

사용자 어휘지능망과 자동문제생성기술을 이용한 한국어 어휘학습시스템[†]

최수일*, 임지희*, 최호섭**, 옥철영*

*울산대학교 컴퓨터정보통신공학과 한국어처리연구실

**한국과학기술정보연구원 지식정보센터 시스템개발팀

our1223@mail.ulsan.ac.kr*, {arisu80, okcy}@ulsan.ac.kr*, hschoe@kisti.rc.kr**

Korean Word Learning System Using User-Word Intelligent Network and Automatic Question Generation Technique

Su-Il Choc*, Ji-Hui Im*, Ho-Scop Choc**, Cheol-Young Ock*

*Korea Language Processing Laboratory,

Dept. of Computer Engineering and Information Technology, University of Ulsan

**System Development Team, Knowledge Information Center,

Korean Institute of Science and Technology Information

요 약

본 논문에서는 올바른 한국어 생활과 한국어 실력 향상을 위하여, 한국어 어휘에 대한 풍부한 정보를 담고 있는 한국어사전, 사용자 어휘지능망(User-Word Intelligent Network : U-WIN) 등의 언어자원을 이용한 자동문제생성기술을 소개하고, 이를 이용한 한국어 어휘학습시스템을 제시한다. 대부분의 학습시스템에서 사용하는 문제 은행식 출제 방식의 문제점을 해소할 수 있는 하나의 방법으로서, 기존의 한국어 어휘 문제의 문항을 분석하여 8가지 문제 유형으로 재편성하고, 각 유형별 자동 문제 생성 패턴에 따라 언어자원이 가지고 있는 한국어 어휘의 형태적 정보, 의미적 정보를 이용하여 한국어 어휘 문제를 자동 출제하는 한국어 어휘학습시스템을 구현하였다.

1. 서 론

잘못된 언어 사용은 국민에게 언어생활 자체의 파괴뿐만 아니라 어휘 학습 능력의 저하 등의 악영향을 줄 수 있다. 어휘학습은 어휘를 이해하고 구사하는 데 필요한 전반적인 언어 능력을 기르는 과정을 말하고, 이러한 어휘 학습 과정을 통해서 습득할 수 있는 어휘력은 언어 활동의 전반에 걸쳐 작용하는 능력으로서, 말하기, 듣기, 읽기, 쓰기 등의 의사소통 과정에서, 이러한 어휘에 관련된 지식은 모든 언어 능력에 필요한 기반 지식을 제공한다. 그러므로 어휘 학습은 언어 학습 및 한국어 학습에서 가장 기본이 되는 중요한 과정이고, 효율적인 어휘학습시스템의 개발은 학습자의 어휘 학습 능력뿐만 아니라 전반적인 언어 능력을 증대시킬 수 있는 하나의 방법이 될 수 있다.

기존의 언어 학습 시스템에서는 문제 은행을 이용한 문제 출제 방식을 사용하고 있으며, 이와 같이 한정된 데이터베이스에 저장된 문제에 의존하는 문제 출제 방식은 문제의 다양성이 떨어지고 동일한 문제의 반복 출제로 인한 학습 능률 저하, 학습 평가의 신뢰성 저하, 새로운 문제 추가 시 별도의 문제 출제 인력 및 문제 출제 비용의 소요 등과 같은 문제점이 있다.

따라서 이러한 문제점을 해소하기 위해서는 다양한 언어 자원과 자연언어처리 기술을 기반으로 한 어휘학습시스템의 개발이 이루어져야 한다. 특히 최근에는 어휘의 의미 정보, 어휘의 형태적·통사적·의미적 구조 정보, 어휘의 관계 정보 등의 다양한 어휘정보를 담고 있는 의미적 언어 자원의 개발 및 구축 작업이 활발하게 진행되고 있으며, 이 중 울산대 한국어처리연구실에서 구축하고 있는 사용자 어휘지능망(User-Word Intelligent Network, 이하 U-WIN)은 한국어를 대상으로 한 대규모 어휘 데이터베이스로서 현재 다양한 기술에서 활용되고 있다.

[†] 본 연구는 정보통신부와 정보통신연구진흥원의 대학 IT 연구센터 육성지원사업의 연구결과로 수행되었습니다.

