

워드넷 기반의 임의 추출 분할 방식을 이용한

동적 문제 출제 시스템 설계*

추승우^o 오정석 김유섭 이재영

한림대학교 컴퓨터공학과

{teacher^o, bil78, yskim01, jylee}@hallym.ac.kr

A Design of Dynamic Question Generation System using a Voluntary Extraction and Division Methodbased on WordNet

SeungWoo Chu^o JungSeok Oh YuSeop Kim JaeYoung Lee
Dept. of Computer Engineering, Hallym University

요 약

문제 은행 방식을 사용하는 웹 기반 학습 시스템의 문제점으로 지적되었던 문제 유출에 따른 평가의 공정성 문제를 해결하고자 임의 추출 분할 방식을 이용한 동적 문제 출제 시스템이 제안되었다. 하지만 이 시스템 또한 문제 은행 방식을 사용하여 위의 문제를 해결하려고 하였다. 본 논문에서는 이러한 문제점을 해결하기 위하여 단어간의 관계를 계층적으로 표현한 어휘 데이터베이스인 한국어 워드넷을 활용한 방법을 적용하였다. 먼저 임의 추출 분할 방식으로 출제된 문제의 예제 문항을 형태소 분석기를 이용하여 명사들을 추출한다. 이 명사들을 이용하여 한국어 워드넷에서 해당 명사의 상위 개념 또는 동일 개념의 Synset을 추출한다. 이렇게 추출된 Synset으로 다른 예시 문항이지만 의미적으로 유사한 다양한 예제 문항을 생성하려는 시스템을 제안한다. 제안된 시스템의 사용으로 평가의 공정성 문제를 해결하고자 한다.

1. 서 론

최근 웹을 기반으로 하는 평가 방법들에 대한 연구가 활발히 진행 중이다[1]. 하지만 [1]처럼 대부분의 웹 기반 학습 시스템들은 출제자가 매번 새로운 문제를 출제하는 것이 아니라 문제은행에 저장되어 있는 문제를 검색하여 재사용할 수 있도록 하였다. 이런 환경은 학습자의 학습능력을 평가할 때, 같은 문제로 여러 학습자를 평가하면 문제가 유출되어 공정하고 정확한 능력을 평가하기 어렵다[2]. 이러한 문제점을 해결하기 위해서 임의 추출 분할 방식을 이용한 동적 문제 출제 시스템이 제안되었다[2].

임의 추출 분할 방식을 이용한 동적 문제 출제 시스템은 임의 추출 알고리즘을 사용하여 비슷한 난이도의 문제와 예제 문항을 다양하게 출제하도록 하는 것이다[2]. 하지만 이 시스템 또한 문제은행에 저장된 문제와 예제 문항을 임의 추출하여 재 정렬하는 방식으로 위의 문제점을 해결하려고 하였다.

이전 절에서 언급된 동적 문제 출제 시스템을 구현하기 위해서는 하나의 문제에 대하여 정답에 대한 예시와 오답에 대한 예시문이 풍부해야 한다. 객관식 검사의 단점으로 제작의 어려움과 제작시간이 많이 걸린다는 점이 있는데 하나의 문항에 대하여 많은 예시문을 작성하기 위해서는 많은 비용이 들게 된다. 본 논문에서는 이러한

문제점을 해결하기 위하여 임의 추출 방식으로 출제된 예제 문항을 다양하게 생성하여, 다른 예제 문항이지만 의미적으로 비슷한 새로운 예제 문항이 생성되도록 하는 시스템을 제안한다. 다양한 예제 문항을 새롭게 작성하기 위하여 단어간의 관계를 표현하는 어휘 데이터베이스인 한글 워드넷을 사용하였다.

본 논문에서 제안된 시스템은 임의 추출 분할 방식을 이용한 동적 문제 시스템에서 생성된 예제 문항 중에서 일부를 형태소 분석기를 이용하여 명사만을 추출한다. 이 명사들을 이용하여 한국어 워드넷에서 각 명사들의 상위 개념 또는 동일 개념의 Synset을 추출한다. 이 추출된 Synset으로 형태소 분석기를 통하여 분석된 예제 문항의 명사들의 자리에 삽입하여 재구성함으로써 의미적으로 유사한 여러 개의 예제 문항을 생성하려는 시스템을 제안한다.

