

연관값을 이용한 XML 공통 VIEW의 설계 및 구현

박진만⁰ 김태우 홍동완 윤지희
한림대학교 컴퓨터공학과
(pjm⁰, dataminer, dwhong, jhyoon)@hallym.ac.kr

Design and Implementation of Common View for XML Documents using the Associative Value

Jin-Man Park⁰, Tae-Woo Kim, Dong-Wan Hong, Jee-Hee Yoon
Dept. of Computer Engineering, Hallym University

요 약

인터넷 상에 산재하는 분산이질 형태의 자료들에 대한 효율적인 검색을 지원하기 위해서는 이들 문서들의 구조를 새로운 정보 구조로 통합하여 사용자에게 제시하여야 한다. 본 논문에서는 DTD 또는 XML Schema를 기반으로 하는 XML 문서의 공통 뷰 설계를 위한 공통 데이터 모델과 처리 방식을 제안한다. 계층, 상속, 복합 등의 다양한 구조 처리를 위한 XML 문서의 경로 정보, 관련성이 있는 요소들의 연결에 이용되는 연관값, 소스 매핑 정보 등으로 구성된 복합 필드 구조를 갖는 트리형의 새로운 공통 데이터 모델을 정의한다. 또한 XQuery 기반의 공통 뷰 정의의 질의 기능과 공통 데이터 구조 상의 질의 처리 방식을 보인다.

1. 서 론

IT(Information Technology) 산업의 발전으로 여러 사업 분야의 전산화와 함께 방대한 양의 데이터가 산출되고 있다. 이러한 데이터들은 각 응용 시스템에 의존적인 분산이질 구조를 가지며, 산재되어 있는 이들 데이터를 유기적으로 공유할 수 있다면 새로운 정보 창출로 인한 파급 효과를 얻을 수 있다. 최근 인터넷 상의 분산된 자료에 대한 효율적인 정보통합 방법으로 미디어이터 시스템[1] 형태가 활발히 연구, 개발되고 있다.

본 연구실에서는 이질 분산 구조의 자료에 대한 통합 검색 기능을 제공하는 HMS(Hallym Mediator System)[2]를 개발하였다. HMS는 DTD(Document Type Definition)를 기반으로 공통 뷰 구조를 제공하는 미디어이터 시스템이다. DTD 기반의 문서는 데이터 형(Type)의 지원 및 문서 구조 표현에 있어 제한적이므로 다양한 구조의 문서를 정의하는데 불편함이 따른다. 최근 다양한 문서 구조 및 데이터 형을 지원하는 XML Schema의 사용이 확산되고 있다. XML Schema는 다수의 데이터 형, 계층 구조 표현, 복합(Composition), 상속(Inheritance) 구조를 이용하여 다양한 데이터 구조를 표현할 수 있다[3]. 따라서 미디어이터 시스템의 공통 데이터 구조는 기존에 축적되어 있는 DTD 기반 문서와 최근 사용이 증가하고 있는 XML Schema 문서를 동시에 지원 가능한 형태로 확장되어야 한다. 이와 같은 기능이 제공될 경우 XML 문서 뿐 아니라 관계형 데이터베이스(Relational Database), 객체관계형 데이터베이스(Object Relational Database), 객체 지향 데이터베이스(Object-Oriented Database) 등의 데이터 구조를 갖는 정보의 통합이 더욱 용이해진다. 다양한 구조의 문서를 통합하는 과정에서 발생할 수 있는 문제점은 문서들을 구성하는 요소사이의 이름, 의미, 구조충돌 등을 들 수 있다[4]. 이러한 문제점을 해

결하기 위해서는 요소들 간의 관계를 파악하여 같은 의미를 갖는 요소들의 연결이 필요하다. 본 논문에서는 이들 다양한 구조의 문서에 대한 공통 뷰를 정의하는데 필요한 공통 데이터 구조(Common Data Structure)를 제안한다. 문서의 경로 정보, 연관값, 소스매핑 정보 등의 복합 필드로 구성된 공통 데이터 구조를 이용하여 DTD/XML Schema 구조의 문서를 효율적으로 통합 처리할 수 있으며, 이 때 의미적으로 관련된 유사 엘리먼트들은 연관값을 통하여 반자동적으로 연결, 통합된다.

