XML Schema를 이용한 스키마 통합시 충돌 문제의 분류

이승원, 권석훈, 김미례, 이경하, 이규철

충남대학교 컴퓨터공학과
(swlee, skhkwon, mbkim, bart, kclee)@ce.cnu.ac.kr

The classification of conflicts on schema integration with XML Schema

Seung-Won Lee, Seok-Hun Kwon, Mi-Hye Kim, Kyong-Ha Lee, Kyu-Chul Lee

Dep't of Computer Engineering, Chungnam National University

요 약

 컴퓨터 시스템의 발달과 WWW(World Wide Web) 기술의 보편화는 수많은 정보 시스템의 출현과 다양한 정보 서비스로 인한 기대와 인터넷 정보 자원을 구축하게 되었고 이로 인하여 분산되어 있는 다양한 형태의 정보 자원들에 대한 통합의 요구가 발생하게 되었다.

이러한 요구사항의 해결방안으로 DTD를 데이터의 스키마 정보로 활용한 데이터 통합에 대한 연구가 진행되어 왔으나 DTD 자체가 가지고 있는 단점들로 인해 많은 문제점들이 지적되어 왔다.

 이를 해결하기 위해 W3C에서 새로운 스키마 정의로 XML Schema를 표준화하였고, XML Schema는 DTD에 비해 방대한 데이터 타입 부분을 고려할 때 상당히 묻지 않은 해결책이 될 많은 충돌 문제들이 존재한다. 이에 따라 XML Schema를 이용해서 스키마 통합을 했을 때 발생할 수 있는 충돌 문제들에 대해 체계적인 분류를 통해 향후 XML Schema를 이용한 스키마 통합 연구에 있어 기반을 제시한다.

1. 서론

기존의 DTD(Document Type Definition)를 스키마 정보로 활용하는 데이터 통합(Data Integration) 방법[1,2]은 다양한 이점 정보의 데이터들을 인터넷상에서의 구조화된 문서를 표현하기 위한 언어인 XML(extensible Markup Language)이라는 형식으로 통합하려고 사용자들로 하여금 데이터가 작성된 시스템의 환경이나 저장 형태에 무관하게 동일한 결의를 통해 데이터들을 통합 검색할 수 있는 장점이 있었다. 하지만 DTD를 사용한 통합 방식은 이점 정보의 데이터 통합이 가능한 장점에도 불구하고, DTD가 표현할 수 있는 데이터 태그의 제한과 XML 문서 파서(parser) 이외에 DTD 파서(parsing)를 위한 별도의 파서가 필요하다는 단점 등이 지적되어 왔다[3].


따라서 본 논문에서는 XML Schema를 이용한 스키마 통합의 방향 연구 기반을 제시하고자 통합시 발생할 수 있는 다양한 충돌문제에 대해 체계적으로 분류한다.

본 논문의 구성은 다음과 같다. 2장은 XML Schema를 사용하여 스키마 통합을 했을 때 발생할 수 있는 충돌문제에 대해 체계적으로 분류해서 각각의 경우에 대해 알아보고, 3장에서는 정론 및 향후 연구 과제를 기술한다.

2. XML Schema 통합시 충돌 문제의 분류

XML Schema를 이용한 스키마 통합에서 발생할 수 있는 충돌문제는 우선 크게 데이터들의 스키마 입점성에 의해 발생할 수 있는 스키마 충돌(schematic conflict)과 데이터간의 초점성에 의해 발생할 수 있는 의미 충돌(semantic conflict)으로 나누어 볼 수 있다.
2.1 스키마 충돌
2.1.1 이름 충돌 (naming conflict)
이름 충돌은 XML Schema에서 쓰인 element, attribute 등의 이름에 관한 충돌로, 다시 두 가지 경우로 나▒다.

(1) 같은 이름이 다른 의미로 쓰인 경우 (synonym)

(2) 같은 의미로 다른 이름이 쓰인 경우 (homonym)
상품 이름을 표현하기 위한 이름으로 [그림 1]에서는 <ProductNum> 을, [그림 2]에서는 <ProductName>을 사용하였다.

