

XML 스키마 매칭 기법을 이용한 구조설계 문서구조 표준화 방법론

A methodology for the standardization of structural design document structure using XML schema matching technique

김 봉 근* · 정 연 석** · 김 동 현*** · 이 상 호****

Kim, Bong-Geun · Jeong, Yeon-Suk · Kim, Dong-Hyun · Lee, Sang-Ho

ABSTRACT

A new formal standardization methodology of the structural design document information is proposed in this paper. The standardization process is divided into three steps: pre-process of the collected sample document (CSD), construction of the document structure, and definition of the occurrence of each element in the document. During the pre-process, the detail document contents in the CSD are indexed with templates defined in this study, and the indexed CSD is translated into XML Schema (XSD) format. Afterwards the degree of confidences of all elements between the temporary standard document (TSD) and the translated CSD are calculated by using the XML schema matching algorithm; the TSD is then updated. This second step is repeated until all of the CSD are compared. In the final step, the common elements and unbounded elements are extracted by determining the occurrence of the temporary document elements, and the standardized document schema is exported in the XSD format. The case study dealing with the structural calculation documents show that the proposed methodology can be effectively used to build a XML-based information model of structural design documents.

Keywords: *standardization, structural design document, XML, schema matching*

1. 서론

구조계산서와 같은 사회기반시설물의 설계문서정보는 구조물을 설계한 조건과 해석한 결과 등 구조물의 역학적 거동에 대한 세부적인 정보를 포함하고 있기 때문에 구조물의 안전성평가와 같은 유지관리단계에서 발생하는 각종 업무에 기본적인 기준이 된다. 1994년 성수대교의 붕괴 이후 제정된 '시설물의 안전관리에 관한 특별법'(건설교통부, 1995)에 의해 1종 및 2종 교량에 대해서는 한국시설안전기술공단에 이미지 파일이나 마이크로 필름과 같은 체계로 설계도서 정보를 보관해 왔다(건설교통부, 1996). 그러나 이러한 정보운영체계는 유지관리 업무에서 필요한 세부정보를 즉시 활용할 수 없기 때문에 최근 STEP과 XML 표준기술을 채택

* 정회원 · 연세대학교 토목공학과 박사과정 E-mail: bgkim@csem.yonsei.ac.kr

** 정회원 · 연세대학교 토목공학과 박사후과정 E-mail: ysjstep@csem.yonsei.ac.kr

*** 연세대학교 토목공학과 석사과정 E-mail: kdh@csem.yonsei.ac.kr

**** 정회원 · 연세대학교 토목공학과 교수 E-mail: lee@yonsei.ac.kr

한 정보시스템이 도입되었다(한국시설안전기술공단, 2004).

건설분야에서 설계문서정보체계 구축에 XML을 이용하는 방법은 크게 두 가지로 구분할 수 있다. 첫째는 메타데이터체계를 정의하는데 XML 정보모델을 활용하는 방법으로서, 이는 박재원과 최재황(2002), Popova 등(2002)의 연구 사례와 같이 설계도면파일을 포함한 대량의 각종 문서들을 통합적으로 운용하는데 주로 이용되어 왔으며, 설계단계에서 산출되는 각 문서의 주요 내용을 저장하고 새로운 프로젝트 진행 시 지식정보로 활용할 수 있도록 문서원본을 검색하여 찾아주는데 주안점을 두고 있다. XML 기반의 Electronic Document Management System (EDMS) 상품들도 이러한 범주에 속하며, 건설교통부(2004), 한국도로공사(2001), 한국시설안전기술공단(2004) 등 교량을 관리하는 주요 공공기관에서 운영하고 있는 납품체계의 XML 마스터 문서와 인덱스 문서도 XML을 일종의 메타데이터체계로 이용하고 있는 대표적인 사례이다. 그러나 이와 같은 설계도서정보의 운용체계는 설계문서의 세부적인 정보를 파일원본 수준에서 관리하므로 워드프로세서와 같은 응용 프로그램에 의존적인 형태를 탈피하지 못하고 있다. 이는 웹 상에서 운영된다 하더라도 서로 다른 응용프로그램 환경을 가지는 기관사이의 문서정보를 공유 및 교환하는데 제약을 줄 수 있으며, 사회 기반시설물의 유지관리단계에 필요한 각종 지능형 정보시스템을 구축하는 경우 사용자마다 다른 형식으로 작성된 설계문서의 세부적인 데이터를 의사결정 프로그램에서 직접 활용하는 것이 거의 불가능 하다.