본 논문에서는 기존의 문제 은행식 출제 방식의 문제점을 해소하기 위해, 사용자 어휘지능망과 한국어사전 등의 언어자원을 이용한 자동문제생성기술을 소개하고, 이를 이용한 한국어 어휘학습시스템을 제시한다. 먼저 기존의 한국어 어휘 문제 분석 결과를 바탕으로 8가지 한국어 어휘 능력 평가 문제 유형을 설정하고, 각 문제 유형별 자동 문제 생성 패턴을 구축하였다. 그리고 구축된 문제 생성 패턴에 따라, 사용자 어휘지능망, 한국어사전 등의 언어자원을 이용하여 형태적 정보·의미적 정보를 반영한 자동문제생성기술을 소개하고, 이를 이용한 한국어 어휘학습시스템을 제시한다.

2. 관련 연구

국외에서는 영어를 중심으로 한 자동 어휘학습시스템 및 자동 언어학습시스템이 개발되고 있으며, 자연언어처리 기술을 이용하여 영어사전, WordNet, 유의어·반의어사전 등의 다양한 언어자원으로부터 형태적 정보, 구문적 정보, 의미적 정보 등을 자동으로 추출하고, 이러한 정보들을 활용하여 어휘 능력, 독해 능력, 문법 능력 등을 비롯한 학습자의 다양한 언어 능력을 평가하기 위한 문제가 자동 출제되고 있다.

[1]은 올바른 독서 습관 조성과 독서지도를 통한 어휘력 신장 및 독서력 향상을 위해 개발된 독서 훈련 시스템(READer-specific Practice system; 이하 REAP 시스템)으로, 학습자 개개인에 대한 어휘 능력 평가와 관심 분야 조사를 사전에 실시하고, 학습자의 흥미와 어휘 수준에 가장 적합한 영어 문서를 제공해 주는 언어학습시스템이다. 이 REAP 시스템에서는 WordNet과 자연언어처리 기술을 기반으로 한 자동문제생성기술을 이용하여 6개 문제 유형과 2개의 문제 패턴을 조합한 다양한 영어 어휘 능력 평가 문제를 자동 출제하고 있다. [3]은 WordNet에서 추출한 유의어, 반의어, 하의어 등을 이용하여 간단한 어휘 문제뿐만 아니라 학습자의 독해력 평가를 위한 질의문을 생성하였고, [4]은 소설책에 서술된 문장을 직접 이용하여, 각 문장으로부터 구문 정보와 의미 정보를 추출하여 특정 어휘를 의문사로 교체하는 방법으로 질의문을 생성하였다. 또는 유의어, 반의어 사전을 이용하여 "Jane is busy."의 문장으로부터 "Is Jane free?"와 같은 질의문을 생성하였다.

국내에서도 최근 들어 어휘학습시스템 및 한국어학습시스템에 대한 연구 개발이 활발히 진행되고 있다. 그러나 한국어 어휘학습시스템이나 자동문제생성기술을 이용한 한국어학습시스템에 대한 연구 개발은 미흡한 상태이다. 이는 기존에 나와 있는 어휘학습시스템의 대부분이 영어를 중심으로 한 외국어 학습시스템들이고, 한국어 학습시스템의 경우 글쓰기 학습시스템, 한글 맞춤법 검사시스템, 한글 받아쓰기 학습시스템 등과 같이 문법에 기초한 한국어의 표현정보를 습득하는 시스템만이 개발된 상태이기 때문이다. 또한 자동 문제 생성에 관한 기술 개발은 아직 초기 연구 단계에 불과하고, 이 또한 특

정 분야에 한하여 연구 개발이 진행되고 있는 실정이기 때문에, 사실상 한국어 학습시스템을 위한 자동문제생성 기술의 개발은 거의 전무한 상태라고 할 수 있다.

3. 한국어 어휘 능력 평가 문제 유형

한국어 어휘 능력 평가 문제를 자동 생성하기 위해서, 먼저 기존 한국어 어휘 문제의 문항 유형을 9가지로 구분하여 분석하고, 이러한 분석 결과를 바탕으로 본 논문에서는 한국어 어휘 능력 평가 문제 유형을 8가지로 재편성하였다.

3.1 기존 한국어 어휘 문제의 문항 분석

한국어 어휘 문제의 문항 분석을 위한 자료로는 국립국어원이 운영하는 '우리말 실력 알아보기'에서 출제된 문항들을 사용하였으며, 2000년부터 2005년까지 출제된 한국어 어휘 문제를 유형별로 분석하였다.

3.1.1 어휘의 기본 의미 파악 유형

고유어, 한자어, 고사성어의 기본 의미를 정확하게 알고 있는지를 묻는 유형이다. 이러한 문항들은 주어진 뜻풀이에 부합하는 어휘 찾기, 주어진 어휘의 정확한 뜻풀이 찾기, 제시된 어휘와 뜻풀이가 잘못 짝지어진 것 찾기, 유사한 의미를 가진 고사성어와 속담 연결하기 등의 세부 유형으로 분류할 수 있다.