[그림 1] XML Schema 1

[그림 2] XML Schema 2

2.1.2 구조적 충돌 (structural conflict)
데이터의 스키마 구조가 다른 경우에 발생하는 충돌로 한쪽에서의 attribute로 쓰인 것이 다른쪽에서는 element로 표현될 경우를 들 수 있었다. [그림 1], [그림 2]를 보면 상품 원본번호를 나타내기 위해 PID를 [그림 1]에서는 attribute로, [그림 2]에서는 element로 나타냈다.

2.1.3 타입 충돌 (type conflict)
서정되어 있는 데이터의 타입이 서로 다른 경우에 발생하는 충돌로, DTD에서의 타입은 XML Schema의 타입과 다르게 정의되어서 충돌 문제가 많게 나타났다. XML Schema에서 제공하는 데이터 타입은 크게 XML Schema 자체에서 제공하는 기본형(Built-in Type)과 사용자가 기본형에서 유도해서 계층화할 수 있는 유도형(Derived Type)의 2가지로 나누어 볼 수 있는데, 이에 따라 발생할 수 있는 타입 충돌은 다음과 같다.

(1) 기본형과 기본형 사이의 충돌

[그림 3] XML Schema 3

[그림 4] XML Schema 4

(2) 기본형과 유도형 사이의 충돌

[그림 5] XML Schema 5

[그림 6] XML Schema 6

(3) 유도형과 유도형 사이의 충돌

[그림 7] XML Schema 7

[그림 8] XML Schema 8
2.2 의미 충돌
2.2.1 단위 충돌 (measurement conflict)
데이터의 단위가 달라 측정에 발생하는 충돌이다.

[그림 9] XML 문서 1

[그림 10] XML 문서 2

2.2.2 표현 단위 충돌 (granularity conflict)
단위 단위와 같이 단위가 다른 경우이나, 여기서의 단위들은 실제 포함 관계가 있는 경우이다.

2.2.3 표현 충돌 (representation conflict)
이용한 의미의 데이터를 달리 표현하는 경우에 발생하는 충돌이다.

위에서 나타낸 각 충돌의 경우를 XML Schema를 구성하는 요소를 이용하여 분류하면 표 1과 같다.

<table>
<thead>
<tr>
<th>Element-to-Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-to-One Element</td>
</tr>
<tr>
<td>- Element Name</td>
</tr>
<tr>
<td>- different names for equivalent elements</td>
</tr>
<tr>
<td>- same name for different elements</td>
</tr>
<tr>
<td>- Element Constraints</td>
</tr>
<tr>
<td>- Element occurrence</td>
</tr>
<tr>
<td>- Element's content model(+Data Type)</td>
</tr>
<tr>
<td>- Element default value</td>
</tr>
<tr>
<td>- Element Nullable</td>
</tr>
<tr>
<td>- Element structure</td>
</tr>
<tr>
<td>- missing attributes</td>
</tr>
<tr>
<td>- missing but implicit attributes</td>
</tr>
<tr>
<td>Many-to-Many Element</td>
</tr>
<tr>
<td>- Element Name</td>
</tr>
<tr>
<td>- Element Constraints</td>
</tr>
<tr>
<td>- Occurrence</td>
</tr>
</tbody>
</table>

[표 1] 스키마 통합시 발생하는 충돌의 분류

3. 결론 및 향후 연구 과제
본 논문에서는 기존 DTD를 이용했던 때의 단점을 극복하기 위해 오던 문제를 반영해서 새롭게 선보인 XML Schema를 두 표준을 통해 스키마 통합시 발생할 수 있는 충돌 문제에 대해 분류하고 보았다. 특히 단일 충돌 부분은 많은 기존과 다른 형태로 이루어질 수 있는 유의문으로 인해 많은 문제를 발생하게 한다.

본 논문에서 분류한 충돌 문제는 향후에 있을 XML Schema를 이용한 스키마 통합 시스템의 연구에 있어 중요한 부분이기 때문에 많은 연구가 필요하다. 이에 따라 XML Schema를 이용한 스키마 통합시 충돌 문제의 해결이 향후 연구가 진행중에 있다.

4. 참고 문헌