XML을 문서정보체계에 활용하는 두 번째 방법은 문서자체에 대한 정보모델을 정의하여 XML 문서로 저장 모델을 개발하여 사용하는 방법이다. 건설교통부가 운영중인 건설사업정보교환시스템(한국건설CALS협회, 2000), 건설인허가·민원전자처리시스템(한국건설기술연구원, 2001) 등에서 서식문서의 교환에 사용하는 XML 문서가 이에 속하는 사례이며, 특히 Zhiliang 등(2004, 2005)이 표준화된 정보모델을 기반으로 한 XML 문서를 프로젝트 수행을 위한 의사결정지원 시스템에 활용한 사례와 같이 XML을 이용한 정보체계는 차세대 지능형 시스템 구축에도 효과적으로 활용될 수 있다. 그러나 구조계산서와 같이 복잡한 구조설계 문서를 표준화된 구조로 정의하기 매우 어렵기 때문에 대부분 건설 프로젝트에 사용되는 문서정보에만 한정되어 실무에 적용되어왔다.

본 연구에서는 대표적인 설계문서인 교량의 구조계산서 문서정보를 표준화하기 위해 수행한 이전 성과들을(Lee 등, 2004; 이상호 등, 2006) 바탕으로 설계문서정보를 표준화하는데 있어 객관적이며 정량화된 평가가 가능한 XML 모델링 방법론을 제시하였다.

2. 구조설계 문서정보 모델링 프로세스

본 연구에서는 구조설계 문서정보의 표준화된 정보항목을 도출하기 위한 프로세스를 그림 1과 같이 크게 3부분으로 구성하였다. 첫째는 실무에서 작성된 설계문서의 비교평가를 자동화하기 위해 수행되는 전처리과정이다. 전처리 과정은 표준화를 수행하고자하는 대상이 되는 영역의 문서를 수집하여 이를 컴퓨터에서 인식이 가능하도록 하는 인덱스를 표기한다. 이때 문서의 세부목차와 함께 구조설계에 있어 중요한 정보를 표기하며, 표기가 완료된 문서는 텍스트 파일로 변환한다. 둘째는 텍스트 파일로 변환된 문서를 읽어들이 XML Schema 파일로 변환하고 3.2절과 3.3절에서 설명한 XML 스키마 매칭 기법을 이용해 임시표준문서와 비교·평가한다. 임시표준문서는 최초로 읽어 들인 문서파일을 기

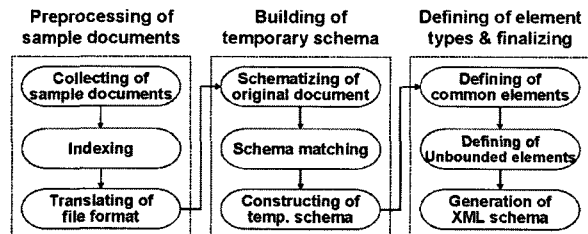


그림 1 XML 문서모델 표준화 프로세스

준으로 3.4절에서 제시된 방법론에 따라 수집된 다른 문서와 순차적으로 비교·평가한 결과를 이용하여 재구성되며, 이와 같은 과정은 수집된 문서를 모두 비교·평가할 때까지 반복된다. 셋째는 최종적으로 생성된 임시표준문서를 XML Schema 요소범주에 따라 각 요소가지는 출현빈도를 정의하는 과정이다. 이 과정에서는 임시표준문서의 각 요소가 공통된 요소인지 혹은 경우에 따라 필요한 요소인지를 구분하고, 반복적인 패턴을 가지고 나타나는 세부요소를 규명하여 최종 XML Schema 파일을 생성하는 과정으로 이루어져 있다.

