

비정형 기술문서에 대한 XML 스키마 개발

정성윤*, 김성진**

한국건설기술연구원 건설CALIS연구센터
e-mail:syjeong@kict.re.kr

A development XML schema for non-formal technical documents

Seong-Yun Jeong*, Seong-Jin Kim**

Construction CALS Research Center, Korea Institute of
Construction Technology

요 약

전자거래기술이 발전되면서 인터넷을 통해 다양한 포맷의 전자문서가 유통되기 시작하였다. 이로 인해 상이한 문서 포맷간의 호환성 결여와 문서 데이터의 교환 등의 문제가 대두되기 시작하였다. 이를 위해 많은 전자문서가 XML 포맷으로 작성, 유통되기 시작하였다. 하지만 대부분의 XML 전자문서는 일정한 형식을 가지면서 분량이 적은 서식문서를 대상으로 개발되고 있으며 분량이 방대하고 비정형 구조를 갖는 기술문서에 대한 XML 전자문서의 연구 개발은 많지 않은 실정이다. 본 연구는 이러한 기술문서를 XML 전자문서로 할 때 공통으로 표현될 수 있는 정보요소와 정보로서 가치가 있는 구성 항목 등을 분석하여 35종의 공통 정보요소에 대한 XML 스키마를 개발하였다.

1. 서론

90년대 말부터 인터넷을 활용한 전자상거래 기술이 발전하면서 기업과 소비자간(B2C)의 전자 거래 체계에서 기업과 기업간(B2B), 기업과 정부간(B2G) 상거래로 점차 그 영역이 확대되고 있는 추세이다. 건설분야에서도 aec(architecture, engineering, construction)-XML, gb(green building)-XML, Land-XML 등과 같이 인터넷을 통해 이해관계자들이 서로의 정보를 교환하거나 공동으로 활용하기 위한 응용 기술 개발이 다양해지고 있다. 이처럼 전자상거래의 영역 확대와 응용 분야가 다양해짐에 따라 여러 유형의 전자문서의 유통이 활발해지고 있다. 이에 따른 표준화된 전자문서 정보체계의 필요성이 강조되고 있다. 특히 타 산업분야에 비해 건설분야는 도로, 댐, 건축 등 하나의 시설물을 기획하여 완공하기까지 짧게는 몇 년에서 길게는 수십년에 이르는 사업기간에 엄청난 분량의 문서가 생성, 유통되고 있다. 이 때 공사유형에 따라 법정 서식문서뿐만 아니라 각종 계획서, 보고서, 계산서, 산출서, 기술서

등 다양한 종류의 기술문서가 만들어진다. 이들 기술문서는 대체로 일정한 형식을 갖추어져 있지 않고 방대한 분량으로 이루어져 있다. 또한 비록 동일한 기술문서일지라도 건설공사 유형이나 발주방식에 따라 서로 다른 체계로 문서 내용이 구성되어 있으며, 동시에 설계도면이나 공사에 필요한 시방서, 지침서, 기술규격 등의 문서를 동시에 열람하는 경우가 많다. 본 연구는 이러한 특성을 갖는 건설관련 기술문서를 인터넷을 통해 관련주체들간에 건설정보를 원활하게 교환 및 공유하기 위한 목적으로 W3C에서 개발한 XML 스키마 규격을 준용하여 비정형 기술문서에 대한 XML 스키마를 개발하고자 하였다. 이를 위해 기술문서를 XML 전자문서 포맷으로 작성하기 위한 설계문서와 준공문서 및 공사시방서 등의 기술문서를 대상으로 공통논리구조를 분석하였다. 분석된 논리구조를 블록다이어그램으로 모델링한 후, W3C의 규칙에 따라 XML 스키마를 개발하였다. 끝으로 표준시방서를 대상으로 개발된 스키마에 따라 XML 전자문서를 개발하였다.

