

# 웹 기반 정보 검색 시스템에서 의미 링크 구현 방안

엄재임\*    변정용  
동국대학교 전산정보학부

## Implementation Method of Semantic Link in Web-Based Information Retrieval System

Chae Im Um\*                      Jeong Yong Byun  
Dept. of Computer and Information Science, Dongguk University  
yumchi@wonhyo.dongguk.ac.kr                      byunjy@mail.dongguk.ac.kr

### 요 약

인터넷의 발전으로 많은 웹 기반 정보 검색 시스템의 구축은 사용자들에게 다양한 정보를 제공하고 있으나, 방대한 정보의 양을 사용자들은 효율적으로 이용하지 못하고 있다. 그러므로 본 논문에서는 웹 기반 정보 검색 시스템에서 의미에 따른 문서 검색 기법을 제공하기 위해 의미 링크에 대해 알아보고, 의미 링크를 이용한 웹 기반 정보 검색 시스템인 불교 문화재 답사시스템을 웹과 데이터베이스 연동 방식에 따라 분류하여 구현해 본다. 앞으로의 연구 방향은 일반적인 정보 검색 시스템에서 로봇을 통하여 데이터들의 의미를 파악함으로써 의미에 따른 검색 기법을 제공하는 것이다.

### 1. 서론

인터넷이 용량의 한계에 도달한 요즘, HTML 표준안의 버전 업 과정은 기술적인 한계 극복과 디자인 면에서 확장된 융통성을 제시하기 위해 노력하고 있다. 뿐만 아니라, 여러 정보 검색 시스템들도 키워드간의 연산자를 적용하거나 검색 범위를 제한함으로써 검색 효율을 높이고 있다.

그러나, 기존의 정보 검색 시스템들은 키워드 중심의 검색을 하고 있기 때문에 문서 내에서 키워드의 발생 빈도에 따라 검색[5]을 하게 된다. 이러한 검색 기법은 문서에 키워드가 나타나더라도 문서 전체를 의미하는 내용이 키워드와 일치하지 않을 경우 많은 정보 검색 시스템 사용자에게 불편을 느끼게 한다.

그러므로, 문서에는 키워드가 나타나지 않더라도 문서 전체가 나타내고 있는 의미에 따른 검색 기법

이 요구된다. 의미에 따른 검색 기법의 적용은 일반적인 정보 검색 시스템과 특정 정보 검색 시스템으로 나누어 생각해 볼 수 있다. 일반적인 정보 검색 시스템에서는 검색 엔진이 문서의 의미를 추론하게 함으로써 가능하고, 특정 정보 검색 시스템에서는 데이터베이스 구성시 문서를 의미에 따라 분류함으로써 가능하다.

본 논문에서는 특정 정보 검색 시스템에서 한 가지 대상에 대하여 다양한 의미에 따라 문서를 세분화하고, 사용자 검색 요구시 사용자가 요구하는 의미에 따라 세분화된 문서에 대한 링크 경로를 결정하는 방법을 알아본다. 논문의 구성은 2장에서 의미 링크에 대해 알아보고, 3장에서는 CGI, JDBC를 이용한 의미 링크 기반 정보 검색 시스템의 구현에 대해 설명하고, 4장에서 결론을 맺는다.

## 2. 의미 링크

본 논문에서는 문서의 의미에 따른 검색 기법을 제공하기 위해서 의미 링크(semantic link)를 이용하여 정보 검색 시스템[2,3,4]을 구현했다. 의미 링크 또는 동적 링크(dynamic link)는 웹 기반 정보 검색 시스템에서 데이터베이스 구성시 문서 또는 멀티미디어 데이터를 의미에 따라 분류하고, 사용자 검색 요구시 사용자가 요구한 키워드와 의미에 따라 키워드에 대한 연결 경로를 동적(dynamic)으로 설정한다. 즉, 사용자가 요구하는 의미에 따라 하나의 대상에 대해 다양한 문서가 연결될 수 있다.