3. 구조설계 문서의 임시 표준문서 구조화 방법

3.1. 문서의 특성정의

본 연구에서는 문서구조를 트리형식으로 특성화하였으며, 문서의 각 목차와 세부항목에 인덱스된 표본문서는 문서요소의 객체가 트리 형태로 컴퓨터 메모리상에 등록된다. 이와 같은 트리 형태로 문서를 이루는 각 요소의 특성은 요소명, 부모요소, 자식요소, 형제요소로 구분할 수 있으며, 이를 각 요소의 특성집합으로 표현하면 식 1과 같다.

$$E = \{E_N, E_P, E_B, E_C\} \tag{1}$$

여기서 E 는 각 요소가 가지는 특성집합을 의미하며, E_N 은 요소명집합, E_P 는 부모요소집합, E_B 는 형제요소집합 그리고 E_C 은 자식요소집합을 의미한다.

3.2. 문서항목간의 유사성 측정

본 연구에서는 Tversky와 Shafir(2004)이 제안한 특성 매칭 모델을 기반으로 비교대상이 되는 문서의 각 요소간의 유사성을 측정하였다. 예로 임시표준문서의 i 번째 요소와 표본문서의 j 번째 요소에 대한 요소명집합 E_N^i 와 E_N^j 사이의 유사도 $S_M(E_N^i, E_N^j)$ 는 식 2와 같이 나타낼 수 있다.

$$S_M(E_N^i, E_N^j) = \frac{N_{E_N^i \cap E_N^j}}{N_{E_N^i \cap E_N^j} + \alpha N_{E_N^i \setminus E_N^j} + (1-\alpha)N_{E_N^j \setminus E_N^i}} \tag{2}$$

여기서 N 은 연산된 집합성분의 개수를 나타낸 것이며, $E_N^i \cap E_N^j$ 는 E_N^i 과 E_N^j 의 교집합을 의미하고, $E_N^i \setminus E_N^j$ 와 $E_N^j \setminus E_N^i$ 는 각각 E_N^i 에서 E_N^j 를 뺀 차집합과 E_N^j 에서 E_N^i 를 뺀 차집합을 의미한다. 그리고 α 는 비교대상에 대한 가중치를 나타내는 계수로서 0에서 1 사이의 값을 가질 수 있으며, 본 연구에서는 두 요소에 대한 가중치가 같다고 보고 0.5를 사용하였다. 이와 같은 방법으로 해당 요소의 다른 특징인 부모요소집합(E_P), 자식요소집합(E_C), 형제요소집합(E_B)에 대한 유사도를 각각 얻을 수 있으며, 각 요소의 특징별로 산출된 유사도는 식 3과 같이 최종적으로 두 요소간의 유사도를 정하는데 이용된다.

$$S(E^i, E^j) = w_n S_M(E_N^i, E_N^j) + w_p S_M(E_P^i, E_P^j) + w_b S_M(E_B^i, E_B^j) + w_c S_C(E_C^i, E_C^j) \tag{3}$$

여기서 w_n, w_p, w_b, w_c 는 각각 요소명, 부모요소, 형제요소 및 자식요소에 대한 가중치를 의미하며, 각 가중치의 합은 1이 된다.

3.3. 문서구조를 고려한 릴렉세이션 레이블링

$B = \{b_1, \dots, b_m\}$ 및 $A = \{\lambda_1, \dots, \lambda_n\}$ 가 각각 객체 집합과 레이블 집합이라 하면 정량화된 제약조건은 $R_{ij} = \{r_{ij}(k, l)\}$ 과 같이 실수를 가지는 적합강도로 이루어진 4차 매트릭스로 표현된다(Yi 등, 2005). 여기서 $r_{ij}(k, l)$ 은 레이블 λ_k 가 객체 요소 b_l 로 할당되고 레이블 λ_l 이 객체 요소 b_j 로 할당되기 위한 적합강도를 말

