

# 한글을 지원하는 온라인 컴퓨터 용어 사전의 개발

황병연                      박성철

가톨릭대학교    컴퓨터전자공학부

byhwang@www.cuk.ac.kc, schpark@hotmail.com

## Development of On-Line Computer Dictionary Supporting Hangul

Byung Yeon Hwang      Sung-cheol Park

School of Computer and Electronics Engineering, The Catholic University of  
Korea

### 요 약

본 논문에서는 컴퓨터 신조어를 빠른 시간 내에 제공하고 한국적으로 용어를 재정의 할 뿐만 아니라 효율적인 검색 인터페이스를 갖춘 온라인 컴퓨터 용어 사전을 개발하였다. 신조어를 빨리 서비스하기 위해서 FOLDOC(Free On-Line Dictionary Of Computing)의 사전을 이용하여 영문 해설을 우선적으로 제공하고, 각 용어를 한 명 이상의 번역자가 한국어로 재정의 하도록 하였다. 또한 SQL과 MS-SQL Server를 이용해서 다양한 검색 인터페이스를 제공하여 사용자가 적은 정보만으로도 원하는 용어를 손쉽게 찾을 수 있게 하였다.

### 1. 서론

컴퓨터 산업의 급속한 발전으로 컴퓨터 용어 사전은 타 사전류에 비해 사용자가 항상 최신의 용어 사전을 이해해야 할 필요성이 매우 크다. 그래서 본 논문에서는 수시로 갱신이 되는 FOLDOC[1]의 사전을 이용하여 우선 영어로 최신 용어를 서비스하고, 한 명 이상의 번역자가 새 용어를 영문 해설을 참조하여 한국어로 정의할 수 있도록 하였다.

### 2. 기존 컴퓨터 용어 사전과의 비교

본 컴퓨터 용어 사전은 타 용어 사전이 제공하지 않는 특징적인 기능이 3가지 있다. 입력 문자로 시작하는 용어 검색, 입력문자가 해설에 들어간 용어 검색, 그리고 한글화 기능이 바로 그것이다.

입력 문자로 시작하는 용어 검색과 입력문자가 들어간 용어 검색은 검색하고자 하는 용어의 정확한 스펠링을 모를 때 사용할 수 있는 기능이다. 스펠링의 일부를 입력하면 이 부분이 들어간 용어들을 검색해준다. 입력문자가 해설에 들어간 용어 검색 기능은 어떤 용어나 주제와 관련된 용어를 찾고 싶을 때 유용하게 사용할 수 있다. 이 기능을 이용하면 용어의 카테고리 영역에 제한을 받지 않고 경우에 따라서는 검색 범위도 많이 줄일 수 있다.

한글화 기능은 영문으로만 정의되어있는 신조어를 영문 정의를 참고하여 한국어로 번역하거나 재정의 할 수

있는 기능이다. 한국어로 정의되어 있는 용어도 한 명 이상의 사람이 재정의 할 수 있다. 이렇게 함으로써 다양한 관점에서의 용어 정의를 유도하고 용어정의를 한국의 상황에 맞도록 많은 사람들이 수정할 수 있도록 하였다. 입력할 때에는 입력자의 성명, 소속, 메일 주소도 기입하도록 하여 되도록 신중을 기하도록 하였다. 그리고 각 해설에 점수를 매김으로써 한 용어에 여러 개의 해설이 존재할 경우 우수한 해설을 식별하기 쉽게 하였다.

이 외에 눈여겨볼 사항은 매칭 에이전트 개념을 이용한 통계 검색 서비스이다. 본 컴퓨터 용어 사전에서는 사용자가 현재 검색한 용어에 대해 해당 용어를 검색한 사람들이 당일 하루동안 검색한 다른 용어들을 찾아준다. 이를 통해 실질적으로 관련이 있는 용어를 쉽게 할 수 있게 하였다.

본 컴퓨터 용어 사전에서는 총 9가지의 용어 검색 기능을 가지고 있다. 다양하고 효과적인 검색 기능으로 각 사용자에게 적합한 검색 서비스를 제공한다. <표 1>은 본 논문에서 구현한 사전과 기존 사전의 검색 기능을 비교한 것이다.

