

BRS/Search 시스템을 이용한 XML 문서 검색시스템 설계 및 구현

Design and Implementation of a XML Document Retrieval System Using the BRS/Search System

손 충 범* 이 병 업** 유 재 수***
Chung-beom Son Byoung-Yup Lee Jae-Soo Yoo

요 약

본 논문에서는 상용 검색 엔진인 BRS/Search 시스템을 이용하여 XML 문서에 대한 구조 기반 검색이 가능하도록 XML 문서 검색시스템을 설계하고 구현한다. 본 논문에서 구현한 시스템은 XML 문서의 논리적인 구조를 Unix 파일시스템의 디렉토리 구조로 표현한다. 이렇게 표현된 XML 문서를 BRS/Search 시스템의 데이터베이스에 정보의 손실없이 저장하기 위해 데이터베이스 스키마를 정의하고 내용 검색 뿐만 아니라, 구조 검색, 혼합 검색, 애트리뷰트 검색이 가능하도록 BRS/Search 시스템에 부가적으로 필요한 ETID 추출기, 구조정보 추출기, 저장 관리기, 질의 처리기를 설계하고 구현한다

Abstract

In this paper, we design and implement a XML document retrieval system to support structure-based retrieval using the BRS/Search system that is a commercial search engine. The implemented system in this paper represents the logical structure of XML documents as the directory structure of the Unix file system. In addition, we define the database schema of BRS/Search system to store documents. We also implement a ETID extractor, a structure information extractor, a storage manager and a query processor additionally to support content retrieval, structure retrieval, mixed retrieval and attribute retrieval in the BRS/Search system.

1. 서 론

최근의 웹(Web) 또는 디지털 전자 도서관 시스템들과 같은 환경에서 많은 문서들이 SGML(Standard Generalized Markup Language) / XML(eXtensible Markup Language)과 같은 마크업 언어로 표현된 문서의 포맷으로 바뀌어 가고 있다. 따라서 이러한 언어로 문서들을 표현함으로써 문서의 논리적인 구조를 표현할 수 있고, 이러한 문서들을 저장 및 검색함으로써 기존의 키워드 기반의 검색과는

달리 문서의 논리적인 구조와 내용에 기반한 검색에 대한 연구가 많이 연구되고 있다[1][2][3][4].

기존의 웹 환경에서의 정보 표현 방법인 HTML(Hyper Text Markup Language)은 인터넷의 급속한 발전의 계기가 되었다. 그러나 HTML의 정보 표현의 한계와 HTML의 모태인 SGML은 문서의 논리적인 구조를 표현할 수는 있지만 그 구성이 너무 복잡한 단점을 가지고 있다. 이에 부합하여 기존의 HTML과 SGML의 단점을 보완하고 장점만을 수용한 1997년 12월에 제안된 XML(eXtensible Markup Language) [5]은 현대 정보화 사회의 다양한 문서들을 기존의 표현 방법보다 더 쉽게 문서의 논리적인 구조와 내용을 구체적으로 표현할 수 있는 차세대 웹 문서 포맷으로서 부각되고 있다[6].

* 준회원 : 충북대학교 대학원 정보통신공학과 박사과정
cbson@trut.chungbuk.ac.kr

** 비회원 : 대우정보시스템 e-솔루션사업팀 차장
bylee@disk.co.kr

*** 정회원 : 충북대학교 정보통신공학과 부교수
yjs@cbucc.chungbuk.ac.kr

이러한 XML에 의해 작성된 문서의 폭발적인 증가가 예상되고 있다. 따라서 XML 문서를 효율적으로 저장하고 XML 문서의 논리적 구조에 기반한 검색에 대한 연구는 매우 중요하다.

본 논문에서는 XML 문서를 효율적으로 저장하고 검색하기 위해 BRS/Search 시스템을 이용한 XML 문서 검색시스템을 설계하고 구현한다. 이를 위해 XML 문서의 논리적인 구조를 ETID, SORD, SSORD의 계층적인 디렉토리 구조로 표현한다. 또한, 이러한 구조 정보를 BRS/Search 시스템의 데이터베이스에 정보의 손실없이 저장하기 위해 데이터베이스에 대한 스키마를 정의하였다. 뿐만 아니라, 구조 정보를 추출하는 구조 정보 추출기, 추출된 구조 정보를 효율적으로 저장하기 위한 저장 관리자, 구조 질의를 처리하기 위한 질의 처리기를 구현함으로써 BRS/Search 시스템의 검색 기능을 확장하여 구조 검색, 혼합 검색, 애트리뷰트 검색 및 내용 검색이 가능하도록 한다.

본 논문의 구성은 다음과 같다. 2절에서는 구조화 문서에 대한 기존 연구에 대해 기술하고 3절에서는 본 시스템의 구현에 있어서 고려사항을 기술하고 본 논문에서 제시하는 XML 문서의 표현 형태 및 스키마 구조를 제시하고 가능한 질의 형태를 분석한다. 4절에서는 XML 문서 검색시스템을 구성하는 핵심 구성 요소인 ETID 추출기, 구조 정보 추출기, 저장 관리자, 질의 처리기 및 사용자 인터페이스에 대해 기술한다. 5절에서는 구현한 시스템의 구현 결과를 제시하고 성능 평가 내용을 기술한다. 마지막으로 결론 및 향후 연구 방향을 제시한다.

2. 관련 연구

최근 들어 SGML/XML 형태로 표현된 문서의 증가로 이러한 문서를 저장하고 검색하기 위한 전문 정보 검색시스템 및 데이터베이스에 대한 연구가 활발히 이루어지고 있다. 이러한 연구들이 데이터베이스와 정보 검색분야에 있어서 효

율적인 저장 관리기와 인덱스 구조에 대한 많은 연구가 진행되고 있다. 이러한 연구에서 다양한 DBMS를 저장 관리기로 이용하고 SGML 문서와 XML 문서를 저장하기 위해 논리적인 구조로 표현된 문서를 분할 모델, 비분할 모델, 혼합 모델로 문서를 저장하고 검색하는 방법이 연구되었다[7][8]. 그러나 DBMS를 저장 관리기로 이용할 경우 DBMS가 제공하는 다양한 기능을 이용하여 문서에 대한 갱신, 삭제 등 문서를 관리하기에 효율적인 장점이 있지만, 일반적으로 길이가 가변적인 문자열에 대한 색인 생성이 비효율적이기 때문에 검색 속도가 느리다는 단점이 있다. 하지만, 기존의 텍스트 기반 검색시스템은 문서의 구조를 다양하게 표현할 수 없어 문서의 관리 측면에서 효율적이지 못하지만 문서 내용에 대한 검색 측면에서 DBMS에 비해 텍스트 기반 검색시스템이 좋은 성능을 보인다[9][16]. 그러나 각 엘리먼트에 대한 관계를 검색하기 위해 문서의 논리적인 구조를 K-ary 트리로 표현할 경우 구조에 기반한 검색 시 연산 오버헤드와 문서의 구조가 변경될 경우 트리를 재구성해야 하는 문제점이 발생한다[1][10].