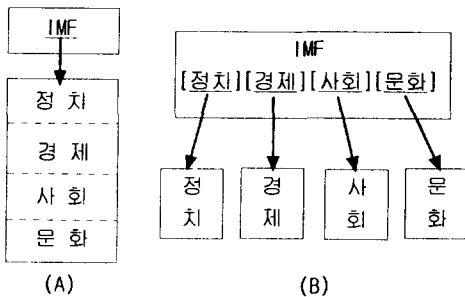


그림 1 고정 링크

현재 많은 웹 문서에서 하나의 대상은 사람의 지식 수준이나 취향에 따라 다양한 의미의 설명문을 가질 수 있다. 이 때 기존의 하이퍼링크를 이용한다면 그림 1(A)와 같이 하나의 대상에 대해 하나의 연결만을 설정할 수 있기 때문에 링크 되는 문서가 혼합된 의미를 가지고 있어야 한다. 그러나 하나의 문서에 모든 내용을 포함시킬 경우 웹 문서 구성이 복잡하고 불필요한 정보가 많아 사용자가 요구하는 문서에 대한 만족도가 저하된다.

```
<html>
정부가 3년 안에 IMF
<a href="path/정치/imf.html">[정치]</a>
<a href="path/경제/imf.html">[경제]</a>
<a href="path/사회/imf.html">[사회]</a>
<a href="path/문화/imf.html">[문화]</a>
체제를 졸업한다는 중기 경제회생 계획을 마련했습니다. </HTML>
```

그림 2 고정 링크를 제공하는 문서

또, 그림 1(B)처럼 문서를 의미에 따라 세분화하

더라도 고정 링크를 이용하게 될 경우, 그림 2와 같이 각 의미에 따라 각각의 연결 경로를 설정해야 하기 때문에 연결 경로가 변경되거나 새로운 의미를 추가할 경우 각각의 HTML문서를 수정해야 하는 문제점이 있다.

고정 링크에 반하여 의미 링크는 문서들을 의미에 따라 분류하였기 때문에 사용자가 요구하는 문서에 대한 만족도를 증가 시키며, 검색 요구시 문서에 대한 연결이 설정되기 때문에 미리 HTML문서로 작성해 놓을 필요가 없다. 즉, 기존의 정보 검색 시스템은 'IMF'와 '경제'라는 키워드를 입력할 때 문서 내에서 이 키워드들을 찾아 검색의 정확도를 낸다. 이것은 문서내에는 해당 키워드가 포함되어 있지만, 문서의 전체적 의미가 키워드와 일치하지 않는 경우 사용자 만족도를 저하 시킨다. 반면, 의미 링크는 그림 3과 같이 데이터베이스 구성시 '경제'라는 의미에 대해 문서를 분류해 두었기 때문에 사용자가 원하는 분야의 문서를 정확하게 찾을 수 있다.

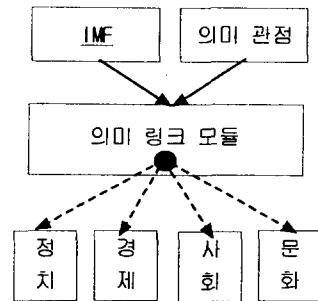


그림 3 의미 링크

의미 링크는 고정 링크의 문제점들을 최소화하고 하이퍼링크의 융통성을 최대화하기 위해 HTML 문서의 고정 링크를 제공하는 앵커(anchor)와 의미 링크 모듈을 사용한다. 의미 링크 모듈은 사용자 요구 시점에 그림 4와 같은 텍스트 문서에서 키워드를 검출하고 고정 링크의 앵커를 이용하여 의미 링크를 제공한다. 그러므로 텍스트 문서 작성자는 그림 2의 고정 링크 경우와는 달리 텍스트 내에서 키워드 인식 표기만 해주면 된다.

```
정부가 3년 안에 [IMF] 체제를 졸업한다는 중기
경제회생 계획을 마련했습니다.
```

그림 4 의미 링크를 위한 문서

의미 링크 모듈은 그림 5와 같이 일반 텍스트 내에서 키워드 인식 표기를 인식하여 키워드를 검출하고 데이터베이스에서 키워드에 대한 기본키값을 검색하여 파일명을 만든다. 초기 검색시 사용자가 요구한 의미에 따라 각각의 의미에 대한 디렉토리에서 검색 대상에 대한 파일명을 찾는다. 의미에 대한 디렉토리와 파일명을 이용해 검색 대상에 대한 연결 경로를 설정함으로써 텍스트 문서를 HTML문서로 변환시켜 연결을 제공하게 된다.