2. 건설분야에서의 기술문서 특징

건설사업은 하나의 시설물에 대하여 기획에서부터 설계, 시공, 유지관리까지 전 과정에 걸쳐 엄청난 분량의 문서가 생성된다. 또한 건설공사의 유형에 따라 공사시방서, 착공보고서, 설계보고서, 시공계획서, 유지관리계획서, 구조계산서, 산출내역서, 작업규정, 지침서, 설계도면, 현장사진, 공문서, 서식문서 등 다양한 종류의 문서가 발생한다. 이들 문서들 중에서 공문서나 서식문서를 제외한 거의 모든 문서들은 다양한 문서구조로 이루어져 있다. 또한 건설관련 기술문서는 설계도면, 건설현장사진, 분석도표, VR(Virtual Reality)과 같은 가상모델, 그림, 수학적 등 여러 종류의 정보요소를 포함하고 있다. 이렇게 다양하고 복잡한 정보요소를 갖는 기술문서마다 독자적인 구성체계와 표현방식으로 디지털화한다는 것은 현실적으로 어려움이 많다. 특히, 기술문서를 발주자, 설계자, 시공자 등 참여주체간에 원활하게 교환 및 공유하기 위해서는 기술문서를 단순히 전자화하기보다는 특정 소프트웨어에 구애를 받지 않으면서 통일된 구성체계와 정보요소에 따라 디지털화할 수 있는 표준화된 기술문서의 정보체계 정립이 필요하다. 이를 위해 본 연구는 기술문서의 포맷으로 W3C에 개발한 XML을 선정하였다.

3. 기술문서의 공통 논리구조정보 분석

기술문서를 정해진 정보체계에 따라 XML 포맷으로 디지털화하기 위해서는 기본적으로 기술문서의 논리구조 분석이 선행되어야 한다. 본 연구는 건설사업을 수행하면서 지속적이면서 빈번하게 참조되는 공사시방서, 기준서, 기술서, 설계보고서, 기술자료집 등을 대상으로 기술문서의 논리구조를 분석하였다. 본 연구는 기술문서의 공통 논리구조정보를 정의하기 위해 다음과 같은 분석과정을 정의하였다. 먼저, 데이터 유형을 분석하기 위하여, 기술문서와 직·간접적으로 관련되는 오디오, 사진, 동영상, 전자지도, CAD 도면 등의 데이터 유형과, 기술문서와 이들 데이터간의 연결정보를 파악한다. 또한 이들 데이터를 전자매뉴얼에서 볼 수 있도록 전용 뷰어 프로그램에 대한 정보를 분석한다. 둘째, 주요 기술정보를 분석하기 위하여 기술문서 내용 중 반복적이면서 빈번하게 나타나는 데이터, 정보로써 가치가 있는 데이터, 타 전자매뉴얼이나 정보시스템과의 상호 연계되는 정보, 업무단계별 또는 참여주체별로 공유 및 재활용되는 정보 등을 분석한다. 셋째, 기술문서를 정해