본 논문에서는 기존 연구[16]에서 DBMS보다 정보 검색시스템(IRS : Information Retrieval System)이 검색 속도면에서 빠르다는 연구결과를 가지고, 현재 연구소, 기관, 학교에서 많이 사용되고 있는 검색엔진인 BRS/Search 시스템을 이용하여 일반 정보 검색시스템이 가지는 장점을 이용한다. 하지만 기존에 대부분의 검색엔진들이 XML 문서에 대한 구조에 기반한 검색을 초보적인 형태로 지원하고 있다. 따라서 본 논문에서는 BRS 검색엔진에 구조에 기반한 검색을 지원하기 위해 구조 정보 추출기, 구조 질의 처리기를 구현하여 추가함으로써 XML 문서에 대한 다양한 구조에 기반한 색인인 조상, 형제, 부모, 자식관계를 포함하는 구조 검색, 혼합 검색, 애트리뷰트 검색을 지원하도록 설계하고 구현하였다. 또한, 구현한 시스템의 구조에 기반한 검색을 검증하기 위해 성능평

가를 실시하였다.

3. XML 문서 검색시스템 설계

3.1 고려 사항

본 논문에서는 BRS/Search 시스템을 이용하여 XML 문서 검색시스템을 설계할 때 고려 사항은 다음과 같다.

첫 번째는 XML 문서의 각 엘리먼트 간의 관계를 효율적으로 검색하기 위해 문서의 논리적인 구조를 디렉토리 구조로 표현한다.

두 번째는 XML 문서의 구조 정보를 손실없이 저장하고 구조에 기반한 질의가 가능하도록 데이터베이스 스키마를 설계한다.

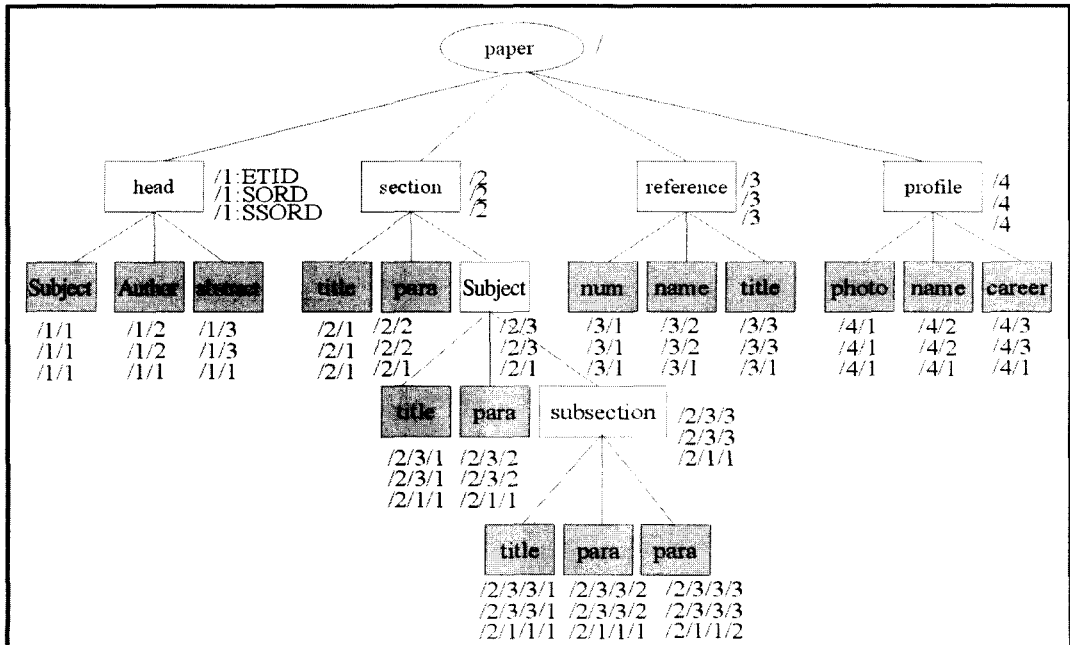
세 번째로 XML 문서와 문서로부터 추출된 구조 정보를 효율적으로 저장하고 관리할 수 있는 저장 관리기가 필요하다.

마지막으로 저장된 XML 문서에 대해 구조 질

의 및 혼합 질의, 애트리뷰트 질의에 대해 질의를 처리할 수 있는 질의 처리기의 설계가 고려되어야 한다.

3.2 XML 문서 표현 방법

본 논문에서 표현한 문서의 논리적인 구조는 특정 엘리먼트 형을 구별할 수 있고 엘리먼트 간의 계층 정보를 표현하는 디렉토리 구조의 ETID (Element Type ID)로 표현된다. 이 ETID는 XML 문서의 DTD를 분석한 후 각 엘리먼트 형(type)에 유일한 값을 할당한다. 또한 엘리먼트 간의 순서 정보(SORD : Sibling Order)의 표현은 XML 문서 인스턴스에 적용함으로써 동일 부모를 갖는 엘리먼트들의 출현 순서와 동일 부모를 갖는 엘리먼트들 중 동일한 형(type) 간의 순서(SSORD : Same Sibling Order)로 표현한다. 제안하는 문서 구조 표현 방법은 XML 문서를 디렉토리 구조로 표현함으로써 문서 본연의 논리적인 계층 구조를 표



(그림 1) XML문서의 논리적 구조 표현 방법

현할 수 있고 문서의 확장성을 고려하여 엘리먼트 간의 관계를 효율적으로 검색할 수 있다. 그림 1은 XML 문서를 트리 형태로 표현하고 각 엘리먼트에 대한 구조 정보인 ETID와 SORD, SSORD를 할당하는 예이다.

3.3 스키마 설계

XML 문서 검색시스템은 BRS/Search 시스템을 이용하여 XML 문서에 대해 효율적으로 저장하고 구조에 기반한 검색을 지원하도록 스키마를 설계하는 것이 중요하다. 특히 엘리먼트 데이터베이스는 각 엘리먼트에 해당하는 구조 정보를 저장한다. 따라서 구조 정보는 BRS/Search 시스템에 의해 자동 색인됨으로써 구조에 기반한 질의에 대해 검색 속도의 성능 향상을 보장 할 수 있다.

따라서, 본 논문에서는 사용자가 작성한 DTD와 XML 문서, 그리고 구조 정보 추출기로부터 추출된 각각의 엘리먼트 정보와 애트리뷰트 정보

를 BRS/Search 시스템의 데이터베이스에 저장하기 위해 DTD(Document Type Definition) 데이터베이스, XML 인스턴스 데이터베이스, 엘리먼트 데이터베이스, 애트리뷰트 데이터베이스의 스키마를 정의한다. 정의된 각각의 스키마 구조는 그림 2와 같다.