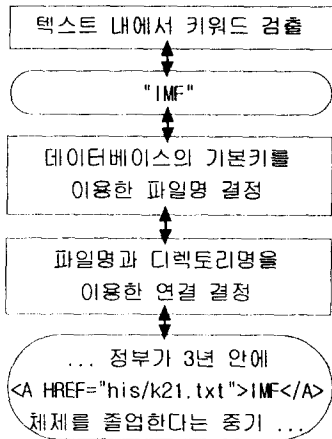


그림 5 의미 링크 모듈 처리 절차

### 3. 구현 방안

의미 링크 모듈을 웹 기반 정보 검색 시스템에 적용하여 구현하는 방안은 데이터베이스와 연동하는 방법에 따라 크게 두 가지로 나누어 볼 수 있다. 첫째, CGI(Common Gateway Interface)를 이용한 구현은 지금까지 웹과 데이터베이스를 연동하는 가장 일반적인 방법으로, HTML FORM을 사용하여 쉽게 사용자 질의를 데이터베이스에 전달할 수 있다. 둘째, JDBC(Java DataBase Connectivity)를 이용한 구현은 동적 웹 구성을 위해 자바가 필수적으로 요구되고, CLI(Call Level Interface)를 따르기 때문에 쉽게 데이터베이스와 연동할 수 있다.

본 논문에서 구현에 직접 적용된 시스템은 불교 문화재 답사 시스템[1,6]이다. 각 문화재는 역사, 미술, 종교 세 가지 의미로 세분화하여 데이터베이스를 구성하였고, 각각 그림, 오디오, 비디오 등의 멀티미디어 정보를 가지고 있다. 또한 효율적인 정보 검색을 위해 기본 정보를 데이터베이스로 구축하고, 데이터베이스의 기본키를 이용하여 각 관점 및

멀티미디어 정보에 대해 디렉토리 파일 시스템으로 구축된 파일들을 검색 하도록 했다.

#### 3.1 CGI를 이용한 구현

CGI를 이용하여 구현된 불교 문화재 답사 시스템은 2-티어 구조로, 클라이언트(client)인 웹 브라우저에서 웹 서버(server), DBMS, ESQ/L/C를 사용하는 CGI 프로그램이 있는 서버로 사용자 요구가 전달된다. 웹 브라우저는 PC상에서 Netscape 4.x를 사용하였으며 Unix System인 서버 측의 웹 서버는 Apache 1.1, DBMS는 Oracle 7.2를 사용하였다.

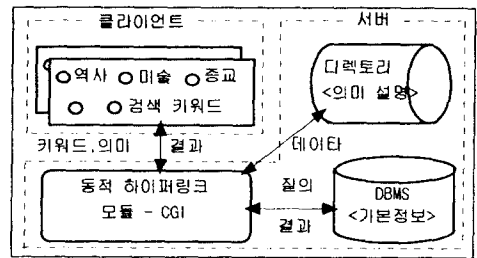


그림 6 CGI 기반 시스템 구성

그림 6의 불교 문화재 답사 시스템은 클라이언트의 HTML 문서에서 사용자 검색 요구시 검색 키워드와 관점을 CGI 모듈에 넘겨주고, CGI 모듈은 데이터베이스 및 디렉토리 파일 시스템과 연동한다. 모듈은 데이터베이스에서 키워드에 대해 검색한 후, 검색 대상에 대한 리스트를 보여주고 사용자 요구시 데이터베이스에서 기본 정보, 디렉토리 파일 시스템에서 검색 대상에 대한 사용자 요구 관점 설명 파일을 검색하여 HTML 문서에 보여주게 된다.

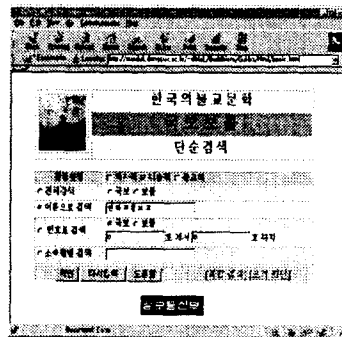


그림 7 초기 화면

그림 7의 초기 화면에서 사용자가 연화교칠보교에 대한 검색을 원할 때 초기화면으로부터 받아들이는 입력 자료는 미술적 관점의 연화교칠보교를 이름으로 검색하는 것이다.