2. 관련 연구

2.1 HMS(Hallym Mediator System)

HMS는 가상 접근기법을 기반으로 하는 웹 정보 통합/검색 시스템이다. 시스템의 기본 기능은 관리자에 의한 공통 뷰 생성 및 관리 기능과 일반 사용자의 정보검색 기능으로 나눌 수 있다.

(1) 공통 뷰 생성/관리 기능 : HMS의 관리자는 웹 상에 이질적인 형태로 산재해 있는 각종 데이터를 통합하여 응용시스템에 적합한 공통 뷰를 정의, 관리한다. HMS는 웹 상의 정보 소스에 대한 각종 정보(소스 내용, 속성 정보, 제약 조건, 소스 신뢰도, 질의처리 능력)를 메타 데이터 형태로 수집, 저장 관리하는데, 관리자는 관리자 인터페이스의 소스 브라우징 기능을 이용하여 소스에 질의 처리를 수행하고 공통 뷰를 정의한다. 공통 뷰의 생성은 추출된 경로 정보를 기반으로 생성된다. 공통 뷰의 생성 과정에서 소스와 공통 뷰와의 매핑 정보가 자동 추출, 관리된다.

(2) 정보 검색 기능 : 사용자는 사용자 인터페이스를 이용하여 자신이 원하는 공통 뷰를 기반으로 검색하고자 하는 정보에 대한 질의를 수행한다. 공통 뷰를 근거로 작성된 사용자의 질의는 공통 뷰 구조에 저장되어 있는 소스매핑 정보를 참조하여 소스에 대한 실제 질의로 변환되어 실행된다.

본 연구는 정보통신부의 정보통신 기초기술연구지원사업(정보통신 연구진흥원)(과제번호 : C1-2002-146-0-3)으로 수행한 연구 결과임.

2.2 HML(Hallym Mediator Language)

공통 뷰의 정의와 정보 검색을 지원하는 HML은 웹 질의어로 기존에 구현되었던 XML-QL[5] 형태에서 확장되어 XQuery[6]의 형태인 FLWR(FOR, LET, WHERE, RETURN) 구문을 따르고 있으며, HMS에서 최적화되어 동작할 수 있도록 개발, 구현되었다.

3. 공통 뷰 생성

분산 이질 구조의 데이터를 효율적으로 검색하기 위해서는 사용자에게 공통 뷰를 제공해야 한다. 이 장에서는 공통 뷰를 제공하기 위해 분산 이질 데이터에 대한 공통 구조 정의, 관련된 엘리먼트를 연결하기 위한 연관값 설정, 공통 뷰 정의 등에 사용되는 XQuery 기반의 확장 질의어에 대해서 언급한다.

3.1 공통 데이터 구조(Common Data Structure)

이질 데이터 구조를 하나의 통합 구조로 변환하여 사용자에게 제공되는 공통 뷰는 트리 구조 형태이다. 공통 뷰 생성 시 시스템 내부에서 처리되는 데이터 구조를 단일 노드 구조로 표현할 경우 복합 문서 간의 연관성 및 상속 등의 구조를 표현하기 어렵다. 이 문제를 해결하기 위해 복합 필드를 갖는 노드로 구성하였다. 사용자에게 제공되는 공통 뷰의 노드는 세부 속성까지 보일 필요는 없다. 그림 1은 구조 통합에 필요한 공통 데이터 구조이다.

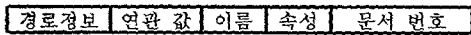


그림 1. 통합 노드 구조

각 필드에 대한 내용은 다음과 같다.

- (1) 경로정보(Path Information) : 경로정보는 임의의 노드가 어느 곳에 위치하는가의 결정과 변경된 정보 구조에 의한 경로를 재설정 하는 경우에 사용된다.
- (2) 연관값(Associative Value) : 엘리먼트 노드에 대한 의미를 수치 값으로 결정한다. 다른 문서와의 통합 시 해당 노드의 연관값을 비교하여 관련 엘리먼트들의 연결을 설정한다.
- (3) 엘리먼트 이름(Element Name) : 엘리먼트 이름을 저장한다.
- (4) 노드 속성(Node Type) : 노드의 속성은 일반 노드들에 대해서 P, 복합 구조에 대해 C, 상속의 경우 부모노드와 자식노드에 대해 각각 I_P, I_C로 정의한다.
- (5) 문서 번호(Document ID) : 노드가 어느 문서에 존재하는지를 알려주는 필드이다. 공통 뷰를 근거로 작성된 일반 사용자의 질의는 해당 필드의 문서로부터 결과값이 추출된다.