2장에서는 기존의 임의 추출 분할 방식을 이용한 동적 문제 출제 시스템의 문제점에 대해 설명한다. 3장에서는 의미적으로 유사한 여러 개의 예제 문항을 생성하기 위해 쓰인 한국어 워드넷에 대한 설명과 이를 적용시킨 시스템의 구성을 상세하게 설명한다. 4장에서는 결론을 밝힌다.

2. 워드넷

워드넷은 어휘의 저장 형태가 어휘 의미를 이용한 계층적 구조로 이루어져 있으므로 어휘의 형태를 활용한 일반적인 어휘 사전에 비해 검색효율이 좋다고 할 수 있다 [3]. 연관성이 높은 언어를 검색할 경우에는 의미 계층

* 이 논문은 2004년도 한국과학재단 지역대학우수과학자 지원 연구(R05-2004-000-10376-0)에 의하여 연구되었음

을 활용하게 되므로 연관어휘 추출 효율이 일반 어휘사전에 비해 탁월하다 할 수 있다[3].

워드넷의 원리는 개념 행렬을 기초로 정의 되었다[7]. 개념 행렬 모델을 [표 1]에 나타내었다. [표 1]은 워드넷의 어휘 의미 관계를 나타내는 기본원리이다.[4] [표 1]에서, Fn은 어휘의 형태를 나타내고 Mn은 의미를 나타낸다. 그러므로 E1,1은 F1의 어휘 형태를 갖고서 M1의 의미를 나타낸다. F1과 F2는 M1의 의미에 있어서 동일한 의미를 가지고 있으므로 F1과 F2는 동의(Synonym) 관계를 나타낸다. F2는 하나의 어휘 형태를 가지면서 어휘의 의미는 M1과 M2를 가지므로 다의(Polysemy)어가 된다. {F1, F2}는 M1의 의미에 있어서 동의어 집합(Set of synonyms synset)이 된다[7].

[표 1] 어휘의 개념 행렬

Word \ Meanings	Word					
	F1	F2	F3	F4	...	Fn
M1	E1,1	E1,2				
M2		E2,2				
M3			E3,3			
⋮					...	
Mn						Em,n

본 논문에서는 주어진 객관식 예시 문항의 각 명사 단 어들을 추출한 후, 워드넷에서 동의어 집합(Set of synonyms synset)을 가져온다. 각 명사들의 동의어 집합을 조합하여 기준 예시 문항과 비슷하지만 다른 의미의 문장을 만들어 내어 매력적인 오답지 작성을 쉽게 할 수 있도록 도와준다.

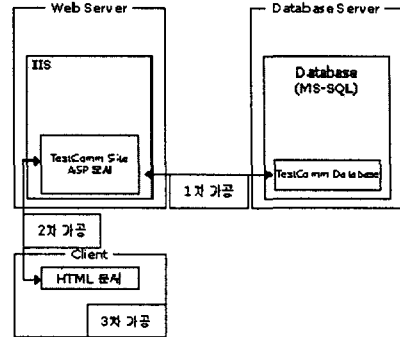
3. 임의 추출 분할 방식을 이용한 동적 문제 출제 시스템(SRSDM)

임의 추출 분할 방식을 이용한 동적 문제 출제 시스템인(SRSDM: A System to Generate Dynamic Test Using the Random Sampling Division Method)은 웹 서버와 데이터베이스 서버, 사용자의 기능을 각각 분할하여 문제를 출제한다.

SRSDM은 사용자가 웹 서버에게 서비스 요청을 보내면 데이터베이스 서버에서는 임의 추출 방식을 이용하여 1차 가공된 문제목록을 웹 서버에게 전달해 준다. 웹 서버에서는 1차 가공된 문제 데이터를 임의 정렬 방식으로 2차 가공하여 수검자에게 전송된다. 마지막으로 수검자 측에서는 2차 가공된 문제 데이터를 마지막으로 임의 정렬 방식을 사용하여 3차 가공된 최종 문제 데이터를 가지고 문제를 해결하게 된다. 수검자는 총 3번에 걸친 문제 데이터의 가공을 통하여 매번 자신의 단계에 적합하지만 이전의 문제와는 다른 문제를 해결하게 된다.