진 정보체계에 따라 작성, 공유 및 재활용될 수 있도록 유사한 의미를 갖는 용어와 약어를 하나의 의미로 사용되도록 용어와 약어를 분석한다. 또한 기술문서의 내용을 관리 및 검색하기 위해 사용될 수 있는 코드정보를 분석한다. 넷째, 검색, 북마크, 변경이력정보 등 전자매뉴얼의 기능을 구현하기 위하여 정보요소에 대한 논리구조를 분석한다. 다섯째, 기술문서를 전자매뉴얼 체계하에서 관리할 수 있도록 필요한 구성항목과 구성체계를 분석한다. 끝으로 개발된 기술문서 정보를 타 전자문서 포맷이나 전자매체로 변환하거나 타 건설정보시스템과의 연계 등 재활용하는데 적용될 수 있는 정보체계를 분석한다. 이와 같이 전제조건에 따라 공통 논리구조를 분석하고 분석된 결과를 기초로 하여 XML 스키마와 XML 전자문서를 개발하기 위해 다음과 같은 절차를 설정하였다. 먼저, 기술문서의 구성요소와 속성을 파악하고 파악된 구성요소들간의 계층적 구성체계를 분석하였다. 이 때 공사명, 과업명, 사업명, 프로젝트명 등과 같이 동일하거나 유사한 의미를 갖는 구성요소들을 하나의 그룹으로 묶어서 대표되는 구성항목을 정의하였다. 또한 용어, 표제(Title), 목차(Index), 표(Table), 도면(Drawing), 목록(List), 그림(Figure), 폰트(Font) 등 기술문서에 표현되는 정보요소들 중 공통으로 사용되는 구성체계와 구성항목을 재구성하였다. 다음으로 주로 전문적인 기술사항을 담고 있는 기술문서 내용을 미숙련된 기술자도 보다 쉽게 이해할 수 있도록 기존의 종이형태에서는 표현하지 못하였던 기능 즉, CAD도면, 디지털 사진, 동영상, 시뮬레이션 등의 멀티미디어 데이터와 하이퍼미디어 기능을 반영할 수 있도록 기술문서의 논리구조를 확대하였다. 이처럼 분석된 논리구조정보를 쉽게 표현하기 위하여 본 연구에서는 최상위 구성항목을 좌측 상단에 위치시키고 우측으로 갈수록 하위 계층의 구성항목을 나타내는 Top-Down 방식 블록다이어그램으로 모델링하였다[1]. 다음으로 W3C에서 설계한 XML 스키마 규칙에서 따라 기술문서의 논리구조정보를 스키마로 정의하였다. 끝으로 기술문서 스키마로 정의한 엘리먼트와 속성, 데이터 형식에 따라 XML 인스턴스와 XSL 스타일시트를 개발하였다. 본 연구는 비정형 구조를 갖는 건설관련 문서를 표준체계에 따라 디지털화하기 위해 미 국방성에서 전자매뉴얼 구축을 위한 기본요건을 명시한 MIL-PRF-87269 표준[2]을 준용하여 건설분야의 기술문서의 논리구조를 분석하였다. 분석 결과, 기술문서의

논리구조는 크게 공통계층과 내용계층 그리고 기술 문서별로 구분된다. 공통계층에는 기술문서에서 반복적이면서 빈번하게 표현되는 정보요소나, 공통으로 사용 가능한 구성항목들을 모아 놓은 집합체를 말한다. 이 공통계층으로 기술문서 내용을 XML 전자문서로 표현하는데 사용되는 “근거참조정보”, “규제”, “이탤릭체”, “글자스타일”, “윗첨자”, “아래첨자”, “단락정보”, “목록정보”, “수학적”, “빈줄”, “새줄”, “표정보”, “그림정보”, “동영상정보”, “음성정보”, “시뮬레이션정보”, “금지_경고_지시정보”, “대화형내용정보” 등의 하위 엘리먼트로 구성되도록 정의하였다. 다음으로 내용계층은 기술문서의 실제 내용을 전자문서 포맷으로 재가공하기 위해 정의한 것이다. 내용계층의 최상위 엘리먼트는 기술문서 내용 변경에 관한 이력정보를 담기 위한 “이력정보” 하위 엘리먼트와, 기술문서의 내용을 XML 전자문서로 표현하기 위한 “표제부정보”, “목차정보”, “설명정보”, “절차정보” 등의 하위 엘리먼트로 구성된다. 특히 “표제부정보”, “목차정보”, “설명정보”, “절차정보” 등의 엘리먼트 중 하나를 선택하도록 논리구조를 정의하였다. 기술문서별로 필요한 정보요소는 공통계층측면과 내용계층측면에서 정의한 논리구조 이외에 기술문서에서 공통으로 사용되거나 정보로서 가치가 있는 내용을 대상으로 논리구조를 정의하였다. 이 계층에서의 논리구조에는, “용어정의정보”, “서식정보”, “검사체크리스트정보”, “질의응답정보”, “절차도정보”, “참고문헌정보”, “시험정보”, “시공사례정보”, “손상정보”, “기호정보”, “코드정보”, “단위정보”, “개정_이력관리정보”, “기관정보”, “교육정보”, “법규정보”, “자재정보” 등의 하위 구성항목으로 이루어진다[3].