3.1.2 어휘의 문맥적 쓰임 파악 유형

주어진 제시문의 빈칸에 적절한 어휘를 찾음으로써, 어휘의 기본 의미를 바탕으로 그것을 문맥에 맞게 사용할 수 있는지를 묻는 유형이다. 특히 제시문이 "(눈)이 높다"와 같은 경우 대상 어휘를 사용하여 관용구를 정확하게 표현할 수 있는지를 묻는 문항이다.

3.1.3 다의어의 의미 구분 유형

한 어휘가 두 가지 이상의 뜻을 가진 다의어의 의미를 구분할 수 있는지를 묻는 유형으로서 문맥 속에서 적절하게 사용되고 있는 다의어의 의미 차이를 분간해 내도록 한다.

3.1.4 어휘 간의 의미 관계 파악 유형

일정한 의미 관계를 맺고 있는 어휘들을 정확하게 파악하고 있는지를 묻는 유형으로 한 어휘가 다른 어휘와 가지는 의미 관계에는 동의/유의 관계, 반의 관계, 상의 관계, 하의 관계 등이 있다. 이 유형의 문항에서는 주어진 어휘와 특별한 관계를 갖는 어휘 찾기, 제시된 어휘 쌍들 중에

한국어 어휘 능력 평가 및 한국어 어휘 문제의 자동 생성을 위해서는 먼저 평가 대상자와 수준별 평가 대상 어휘 목록이 정해져야 한다. 즉 누구를 대상으로 하여 평가할 것인가를 결정해야 한다. 평가를 받는 대상이 내국인인가 외국인인가에 따라, 그리고 학생(초·중·고등학교 및 대학생)인가 일반인인가에 따라서 그들의 어휘력 수준에 맞는 어휘를 이용하여 문제를 생성해야 하기 때문이다.

본 논문에서 제시하는 자동문제생성기술은 한국어 어휘 학습의 기초적이면서 중요한 시기인 초등학교 및 한국어를 배우려는 외국인을 평가 대상으로 설정하였다. 그리고 이를 위한 평가 대상 어휘는 국립국어원에서 발표한 '한국어 학습용 어휘 목록'(총 5,965개)으로 정하였다. '한국어 학습용 어휘 목록'은 한국어 사용 빈도 조사 결과를 바탕으로 하여 일상 언어생활에서 사용 빈도수가 높은 어휘를 중심으로 선정한 것이며, 이 어휘들은 언어 전문가에 의해 3단계로 학습용 어휘의 등급을 분류하고 있다. 그러므로 이는 본 연구에서 정한 평가 대상자의 한국어 어휘 능력 평가를 위한 대상 어휘로 적합하다고 판단되며, 언어 전문가에 의해 분류된 어휘의 등급은 수준별 어휘 능력 평가에 유용하게 이용될 수 있다.

5.2 문제 난이도 조절

한국어 어휘학습시스템은 학습자의 어휘 수준에 따라 적절한 난이도의 문제를 출제함으로써 어휘 학습 능력이나 학습자의 학습 성취도를 높일 수 있고, 학습자의 어휘력을 올바르게 평가할 수 있다. 어휘학습시스템의 문제 난이도 조절은 어휘의 등급을 설정하는 방법으로 이루어질 수 있으며, 본 논문에서는 이를 위해 한국어 어휘의 사용 빈도와 사용자 어휘지능망을 이용하여 기초어휘를 3등급으로 분류하였다.

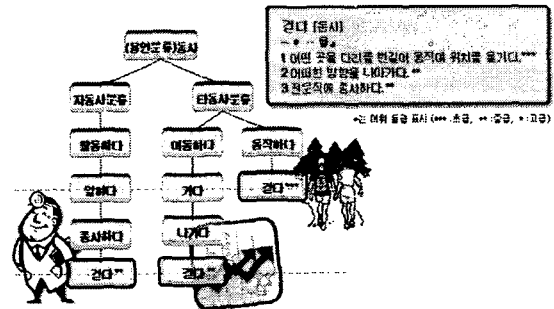
사용자의 계층이나 사용 분야에 관계없이 널리 사용되는 어휘는 한국어 학습의 가장 기초적인 어휘이며 가장 중요한 어휘이다. 그러므로 말뭉치에서 나타난 어휘의 빈도수에 따라 학습용 어휘의 등급을 결정할 수 있다. 국립국어원의 '한국어 학습용 어휘 목록'은 한국어 어휘 사용 빈도 조사 결과를 바탕으로 전문가에 의해 1단계 982개, 2단계 2,111개, 3단계 2,872개로 어휘의 등급을 나누고 있다.

