

웹 기반의 교수 지원 시스템을 위한 XML 문서의 분류 및 검색

김 행 곤*, 김 지 영*, 최 문 경*, 김 성 원**
대구가톨릭대학교 컴퓨터공학과*
안양대학교 컴퓨터공학과**

e-mail: hangkon@cuth.cataegu.ac.kr, kimjy@cuth.cataegu.ac.kr,
ditocmk@hanmail.net, swkim@aycc.anyang.ac.kr

Classification and Retrieval of XML Document for Teacher Support System based on Web

Haeng-Kon Kim*, Ji-Young Kim*, Mun-Kyoung Choi*, Soung-Won Kim**
Dept. of Computer Engineering, Catholic University of Daegu*
Dept. of Computer Engineering, Anyang University**

요 약

최근 인터넷이 급속히 성장함에 따라 웹을 기반으로 한 학습이 활발히 진행되고 있고, 또한 학교 업무의 효율화를 지원하기 위한 분야에서도 웹이 응용되고 있다. 특히 웹에서 교수를 위한 복잡한 학교 업무의 관리와 학습자료 및 업무 자료를 지원하기 위해서는 확장성과 호환성, 편의성을 가진 XML 형태의 문서가 제공되어야 한다. 따라서 교수 업무 지원을 위해 XML 문서의 정보들을 효율적이고 정확하게 이용하기 위해 이들 문서를 적절하게 분류하고 저장, 검색하기 위한 방법이 필요하다.

본 논문에서는 XML로 작성된 교수 업무 지원 문서의 저장과 검색을 위한 선행작업으로서, 일반적인 메타 데이터와 DTD 데이터를 정의하고, 이렇게 정의된 데이터를 이용하여 패시 검색과 구조기반 검색, 키워드 검색을 제공함으로써 사용자는 원하는 문서를 쉽게 검색할 수 있다. 따라서 이를 통해 교수 업무 지원 문서들을 웹 상에서 효율적이고 정확하게 저장하며, 사용자가 원하는 문서를 정확하고 신속하게 검색할 수 있게 하고자 한다.

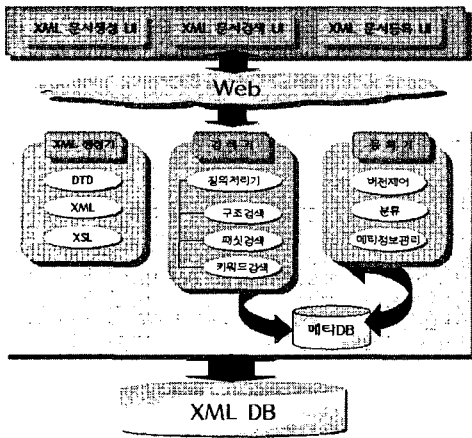
1. 서론

최근 인터넷의 급속한 성장과 교육 환경의 변화로 인해 학습자를 위한 학습 환경뿐만 아니라 교수 업무를 지원하는 문서에서도 웹을 기반으로 한 정보의 공유가 요구되어지고 있다. 또한 인터넷에는 수많은 양의 정보들이 전자 문서의 형태로 존재함으로써 이들 문서들을 보다 효율적으로 저장하고 검색, 이용하는 연구가 필요하게 되었다. 이러한 연구는 다양한 형태의 전자 문서의 표준을 요구하게 되었고, 이 가운데 W3C에서 제안한 XML(eXtensible Markup Language)이 차세대 인터넷 전자 문서의 표준으로 많은 연구 분야에서 사용하고 있다[1].

XML은 웹 상에서 구조화된 문서를 전송 가능하도록 설계된 표준 마크업(Markup)언어이다. 이것은 인터넷 상에서 가장 많이 사용하는 HTML의 단순

함을 극복하고, 복잡한 SGML을 간략화했으며, HTML과 SGML의 표현 사이에서 상호 운용성 및 용이한 구현 환경을 제공한다. 다양한 XML 관련 연구 중에서 디지털 도서관, 전자상거래, 데이터웨어하우징, EDI(Electronic Data Interchange)등과 같은 분야에서는 XML 문서를 효율적으로 저장, 검색, 이용하기 위한 많은 연구를 하고 있으나, XML 문서에 대한 재구성과 재사용이 없는 검색 기능만을 제공하는 경우가 대부분이다[2].

따라서, 본 논문에서는 교수의 효율적인 업무 관리를 위해 작성된 XML 문서를 보다 효과적으로 재사용하기 위한 분류 방법을 제시하고, 이 분류 방법에 따라 저장된 문서들을 효율적으로 검색할 수 있는 시스템을 제안하고자 한다.