본 논문에서는 인터넷을 통해 서비스하고 있는 컴퓨터 용어 사전만을 비교하였고 책과 CD-ROM 형태의 사전은 제외하였다. 책의 경우에는 용어의 갱신 서비스가 이루어지지 않고, CD-ROM의 경우에는 2001년 현재 국내에 판매되는 컴퓨터 용어 사전이 없기 때문이다.

<표 1> 주요 컴퓨터 용어 사전과 본 컴퓨터 용어 사전의  
검색 기능별 비교

사전 항목	가	나	다	라	마	바	사	아
A				√	√	√	√	√
B	√					√	√	√
C		√	√		√	√	√	√
D								√
E		√						√
F								√
G	√				√	√	√	√
H	√				√	√	√	√
I							√	√
J	√	√	√		√	√	√	√
K					√	√	√	√
L	√				√	√		√

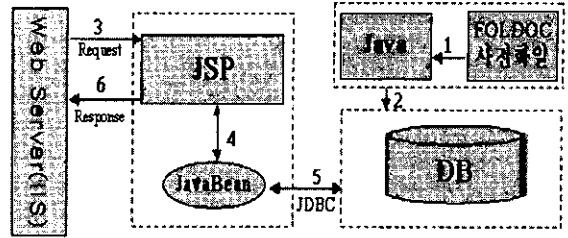
가:팀즈 코리아[2], 나:들도끼[3], 다:아이누리 컴퓨터대사전 [4], 라:EzWord[5], 마:Webopedia[6], 바:what?is[7], 사:FOLDOC, 아:가톨릭대 컴퓨터 용어 사전

A:카테고리별 검색, B:알파벳 순 검색, C:입력용어 검색, D:입력문자로 시작하는 용어 검색, E:입력문자가 들어간 용어 검색, F:입력문자가 해설에 들어간 용어 검색, G:검색용어의 설명에 포함된 컴퓨터 용어 검색, H:검색 용어와 관련된 HTTP 사이트 링크, I:검색 용어와 관련된 FTP 사이트 링크, J: 설명화면과 같은 화면에서 용어 검색, K:검색한 용어와 같은 카테고리에 속하는 용어 검색, L: 검색 통계를 이용한 용어 검색.

3. 시스템 구조

본 컴퓨터 용어 사전은 Windows NT와 R-DBMS인 MS-SQL Server를 기반으로 하여 Java와 JavaBeans 그리고 JSP(JavaServer Pages™)를 사용하였다. 웹 서버로는 IIS(Internet Information Server)를 사용하였다. (그림 1)은 시스템의 구조를 나타내고 있다.

- (1) FOLDOC 사전 파일에서 용어 스트링을 읽는다.
- (2) 읽어들이는 스트링을 분석해서 용어 명, 카테고리, 해설로 분류해서 DB에 저장한다.
- (3) 사용자가 웹 브라우저를 통해 검색을 요청하면 웹 서버는 JSP로 입력정보를 넘겨준다.
- (4) JSP는 검색 요구에 따라 해당 JavaBean에 정보를 요청한다.