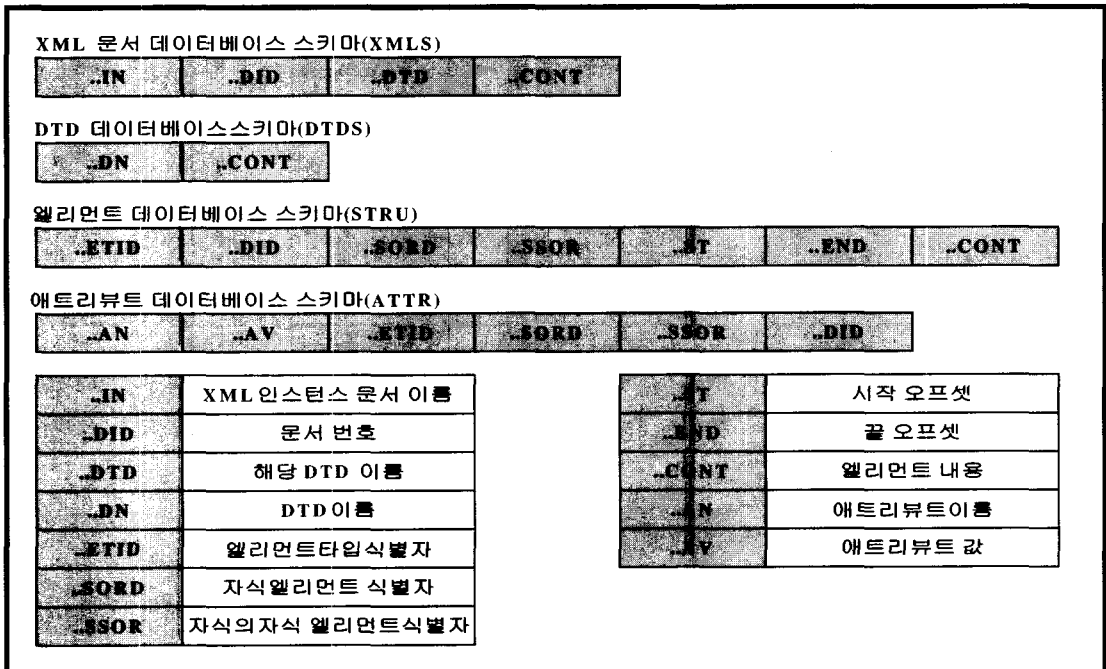
3.4 질의 분석

XML 문서의 특성을 고려할 때 다음과 같은 검색 조건을 사용자가 질의 할 수 있어야 한다. XML 문서에 대해 내용 질의, 구조 질의, 혼합 질의, 애트리뷰트 질의가 가능해야 한다.

다음은 XML 문서에 대한 질의의 예이다.

◎ 구조 질의

- (1) 특정 엘리먼트를 찾으시오.
- (2) 특정 엘리먼트에 대해 부모, 자식, 형제 엘리먼트를 찾으시오.



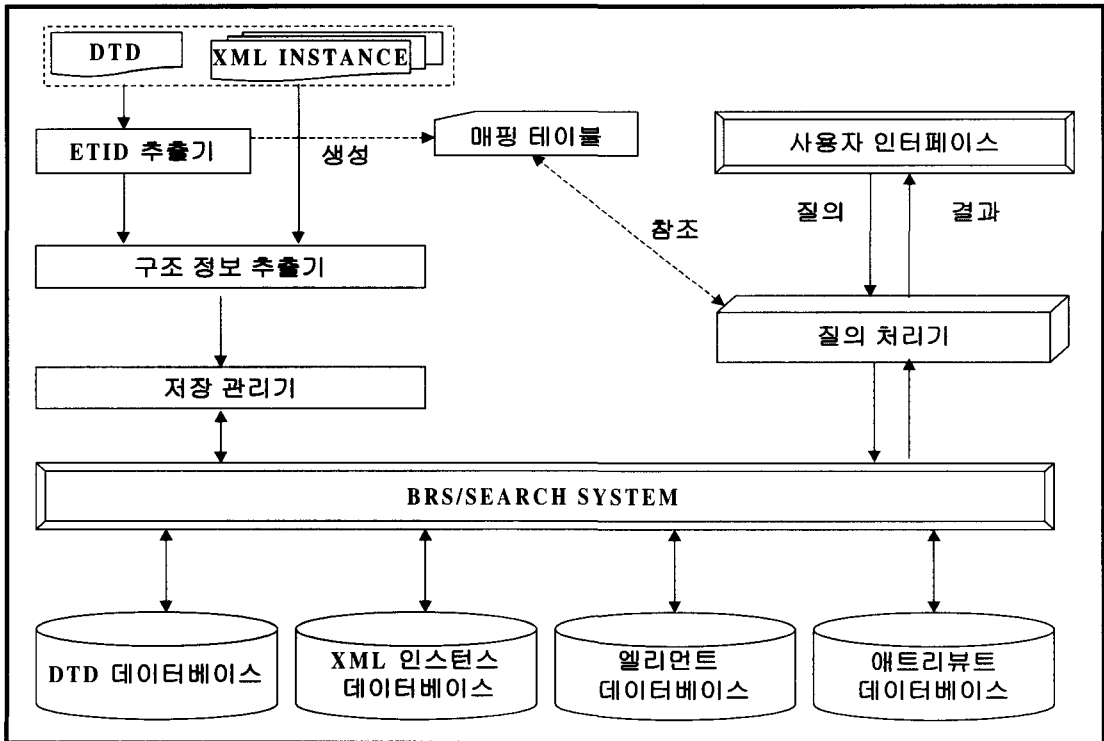
(그림 2) 데이터베이스 스키마 구조

- ◎ 혼합 질의
 - (1) 엘리먼트에 “키워드”라는 단어를 포함하는 특정 문서를 찾으시오.
 - (2) “키워드”라는 단어를 포함하는 엘리먼트의 부모, 자식, 형제 엘리먼트를 찾으시오.
- ◎ 애트리뷰트 질의
 - (1) 애트리뷰트의 이름이 RID인 엘리먼트를 찾으시오.
 - (2) 애트리뷰트의 값이 “yes”인 엘리먼트를 찾으시오.
 - (3) 애트리뷰트의 이름이 RID이고, 값이 “yes”인 엘리먼트를 찾으시오.
- ◎ 일반 질의
 - (1) “키워드”라는 단어를 포함하는 문서를 찾으시오.