다음 그림 1은 본 연구에서 분석한 결과를 토대로 기술문서의 공통 논리구조를 모델링 한 내용이다

4. XML 스키마 개발

본 연구에서는 이와 같이 분석된 공통 논리구조를 기초로 하여 35종의 대표 정보요소에 대한 XML 스키마를 개발하였다. 이 중 대표적인 스키마를 살펴보면 다음과 같다. 먼저 내용측면에서 정의한 표제부 정보에 대한 스키마는 다음과 같다. 기술문서의 표지에 수록된 정보를 XML로 표현하기 위해 문서의 관리를 위한 “문서관리번호”, 문서를 대표하는 “문서명”, “발행일”, 문서를 발행한 기관정보를 기술하는 “발행기관명” 등의 하위 엘리먼트를 가진다.

이 하위 엘리먼트들은 순차적으로 발생하도록 논리 구조를 정의하였다.

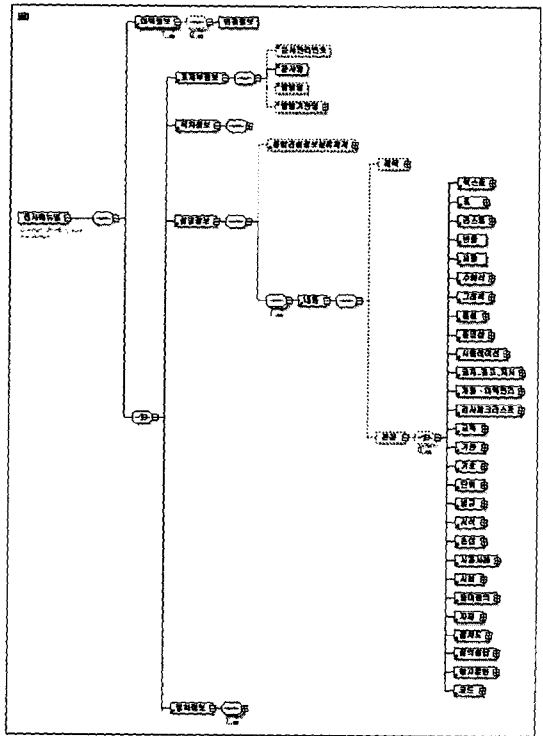


그림 1. 기술문서의 주요 정보에 대한 논리구조모델링 특히 “발행기관명” 엘리먼트는 문서발행기관의 로고를 이미지로 표시하도록 “그림정보” 엘리먼트를 하위 계층으로 구성하였다. 다음은 표제부정보에 대해 XML스키마를 정의한 내용이다.

```

<xs:element name="표제부정보">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="문서관리번호"
        type="xs:string" minOccurs="0"/>
      <xs:element name="문서명" type="xs:string"/>
      <xs:element name="발행일" minOccurs="0">
        <xs:complexType> <xs:simpleContent>
          <xs:extension base="xs:date"/>
        </xs:complexType>
      </xs:element>
      <xs:element name="발행기관명" minOccurs="0">
        <xs:complexType mixed="true">
          <xs:sequence minOccurs="0">
            <xs:element ref="그림정보"/>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
    <xs:attribute name="표제부정보_아이디"
      type="xs:string" use="required"/>
  </xs:complexType>
</xs:element>
  
```

다음으로 기술문서와 관련한 업무단계별 또는 기능별, 숙련수준별로 관련 정보를 상호 연계하여 순차

적인 정보 흐름 또는 사용자가 요구하는 조건에 따라 그 결과를 표현하기 위해 다음과 같이 절차정보" 엘리먼트를 정의하였다.

```
<xs:element name="절차정보">
  <xs:complexType>
    <xs:sequence>
      <xs:element ref="절차"/>
    </xs:sequence>
    <xs:attribute name="절차고유위치" type="xs:string"
      use="required"/>
    <xs:attribute name="절차정보명" type="xs:string"
      use="required"/>
    </xs:complexType>
  </xs:element>... 이하 생략 ...
```