3.2 연관값(Associative Value)

분산되어 있는 다양한 데이터를 공통 뷰로 구성하기 위해서는 엘리먼트들의 연관관계를 파악하여 연관성이 있는 엘리먼트를 연결해야 한다. 이러한 기능을 제공하기 위해 엘리먼트의 연관성을 수치로 표현하여 연관값(Associative Value)을 정의한 후 연관값 저장소(Associative Value Storage)에 저장한다. 상호간에 연관성이 높은 엘리먼트는 같은 값을 가지고 있으므로 서로 연결하여 통합구조를 구성할 수 있다.

본 연구에서는 의료정보 통합을 위한 응용 시스템 개발을 진행하고 있으며[7], 소스정보를 구성하고 있는 엘리먼트에 대한 연관값은 의료 데이터 정의 및 전자적 교환 표준으로 제안된 HL7[8]의 RIM(Reference Information Model)[8,9]을 이용하였다. 표 1은 엘리먼트에 부여된 경로 정보 표현에 사용될 키(Key) 값과 연관값을 보인다.

표 1. 엘리먼트 연관값

Element Name	Key	Associative Value	Element Name	Key	Associative Value
ADDRESS	A	1	KIND	K	14
AMOUNT	AM	2	LICENSE	L	15
AREA	AR	3	MEDICAL RECORD	MR	16
CODE	C	4	MEDICINE	M	17
CONDITION	CN	5	NAME	N	18
DATE	D	6	NUMBER	NB	19
DEPT	DP	7	PATIENT	P	20
DISEASE	DI	8	PNAME	PN	<20,18>
DOCTOR	DT	9	PRESCRIPTION	PR	21
D_NAME	DN	<9,18>	SEX	S	22
ID	ID	10	STUDY CONTENT	SC	23
INJECTION	IN	11	TIME	T	24
INSURANCE	IS	12	TERM	TE	25
JUMIN	J	13	TYPE	TY	26

그림 2, 3은 본 의료정보 시스템에서 사용되는 데이터 예로, 환자의 신상 정보 DTD와 진료내역에 대한 XML Schema 문서를 공통 구조로 변환한 예를 보인다.

```
<!ENTITY % common SYSTEM "common.dtd">
<ENTITY % insurance "(CODE, TYPE, AREA)">
<ELEMENT PATIENT (ID, JUMIN, NAME, SEX, ADDRESS, INSURANCE)+>
%common;
<ELEMENT NAME (#PCDATA)>
```

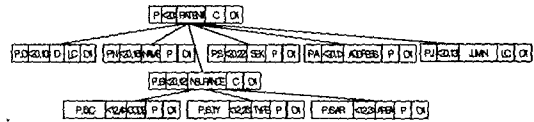


그림 2. PATIENT_INFO.dtd의 공통구조 변환

```
<xs:element name="PATIENT" type="PATIENTType"/>
<xs:complexType name="PATIENTType">
  <xs:complexContent>
    <xs:extension base="common"/>
    <xs:sequence>
      <xs:element name="NAME" type="xs:string"/>
      <xs:element name="MEDICAL_RECORD" type="MEDICAL_RECORDType"
        minOccurs="0" maxOccurs="unbounded"/>
    </xs:sequence>
  </xs:extension>
</xs:complexType>
```

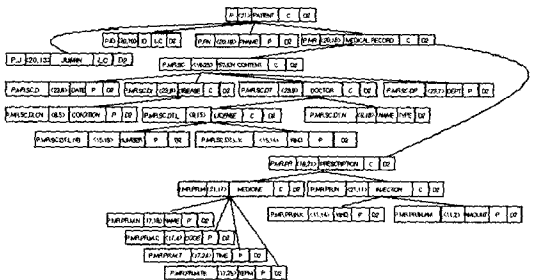


그림 3. MEDICAL_RECORD.xsd의 공통구조 변환

위의 공통 구조 연관값은 다음과 같이 부여된다. 노드는 부모와 자신의 연관값을 가진다. 'NAME'의 경우 'DOCTOR/NAME'과 'PATIENT/NAME'의 의미를 구분해야 한다. 'PATIENT/NAME'은 <20,18>, 'DOCTOR/NAME'은 <9,18> 값을 가짐으로서 동음이의어의 충돌을 해결한다. 또한 'PNAME'의 경우 'PATIENT/NAME'과 같은 <20,18> 값을 가짐으로서 이음동이의어의 충돌을 해결한다.