4. XML 문서 검색시스템 구현

4.1 전체 시스템 구성

본 논문에서 구현한 XML 문서 검색시스템의 전체 구성은 그림 3과 같다. 구현한 시스템은 크게 ETID 추출기, 구조 정보 추출기, 저장 관리기, 질의 처리기로 구성된다. ETID 추출기는 DTD의 엘리먼트에 유일한 식별자를 부여하여 엘리먼트 형과 구조 정보를 매핑 테이블에 저장하고 유지한다. 구조 정보 추출기에서는 추출된 ETID 정보를 갖는 매핑 테이블을 이용하여 각각의 XML 문서에서 각 엘리먼트와 애트리뷰트에 대한 구조 정보를 생성한다. 이렇게 생성된 각각의 정보와 XML 문서는 저장 관리기에서 BRS/Search 시스템의 데이터베이스 스키마 정의에 부합하게 변환되어 해당 데이터베이스에 정보의 손실없이 일괄적으로



(그림 1) 전체 시스템 구성

저장된다. 질의 처리기에서는 사용자 인터페이스를 통하여 XML 문서에 대한 내용 질의, 구조 질의, 혼합 질의, 애트리뷰트 질의를 분석하고 처리하여 질의 결과를 사용자 인터페이스로 반환한다.

구현한 XML 문서 검색시스템의 구현 환경은 BRS/Search 버전 6.3 시스템과 SUN Solaris 2.5.1 환경에서 C 언어로 구현하였다.

4.2 ETID(Element Type ID) 추출기 및 구조 정보 추출기

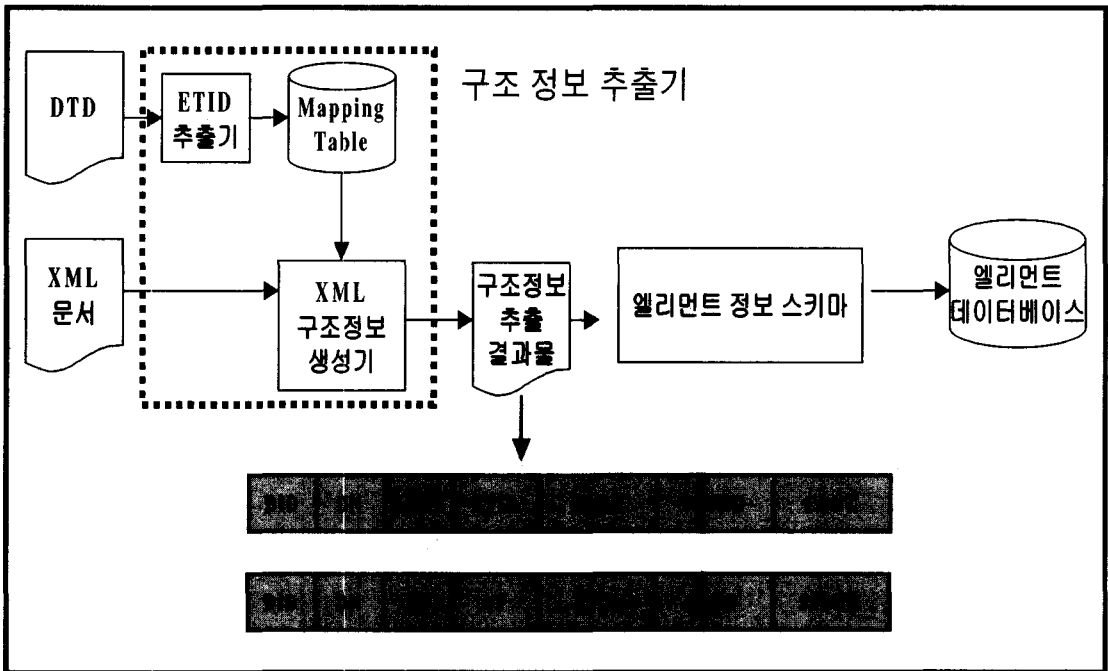
본 논문의 XML 문서 검색시스템은 ETID 추출기와 구조 정보 추출기를 이용하여 XML 문서의 기본 단위인 엘리먼트 정보를 추출함으로써 구조 질의와 혼합 질의의 검색에 유용한 정보를 생성한다. ETID 추출기에서는 작성된 XML 문서의 DTD로부터 각 엘리먼트들의 형과 이에 해당하는 유일한 식별자인 ETID 정보를 추출하여 매핑 테이블을 생성한다. 구조 정보 추출기는 XML 인스턴스와 매핑 테이블을 참조하여 각 문서의 엘리먼트에 대한 구조 정보인 ETID(Element Type ID), SORD와 SSORD를 생성하고, 이에 해당하는 내용과 애트리뷰트 정보를 추출하여 혼합 질의와 애트리뷰트 질의에 대해 효율적인 검색을 가능하게 한다.

그림 4는 DTD와 XML 인스턴스로부터 ETID와 구조 정보가 추출되는 과정을 보여 준다. 또한 구현한 시스템에서 사용되는 구조 정보 추출 결과를 제시한다.

그림 4는 DTD와 XML 인스턴스로부터 ETID와 구조 정보가 추출되는 과정을 보여 준다. 또한 구현한 시스템에서 사용되는 구조 정보 추출 결과를 제시한다.

4.3 저장 관리기

XML 문서 검색시스템의 저장 관리기에서는 사용자가 작성한 DTD와 XML 인스턴스, 구조 정보 추출기에서 생성된 엘리먼트 정보와 애트리뷰트 정보들을 BRS/Search 시스템의 각 데이터베이스의 스키마 정의에 부합하도록 변환하고 해당 필드 구분자를 추가하여 각 데이터베이스에 저장



(그림 4) ETID 추출기 및 구조 정보 추출기 구성

한다.

또한 저장 관리기는 변환된 문서들을 BRS/Search 시스템의 데이터베이스에 저장하기 전에 저장될 각각의 DTD, XML 인스턴스, 엘리먼트 정보, 애트리뷰트 정보가 서로가 관련된 문서라는 것을 식별하기 위해서 동일한 DID를 부여하고 관리한다.

4.4 질의 처리기

본 논문에서는 XML 문서의 구조와 내용에 기반한 검색을 지원하도록 BRS/Search 시스템의 검색 엔진을 이용하여 구조 질의, 혼합 질의, 애트리뷰트 질의 및 내용 질의를 처리하는 질의 처리기를 구현하였다. 구현한 질의 처리기는 질의 유형을 분석하여 각각의 데이터베이스에 저장된 XML 인스턴스 정보, 엘리먼트 정보, 애트리뷰트 정보와 매핑 테이블 정보에 기반하여 질의에 적합한 데이터베이스 및 매핑 테이블 정보를 검색하여 질의에 만족하는 검색식을 생성하여 질의를 처리한다.

다음은 각 질의에 해당하는 질의 처리기의 처리과정을 기술한다.