이 엘리먼트는 논리적인 조건이나 수학적인 단항연산, 이항연산 등의 논리조건에 따라 "선행조건정보"와 절차의 현 단계에서 처리되는 내용을 기술하는 "본문내용정보", "후행조건정보" 등의 하위 엘리먼트로 구성되며, 각 절차별로 피드백할 수 있도록 "절차고유위치"라는 속성을 두었다. 다음은 공통계층 측면에서 "시물레이션정보"를 정의한 내용이다.

```
<xs:element name="시물레이션정보">
  <xs:complexType>
    <xs:sequence>
      <xs:element name="시물레이션제목"
        type="xs:string" minOccurs="0"/>
      <xs:element name="시물레이션">
        <xs:complexType>
          <xs:attribute name="넓이"
            type="xs:integer" use="required"/>
          <xs:attribute name="높이"
            type="xs:integer" use="required"/>
          ... 중간 생략 ...
        </xs:complexType>
      <xs:element name="시물레이션주석" minOccurs="0"/>
    </xs:sequence>
    <xs:choice maxOccurs="unbounded">
      <xs:element ref="단락정보"/>
      <xs:element ref="목록정보"/>
      <xs:element ref="수학적식"/>
      ... 중간 생략 ...
      <xs:element ref="동영상정보"/>
      <xs:element ref="음성정보"/>
      <xs:element ref="시물레이션정보"/>
      <xs:element ref="금지_경고_지시정보"/>
    </xs:choice>
  </xs:complexType>
  </xs:element>
  <xs:attribute name="시물레이션_아이디"
    type="xs:string" use="required"/>
</xs:element>
```

"시물레이션정보" 엘리먼트는 그래픽, 애니메이션, 사운드 등을 이용하여 가상의 환경을 표현하는 항목으로서, 전자매뉴얼에서 VRML 형식으로 표현된다. 이 엘리먼트는 시물레이션의 내용을 대표하는 "시물레이션제목", 실제 시물레이션 파일을 연결하는 "시물레이션", 그리고 시물레이션에 대한 부가설명을 기술하는 "시물레이션주석" 등의 하위 엘리먼트로

구성되어 있으며, "시물레이션_아이디"라는 속성을 가진다. "시물레이션제목" 엘리먼트는 텍스트타입으로 표현된다. 시물레이션 엘리먼트는 시물레이션 데이터를 컴퓨터 화면에서 표시하는데 필요한 "넓이", "높이", "종류", "표시형태"의 속성과 실제 파일명을 지정하는 "원본파일" 속성을 포함하고 있다. "종류" 속성값에는 "VRML"을 사용한다. "표시형태" 속성값에는 "동영상", "음성" 엘리먼트와 같이 "인라인", "아웃라인" 등의 값으로 표현한다.

5. 결론

최근 들어 B2C, B2B, G4C 등 전자거래의 영역이 확대되면서 전자문서의 유통이 중요시 되면서 XML을 이용한 전자문서 개발이 활발히 진행되고 있다. 하지만 대부분이 일정한 형식을 가지면서 분량이 작은 서식문서를 대상으로 XML 전자문서를 개발하고 있으며 분량이 방대하고 비정형 구조를 갖는 기술문서에 대한 XML 전자문서 개발 사례가 많지 않은 실정이다. 특히 건설분야의 특성상 공사시방서, 규격서, 기술서 등 다양한 기술문서를 반복적이면서 빈번하게 열람하는 경우가 많다. 이러한 기술문서를 웹 환경에서 보다 효율적으로 열람 및 교환할 수 있도록 미 국방성에서 개발한 MIL-PRF-87269 표준의 기본 개념을 참조하여 기술문서를 XML 포맷으로 개발할 때 공통으로 표현될 수 있는 정보요소와, 정보로서 가치가 있는 구성항목 등을 분석하여 35종의 공통 정보요소에 대한 XML 스키마를 개발하였다. 본 연구에서 개발한 XML 스키마는 건설분야 뿐만 아니라 타 산업분야의 기술문서를 XML로 개발하는데 적용할 수 있다. 아울러 전문적인 기술문서를 숙련되지 않은 기술자도 쉽게 이해할 수 있도록 기술문서와 의사결정지원시스템 또는 지식관리시스템과의 연계를 위한 스키마에 관한 추가적인 연구가 필요하다.

참고문헌

[1] 한국건설기술연구원, "전자문서교환을 위한 응용 기술연구", 1999.12
 [2] 미국방성, MIL-D-87269 Data Base Revisable : Interactive Electronic Technical Manuals, 1997.
 [3] 한국건설기술연구원, "건설CALS/EC 표준화 개발(II)", 2004. 6