3.3 질의처리

HMS에서 질의는 그래픽 기반의 사용자 인터페이스 환경에서 작성되고, 시스템 내부에서 XQuery 형태로 변환된다. 예제 1은 "임의의 환자에 대한 기초정보(환자번호, 이름, 주민번호)와 과거 진료내역"을 보이는 공통 뷰 생성 질의이다. DTD / XSD문서에서 스키마 정의 변수 \$p, \$m에 로딩된 정보 중 환자번호(ID) 연관값이 일치하는가를 검사하여, 'Patient' 노드에 환자 기초 정보와 진료 내역의 구조를 갖는 공통 뷰를 생성한다.

예제 1.

```
FOR $p IN document("PATIENT_INFO.dtd") ..... (1)
  $m IN document("MEDICAL_RECORD.xsd")
WHERE $p/ID = $m/ID ..... (2)
RETURN
  <PATIENT>
  <$p/ID/>
  <$p/NAME/>
  <$p/JUMIN/>
  (<$m/DATE/>)
  <$m/PRESCRIPTION>
  //PRESCRIPTION/*
  </PRESCRIPTION>
</PATIENT> ..... (3)
```

- (1) FOR : 공통 뷰 생성 대상으로 사용할 PATIENT_INFO.dtd 와 MEDICAL_RECORD.xsd를 로딩한다.
- (2) WHERE : 문서 통합에 사용되는 연결 조건을 기술하는 것으로 두 문서 간 매핑 엘리먼트인 ID의 연관값을 이용한다. WHERE절에 명시적으로 연결 조건을 기술하지 않더라도 각 엘리먼트의 연관값을 이용하여 자동으로 연결할 수 있는데, \$m/PATIENT/PNAME은 \$p/PATIENT/NAME으로 자동 연결된다.
- (3) RETURN : 원하는 엘리먼트(ID, NAME, JUMIN, DATE, PRESCRIPTION)를 추출하여 공통 뷰를 생성한다.

그림 4는 예제 1의 질의처리 과정을 거쳐 생성된 공통 뷰의 문서와 구조이다.

```
<xs:element name="PATIENT" type="PATIENTType" />
<xs:complexType name="PATIENTType">
  <xs:complexContent>
    <xs:extension base="common" />
  </xs:complexContent>
  <xs:sequence>
    <xs:element name="DATE" type="xs:date" minOccurs="1"
      maxOccurs="unbounded" />
    <xs:element name="NAME" type="xs:string" minOccurs="1"
      maxOccurs="unbounded" />
    <xs:element name="PRESCRIPTION" type="PRESCRIPTIONType"
      minOccurs="1" maxOccurs="unbounded" />
  </xs:sequence>
</xs:extension>
</xs:complexContent>
</xs:complexType>
```

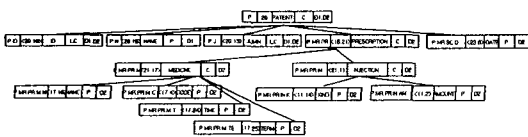


그림 4. 공통 뷰

3.4 시스템 구조

그림 5는 본 시스템의 시스템 구조도로 세부 기능은 다음과 같다.

- (1) 공통구조변환기(Common Structure Converter) : 관리자는 래퍼(Wrapper)로부터 얻은 소스데이터를 연관값 생성 모듈로 보내어 연관값을 할당받는다. 연관값을 할당받은 소스데이터는 공통 구조로 변환되어 관리자에게 보여지고, 관리자는 변환된 공통 구조를 참조하여 공통 뷰 생성 질의를 한다.
- (2) 연관값 생성 모듈(Associative Value Generator Module) : 입력된 소스데이터는 연관값 저장소를 참조하여 연관값을 받게 된다. 만약 입력된 소스 데이터가 저장소의 연관값에 매핑되지 않을 경우 관리자는 직접 소스데이터의 연관값을 부여한다.