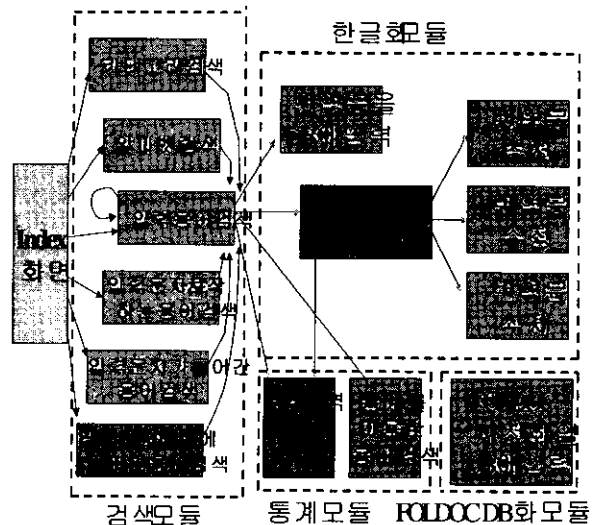
(그림 1) 시스템 구조

- (5) JavaBean은 JDBC를 통해 DB에 검색 SQL문을 보내거나 저장 프로시저(stored procedure)를 실행시킨다. DB에서 검색 결과를 전달해주면 이 결과를 정리해서 JSP로 넘겨준다.
- (6) JSP는 JavaBean으로부터 넘겨받은 검색 결과를 사용자 인터페이스 형식에 맞게 구성해서 Web Server로 보내준다. 그러면 Web Server는 사용자 웹 브라우저로 결과를 전송한다.

3.1 기능별 구성 요소

본 컴퓨터 용어 사전은 크게 네 부분으로 구성된다. FOLDOC 사전 파일 데이터베이스화 모듈, 데이터베이스에 저장된 용어 검색 모듈, 한글화 모듈, 검색 통계 모듈로 나뉜다. (그림 2)는 이 구성 요소간의 관계를 보여준다.

FOLDOC 사전 파일 데이터베이스화 모듈은 FOLDOC의 사전 파일을 데이터베이스에 입력시키는 모듈이다. 데이터베이스를 생성하거나 갱신할 때 사용한다.



(그림 2) 기능별 구성 요소

검색 모듈은 카테고리 별 검색, 알파벳 순 검색, 입력 용어 검색, 입력문자로 시작하는 용어 검색, 입력문자가 들어간 용어 검색, 입력문자가 해설에 들어간 용어 검색으로 구성된다. 입력 용어 검색은 입력된 용어 하나를 검색하고 세부 내용을 보여준다. 입력 용어 검색을 제외한 나머지 검색은 찾은 용어 리스트를 보여준다. 이 리스트에서 한 용어를 선택하면 선택된 용어 명을 입력 용어 검색 모듈로 넘겨서 해당 용어의 세부 내용을 보여준다. 세부 내용이 나온 페이지에서는 검색 용어의 번역문 검색, 해당 카테고리에 속하는 용어 검색, 해설에 포함된 컴퓨터 용어 검색을 할 수 있다. 기본적으로 세부 내용이 영어로 나오고 해당 용어가 번역되어 있을 때에는 번역문을 볼 수 있다.

한글화 모듈은 입력 용어 검색 기능과 연결되어 있다. 용어를 검색해서 하나의 용어에 대한 세부 내용이 나온 페이지의 아래쪽에는 검색된 내용을 한글로 입력할 수 있는 폼이 있다. 이 폼은 영문 정의를 번역해서 한글로 입력하거나 한국어로 재 정의할 때 사용된다. 이 폼에 각 항목을 입력하면 입력 항목들이 데이터베이스에 입력된다. 한 용어에 대해 복수 개의 한글 용어 정의가 있을 수 있으며, 각 한글 용어 정의는 비밀 번호가 맞을 경우에 한해서 수정 및 삭제가 가능하다. 그리고 한글 용어 정의는 사용자가 추천할 때마다 점수가 증가한다.

검색 통계 모듈은 검색 내역을 데이터베이스화하고 이것을 통계 처리를 하여 사용자가 검색한 용어와 관련된 용어를 사용자에게 보여준다.

#### 4. 주요 모듈의 구현

사용자의 검색 요구를 처리할 때 가장 시간이 오래 걸리는 부분이 데이터베이스와 커넥션 할 때이고 그 다음이 질의문을 전송하고 받을 때이다[8]. 그래서 본 컴퓨터 용어사전에서는 커넥션을 미리 만들어놓고 필요할 때 사용하는 DBPool이라는 기법을 썼다. 그리고 질의문 전송을 줄이기 위해서 통계 모듈에서는 저장 프로시저(stored procedure)를 호출하여 한 번의 질의문 전송으로 해결했으며 한글화 모듈에서는 엔터티 삽입시 중복 확인과 비밀번호 확인을 위한 질의문 전송을 없애는 기법을 개발해서 사용하였다.