한다. 본 연구에서는 두 문서 중에서 기준이 되는 임시표준문서의 구조와 비교 대상이 되는 표본문서의 구조를 각각 T 및 T' 으로 정의하고, 객체 집합 B 는 T 를 의미하며, 레이블 집합 A 는 T' 를 의미하는 것으로 가정하였다. 앞에서 언급한 유사도 평가 결과는 T 및 T' 에 포함되는 비교 대상 요소만을 고려되는 반면에 릴렉세이션 레이블링 기법을 적용하면 전체 문서구조와 이를 이루고 있는 연결관계가 스키마 매칭에 반영된다. 본 연구에서는 Yi 등(2005)이 제안한 방법에 따라 3.2절에서 언급한 유사도가 곧 두 문서 요소 사이에 매칭이 될 수 있는 초기신뢰도를 나타내는 것으로 가정하였으며, 릴렉세이션 레이블링 기법을 적용하여 개선되는 신뢰도는 식 4와 같다.

$$p_i^{(t+1)}(k) = \frac{p_i^{(t)}(k)q_i^{(t)}(k)}{\sum_{l=1}^m p_i^{(t)}(l)q_i^{(t)}(l)} \tag{4}$$

여기서 i 는 반복되는 스텝, i 와 j 는 임시 표준문서 T 에 속하는 요소를 나타내며, k 와 l 은 비교 대상이 되는 표본문서 T' 에 속하는 요소를 의미한다. 그리고 $p_i(k)$ 및 $p_i(l)$ 는 각각 k 가 i 로 매칭될 수 있는 신뢰도 및 l 이 i 로 매칭될 수 있는 신뢰도를 의미한다. 또한 $q_i(k)$ 및 $q_i(l)$ 은 지지정도로서 $q_i(k)$ 의 경우 식 5와 같이 나타낼 수 있으며, 앞서 가정한 사항에 따라 k 가 i 로 할당될 초기 신뢰도 $p_i^{(0)}(k)$ 는 유사도 평가를 통해 최종적으로 얻은 SE^i, E^k 를 이용하였다.

$$q_i^{(0)}(k) = \sum_{j=0}^{m-1} \sum_{l=0}^{n-1} r_{ij}(k, l)p_j^{(0)}(l) \tag{5}$$

여기서 $r_{ij}(k, l)$ 은 식 6과 같이 각 요소의 경로 특성에 따른 값을 사용한다.

$$r_{ij}(k, l) = \begin{cases} 1 & \text{if } e(i, j) = 1 \wedge e(k, l) = 1 \\ \left(\frac{1}{d_i + d_j} \right) & \text{if } e(i, j) = 1 \wedge e\sim(k, l) = 1 \text{ or } e\sim(i, j) = 1 \wedge e(j, k) = 1 \\ & \text{or } e\sim(i, j) = 1 \wedge e\sim(k, l) \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

여기서 e 는 에지(edge)로서 $e(i, j)$ 와 같이 2요소 쌍으로 정의되며, 요소 i 와 j 가 서로 직접적인 연결관계에 있다면 $e(i, j)=1$ 과 같이 정의하고 그렇지 않은 경우는 $e(i, j)=0$ 이다. 또한 $e\sim(i, j)$ 는 경로를 의미하는 것으로 임의의 요소 i 와 j 사이의 요소들이 연속되게 존재하는 경우 $e\sim(i, j)=1$ 과 같이 정의하고 그렇지 않은 경우에는 $e\sim(i, j)=0$ 이 되며, d_i 는 $e\sim(i, j)=1$ 인 경우 요소 i 와 j 사이에 존재하는 에지의 수를 의미한다.

3.4. 임시 표준문서의 구조화 방법

본 연구에서 의미하는 임시표준문서는 XML 문서구조가 정의되기 이전 표본문서에 포함된 문서항목을 모두 포함하는 문서모델을 의미한다. 이를 정의하는 방법은 3.2절과 3.3절에서 설명한 두 기법을 이용하여 상호 비교된 두 문서의 스키마 매칭 결과를 이용한다. 그러나 스키마 매칭을 통한 두 문서의 비교·평가과정은 단순히 두 문서에 포함된 각 요소간의 매칭 신뢰도만을 나타낸다. 따라서 평가결과 수치에는 크게 나타나는 경우와 작게 나타나는 경우가 존재하게 된다. 평가결과가 크게 나타난 요소의 경우 임시표준문서의 구조가 표본문서를 적절히 반영하는 것으로 볼 수 있으나 수치결과가 낮게 나타난 경우 표본문서의 내용을 반영하여 임시 표준문서를 새로이 구성해야 한다. 본 연구에서는 이러한 과정을 수행하는데 있어 임시 표준문서는 실무에서 작성한 모든 요소를 포함할 수 있는 일종의 데이터요소 풀이 되도록 구성하는 것으로 가정하였다. 이러한 임시표준문서를 갱신하기 위해 본 연구에서는 2개의 규칙을 정의하였으며 다음과 같다.

