온톨로지 기반의 효과적인 영어 학습 웹 어플리케이션

매튜 오클리*, 기창진*, 송은지*
*남서울대학교 컴퓨터학과

An Web Application for Effective English Leaning based on Ontology

Matthew Oakley* ,Chang-jin Kee* , Eun-jee Song*
*Dept. of Computer Science, Namseoul University
E-mail : samphire@nsu.ac.kr

요 약

우리나라에는 영어 학습을 위한 수많은 도구들이 있다. 특히, 최근 스마트폰의 출현으로 영어 학습을 위한 어플리케이션 개발이 활발하다. 그러나 대부분 영어단어를 단순히 암기하는 방법에 관한 것으로 텍스트 분석에 대한 어플리케이션은 부재하다. 본 논문에서는 사용자가 텍스트를 입력하면 단어가 갖고 있는 희소성에 따라 온톨로지 기반으로 분석되어져서 등급을 매기는 어플리케이션 시스템을 제안한다. 결과물은 파일 시스템과 데이터베이스에 저장되어 사용자는 텍스트의 희소성에 따라 선택한 텍스트를 읽음으로서 사용자의 눈높이에 맞게 신속하고 효율적인 어휘획득 능력을 갖게 된다.

키워드

Java EE web application, Java Server Faces, Ontology, English Learning Application

Ⅰ. 서 론

최근 스마트폰의 출현으로 영어 학습을 위한 어플리케이션 개발이 활발하다. 그러나 대부분 영어단어를 단순히 암기하는 방법에 관한 것으로 텍스트 분석에 대한 어플리케이션은 부재하다.

본 논문에서는 영어 텍스트의 내용을 분석하고 단어들의 빈도에 따라 등급을 매길 수 있는 도구를 제공하는 어플리케이션 시스템을 제안한다. 입력된 텍스트는 온라인 사전 서비스로 모든 단어를 하이퍼링크를 걸어 변환되고 배워야 할 단어들이 있는 학습 목록에 클릭된 모든 단어를 저장한다. 온라인 사전으로 빠르고 쉬운 접근을 제공함으로써 집중적인 독해의 경험은 광범위한 독해의 경험으로 더 가깝게 가게 된다. 또한 텍스트를 위한 온톨로지를 제공함으로써 유저가, '광범위하게' 또는 '집중적으로' 공부하기를 원하느냐에 따라, 자신에게 맞는 읽을 텍스트를 선택할 수 있다. 제안하는 애플리케이션은 텍스트에 나오는 어휘의 모든 아이템들에 대해 유저가 사용한 어휘를 추적함으로써 어휘를 효과적으로 습득할 수 있는 도구가 된다.

Ⅱ. 온톨로지 기반 영어 학습 어플리케이션

1. 시스템 환경 및 데이터베이스

선택된 플랫폼은 GlassFish Server에서 동작하는 Java Server Faces를 사용하는 Java EE6이다. 자바의 객체지향적 성격은 자바에게 텍스트 엔티티의 조작을 부여한다 그리고 유니코드의 전면적인 구현은 텍스트 요소의 파싱에 가장 유용하며 자바 EE 아키텍처는 견고하고 확장성이 있다. 자바는 안드로이드를 운행하는 모바일 디바이스에로의 애플리케이션 이식 수월하다. Java Server Faces는 디자인의 수월함과 역동적인 유저 경험을 위한 Ajax-enabled 컴포넌트를 제공하기 때문에 프레임워크로 선택되었다. 사용된 컴포넌트 라이브러리는 PrimeFaces 3.0 이다. 모든 컴포넌트는 http 세션으로 향하는 객체로서 존재하는 'view'에 제한되어 있다. 'view'는 자바 메서드가 실행되는 컨테이너 관리 지원 빈에 제한된다. 이 모든 것은 JSF 아키텍처에 의해 제공된다.

어휘 데이터는 British National Corpus (BNC)의 천만 토큰 구어체 세션으로부터 가져왔으며 단어들은 '표제어'라 불리는 페밀리에 정렬된다. 예를 들어 그림1과 같이 단어 'AGENT'가, 복수형과 명사 'AGENTCY'의 단수형과 복수형을 포함하여, 3개의 파생어 유형(forms)을 가진다는 것을 보여준다.

headword	frequency	id	word	headword_id
AGENT	1	131	AGENCY	23
AGENT	1	132	AGENCIES	23
AGENT	1	133	AGENTS	23
AGENT	1	36623	AGENT	23