##### 4.1 FOLDOC 사전 파일 데이터베이스화 모듈

FOLDOC 사전의 경우에는 offset방식을 이용하기 때문에 순서화의 제약을 받는다. 반면에 본 컴퓨터 용어 사전은 FOLDOC의 사전 정보를 용어명, 카테고리, 해설 이렇게 세 부분으로 데이터베이스화하였기 때문에 검색이 나 사전 업데이트 및 수정이 용이하다.

##### 4.2 한글화 모듈

한글화 모듈은 한글 용어 정의 입력 기능, 번역된 용어 검색 기능, 수정 기능, 삭제 기능, 추천 기능 등으로 구성되어있다.

한글화 모듈에서는 데이터베이스로의 SQL문 전송과 결과 값을 호출하는 쿼리를 줄이는 기법을 개발하여 처리 속도를 향상 시켰다. 한글로 정의된 용어 수와 같이 매번 사용되고 DB와는 독립적으로 계산 가능한 변수는 JSP에서 scope를 application으로 정의 해놓음으로써 JSP/Servlet 엔진이 꺼질 때 까지 계속해서 사용할 수 있게 하였다. 그리고 기본키 중복확인, 기본키 Null 확인 그리고 외래키 확인 등과 같이 데이터베이스에서 자체적으로 검사하는 사항에 대해서는 미리 SQL문을 이용 한 확인작업을 하지 않고 데이터베이스에 바로 처리 명령을 내린 다음 수행 여부로 확인을 하였다. 비밀번호 확인도 이 방법을 응용하여 수행 속도도 높이고 사용자 로 하여금 로그인하는 작업을 없앴으로써 이용의 편리성도 증대시켰다.

#### 4.3 검색 통계 모듈

사용자에 대한 정보와 검색 정보는 사용자가 검색을 할 때마다 데이터베이스에 기록해야 한다. 게다가 한번에 여러 번의 질의문 전송이 이루어져야 한다. 이 과정을 저장 프로시저로 만들어 놓음으로써 질의문 전송을 한 번으로 해결했다.

#### 5. 결론 및 향후 과제

본 논문은 Windows NT 환경에서 온라인으로 서비스되는 컴퓨터 용어 사전의 개발에 대해서 기술하였다.

본 논문의 구현 내용 및 장점은 다음과 같다. 다양하고 효과적인 검색 인터페이스를 지원함으로써 사용자의 편의를 도울 뿐만아니라 검색하려는 용어명을 정확히 알지 못해도 쉽게 찾을 수 있게 하였다. 그리고 신조어를 신속하게 서비스하기 위해서 영문으로 우선 검색할 수 있게 한 다음 한글로 번역할 수 있게 하였다. 한 용어에 대해서 한 명 이상의 사람이 정의할 수 있게 함으로써 다양한 관점에서 재정의 할 수 있도록 하였다.

앞으로 구현하거나 연구해야할 부분은 본 사전의 구현 과정에서 개발한 기법들이 효과적인지, 즉, 성능 향상에 기여를 어느 정도 하는지 확인한 후에 보완하는 작업도 필요하다. 이를 위해서는 본 사전에 적합한 성능 평가 방법에 대해서도 연구해야 할 것이다.

#### 6. 참고 문헌

- [1]. FOLDOC, <http://foldoc.doc.ic.ac.uk/foldoc/index.html>.
- [2]. 텃즈 코리아, <http://www.terms.co.kr/>.
- [3]. 돌도끼, <http://www.doldoki.org/dic/>.
- [4]. 아이누리 컴퓨터 대사전, <http://dic.inuri.com/>.
- [5]. ezWord, <http://www.ezword.co.kr/>.
- [6]. Webopedia, <http://www.pcwebopaedia.com/>.
- [7]. Whatis?com, <http://whatistechtarget.com/>.
- [8]. Danny Ayers의 14명 저, 하수정 역, Professional Java Server Programming, pp.231-304, 2000.