[정의 1] 임시 표준문서 집합 TS 와 표본문서 집합 CS 가 주어진다면, 새로운 임시 표준문서 집합 TS_{update} 는 TS 와 CS 의 상용요소를 제외한 표본문서의 새로운 문서요소 집합 $CS_{\neq w}$ 와 기존 임시 표준문서 집합 TS 의 합집합으로 이루어지며 식 7과 같다.

$$TS_{update} = TS \cup CS_{\neq w} \quad (7)$$

여기서 $CS_{\neq w} = \{j | i \in TS \wedge j \in CS \wedge \exists j \forall i (\max(p, (j)) < x)\}$ 이며, x 는 새로운 문서요소를 결정하기 위해 설정하는 변수값으로서 5장의 수치해석 예제에서는 0.8을 이용하였다.

[정의 2] 임시 표준문서 집합의 요소 i , 새로운 문서요소 j 그리고 j 의 자식요소 k 가 주어졌을 때, i 와 j 의 어미요소가 같고 i 의 이름특성과 k 의 이름특성 집합이 서로 상용하는 경우 새로운 임시 표준문서 집합에서 i 는 소거하며, 식 8과 같다.

$$E^i = \emptyset, \text{ if } (E_N^i = E_N^j) \wedge (\exists k (E_N^i = E_N^k)) \quad (8)$$

4. 구조설계 문서의 임시 표준문서 구조화 방법

4.1. 설계문서의 공통 요소 추출 방법

공통요소는 설계문서를 작성하는데 있어서 필수적으로 작성되어야 하는 요소를 말한다. 즉, 설계문서를 작성하는데 적어도 1번은 반드시 기입되어야 하는 문서요소를 말한다. 이는 XML Schema에서 요소의 출현회수를 정의하는 과정에 해당된다. 각 요소는 출현횟수에 대한 최소값과 최대값에 따라 크게 4개의 형태로 구분된다. 최소값이 1인 경우 공통요소로 구분할 수 있으며, 최소값이 0인 경우 선택요소로 구분할 수 있다. 본 연구에서는 최종 생성된 임시 표준문서구조를 다시 표본문서와 비교하여 각 요소의 최소 출현빈도를 산정하는 과정으로 설계하였으며, 각 요소의 최소빈도를 정하기 위해 다음과 같은 규칙을 정의하였다.

[정의 3] 최종 생성된 임시표준문서에 포함된 요소를 i 가 주어졌을 때 수집된 표본문서와 매칭되는 신뢰도가 일정비율 이상 출현되는 경우 i 는 공통요소이며, 식 8과 같다.

$$\text{minimum occurrence of element } i = \begin{cases} 1, & \text{if } \frac{\text{No. of } \max(p, (j)) \geq x}{\text{No. of sample document}} \geq \lambda \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

여기서 λ 는 공통요소를 결정하기 위해 설정하는 매개변수값으로서 0.9인 경우 최종 생성된 임시표준문서에 포함된 요소 i 가 전체 수집된 표본문서를 대상으로 90%이상 매칭되는 경우를 의미한다.

4.2. 문서요소의 최대 출현횟수의 결정 방법

요소 출현횟수의 최대값을 결정하는 방법은 정규화과정을 포함하고 있기 때문에 최소값을 결정하는 방법에 비하여 복잡하다. 표본문서를 통해 생성된 임시 표준문서의 경우 구조물 설계문서의 특성에 따라 자식요소가 비슷한 항목을 가진 형제요소가 많아진다. 이러한 일정한 패턴을 가지는 자식요소들을 포함하는 형제요소들은 새로운 요소명으로 대체하고 최대 출현횟수를 '∞'로 설정함으로써 XML Schema를 이용한 문서구조를 보다 단순화 시킬 수 있다. 본 연구에서는 이러한 과정을 수행하기 위해 다음과 같은 규칙을 정의하였다.