[10]은 한국어 교육에서 의미관계에 따른 어휘 교육에 관하여 언급하고 있는데 이는 어휘 학습에도 적용될 수 있다. 즉, 어휘는 기본적으로 다의어의 특징을 가지고 있으며, 빈도수가 높은 기본 어휘의 경우 이러한 성격이 두드러지게 나타난다. 그러므로 어휘 학습 단계에 따라 다의어 중 어떤 의미부터 학습할 것인가를 고려하고, 중심 의미에서 주변 의미로 확대하는 방향으로 학습이 이루어져야 한다. 이는 어휘의 등급 설정에 있어서도 사용 빈도뿐만 아니라 다의어의 의미에 따른 어휘 등급 설정

이 필요함을 의미한다.

다의어의 의미에 따른 어휘 등급을 구분하기 위한 방법으로는 U-WIN을 활용하는 방법을 사용하였다. U-WIN의 계층 구조는 보편적인 개념화과정으로 인식되는 구조로서 상하관계를 통한 추상화·일반화 및 구체화·특수화가 형성되는 구조이다. 즉, U-WIN에서 상위어는 보편적이고 일반적인 의미적 특성을 지니는 반면, 하위어는 구체적이고 전문적인 의미적 특성을 지닌다. 이러한 특성을 이용하여 U-WIN에서 나타나는 깊이(depth)를 어휘 등급 설정을 위한 하나의 기준으로 사용하였다.

따라서 본 논문에서 제시한 한국어 어휘학습시스템에서는 수준별 어휘 능력 평가를 위한 문제 난이도 조절을 위하여 '한국어 학습용 어휘 목록'의 3단계 어휘 등급을 바탕으로 하고, 다의어가 U-WIN에서 내부적으로 가지는 깊이(depth)가 다르게 나타나는 경우, 그 의미에 따라 어휘의 등급을 다르게 설정하였다. [그림 2]는 동사 '걷다'를 다의어의 의미에 따라, U-WIN을 이용하여 상이한 등급으로 나눈 예이다.



[그림 2] 동사 '걷다'의 등급을 나눈 예

5.3 자동 문제 생성을 위한 정보

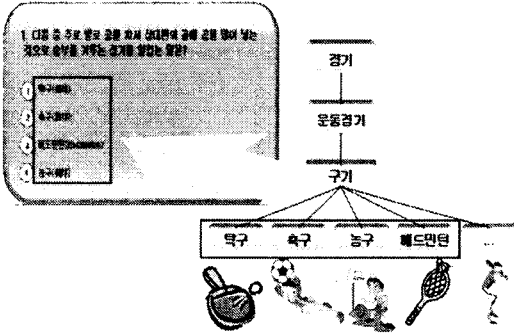
한국어 어휘학습시스템의 자동 문제 생성을 위한 정보는 한국어 어휘에 대한 풍부한 정보를 담고 있는 한국어 사전과 어휘 관계에 대한 다양한 정보를 담고 있는 U-WIN으로부터 자동 추출한 정보를 이용하였다.

자동 문제 생성 과정의 질의문 생성, 예문 생성, 보기 생성 등에서 사용되는 정보들은 한국어 사전으로부터 추출한 표제어, 뜻풀이, 발음, 원어, 용례 등을 이용하였으며, 이를 위해 각종 정보의 추출 및 검색이 용이하도록 한국어 사전 DB를 재구성하였다. 특히 어휘 등급의 설정, U-WIN과의 사상(mapping), 어휘 정보 처리의 용이성 등을 고려하여 어휘와 뜻풀이의 관계가 1:1의 관계가 되도록 하였다.

자동 문제 생성을 위해서 대상 어휘와 특정한 어휘 관계를 맺고 있는 어휘를 추출하기 위하여 U-WIN을 활용하였다. 자동 문제 생성 과정에서 사용하는 U-WIN의 관계 정보는 상위관계(hyperonym), 하위관계(hyponym), 동의관계(Csynonym), 유의관계(Psynonym), 반의관계(antony

m), 부분전체관계(meronym/holonym), 높임말관계, 낮춤말관계, 준말본말관계 등이 있다.

오답어휘의 개수가 달라질 수 있다. 따라서 N지선다형의 보기생성을 위한 전체 어휘의 개수는 2개/N개/N+1개/2N-1개로 나눌 수 있다.