다음의 [그림 1]은 SRSDM의 구성도이다. 이전 절에서 설명한 총 3차에 걸친 문제 데이터의 가공이 각각 분할된 데이터베이스 서버와 웹 서버 그리고 클라이언트에서 이루어지게 된다. 이렇게 데이터의 임의 추출을 위한 수

행 작업이 총 3차에 걸쳐 이루어짐에 따라 이전의 집중된 가공 방식보다 작업 효율을 높일 수 있고 또한 이전에 시스템에 비해 향상된 수검자에 최적화된 문제 데이터의 임의 가공이 가능해진다.



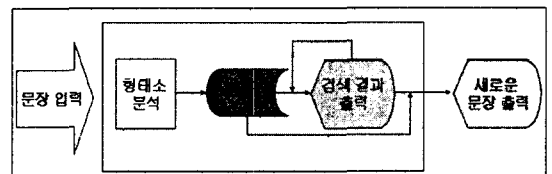
[그림 1] SRSDM의 시스템 구성도

SRSDM은 1차 가공 단계에서 문제 및 문제에 해당하는 답안 예시를 모두 검색하여 2차 가공 단계에서 임의로 추출을 하는 방식을 사용하는데 이 때 답안 예시가 풍부할수록 평가가 공정하고 효과적으로 이루어진다.[9] 따라서 문항 작성시 답안 예시문을 다양하게 작성하여야 하는데 객관식 검사의 단점 중 하나로 제작의 어려움과 제작시간이 많이 걸린다는 점이 있다.

4. 시스템의 구성

본 논문의 시스템은 임의 추출 분할 방식을 이용한 동적 문제 출제 시스템에서 출제자의 문제 출제 과정 중 예시를 자동으로 생성해주기 위한 보조 시스템이다. 따라서 [그림 1]의 과정 중에서 출제자가 문제의 모든 예시를 수동으로 작성하였던 방식에서 하나의 예시 문장을 입력하면 자동으로 여러 개의 예시 문장을 생성하도록 하였다.

다음의 [그림 2]는 객관식 문항 작성 보조 시스템의 흐름도이다.



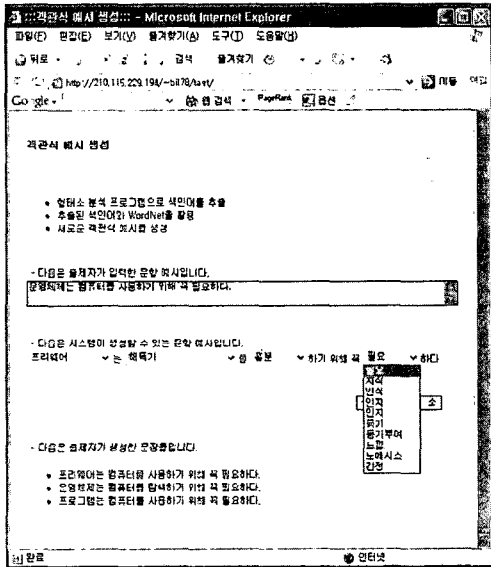
[그림 2] 객관식 문항 작성 보조 시스템 흐름도

출제자가 문제를 생성하기 위하여 예시 문장을 입력하면 시스템에서는 입력된 문장을 형태소 분석 프로그램인 'HAM'[10]을 이용하여 명사 용어를 추출한다. 이 추출된 명사 용어를 가지고서 적절한 쿼리를 작성하여 데이터베이스 서버에 저장된 '한글 워드넷'[11]을 검색한다. 이때 추출된 명사 용어와 관련된 Synset이 존재하지 않

을 경우에는 시스템이 종료하게 된다. 하지만 관련된 Synset이 존재할 경우에는 시스템은 학습자의 화면에 추출된 명사 용어의 위치에 Synset을 출력함으로써 출제자로 하여금 새로운 문장들을 생성할 수 있도록 한다.

5. 시스템의 구현 및 검토

본 시스템은 SRSDM에서 출제자가 문제를 출제 할 때에 출제가 용이 하도록 보조해주는 기능을 추가하였다. [그림 3]은 출제자 인터페이스 화면으로 출제자가 문제 생성에 필요한 하나의 예시 문장을 입력하게 된다.