[정의 4] 최종 임시 표준문서 집합내의 요소 i 와 j 가 주어졌을때, i 와 j 가 형제요소이면서 자식요소수가 같고 i 와 j 의 자식요소들이 일대일의 상응관계를 가지는 경우 i 의 최대 출현빈도는 '∞'이며, 임시 표준문서 집합에서 j 는 소거하며, 식 9와 같다.

$$\text{maximum occurrence of element } i = \begin{cases} \text{unbounded, if } \mathcal{O}(i, i) = 1 \\ 1, \text{ otherwise} \end{cases} \quad (9)$$

여기서 $\mathcal{O}(i, j) = N_c(i, j) \times B(i, j) \times C(i, j)$ 이며, $N_c(i, j)$ 는 i 요소와 j 요소의 자식요소 수의 일치여부를 나타내는 값으로 자식요소 수가 같으면 1이고 자식요소 수가 같지 않으면 0의 값을 갖는다. $B(i, j)$ 는 각각 형제요소 집합의 일치여부를 나타내는 값으로서 i 요소와 j 요소를 제외한 형제요소의 집합이 같으면 1이고 같지 않으면 0의 값을 갖으며, $C(i, j)$ 는 자식요소 집합의 일치여부를 나타내는 값으로서 i 요소와 j 요소의 자식요소 집합이 같으면 1이고 같지 않으면 0의 값을 갖는다.

5. 교량 구조계산서를 대상으로 한 실험

본 연구에서 개발한 방법론의 적합성을 검증하기 위해 그림 2와 같이 교량 바닥판 설계서 작성된 구조계산서의 일부항목을 대상으로 실험을 실시하였다. 실험에 사용된 구조계산서의 문서항목은 각각 다른 회사에서 작성된 실제 구조계산서의 일부항목을 재편성한 것이다. 그림 2의 왼쪽은 표본문서를 나타낸 것으로 바닥판 설계와 관련하여 문서에 작성된 항목은 총 16개의 요소로 이루어져 있으며, 중앙의 임시표준문서로 가정된 문서는 총 13개의 요소로 이루어져 있다. 표 1은 3.2와 3.3절에서 설명한 스키마 매칭 기법에 의해 산정된 각 요소간의 매칭 신뢰도를 나타낸 것이다. 표 1에 나타난 결과에서 S3, S9-10, S13 및 S15-16은 매칭 신뢰도가 3.4절에서 정의한 '0.8'을 넘지 않아 새로운 요소로 등록된다. 이때 [정의 2]에 의해 T3-T7에 해당하는 요소는 임시표준문서의 해당 위치에서 삭제되고 새로이 포함된 S3과 S4-T8까지 매칭되는 자식요소가 임시표준문서를 갱신할 때 삽입된다.

그림 3은 4장에서 제시한 [정의 3]과 [정의 4]가 반영되어 생성된 XML 문서모델을 나타낸 것이다. 그림 3의 문서모델은 7종의 강교량 구조계산서에 포함된 바닥판 쉐일레버부를 대상으로 실험한 결과를 나타낸 것이다. 그림 3의 (a)는 [정의 3]까지 반영된 결과를 나타낸 것으로서 실선박스로 나타낸 요소들은 비교대상이