[그림 3] U-WIN을 이용한 자동 문제 생성의 예

객관식 문제에서 오답은 문제의 난이도 및 학습자의 대상 어휘에 대한 이해도 평가에 중요한 역할을 한다. 그러므로 오답을 생성하기 위한 오답어휘를 선택할 때에는 정답어휘와의 관계를 충분히 고려하고, 이전에 선택한 정답어휘의 정보를 적절히 활용하여야 한다. 오답어휘는 정답어휘와 형태적·의미적으로 유사할수록 매력적인 오답을 생성할 수 있으며, 정답을 선택해야 하는 학습자에게 혼돈을 줄 수 있다. 따라서 본 연구의 자동 문제 생성 과정에서는 오답어휘 선택 이전에 정답어휘 선택을 수행하고 정답어휘로부터 추출한 품사, 원어, 어휘의 길이, 어휘의 첫음절 및 끝음절 등의 정보를 오답어휘 선택 과정에서 활용하였다. 또 어휘간의 의미적 관계를 고려하기 위해서 U-WIN의 상하관계 정보를 이용하였다. [그림 3]은 자동 문제 생성 과정에서 U-WIN을 이용하여 같은 상위어를 갖는 어휘들을 오답어휘로 선택하여 문제를 출제한 예이다.

5.4 자동 문제 생성 패턴

자동 문제 생성 패턴은 3장에서 분류한 8가지 한국어 어휘 능력 평가 문제 유형을 기반으로 각 문제 유형별로 하나 이상의 문제 생성 패턴을 구축하고, 각 패턴별로 자동 문제 생성과정에서 적용되는 규칙을 정의하였다. 문제 유형별 자동 문제 생성 패턴을 정리하면 [표 2]와 같다.

문제 유형별 자동 문제 생성 패턴은 자동 문제 생성과정에서 사용되는 질의패턴에 따라 분류함으로써, 동일한 패턴의 문제는 질의 테이블로부터 동일한 질의패턴을 추출할 수 있도록 하였다.

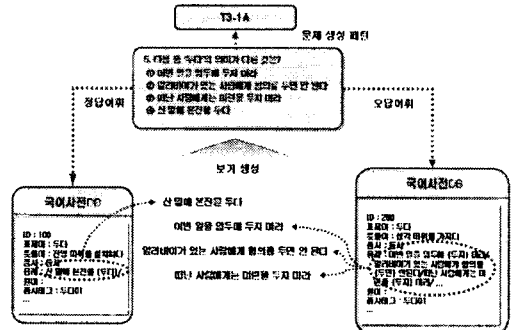
본 논문에서 자동 생성한 한국어 어휘 능력 평가 문제는 모두 객관식으로 만들어지며, 보기문항의 개수가 N개인 N지선다형으로 출제된다. 또한 N개의 보기문항은 1개의 정답과 N-1개의 오답으로 구성되며, 각 문제 생성 패턴에 따라 보기생성을 위한 어휘의 개수 및 정답어휘와

[표 2] 문제 유형 및 자동 문제 생성 패턴과 질의문의 예

문제 유형	질문 패턴	질의문 예
어휘의 사전적 의미 파악하기	T1-1A	하나의 단위로서 처리되는 서로 만원있는 레코드의 집합을 일컫는 말은?
	T1-2A	다음 중 '학교'의 사전적 의미는?
	T1-3B	단어의 뜻을 잘못 풀이한 것은?
어휘의 문맥적 쓰임 파악하기	T2-1A	다음 문장의 빈칸에 가장 알맞은 말은?
	T2-1B	다음 문장의 빈칸에 가장 알맞은 말은?
다의어 및 동형다의어의 의미 구분하기	T3-1B	다음 중 '두다'의 의미가 다른 것은?
	T4-1A	같은 뜻을 갖는 단어가 틀바르게 연결된 것은?
	T4-1B	다음 중 의미가 같은 말로 연결되지 않은 것은?
	T4-2A	단어의 의미가 반대인 것은?
	T4-2B	다음 중 반대말로 학자되지 않은 것은?
	T4-3B	다음 중 조달본말의 관계가 아닌 것은?
	T4-4B	다음 단어의 높임말이 잘못 연결된 것은?
	T4-4C	다음 단어의 높임말이 잘못 연결된 것은?
표준 발음 알기	T5-1A	다음 발음 기호의 긴소리 표기가 옳은 것은?
	T5-1B	다음 발음 기호의 긴소리 표기가 틀린 것은?
	T5-2A	다음 중 발음 표기가 옳은 것은?
	T5-2B	다음 중 발음 표기가 틀린 것은?
순화어로 바꿔 쓰기	T6-1B	다음 중 순화한 표현이 잘못된 것은?
	T7-1A	다음 중 표준어는?
표준어규정에 맞게 표기하기	T7-1B	다음 중 표준어가 아닌 말은?
	T8-1B	다음 중 외래어 표기가 잘못된 것은?