(그림 1) 시스템 구조

2. 관련연구

2.1 교수 지원 시스템

교수 지원 시스템은 웹을 기반으로 한 클라이언트/서버 시스템으로 학생 생활, 성적, 공부, 교무 업무 등으로 이루어진 학사정보시스템이다. 이 시스템은 복잡한 교무 업무를 체계적으로 전산화하고 네트워크를 이용하여 교수-학습 자료를 공유함으로써 자료 이용의 효율성을 극대화 할 수 있다. 또한 이들 각각의 시스템은 XML을 기반으로 구축되어지고, 다양한 문서들을 표준화함으로써 문서 작성의 용이성과 융통성을 제공한다. 따라서 웹에서의 문서 공유뿐만 아니라 효율적인 문서 분류 방법을 통해 정확하고 용이한 검색을 제공할 수 있다. (그림 1)은 본 논문에서 연구된 전체 시스템 구조를 나타낸다[3,4].

2.2 기존 분류 방법

계층 열거 분류 방법은 분류 대상 전체에 대한 정보를 모두 합성하여 표현한 가장 포괄적인 정보를 기점으로 하여, 점차로 좁은 합성정보의 군으로 나누어서 그들 사이의 관계를 계층적으로 나타내는 표현 방식이며, 가장 전통적인 분류법으로 도서 분류에 흔히 이용되고 있다.

중심 단어 분류 방법은 분류 대상물 각각의 정보 표현 중에서 중심적인 단어들을 나열하여, 개개의 대상물을 분류하는 방법으로 어떠한 형태의 새로운 것도 그것의 특성이 어휘이고 표현될 수만 있다면 분류할 수 있다는 장점을 가지고 있다.

패킷 분류 방법은 계층 분류 방법의 경직성을 보완하여 확장유연성을 신장시킨 방법이며, 제어형 계층 열거 분류법과 중심단어 분류법을 혼합한 형태이

<표 1> 메타데이터 정의 원칙

- ① 공인된 표준 메타 형식에 기반 한다. (일반성)
- ② 단순하면서도 구조화된 형식을 갖춘다. (단순성, 형식성)
- ③ 아키텍처 기반의 분류 정보를 사용한다.
- ④ 메타데이터 요소는 함축적 의미를 내포한다.(함축성)

다. 기존의 계층 열거 분류법으로 표현된 지식들이 분류법으로 쉽게 변환이 가능하다는 장점 또한 가지고 있다.

2.3 메타 데이터

메타데이터(MetaData)는 데이터가 가지는 속성과 구조, 물리적인 위치의 변화 이력, 연결성 등 하나의 데이터가 가지는 가장 기본적인 사항을 포함하고 있는 데이터의 데이터를 말한다. 또한 메타데이터는 애플리케이션 시스템에서의 저장은 물론 검색, 관리, 운영, 유지보수 등에 결정적인 역할을 수행한다. 메타데이터는 각종 데이터 즉, 문서, 파일, 미디어 등을 효과적으로 운영 관리할 수 있게 하여 정보의 고효율도, 시스템의 용이한 유지 보수를 가능케 한다 [4]. 따라서, 사용하고자 하는 정보의 메타데이터 정의는 시스템의 원활한 운영과 소기의 목적 달성을 위해 필수적이며 신중하여야 할 것이다<표 1>.

3. XML 문서 분류 방법

3.1 메타데이터의 정의 및 분류

메타 데이터는 데이터가 가지는 속성과 구조, 연결성 등의 기본적인 사항을 포함하고 있는 데이터의 데이터로 대상의 저장, 검색, 관리, 운영, 유지보수 등에 결정적인 역할을 수행함으로써 정보의 높은 활용을 가능케 한다[5].

본 논문에서는 단순히 XML로 표현된 교수 지원 문서의 효율적인 저장과 검색, 관리가 용이하도록 교수 지원 문서를 지원할 수 있는 메타 데이터를 분류하고자 한다. 이를 위해 교수 지원 문서의 공통적이고 핵심적인 항목으로 <표 2>와 같이 메타 데이터를 정의하였다. 정의된 메타 데이터는 데이터베이스에 저장되기 위한 핵심 데이터가 될 뿐만 아니라 사용자들에 의한 검색 및 이해 정보 획득을 위한 표준 포맷을 제공한다.

3.2 DTD를 통한 데이터 분류

<표 2> 메타 데이터 분류

메타데이터	설명
도메인	교수지원시스템의 도메인 이름
이름	문서 이름
분반	분반 코드
과목명	과목 이름
날짜	문서를 작성한 년/월/일
작성자	문서를 작성한 사람의 이름
서술	문서 내용을 간단히 기술

XML 문서의 DTD를 통해서 의미있는 정보 블록의 데이터를 기술한다.