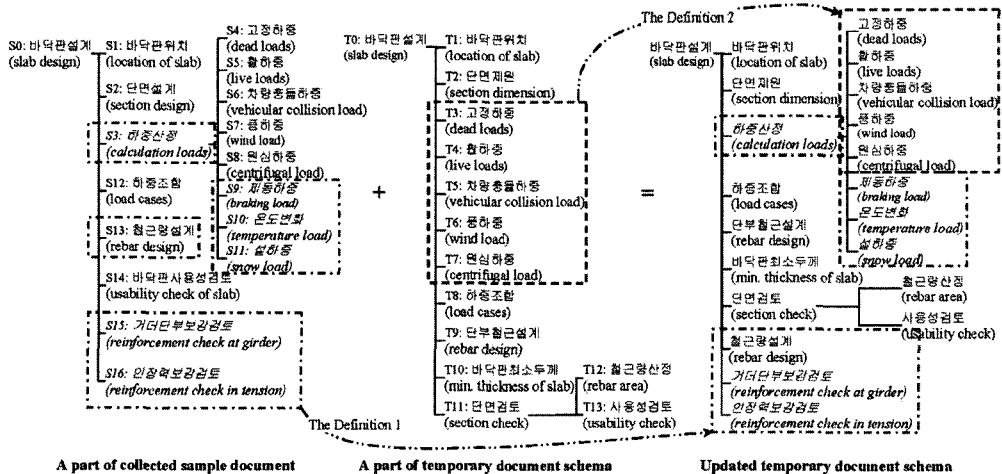
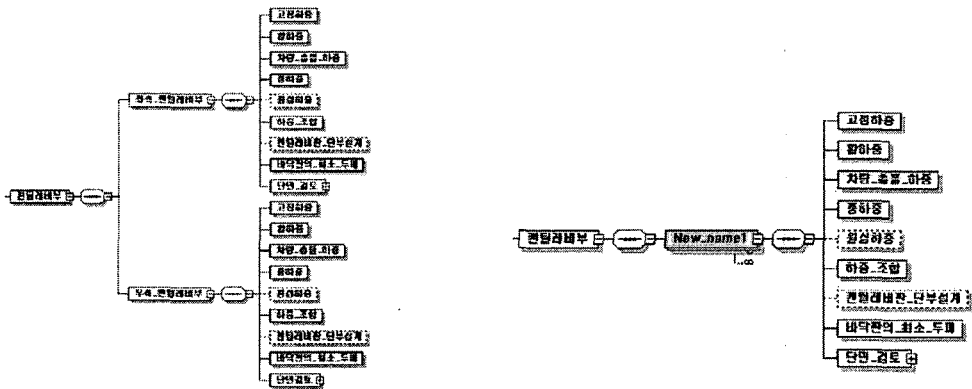


그림 2 실험대상 두 스키마와 임시문서구조의 갱신 결과

되는 모든 표본문서에 포함되어져 있는 공통의 문서항목을 의미한다. 그림 3의 (b)는 [정의 4]까지 반영되어 최종적으로 생성된 XML 문서모델을 나타낸 것으로서 그림 3의 (a)에 나타난 동일한 자식요소를 가지는 두 형제요소 '좌측_켄틸레버부'와 '우측_켄틸레버부'가 합쳐져 반복적으로 나타날 수 있도록 최대 출현빈도가 설정된 모습을 확인할 수 있으며, 본 연구에서는 사용자가 해당 요소명을 정의할 수 있도록 'New_name#' 형식으로 자동 생성되도록 하였다.

표 1 실험대상 두 스키마의 스키마 매칭 결과

	S0	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	S16
T0	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T1	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T2	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T3	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0
T4	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0	0
T5	0	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0	0
T6	0	0	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0	0
T7	0	0	0	0	0	0	0	0	1.00	0	0	0	0	0	0	0	0
T8	0	0	0	0	0	0	0	0	0	0	0	0	1.00	0	0	0	0
T9	0	0	0.50	0	0	0	0	0	0	0	0	0	0	0.50	0	0	0
T10	0	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T11	1.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T12	0	0	0	0.50	0	0	0	0	0	0	0	0	0	0.50	0	0	0
T13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1.00	0	0