(1) 보기생성을 위한 어휘가 2개인 경우

자동 문제생성 패턴이 (T3-1A)인 경우에는 보기생성을 위한 어휘가 2개이다. 예를 들어, [그림 4]와 같이 2개의 서로 다른 의미를 갖는 '두다'가 보기 생성을 위한 어휘로 선택되고, 1개의 정답어휘를 기반으로 정답 보기 항목이 생성되고, 다른 나머지 1개의 오답어휘를 기반으로 N-1개의 오답 보기 항목이 생성된다. 즉, 복수개의 오답 보기 항목을 생성하는데 오직 한 개의 오답어휘가 이용되는 경우이다.



[그림 4] 보기생성을 위한 어휘가 2개인 경우 (N=4인 사지선다형의 예)

(2) 보기생성을 위한 어휘가 N개인 경우

보기생성을 위한 어휘가 N개인 경우에는 1개의 정답어휘를 기반으로 1개의 정답 보기 항목이 생성되고, 나머

지 N-1개의 오답어휘를 기반으로 N-1개의 오답 보기 항목이 생성될 수 있다. 즉, 이러한 경우는 보기생성을 위해 선택된 N개의 어휘에 대한 각각의 정보가 별개의 보기 항목 생성에 이용되는 경우이다. 예를 들어, [표 2]의 자동 문제 생성 패턴 (T1-2A)의 경우, 정답어휘 '학교'를 포함한 N개의 어휘가 임의 방식으로 선택되며, 이 문제 패턴에서 생성 가능한 "다음 중 '학교'의 사전적 의미는?"에 대한 정답 보기 항목은 정답어휘 '학교'의 뜻풀이 정보를 이용하여 생성되고, N-1개의 오답 보기 항목은 나머지 N-1개의 오답어휘들의 뜻풀이 정보를 이용하여 각각 생성된다.

[표 3] 자동 문제 생성 패턴별 보기생성을 위한 어휘 개수

자동 문제 생성 패턴	보기생성을 위한 어휘 개수
T3-1A	2
T1-1A, T1-2A, T2-1A, T3-1B, T5-1A, T5-1B, T5-2A, T5-2B, T7-1A, T7-1B, T8-1A, T8-1B	N
T1-3A, T4-1A, T4-2A, T4-3A, T6-1A	N+1
T1-3B, T4-1B, T4-2B, T4-3B, T6-1B	2N-1

(3) 보기생성을 위한 어휘가 N+1개인 경우

보기생성을 위한 어휘가 N+1개인 경우에는 2개의 정답어휘를 기반으로 1개의 정답 보기 항목이 생성되고, N-1개의 오답어휘를 기반으로 나머지 N-1개의 오답 보기 항목이 생성될 수 있다. 즉, 이러한 경우는 정답 보기 항목 1개당 2개의 정답어휘가 필요하고 오답 보기 항목 1개당 1개의 오답어휘가 필요한 경우를 의미한다. 예를 들면, (T1-3B)의 문제 패턴과 같이, 개별 보기 항목은 '단어:뜻풀이'관계의 연결 구조를 가지며, "단어의 뜻을 잘못 풀이한 것은?"의 질의문에 대하여 총 N+1개의 어휘가 필요하게 된다. 즉, N=4인 사지선다형인 경우, N+1개인 A,B,C,D,E의 5개의 어휘가 선택되고, 선택된 어휘의 뜻풀이 정보로부터 다음과 같이 ①~④의 보기 항목이 생성될 수 있다. (① A표제어:A뜻풀이, ② B표제어:B뜻풀이, ③ C표제어:C뜻풀이, ④ D표제어:E뜻풀이; 정답은 ④번 보기 항목)

(4) 보기생성을 위한 어휘가 2N-1개인 경우

보기생성을 위한 어휘가 2N-1개인 경우에는 1개의 정답어휘를 기반으로 1개의 정답문항이 생성되고, 2(N-1)개의 오답어휘를 기반으로 N-1개의 오답 보기 항목이 생성될 수 있다. 즉, 이러한 경우는 정답 보기 항목 1개당 1개의 정답어휘가 필요하고 오답 보기 항목 1개당 2개의 오답어휘가 필요한 경우를 의미한다.