(그림 2)는 교수 지원 문서 중 하나인 학습지도안 문서를 XML DTD로 표현한 것으로써, 엘리먼트와 애트리뷰트를 포함하는 정보블록 단위로 (그림 3)과 같은 데이터 테이블을 생성할 수 있다.

Document Name	Element Name	Attribute Name	Attribute Type	Value
---------------	--------------	----------------	----------------	-------

(그림 3) 데이터 테이블

(그림 3)의 데이터 테이블을 구성하는 각각의 요소를 간단히 설명하면 다음과 같다.

- Document Name: 각각의 문서를 통합하는 교수 지원 문서
- Element Name : 교수 지원 문서를 구성하는 하나의 문서
- Attribute Name : 하나의 문서에 포함되는 항목
- Attribute Type : 항목의 형식
- Value : 애트리뷰트의 실제 값

(그림 4)는 (그림 2)에서 기술한 XML DTD를 계층적인 트리 형태로 표현함으로써, 엘리먼트와 애트리뷰트의 식별이 용이하도록 하고, 크게 레벨 2까지의 정보를 제공한다.

```

<?xml version="1.0" encoding="euc-kr"?>
<ELEMENT DOCUMENT (학습지도안)>
<ELEMENT 학습지도안
(단원명, 지도일시, 학습목표, 수업단계)>
<ELEMENT 단원명 (#PCDATA)>
<ELEMENT 지도일시 (#PCDATA)>
<ELEMENT 학습목표 (#PCDATA)>
<ELEMENT 수업단계
(도입, 전개, 정리, 형성평가)>
<ELEMENT 도입 (교수-학습활동)>
<ELEMENT 교수-학습활동 (교사, 학생)>
<ELEMENT 교사 (#PCDATA)>
<ELEMENT 학생 (#PCDATA)>
<ELEMENT 전개 (교수-학습활동)>
<ELEMENT 정리 (교수-학습활동)>
<ELEMENT 형성평가 (#PCDATA)>
    
```

(그림 2) 학습지도안 DTD

4. XML 문서 검색 방법

교수지원 시스템에서 제공하는 검색 방법은 다음과 같이 3가지를 제공한다. 사용자가 원하는 문서를 DTD의 구조에 따라 매핑되는 값을 입력하거나, 혹은 제시되어진 패시 항목의 선택에 의한 조합, 또는 문서에 대한 명확한 이름을 입력함으로써 가능하다.

4.1 DTD 구조 기반 검색

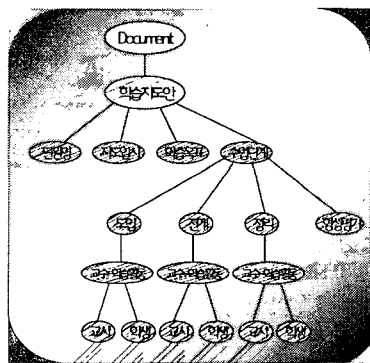
구조 기반 검색은 단순한 문서 이름 검색뿐만 아니라 엘리먼트의 이름, 애트리뷰트 이름, 부모·자식 관계, 그리고 내용이 혼합된 형태의 질의를 통해 XML 문서 내에서 검색하는 기능을 말한다. 예를 들어 “국어 학습지도안이라는 엘리먼트의 애트리뷰트인 학습목표 엘리먼트에서 소설이라는 단어를 포함하는 XML 문서를 찾아라”는 질의를 통해 검색을 실시하는 것을 말한다. 즉, 이것은 DTD를 구성하는 엘리먼트와 애트리뷰트의 결합된 검색 기능을 제공한다[5].

본 논문에서 제시하는 구조 기반 검색은 교수 지원 문서에 대한 Document Name, Element Name, Attribute Name, Attribute Type, Value의 기준에 따라서 분류하고, 이 분류 기준을 조합해서 검색할 수 있도록 하였다.

4.2 패시 검색

패시 검색은 일반적인 메타 데이터의 정의를 기반으로 패시 항목을 6개(도메인, 이름, 분반, 과목명, 날짜, 작성자)를 설정하여 이들의 조합에 의해 필요한 문서를 검색한다.