(a) [정의 3]에 따른 XML Schema (b) [정의 4]에 따라 재생성된 XML Schema

그림 3 교량 구조계산서의 바다판 설계 켄틸레버부를 대상으로 한 실험 결과

6. 결론

본 논문에서는 복잡한 구조를 지니고 있는 사회기반시설물의 설계문서의 표준문서모델을 보다 효과적으로 개발할 수 있는 방법론을 제시하였다. 본 연구에서는 설계실무에서 작성된 문서항목간의 유사성을 정량적으로 평가하기 위해 스키마 매칭 기법을 사용하였으며, 5장의 실험예제에서 나타난 바와 같이 본 연구에서 개발한 스키마 매칭 알고리즘은 동일한 구조물을 다루고 있지만 작성자에 따라 서로 다르게 만들어진 문서의 각 항목간에 일치성 여부를 판단하는데 효과적으로 적용될 수 있었다. 또한 본 연구에서는 스키마 매칭 기법

을 통해 상호 비교·평가된 신뢰도를 바탕으로 새로운 문서구조를 생성하는데 필요한 방법을 개발하였으며, 이를 통해 대량의 표본문서들을 자동으로 비교·평가하여 표준화된 문서구조를 생성할 수 있는 기반을 제공하였다. 또한 이러한 과정에 의해 생성된 문서구조는 여러 형태의 데이터베이스나 응용 프로그램에서 활용될 수 있도록 플랫폼에 의존적이지 않은 XML Schema 형태로 자동으로 생성되도록 하였다. 본 연구에서 제시한 방법론은 범 국가적 차원에서 사회기반시설물에 대한 재난관리나 경제적인 시설물관리에 필요한 통합 데이터베이스 구축을 위해 여러 기관에 산재한 건설분야의 문서를 통합하여 표준화하는 경우 효과적으로 활용될 수 있을 것으로 기대된다.

감사의 글

본 연구는 건설교통부에서 실시한 건설핵심기술연구개발사업(교량설계핵심기술연구단)의 연구비 지원에 의해 수행되었으며, 이에 깊은 감사를 드립니다.

참고문헌

- 건설교통부 (2004) 도로 및 하천공사의 설계·준공도서 전자납품 편람 (V. 1.0), 건설교통부.
- 건설교통부 (1996) 준공도서사본작성지침, 건설교통부.
- 건설교통부 (1995) 시설물의 안전관리에 관한 특별법, 건설교통부.
- 이상호, 정연석, 김봉근 (2006) 교량 유지관리 지원을 위한 CAD/CAE 정보와 엔지니어링 문서정보의 통합 데이터베이스, 한국CAD/CAM학회논문집, accepted.
- 박재원, 최재황 (2002) 건설공사정보. 메타데이터의 XML 스키마 설계에 관한 연구, 한국문헌정보학회지, 36(3), pp. 155-179.
- 한국건설기술연구원 (2001) 건설인허가·민원업무 전자처리체계 시범운영, 건설교통부.
- 한국건설CALS협회 (2000) 건설CITIS 시스템 사용자 지침서, 한국건설CALS협회.
- 한국도로공사 (2001) 전산설계도서 작성지침서, 한국도로공사.
- 한국시설안전기술공단 (2004) 설계도서 등의 사본작성 및 관리지침, 한국시설안전기술공단.
- Lee, S.-H., Kim, B.-G., Jeong, Y.-S., and Kang, H.T. (2004) The Bridge Design Process with Web-based Documents, *Proceedings of the Third International Conference on Advanced in Structural Engineering and Mechanics (ASEM'04)*, CD-ROM paper, pp. 1198-1204, Seoul.
- Popova, M., Peter, J. and Lindgren, H. (2002) An Integrated Platform for Case-based Design, *CIB W78 conference 2002*, Arhus, Denmark.
- Tversky, A. and Shafir, E. (2004) *Preference, Belief, and Similarity*, A Bradford Book, The MIT Press Cambridge.
- Yi, S., Huang, B., and Chan, W.T. (2005) XML application schema matching using similarity measure and relaxation labeling. *Information Sciences*, 169(1), pp. 27-46.
- Zhiliang, M., Heng, L., Shen, Q.P., and Jun, Y. (2004) Using XML to support information exchange in construction projects. *Automation in Construction*, 13(5), pp. 629-637.
- Zhiliang, M., Heng, L., Shen, Q.P., and Jun, Y. (2005) Utilizing exchanged documents in construction projects for decision support based on data warehousing technique. *Automation in Construction*, 14(3), pp. 405-412.