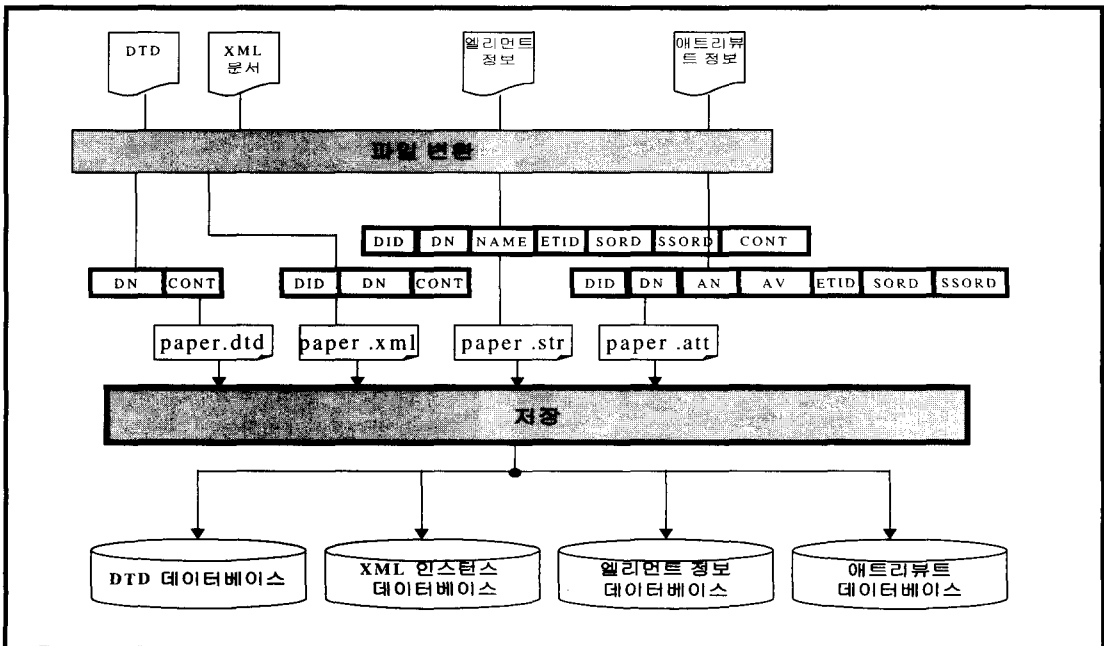
1) 구조 질의

구조 질의는 매핑 테이블의 엘리먼트 이름과 ETID 정보를 참조하여 엘리먼트 상호간의 관계를 검색하고 결과를 사용자에게 제시한다.

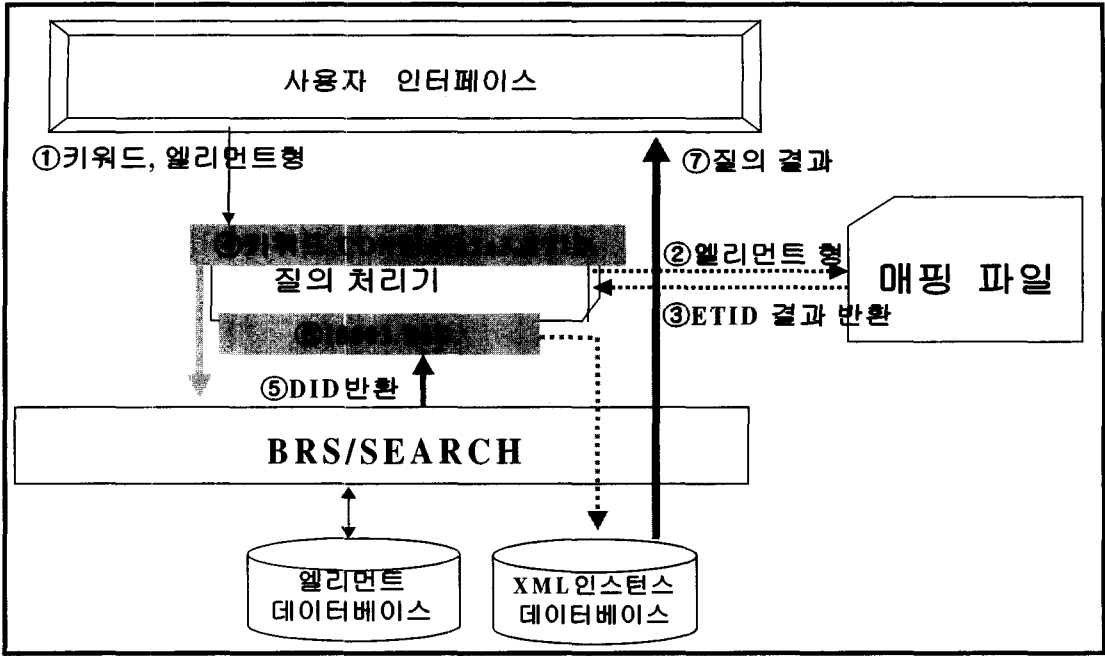
2) 혼합 질의

혼합 질의는 엘리먼트 형과 키워드를 파라미터로 특정 문서에 대한 검색을 위한 질의와 특정 키워드를 포함하는 특정 엘리먼트의 관계에 대해 질의 할 수 있는 경우의 2가지로 분류하여 처리할 수 있다.

첫 번째 경우는 특정 키워드와 엘리먼트 이름을 사용하는 질의로써 먼저 엘리먼트 이름에 대해서 매핑 테이블의 해당 엘리먼트의 ETID 정보를 검색하고 이 정보를 질의 처리기에서 ETID와 키워드에 해당하는 검색식으로 변환하여 BRS/



(그림 5) 저장 관리기 구성



(그림 6) 혼합 질의 처리 1

Search 시스템이 처리하고 만족하는 결과에서 구조 정보에 대한 DID 필드의 정보를 이용하여 XML 인스턴스 데이터베이스에 접근하여 해당하는 문서를 출력한다. 그림 6은 첫 번째 혼합 질의의 처리과정을 보여 준다.

그림 7은 두 번째 혼합 질의의 처리과정을 보여 준다. 처리과정을 보면 특정 키워드를 포함하는 엘리먼트의 부모, 자식, 형제에 대한 검색으로서 먼저 키워드를 포함하는 엘리먼트를 검색하고, 질의에 만족하는 엘리먼트의 구조 정보 중에 SORD의 계층적인 관계를 처리하여 해당하는 SORD 정보를 생성하고 이 SORD 정보와 키워드를 포함하는 검색식으로 변환하여 BRS/Search 시스템이 처리하고 전문을 요구할 경우 DID를 반환하고 해당 엘리먼트를 요구할 경우 DID와 SORD를 반환하여 사용자에게 결과를 제시한다.