[그림 2] 출제 인터페이스

문장의 예시로 “운영체제는 컴퓨터를 사용하기 위해 꼭 필요하다.”를 선택해 'HAM'을 이용하여 명사 용어를 추출하였다. 추출된 명사 용어는 {운영체제, 컴퓨터, 사용, 필요}가 추출 되었다. 추출된 명사 용어를 가지고서 데이터베이스 서버에 저장된 '한글 워드넷'을 검색한다. 각 명사 용어에 대해 추출된 Synset은 다음과 같다.

- 운영체제={프리웨어, 통제프로그램, ...}
- 컴퓨터={해독기, 시뮬레이터, 계산기, ...}
- 사용={탐색, 통제, 준비, ...}
- 필요={동기, 동기부여, 느낌, ...}

추출된 명사 용어에 해당하는 Synset들은 출제자가 입력한 문장에서 해당하는 명사 용어의 위치에서 선택하도록 구성하였다. 출제자는 선택 버튼을 이용하여 각 Synset에서 적절한 단어를 선택하고, 생성버튼을 누르면 시스템은 선택된 단어들을 본래의 예시문장과 같은 형태로 생성해 준다. 출제자는 여러 개의 새로운 문장들을 만들어 낼 수 있도록 구성하였다. 예시 문장으로 시스템을 이용하여 만들어낸 문장은 다음과 같다.

- 프리웨어는 컴퓨터를 사용하기 위해 꼭 필요하다.
- 운영체제는 컴퓨터를 탐색하기 위해 꼭 필요하다.
- 프로그램은 컴퓨터를 사용하기 위해 꼭 필요하다.

5. 결론 및 향후 연구과제

본 논문에서는 동적 문제 출제 시스템을 구현하기 위해서는 하나의 문제에 대하여 정답에 대한 예시와 오답에 대한 예시문이 풍부해야 하는데 출제자가 여러 개의 예시문을 만드는 것에 많은 시간과 비용을 들여야 한다는 것에 착안하여 형태소 분석기와 워드넷(WordNet)을 이용하여 인터페이스를 설계하였다. 출제자에게 인터페이스를 이용하게 함으로써 하나의 예시문으로 다양한 새로운 문장을 짧은 시간에 만들어 낼 수 있었다.

본 실험에서는 정량적 실험을 할 수 없는 관계로 임의 출제자를 선택하여 실험한 결과에 의존하여야 했지만, 출제 인터페이스의 설계로 문제 출제 시간 및 비용을 줄일 수 있었다. 향후 과제로서 문장에서 명사 외에 형태소를 사용하여 다양한 문장을 만들어 낼 수 있을 것이다. 또한 Synset 외에 반의어 등을 사용하여 시스템을 구성하면, 더욱 향상된 시스템을 구성할 수 있을 것이다.

참고문헌

- [1] 전우천, 홍종기, “수준별 평가를 위한 문제은행 시스템”, 정보교육학회논문지, 한국정보교육학회, 2002
- [2] 원대회, 강태호, 김원진, 방훈, 이재영, “임의 추출 분할 방식을 이용한 동적 문제 출제 시스템”, 추계 학술대회, 한국정보과학회, 2001.10
- [3] 김형일, 김준태, “워드넷 기반 협동적 평가와 하이퍼링크를 이용한 검색엔진의 성능 향상”, 정보처리학회, Vol11, June, 2004
- [4] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment", The Journal of the ACM, Vol46, Issue 5, 1999
- [5] C. Fellbaum, "WordNet : An Electronic Lexical Database", MIT Press, 1998
- [6] S. Scott, and S. Matwin, "Text Classification Using WordNet Hypernyms", Colling-ACL '98Workshop, 1998
- [7] G. A. Miller, "WordNet : An On-Line Lexical Database", International Journal of Lexicography, 1990
- [8] Eric Siegel, "Disambiguating Verbs with the WordNet Category of the Direct Object", Colling-ACL '98workshop, 1998
- [9] 최돈은, 서현진, 박기석, 이재영 “동적인 문제출제 시스템의 설계 및 구현”, 정보과학회, 학술발표논문집 제27권 1호, 2004. 4
- [10] 강승식, 범용 형태소 분석기 “HAM Ver 6.0.0”, 국민대학교 자연언어 정보 검색 연구실, <http://nlp.kookmin.ac.kr>
- [11] 이은영 임성신, WordNet2.0의 한국어 번역 작업과 결과물, 부산대학교 한국어정보처리연구실