생 답안에 대하여 관찰하고자 하는 개념을 정의하고, 해당 학생 답안에 개념들이 출현하면 부여할 점수를 기입할 수 있도록 구성하였다.

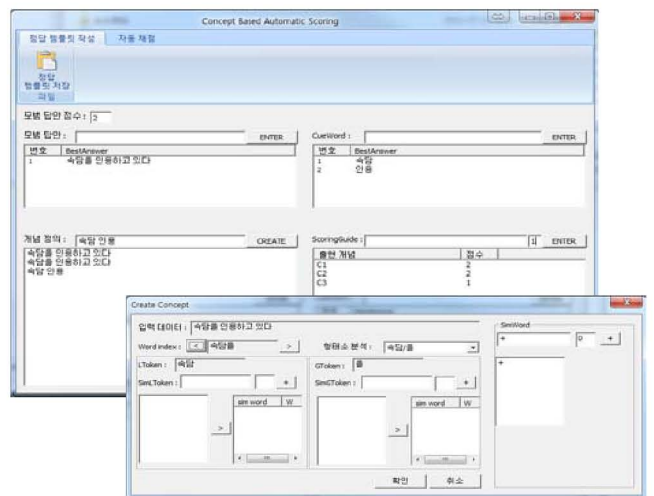


그림 5. 정답 템플릿 작성 도구

하였고 예제에서는 점수 가이드 방법이 3가지가 존재함을 나타낸다. <ScoringGuide ord="1">은 개념 1이 출현하면 그 답안은 2점을 부여한다는 것을 의미하고 ord="3"은 개념 3과 개념 4가 출현하면 1점을 부여한다는 것을 의미한다.

채점 기준표의 내용들을 컴퓨터가 이해하기 위한 최종적인 정답 템플릿 구성은 그림 5와 같다. ‘모범 답안’ 기입을 위하여 ModelAnswer 태그를 이용하였으며, 여러 개의 모범 답안을 기입할 수 있다. 또한 단서어 존재 여부에 따라 오답/미채점 답안 분류를 수행하기 위하여 CueWord 태그를 사용하여 단서어를 기입할 수 있도록 하였다. 단서어는 형태소 분석 결과 오류로 인하여 오답 처리 되는 답안에 대해 해당 단서어가 존재하면 오답 처리를 하지 않고 한번 더 확인해 볼 필요가 있다는 것이다. 그리고 ‘개념’에 대하여 정의할 수 있도록 하였으며 예제에서는 7개의 개념이 정의되어 있음을 알 수 있다. 마지막에는 출현한 개념에 대하여 부여할 점수를 기입하는 ‘점수 가이드’를 배치하였다.

3.2 정답 템플릿 기술 도구

다양한 문항에 대한 정답 템플릿을 작성해 보면 문제에서 요구하는 정답은 2-3 단어로 간단한 것처럼 보이지만 유사한 표현이 다양한 경우에 정답 템플릿의 크기가

개념 정의 부분은 문장 또는 구, 단어를 입력하면 앞서 템플릿 구성에서 설명한 Word 단위로 자동으로 나누어 작성되며 형태소 분석을 통해 LToken, GToken이 자동으로 기입되도록 하였다. 형태소 분석기는 국민대학교 KLT 라이브러리를 사용하였다.[5] 형태소 분석된 해당 토큰들에 대하여 SimLToken, GToken, SimWord를 작성할 수 있도록 작성 도움 창을 제공하여 사용자는 유사어휘/문법토큰 또는 Word 단위에 대하여 쉽게 작성할 수 있다. 그리고 load/save 기능을 제공하여 작성한 정답 템플릿을 XML 파일로 저장이 가능하고 이를 이용하여 자동채점을 수행한 뒤 정답 템플릿 내용에 대하여 수정 및 보완을 용이하게 하였다.

4. 결론 및 향후 연구

본 논문은 컴퓨터에게 채점 기준표를 인식시키기 위해 정답 템플릿을 정의하였으며, 개념 단위를 사용하여 채점에 필요한 정보를 기술하고 출현 개념에 따라 점수를 부여할 수 있도록 정답 템플릿을 설계하였다. XML 형식의 정답 템플릿에 따라 템플릿 작성 도구를 이용하여 학업 성취도평가 각 문항에 대한 채점 기준표를 정답 템플릿으로 작성하여 채점기준표를 작성하는 실험을 수행하였다. 본 연구에서 정의한 정답 템플릿과 템플릿 작성 도구는 실제로 서답형 자동채점 시스템에서 활용하는 실

험을 진행하고 있다.

참고문헌

- [1] 노은희, 심재호, 김명화, “대규모 평가를 위한 한국어 서답형 자동채점 프로그램 개발의 전망”, 한국교육과정평가원 연구자료 ORM, 2012-92, 2012.
- [2] 권오용, 웹 기반 주관식 평가문항 채점 알고리즘 설계 및 구현, 한서대학교 교육대학원 석사학위 논문, 2004.
- [3] C. Leacock and M. Chodorow, “C-rater: Automated Scoring of Short-answer Questions,” Computers in the Humanities, Vol. 37, pp.389-405, 2003.
- [4] J. Sukkariéh and J. Blackmore, “C-rater: Automatic Content Scoring for Short Constructed Response,” Proceedings of the Twenty-Second International FLAIRS Conference, pp.290-295, 2009.
- [5] 강승식, 한국어 형태소 분석과 정보 검색, 홍릉과학출판사, 2002.