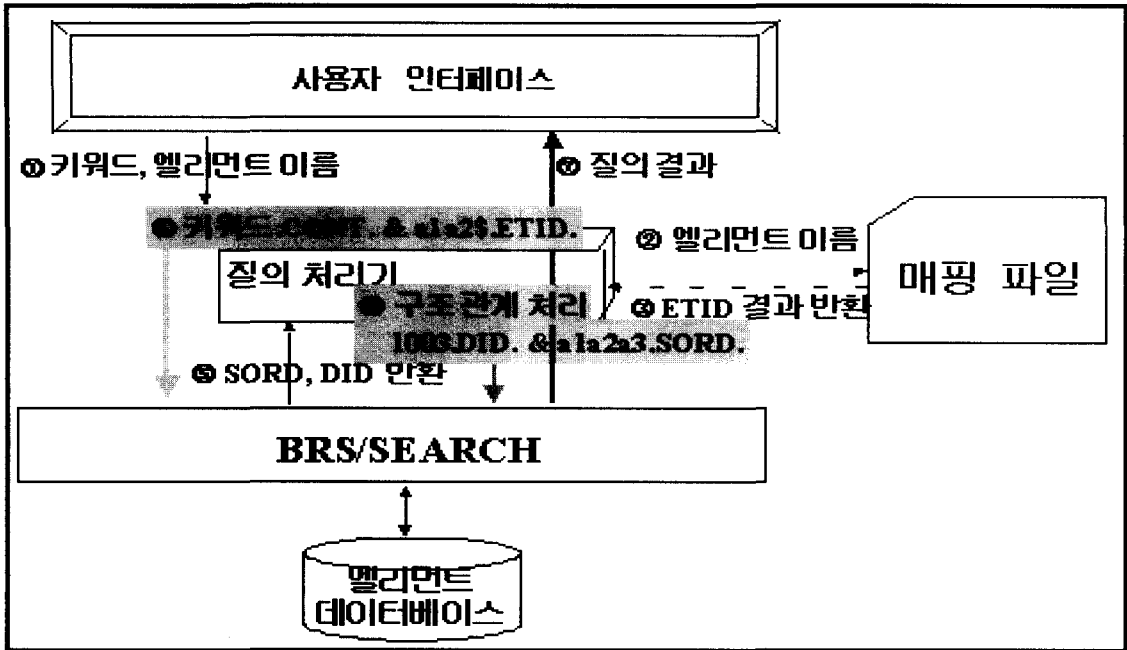
3) 애트리뷰트 질의

특정 엘리먼트에 해당하는 애트리뷰트 질의

는 질의 처리기에서 질의에 대한 검색식을 생성하여 BRS/Search 시스템의 검색엔진에 반환하여 애트리뷰트 데이터베이스에 접근하여 특정 엘리먼트에 대한 애트리뷰트의 값과 이름 및 혼합된 검색이 가능하다.

4.5 사용자 인터페이스

사용자가 XML 문서에 대한 검색을 하기 위해서는 사용자 인터페이스가 필요하다. 따라서 사용자 인터페이스는 기본적으로 구조 질의와 혼합 질의, 애트리뷰트 질의 및 내용 질의에 만족하는 질의 처리 결과를 사용자에게 충분히 반영할 수 있어야 한다. 이에 구현한 XML 문서 검색시스템은 사용자 인터페이스에 있어서 XML 문서의 기본 단위인 엘리먼트와 해당 엘리먼트의 내용만을 보여주는 기능과 해당 엘리먼트의 DTD 문서를 XML 인스턴스 문서를 보여주는 기능을 제공한다. 또한 본 논문에서는 검색시스템을 테스트하기



(그림 7) 혼합 질의 처리 2

위해 검색 인터페이스로서 웹 상에서 검색 가능하도록 CGI로 구현하였다.

내용 검색은 XML에 표현되어 있는 텍스트에 대한 스트링 매칭 검색을 하고, 구조 검색에서는 계층 간의 관계를 위해 조상, 자손의 관계가 있으며, 같은 계층 내의 관계를 위해 형제 관계가 있다. 또한 각 엘리먼트들 간의 순서에 따라 선후 관계(+ , - 숫자)를 가질 수 있도록 하였으며, 검색 결과로는 문서 전체 또는 엘리먼트를 지정할 수 있다.

그림 9와 그림 10은 설계된 사용자 인터페이스를 이용하여 “이동전화를 포함하는 첫 번째 chapter의 자식 엘리먼트를 찾아라” 라는 구조와 내용이 혼합된 혼합 질의와 질의 처리 결과의 예를 보여주고 있다.

5. 성능 평가

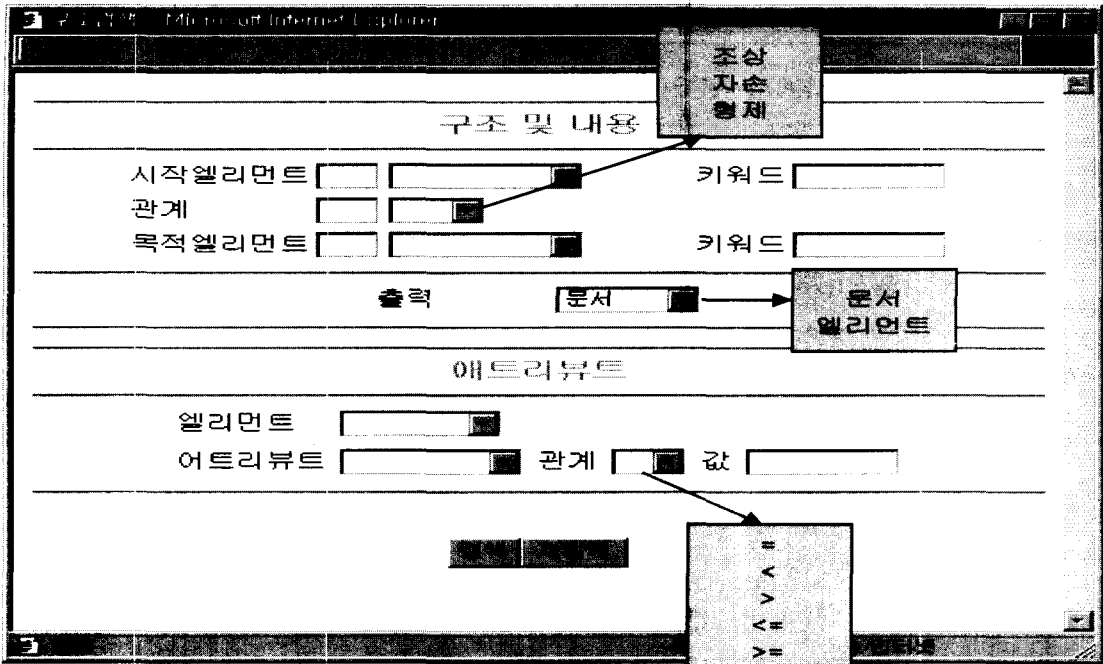
본 논문에서는 BRS 검색엔진을 이용하여 구조 기반의 검색을 지원하는 XML 검색시스템을 테스트

한 결과를 제시한다. 사용한 DTD는 석·박사 학위논문의 형식에 맞게 설계한 thesis.dtd이고, 테스트 데이터로서 학위문서 500건을 사용하였다. 테스트 환경은 웹기반으로 웹 브라우저를 이용하여 구현한 검색시스템을 테스트하였다.