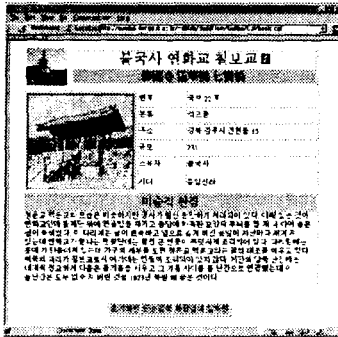


그림 8 검색 결과

초기 화면에서 사용자로부터 얻은 정보중 '이름으로 검색'과 '연화교칠보교'를 이용하여 데이터베이스 테이블에서 키워드 검색을 하고 검색 결과 '불국사 연화교칠보교'에 기본 정보를 그림 8의 상단과 같이 보여준다. 그리고 미술적 관점의 디렉토리에서 '불국사 연화교칠보교'에 대한 미술적 관점의 설명문과 함께 사진을 보여주며, 오디오, 비디오 파일은 아이콘으로 표기한다.

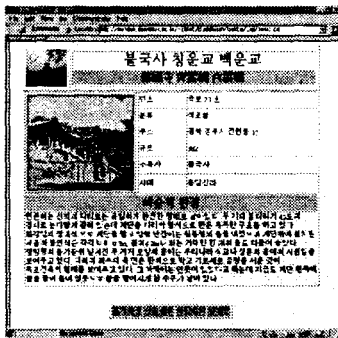


그림 9 텍스트내 연결에 의한 검색 결과

그림 8의 미술적 관점 설명문 내에는 '청운교백운교'에 대한 연결이 다시 설정되어 있는데, 이것이 CGI 의미 링크 모듈에서 키워드를 인식하여 연결을 동적으로 설정해 주는 것이다. 즉, '불국사 청운교백

운교'에 대한 링크가 '불국사 연화교칠보교'에 대한 미술적 관점의 설명문이 작성될 때에 설정되는 것이 아니라 검색에 의해 설명문이 보여질 때 설정되는 것이다. 또한 텍스트 내의 키워드 인식에 의해 설정되는 링크는 초기 화면에서 받아들인 미술적 관점에 의해 그림 9와 같은 결과를 보인다.

### 3.2 JDBC를 이용한 구현

JDBC를 이용하여 구현된 불교 문화재 답사 시스템은 3-티어 구조로 클라이언트, 미들 티어, 서버로 구성된다. 클라이언트의 웹 브라우저는 PC상에서 Netscape 4.x를 사용하였으며, 미들 티어의 웹 서버는 Apache 1.1, 모듈은 Java 1.3, JDBC Driver는 Oracle사의 JDBC Driver 7.3 thin driver를 사용하였다. 서버의 DBMS는 Oracle 7.2를 사용하였다.

그림 6과 같이 일반 사용자에게 해당하는 클라이언트는 사용자 인터페이스의 향상을 위해 웹 브라우저에서 실행할 수 있는 애플릿(applet)으로 개발하며 애플릿에서 사용자로부터 의미와 키워드를 받아들여 미들 티어에 보낸다.

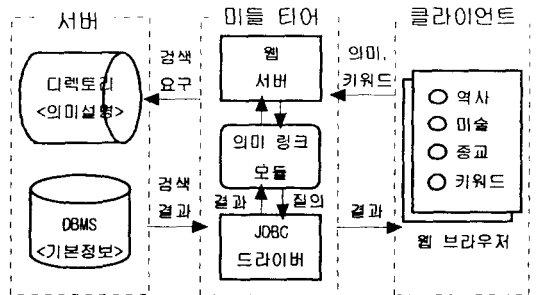


그림 6 JDBC 기반 시스템의 구성

미들 티어는 실제 데이터 처리에 대한 의미 링크 모듈이 수행되는 부분으로 의미 링크 모듈은 자바 어플리케이션으로 개발한다. 이는 사용자에게 실제 수행 모듈을 감춤으로써 CGI에 비해 보안의 문제점들을 해결할 수 있다. 의미 링크 모듈은 직접 JDBC 드라이버와 질의에 대한 정보를 주고받음으로써 다양한 데이터베이스와의 연동이 가능해진다.

서버에는 데이터베이스와 디렉토리 파일 시스템이 있으며 DBMS는 질의에 대한 모든 정보를 미들 티어의 JDBC 드라이버에게 넘겨주고 디렉토리 파일 시스템은 의미 링크 모듈과 직접 연동 한다.