검색시에 도메인과 과목명은 필수 조건으로 AND 조건을 수행하고, 나머지 항목들은 OR 조건을 만족



(그림 4) 학습지도안 DTD 정보블록 단위

<표 3> 패시 항목

도메인	이름	분반	과목명	날짜	작성자
학습지도안	집합과 멱제	1학년 1반	수학	1998/1	홍길동
과제물평가	도 양	2학년 3반	과학	1999/2	김철수
성적 관리	글과 표현	3학년 1반	국어	2000/1	심은하
강의 평가	인간과 윤리	3학년 3반	윤리	2001/2	이영애

한다. 예를 들면, “도메인이 학습지도안이고, 과목명이 수학이며, 홍길동이 작성한 문서”를 찾을 수 있도록 한다. <표 3>은 교수 지원 문서에 대한 패시 항목과 값의 예를 나타낸 것이다. 이렇게 일반적인 메타 데이터를 패시로 분류해 놓음으로써, 사용자는 문서에 대한 정확한 이해 없이도, 원하는 문서를 손쉽게 빠르게 검색할 수 있다.

4.3 키워드 검색

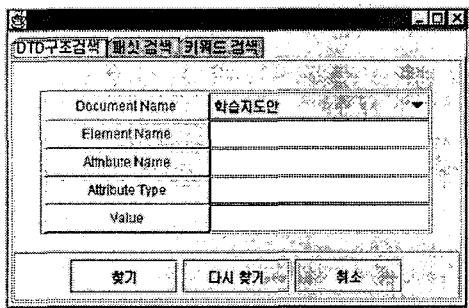
키워드 검색은 사용자가 원하는 문서 이름을 직접 키 필드에 입력함으로써 문서를 획득할 수 있는 검색 방법이다.

5. XML 문서 검색 UI

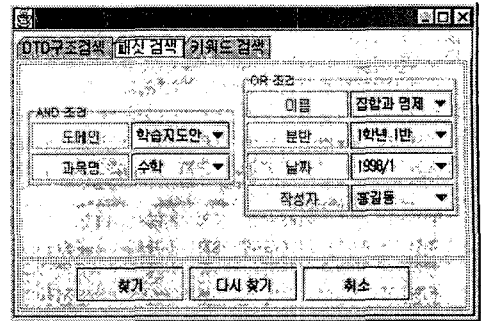
다음 그림은 교수지원 시스템에서 제공하는 XML 문서 검색 UI를 나타내고 있다. 검색은 3가지 방식을 제공하는데 DTD 구조를 통한 검색과 메타데이터를 기반한 패시 검색, 키워드 검색을 제공한다.

(그림 5)는 미리 정의된 DTD 구조의 정보를 기반으로 한 검색을 나타낸다. Document Name은 미리 정의된 Document 리스트 중에서 선택되어지고, 나머지 항목은 사용자가 검색하고자 하는 단어를 직접 입력함으로써 원하는 문서를 획득할 수 있다.

(그림 6)은 <표 3>의 패시 항목을 기반으로 검색



(그림 5) DTD 구조 검색 UI



(그림 6) 패시 검색 UI

이 가능하다. 도메인과 과목명은 AND 조건으로 반드시 선택되어야하는 항목이고, 다른 항목은 OR 조건으로 사용자가 원하는 항목만을 선택하여 검색할 수 있도록 하였다.

6. 결론 및 향후연구

본 논문에서는 교수 업무 지원을 위해 작성된 XML 문서를 보다 효율적으로 재사용하기 위한 분류 방법을 제시하고, 이를 기반으로 문서를 등록하고 검색할 수 있는 방법을 제공한다. 즉, 이들 문서를 검색하기 위한 선행작업으로 문서에 대한 메타데이터와 DTD 구조 정보를 정의함으로써 좀더 쉽게 일반적인 정보의 검색에서부터 복잡한 구조 정보의 검색까지 제공함으로써 원하는 문서를 쉽고, 다양하게 획득할 수 있도록 하였다.

향후, 교수 업무를 지원하는 XML 형태의 다양한 문서를 UI를 통해 자동으로 생성할 수 있는 XML 생성기의 개발과 사용자의 요구에 맞는 좀더 정확하고 확장된 검색 방법과 시스템 구현이 요구된다.

[참고문헌]

- [1] "Extensible Markup Language(XML) 1.0", <http://www.w3.org/TR/REC-xml/>
- [2] 이계준외, "XML문서의 검색을 위한 효율적인 색인 기법과 질의언어(TQL)의 설계", 정보과학회 학술발표논문집 Vol.26, No.2, 1999
- [3] 김행곤외, "웹 기반의 교수 지원 시스템에서 XML형식의 학습지도안 설계", 정보처리학회 학술발표논문집, 제8권 제1호, 2001.
- [4] Sami Rollins, "A framework for creating customized multi-modal interfaces for xml documents", IEEE, 2000.
- [5] Bearman, David. "Relations Working Group" http://purl.oclc.org/metadata/dublin_core/, 1997.
- [6] 김행곤외, "컴포넌트 메타데이터의 정의를 통한 효율적인 검색 방법에 관한 연구", 정보과학회 학술발표논문집 Vol.27, No.2, 2000