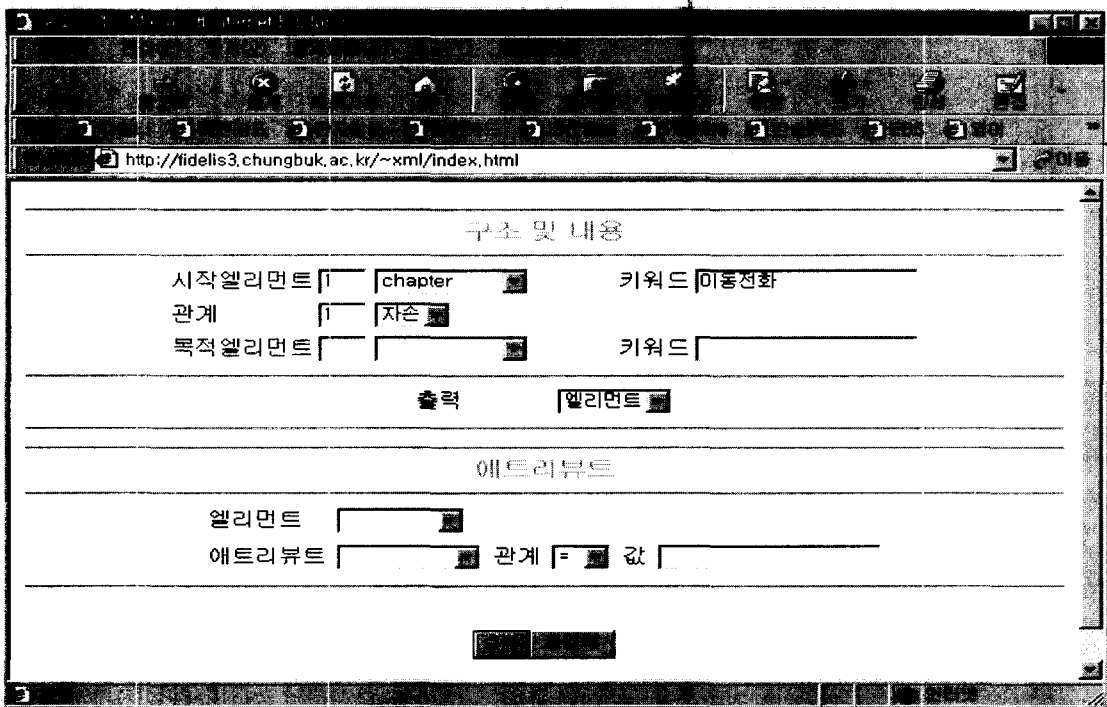
정보 검색시스템의 검색 효율의 평가 방법으로 일반적으로 재현율(recall)과 정확율(precision)을 이용된다. 재현율은 문서 집합에서 사용자가 원하는 문서를 어느 정도 검색하였는가를 나타내고, 정확율은 검색된 문서들 중에서 사용자가 원하는 문서가 얼마나 포함되어 있는가를 나타낸다.

$$\text{재현율} = \frac{\text{검색된 적합 문서 수}}{\text{적합 문서 총수}}$$

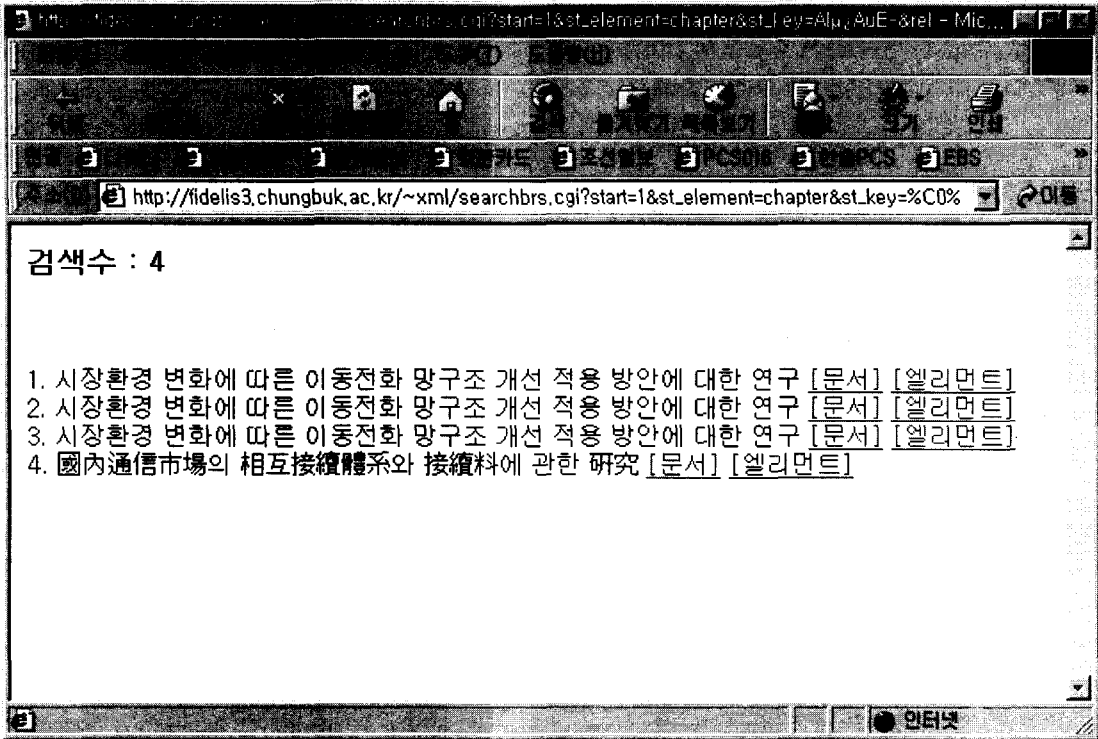
$$\text{정확율} = \frac{\text{검색된 적합 문서 수}}{\text{검색된 문서 총수}}$$



(그림 8) 사용자 인터페이스



(그림 9) 질의 인터페이스



(그림 10) 질의 결과

본 논문에서는 BRS/Search 검색엔진을 이용하여 구조에 기반한 검색의 검색 효율을 평가하기 위해 5가지의 구조에 기반한 질의를 선택하였다. 평가에 사용된 질의 예는 다음과 같다.

1. “abstract”라는 엘리먼트에 “정보”라는 단어를 포함하는 문서를 검색하시오.
2. “분산”이라는 단어를 포함하는 “title” 엘리먼트의 상위 엘리먼트를 검색하시오.
3. “논문”이라는 단어를 포함하는 두 번째 “title” 엘리먼트를 검색하시오.
4. “데이터”라는 단어를 포함하는 첫 번째 “section”의 자식 엘리먼트를 검색하시오.
5. 애트리뷰트의 이름이 “lineWidth”이고, 값이 “100”인 엘리먼트를 검색하시오.

위 질의에 대하여 구현한 검색시스템의 구조에

기반한 검색 효율을 평가하기 위해 재현율과 정확율을 각각 계산하였다. 표 1에서 보듯이 BRS/Search 검색엔진을 이용하여 XML 문서에 대하여 구조에 기반한 검색을 정확하게 지원함을 알 수 있다.