예를 들면, (T4-1A)의 문제 패턴과 같이, 개별 보기문항은 '단어:동어어' 관계의 연결 구조를 가지며, "같은 뜻을 갖는 단어가 올바르게 연결된 것은?"의 질의문에

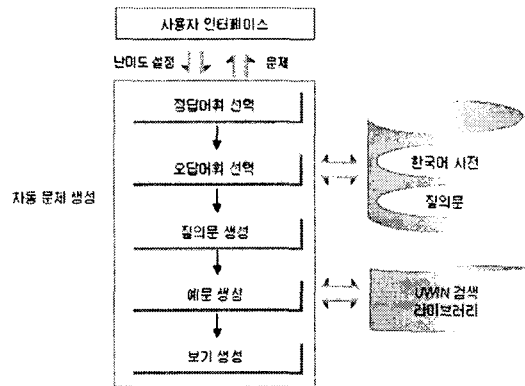
대하여 총 2N-1개의 어휘가 필요하게 된다.

즉, 예를 들어 사지선다형인 경우, A,B,C,D,E,F,G인 7개의 어휘가 선택되고, 선택된 어휘의 뜻풀이 정보로부터 다음과 같은 ①~④의 보기 항목이 생성될 수 있다. (① A표제어:A동어어, ② B표제어:E표제어 ③ C표제어:F표제어, ④ D표제어:G표제어; 정답은 ①번 보기 항목)

5.5 자동 문제 생성 과정

한국어 어휘학습시스템을 위한 자동 문제 생성은 [그림 5]와 같이 5단계의 과정을 거치게 된다. 먼저 학습자의 어휘 수준에 따라 적당한 문제 난이도가 결정되면, 각 문제 유형에 따라 자동 문제 생성 패턴이 결정된다. 자동 문제 생성을 위한 정답어휘와 오답어휘는 [표 3]의 보기생성을 위한 어휘 개수에 따라 자동 선택되며, 정답과 형태적·의미적으로 관련 있는 오답을 생성하기 위해 오답어휘 선택 이전에 정답어휘 선택 과정을 수행하고, 이미 선택된 정답어휘의 정보를 오답어휘의 선택 과정에서 활용한다.

질의문 생성 단계에서 질의문은 질의패턴 테이블에 저장된 질의패턴과 정답어휘의 다양한 어휘 정보를 조합하여 생성될 수 있다. 예를 들어, [표 2]의 자동 문제 생성 패턴 (T1-1A)의 경우, 한국어 사전DB로부터 '파일'을 정답어휘로 선택한 후, 질의패턴 테이블로부터 "다음 중 () 일컫는 말은?"과 같은 질의패턴을 추출한다. 다음으로, 이미 추출된 질의패턴과 정답어휘의 뜻풀이 정보를 조합하여 "다음 중 하나의 단위로써 처리되는 서로 관련 있는 레코드의 집합을 일컫는 말은?"과 같은 질의문을 생성할 수 있다.



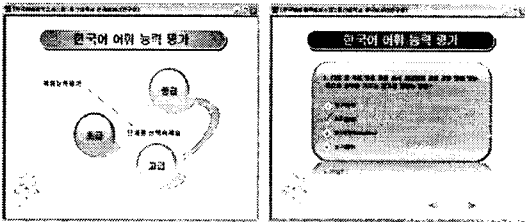
[그림 5] 자동 문제 생성 과정

다른 예로, [표 2]의 문제 생성 패턴 T2-1A의 경우, 한국어 사전DB로부터 '그래서'라는 정답어휘를 선택하고, 질의패턴 테이블로부터 "제시된 문장의 빈칸에 들어갈 알맞은 말은?"이라는 질의문을 추출한다. 다음으로, 정답어휘의 용례 정보를 이용하여 "어제는 많이 아팠어

요. {그래서} 결석 했어요."로부터 "어제는 많이 아팠어요. () 결석 했어요."와 같은 예문을 생성할 수 있다.

6. 한국어 어휘학습시스템

본 논문에서 제시한 자동문제생성기술은 한국어 어휘 능력 평가에 이용될 수 있고 나아가 국민들의 한국어 실력 향상에 도움이 될 수 있는 한국어 어휘학습시스템에도 이용될 수 있다. 본 연구를 통해 구현한 자동문제생성기술을 이용한 한국어 어휘학습시스템에서는 문제의 난이도를 초급, 중급, 고급의 3등급으로 구분하여, 학습자의 어휘 수준에 적당한 어휘력 평가가 이루어 질 수 있도록 하였다. [그림 6]은 한국어 어휘학습시스템의 구현 예이다.