그림1. 표제어 페밀리의 예

전산 해석(computational analysis)에 의해 표제어는 BNC에 나타나는 빈도수에 따라 1000개중 14그룹으로 정렬되었다. 그러므로 표제어의 전체 숫자는 그래서 파생어를 하여 50606 단어를 가진 14,000이다. 본 연구의 목적을 위해 이들 목록은 ISAM 테이블 'headwords'와 'words'로서 MySQL 데이터베이스에 포함되었다. 테이블 'headwords' (표제어)는 단어의 빈도수 값을 나타내는 정수이다. 표제어의 빈도는 텍스트의 오토로컬 분석에서 사용된다. 기본 데이터베이스 스키마는 그림2에 나타나 있다. 'text'는 온톨로지 데이터를 담고 있으며 'learning list' (학습 목록)는 각 유저가 사용하고 암기하려고 하는 단어들을 저장하기 위해 사용된다. 'learned'는 유저가 접한 단어들의 전체 목록을 포함한다. 이것은 유저가 사용을 한 'readers'에 포함되어 있는 모든 단어를 의미한다.

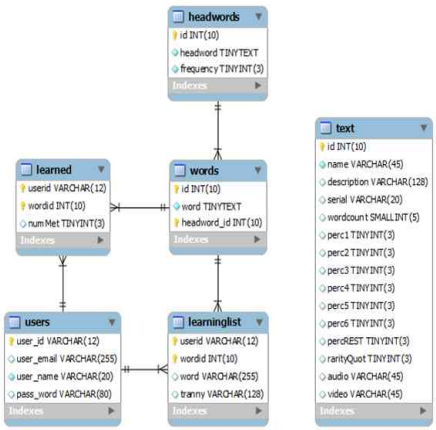

그림2. 데이터베이스 스키마 다이어그램

2. 온톨로지 분석

학습자가 사용하여 발생하는 빈도에 따라 단어의 숫자는 온톨로지 분석에 의해 얻어진다.

일반적인 'rarity quotient'는 단어들의 빈도의 합을 단어 개수로 나누어서 얻어진다. 복제된 목록이 사용되는 것은 이 이유 때문이다. 높은 빈도를 가진 단어는 같은 텍스트에서 여러 번 발생 할 수 있다 그러나 그것은 단지 difficulty의 하나의 단순 요소를 나타낸다. 왜냐하면 그것의 의미는 처음으로 발생했을 때 알려질 것이거나 look up 될 것이기 때문이다. 기본적인 온톨로지는 그림3에서 보이는 것처럼 유저가 공부하기 위해 리더를 선택할 때 유저가 사용할 수 있다. 녹색부터 빨강색까지의 컬러 스케일은 텍스트의 'rarity quotient'를 나타낸다. 유저가 광범위하게 공부할 분위기에 있을 때 유저는 스택트럼의 녹색 끝을 향해 선택하며 유저가 집중적으로 공부하여 어휘 습득하고 싶을 때는 스택트럼의 빨강색 끝을 향해 선택한다.

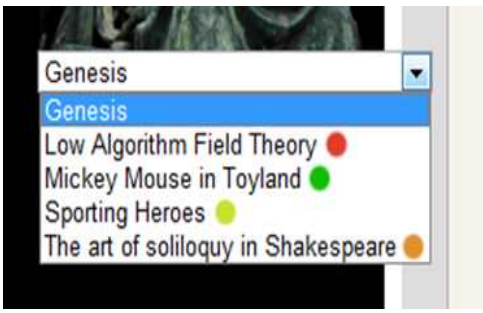

그림 3. 리더 선택 화면

III. 결 론

최근 스마트폰이 보편화 되면서 영어 학습을 위한 어플리케이션 개발이 활발하다. 그러나 대부분 영어단어를 단순히 암기하는 것으로 텍스트 분석에 대한 어플리케이션은 부재하다. 본 논문에서는 사용자가 텍스트를 입력하면 단어가 갖고 있는 희소성에 따라 온톨로지 기반으로 분석되어져서 등급을 매기는 어플리케이션 시스템을 제안하였다. 이 시스템을 통해 학습자는 자신이 관심이 있고 자신에게 맞추어진 어휘 목록에서 테스트 될 수 있는 리더를 선택할 수 있다. 학습자의 흥미에 맞추어진 구체적인 단어목록에 기초한 온톨로지컬 분석을 하며 학습해야 될 주제에 연관된 텍스트가 그러한 단어목록을 시스템에 의해 만들어내기 위해 사용될 수 있다.

참고문헌

[1] 서인영외,"영어 어휘 학습에 관한 소프트웨어 설계 지침",한국멀티미디어 언어 교육학회 논문지,제6권 2호,pp.119-136,2012.

[2] 미소인터렉티브 출판사, "자바플렛 폼", 마이크로소프트웨어 4월호,2011.