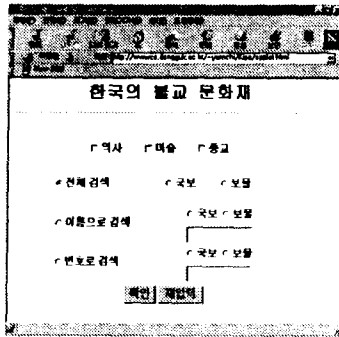


그림 10 초기 화면

그림 10의 초기 화면에서 그림 7의 경우와 같은 사용자 요구를 받아들일 때 애플릿은 객체화된 '미술', '연화교칠보교', '이름으로 검색' 데이터를 미들-tier의 어플리케이션으로 보낸다. 미들-tier의 자바 어플리케이션 의미 링크 모듈에서는 데이터베이스와 디렉토리 파일 시스템에 연동한 결과를 그림 8과 같은 결과 화면의 HTML문서를 만들고, 이 문서를 클라이언트에서 로드할 수 있게 한다. 클라이언트에서 텍스트내 '칭운교백운교'에 대한 설명문을 요구할 시 그림 9와 같은 화면을 보여준다.

#### 4. 결론

지금까지 문서의 의미에 따른 검색 기법을 특정 정보 검색 시스템에 적용함으로써 사용자가 요구하는 의미의 문서를 정확하게 검색할 수 있는 방안을 보았다. 구현에는 하나의 문화재에 대해 사용자가 여러 가지 의미를 요구할 수 있고, 설명문, 사진, 오디오, 비디오 등의 다양한 데이터를 가지는 불교 문화재 답사 시스템을 적용하였다.

구현 방안은 웹과 데이터베이스의 연동 방식에 따라 CGI 방식과 JDBC 방식으로 구현하였는데, CGI 방식은 간단한 시스템 구성을 가지지만, 의미 링크 모듈이 서버에서 실행되기 때문에 시스템의 부하를 가져올 수 있다. 즉, 사용자가 클라이언트에서 검색을 한 번할 때마다 서버의 데이터베이스와 연결을 설정, 해제해야 하기 때문에 시스템의 부담이 크다. 반면, JDBC 방식은 자바 언어가 개발 단계이기 때문에 속도가 느리고 3-tier의 복잡한 시스템 구성을 가지고 있지만, 의미 링크 모듈이 미들-tier에서 수행되고 인터페이스는 클라이언트에서 실행되기 때문에 시스템의 부하가 감소되고 보안의 문제도 해결할

수 있다. 즉, JDBC는 데이터베이스와의 연동에 있어서 한 번 설정된 연결을 유지할 뿐만 아니라 분산 환경에서 각각의 tier를 수행하기 때문에 시스템의 부담을 줄여준다.

앞으로의 연구 방향은 문서의 의미에 따른 검색 기법을 문서뿐만 아니라 다른 데이터들 즉, 그림, 비디오, 오디오 등의 의미를 파악하여 검색할 수 있는 기법과 정보 검색 시스템에 대한 적용 방안중 로봇과 같은 검색 엔진이 데이터들의 의미를 파악할 수 있게 하는 것이다.

#### 참고 문헌

- [1] 변정용, 한국의 불교 문화 유물 유지 답사 시스템 개발, 정보통신부, 1997.
- [2] 엄채임, 변정용, "동적 하이퍼링크를 이용한 웹 기반 정보 검색 시스템", 한국정보과학회 데이터베이스연구회 동계학술대회논문집, 제14권, 제1호, pp.209-214, 1998.
- [3] 엄채임, 변정용, 김성욱, "의미 링크를 이용한 JDBC 기반 정보 검색 시스템", 한국정보처리학회 춘계 학술발표논문집 CD.
- [4] 엄채임, 변정용, "동적 하이퍼링크를 이용한 웹 기반 정보 검색 시스템의 성능 평가", 한국정보과학회 봄 학술발표논문집, 제25권, 제1호, pp.3-5, 1998.
- [5] 오종인, 백준호, 최준혁, 이정현, "사용자 중심의 웹 정보검색 시스템 설계", 한국정보과학회 봄 학술발표논문집, 제24권, 제1호, pp.425-428, 1997.
- [6] URL 불교문화재 답사 시스템  
<http://dmlab.dongguk.ac.kr/Buddhism>