[그림 6] 한국어 어휘학습시스템의 예

이 시스템은 향후 좀 더 다양한 정보, 확장된 문제 생성 어휘, 구체적인 문제 난이도 결정방법 등을 이용하여, 다양한 계층의 사람들을 위한 한국어 어휘학습시스템으로 발전해 나갈 것이다.

7. 결론

본 논문에서는 기존 문제 은행식 출제 방식에서 나타나는 문제 출제 인력, 문제 출제 비용, 문제 출제 시간 등과 같은 문제점을 해소하기 위해, 한국어 어휘에 대한 풍부한 정보를 담고 있는 한국어사전, 사용자 어휘지능망 등의 언어자원을 이용한 자동문제생성기술을 소개하고, 이를 이용한 한국어 어휘학습시스템을 제시하였다.

현재 국내의 한국어 어휘학습시스템에 대한 연구 개발은 아직 미흡한 실정이지만, 전산학, 한국어처리, 국어학, 국어교육학 등 다양한 분야에서 한국어 학습시스템 개발을 비롯하여 한국어정보처리, 한국어 능력 평가, 한국어 교육 등에 대한 꾸준한 노력을 기울이고 있고, 특히 최근 들어서는 각종 신문이나 방송의 우리말 관련 연재와 프로그램을 비롯한 국립국어원의 '우리말 실력 알아보기', 한국언어문화연구원의 '국어능력인증시험', KBS의 '한국어능력시험', 외국인들을 위한 한국어교사 양성 프로그램 등과 같이 한국어에 대한 관심을 고조시키고 한국어 교육을 강화하려는 노력이 활발히 진행되고 있다. 그러므로 이러한 노력에 힘입어 국민들의 올바른

한국어 생활과 한국어 실력 향상 및 한국어의 국제 경쟁력을 강화하기 위해서는 한국어 및 한국어 교육 이론 기반의 한국어정보처리, 한국어 교육을 위한 보조학습시스템 개발, 웹기반 한국어 학습시스템 개발 등이 이루어져야 할 것이다.

또한, 본 연구를 통해 제시한 자동문제생성기술 및 한국어 어휘학습시스템은 자동생성된 어휘 능력 평가 문제에 대한 전문가 평가 및 초등학생·외국인 대상의 어휘 능력 평가를 통해 본 연구의 실용성을 검증해 볼 필요가 있으며, 향후 다양한 계층의 사람들이 활용할 수 있는 웹기반 한국어 어휘학습시스템으로 발전해 나가야 할 것이다.

참고문헌

- [1] J. Brown, G. Frishkoff, and M. Eskenazi, "Automatic question generation for vocabulary assessment", *HLT/EMNLP 2005*, 2005.
- [2] J. Brown and M. Eskenazi., "Retrieval of Authentic Documents for Reader-Specific Lexical Practice.", *InSTIL/ICALL Symposium 2004*, 2004.
- [3] G. Aist, "Towards automatic glossarization: automatically constructing and administering vocabulary assistance factoids and multiple-choice assessment", *Interantional Journal of AI in Ed.*, 12(2), pp. 212-231, 2001.
- [4] H. Kunichika, T. Katayama, T. Hitashima, and A. Takeuchi, "Automated question generation methods for intelligent English learning systems and its evaluation", *ICCE2004*, 2004.
- [5] 조남호, 한국어 학습용 어휘 선정 결과 보고서, 국립국어원 보고서, 2003.
- [6] 권혁철 외, 웹에 기반한 '바른 우리글 쓰기' 학습 시스템, 정보통신부, 2000.
- [7] 최호섭 외, "온톨로지 구축 방법과 사례", 정보과학회지, 제24권 제4호, pp. 31-44, 2006.
- [8] 최호섭 외, "대규모 지능형 한국어 어휘망 구축", 609돌 세종날 기념 한글 학회 전국 국어학 학술 대회 발표자료집, 한글 학회, 2006.
- [9] 최수일 외, "한국어 어휘학습시스템을 위한 자동 문제 생성", 2006 한국인지과학회 춘계학술대회 발표 자료집, 한국인지과학회, 2006.
- [10] 조현용, 한국어 어휘의 특징과 어휘교육, 한국어교육 제10권 1호, pp. 265-281, 1999.
- [11] 송창선, "평가 대상 어휘의 선정", 국어교육연구 30, pp. 85-100, 1998..
- [12] 김광해, "어휘력과 어휘력의 평가", 선청어문 25, pp. 1-29, 1997.
- [13] http://www.korean.go.kr/9000/50_quiz_prv.jsp,