(표 1) 질의에 대한 재현율과 정확율

질의	재현율	정확율
1번 질의	1	1
2번 질의	1	1
3번 질의	1	1
4번 질의	1	1
5번 질의	1	1

6. 결 론

본 논문에서는 XML 문서에 대하여 구조에 기

반한 검색을 지원하도록 기존의 상용 검색시스템인 BRS/Search 시스템을 이용하여 검색 기능을 확장하여 XML 문서 검색시스템을 구현하였다. 구현한 시스템은 XML 문서의 DTD를 분석하여 문서의 논리적인 구조 정보 표현으로 각 엘리먼트 이름에 할당되는 유일한 식별자인 ETID와 XML 문서에서 엘리먼트 간의 순서 정보인 SORD, SSORD로 표현하였다. 이와 같이 XML 문서에 대해 구조 정보를 표현함으로써 특정 엘리먼트의 관계를 효율적으로 검색 할 수 있도록 하였다. 뿐만 아니라 XML 문서에 대한 구조 질의, 혼합 질의, 애트리뷰트 질의를 효과적으로 처리 할 수 있도록 BRS/Search 시스템에 데이터베이스 스키마를 정의하고, 부가적으로 ETID 추출기, 구조 정보 추출기, 저장 관리기, 질의 처리기를 구현하였다. 구현한 XML 문서 검색시스템의 검색 효율을 평가하기 위해 재현율과 정확율을 이용하여 평가한 결과를 보면 XML 문서에 대하여 구조에 기반한 질의를 모두 만족함을 알 수 있다.

향후 연구 방향으로 구현한 검색시스템과 DBMS 기반 검색시스템과의 다양한 측면에서의 성능평가를 수행하는 것이다. 또한, 다양한 DTD 문서와 대용량의 XML 문서들을 적용하여 구현한 XML 문서 검색시스템이 어떠한 경우에도 정확하게 동작함을 보일 것이다.

참 고 문 헌

- [1] Dongwook Shin, Hyuncheol Jang, and HongLan Jin. "BUS : An Effective Indexing and Retrieval Scheme in Structured Documents", ACM. pp. 235 - 243, 1998.
- [2] Francois. "Generalized SGML repositories: Requirements and modelling", Computer Standards & Interfaces, 1996.
- [3] Tuong Dao, Ron Sacks-Davis, James A.Thom. "An indexing scheme for structured documents and its implementation", Proceedings of the 4th International Conference on DATABASE Systems for Advanced Applications, Melbourne, Australia. pp.125 - 134, 1997.
- [4] 맹성현, 주종철. "문서 구조화와 정보 검색", 정보과학회지, 제 16권, 제 8호, 1998. 8.
- [5] W3C. Extensible Markup Language(XML) 1.0, <http://www.w3.org/TR/#Recommendations>, 1998.
- [6] 손정환, 이회주, 장재우, 심부성, 주종철. "구조화된 문서를 위한 정보검색시스템의 설계 및 구현", '98 동계 데이터베이스 학술대회 논문집, 제 14권, 1호, pp.102-106, 1998.
- [7] 이용배. "SGML문서 저장을 위한 스키마생성기 및 자동삽입기의 설계 및 구현", 석사학위논문, 충남대학교, 1998.
- [8] 연제원, 장동준, 김용훈, 이강찬, 이규철. "효율적인 검색 지원 SGML 저장 관리기의 설계 및 구현", '99 한국 데이터베이스 학술대회 논문집, 15권, 1호, pp.136-143, 1999.
- [9] V. Christophides, S. Abiteboul, S. Cluet and M. Scholl. "From Structured Documents to Novel Query Facilities", ACM. pp.314-325, 1994.
- [10] Yong Kyu Lee, Seong-Joon Yoo, Kyoungro Yoon. "Index Structures for Structured Document", ACM. pp.91 - 99, 1996.
- [11] T. Arnold-Moore, M.Fuller, B.Lowe, J.Thom, R. Wilkinson. "The ELF data model and SGQL query language for structured document databases", Proc. 6th Australian Database Conference, pp. 17-26, 1995.
- [12] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Extensible Markup Language (XML) 1.0, REC-xml-19980210,
- [13] V.Christophides. et al. "From Structured Documents to Novel Query Facilities", ACM SIGMOD, pp.313-324, Minesota, USA, 1994.
- [14] W3C. A Query Language for XML, <http://>

www.w3.org/TR/#Notes, 1998.

- [15] 김현기, 이상기, 주종철. “SGML/XML 문서 관리 시스템의 설계 및 구현”, 정보처리학회 추계 학술발표논문집, 5권, 2호, pp.1251-1254, 1998.
- [16] 박철현, 정재현, 심대익, 이상구. “구조화된 문서에 대한 DBMS와 IRS의 성능비교”, 한국 데이터베이스 학술대회 논문집, 15권, 1호, pp.218-225, 1999.
- [17] 유재수의 8인. “전자도서관 표준문서관리를

위한 XML 저장 관리기 기술 개발”, 케이오티크 최종보고서, 1999.

- [18] 이원석. “대량의 문서 관리를 위한 SGML 저장관리기의 설계 및 구현”, 석사학위논문, 충남대학교, 1998.
- [19] Charles L. A. Clarke, Gordon V. Cormack, Forbes J. Burkowski, “An Algebra for Structured Text Search and a Framework for its Implementation”, The Computer Journal 38(1), pp.43-56, 1995.

● 저 자 소개 ●



손 충 범 (孫忠範)

1997년 충북대학교 정보통신공학과(공학사)
 1999년 충북대학교 대학원 정보통신공학과(공학석사)
 1999년~현재 충북대학교 대학원 정보통신공학과 박사과정
 관심분야 : 데이터베이스 시스템, XML, 정보검색 프로토콜, 분산 객체 컴퓨팅 분야, etc.
 E-mail : cbson@trut.chungbuk.ac.kr



이 병 엽 (李秉燁)

1991년 한국과학기술원 전산학과(공학사)
 1993년 한국과학기술원 전산학과(공학석사)
 1997년 한국과학기술원 경영정보공학(공학박사)
 1997년~현재 대우정보시스템 e-솔루션사업팀 차장
 관심분야 : 데이터마이닝, XML, 인공지능, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅 분야, etc.
 E-mail : bylee@disk.co.kr



유 재 수 (柳哉秀)

1989년 전북대학교 컴퓨터공학과(공학사)
 1991년 한국과학기술원 전산학과(공학석사)
 1995년 한국과학기술원 전산학과(공학박사)
 1995년~1996년 8월 목포대학교 전산통계학과 전임강사
 1996년 8월~현재 충북대학교 정보통신공학과 부교수
 관심분야 : 데이터베이스 시스템, XML, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅, etc.
 E-mail : yjs@cbucc.chungbuk.ac.kr