- (3) 질의처리기(Query Processor) : 관리자에 의해 작성된 공통 뷰 생성 질의는 질의어 분석기(View Definition Query Parser)에 의하여 분석된다. 분석된 질의는 공통 뷰 생성 모듈(Common View Generation Module)에 의해 일반 사용자에게 보여질 공통 뷰로 작성된다.
- (4) 실행엔진(Execute Engine) : 실행 엔진은 실행 전략에 따라 서버 질의 형태로 분할하고, 래퍼에게 각 서버 질의의 결과를 요청한다. 래퍼는 서버 질의의 결과를 XML 문서 형태로 실행 엔진에 반환하며, 실행 엔진은 이들 결과를 취합하여 일반 사용자에게 보인다.

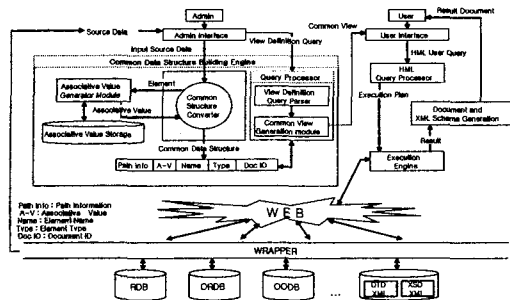


그림 5. 시스템 구조도

4. 결론 및 향후 연구과제

본 논문에서는 분산 이질 데이터에 대한 공통 뷰를 작성하기 위하여 복합 필드를 가진 공통 데이터 구조를 제안하였다. 공통 데이터 구조는 경로정보, 노드의 타입 정보를 가지며, 관련 요소들의 연결을 위한 연관값, 엘리먼트 이름, 소스 매핑 정보 등으로 구성된다. 공통 데이터 구조를 이용하여 공통 뷰를 작성할 경우 DTD 또는 XML Schema 기반의 XML 문서에 대한 처리 뿐 아니라, 다양한 논리 구조를 갖는 데이터베이스 시스템에 대한 처리도 가능하다. 향후 연구과제로 연관값 생성의 자동화 능력을 높일 것이며, 공통 뷰를 이용한 사용자 질의의 시, 절의 결과에 대한 타당성을 검증하기 위해 신뢰도(relevance value)[10] 추출에 대한 연구를 진행할 예정이다.

참고문헌

- [1] Wiederhold, G., "Mediators in the Architecture of Future Information System," IEEE Computer Vol.25, No3, pp. 38-49, 1992.
- [2] 양정욱, 홍동완, 이덕형, 윤지희, "웹 정보 통합 및 검색을 위한 XML기반 미디어 데이터 시스템의 개발," 한국데이터베이스 학회 2001년 춘계 Conference, pp. 281-294, 2001.
- [3] Lee, D., Chu, W. W., "Comparative Analysis of Six XML Schema Languages," ACM SIGMOD Record, 29(3), pp. 76-87, 2000.
- [4] 이승원, 권석훈, 김미혜, 이경하, 이규철, "XML Schema를 이용한 스키마 통합 시 충돌 문제의 분류," 한국정보과학회 학술발표논문집, Vol. 28, No. 2, pp. 31-33, 2001.
- [5] "A Query Language for XML," <http://www.w3.org/TR/NOTE-xml-ql>
- [6] "An XML Query Language," <http://www.w3.org/TR/2002/WD-xquery-20020816>
- [7] 홍동완, 윤지희, 남궁숙, "XML문서를 이용한 병원정보교환 시스템," 대한의료정보과학회지, Vol. 7, No. 2, 2001
- [8] "Health Level 7," <http://www.hl7.org>
- [9] Dollin, R., Biron, P., "HL7 Recommendation : Using XML as a Supplementary Messaging Syntax for HL7 Version 3.0," HL7 XML SIG, 1999.
- [10] Cooper, B., Neal, S., Franklin, M. J., Shadmon, M., "Extensible Data Management in the Middle-Tier," RIDE 2002.