

# XML Schema를 이용한 스키마 통합시 충돌 문제의 분류\*

이승원, 권석훈<sup>0</sup>, 김미혜, 이경하, 이규철  
충남대학교 컴퓨터공학과  
(swlee, shkwon, mhkim, bart, kcllee)@ce.cnu.ac.kr

## The classification of conflicts on schema integration with XML Schema

Seung-Won Lee, Seok-Hun Kwon<sup>0</sup>, Mi-Hye Kim, Kyong-Ha Lee, Kyu-Chul Lee  
Dep't of Computer Engineering, Chungnam National University

### 요약

컴퓨터 시스템의 발달과 WWW(World Wide Web) 기술의 보편화는 수많은 정보 시스템의 출현과 다양한 정보 서비스로 인한 거대한 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 정의 정어로 XML Schema를 표준화하였고, XML Schema는 DTD에 비해 방대해진 데이터 타입 부분을 고려할 때 장점 못지않은 해결해야 할 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 이질 정보의 데이터들을 인터넷 상에서의 구조화된 문서를 표현하기 위한 언어인 XML(eXtensible Markup Language)이라는 형태로 통합함으로써 사용자들로 하여금 데이터가 저장된 시스템의 환경이나 저장형태에 무관하게 동일한 질의를 통해 데이터들을 통합 검색할 수 있는 장점이 있었다. 하지만 DTD를 사용한 통합 방법은 이질 정보의 데이터 통합이라는 장점에도 불구하고, DTD가 표현할 수 있는 데이터 타입의 제한과 XML 문서 파서(parser) 이외에 DTD 파싱(parsing)을 위한 별도의 파서가 필요하다는 단점 등이 지적되어 왔다[2].

최근 W3C에서는 기존 DTD의 단점을 보완한 XML Schema[3, 4, 5]라는 새로운 스키마 정의 언어를 제시하였는데, 기존 DTD가 단순히 문자열만을 표현할 수 있는데 비해 XML Schema에서는 DTD에서 표현할 수 있는 모든 데이터 타입 뿐만 아니라 정수, 부동소수점, 사용자 정의형 데이터 등 다양한 데이터

타입을 표현할 수 있게 되었다. 이로 인해 데이터에 대한 좀 더 정확한 표현이 가능하게 되었고, 또 XML Schema는 그 자체가 XML의 한 응용이기 때문에 XML Schema를 다루는 응용 프로그램내에서는 별도의 파서를 쓸 필요없이 기존의 XML 파서를 이용할 수 있게 되었다. XML Schema의 이런 장점은 스키마 통합시 DTD를 사용할 때 발생하던 많은 문제점들을 해결했으나, 반면에 DTD에 비해 더욱 복잡하고 방대해진 데이터 타입과 제약 조건(constraint) 등으로 인해 데이터 통합시 많은 어려움을 문제점으로 안고 있다.

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

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

### 2. XML Schema 통합시 충돌 문제의 분류

XML Schema를 이용한 스키마 통합에 있어서 발생할 수 있는 충돌문제는 우선 크게 데이터들의 스키마 이질성에 의해 발생할 수 있는 스키마 충돌(schematic conflict)과 데이터간의 이질성에 의해 발생할 수 있는 의미 충돌(semantic conflict)로 나누어 볼 수 있다.

\* 본 연구는 한국전자통신연구원의 위탁연구과제인 이질 자료 모델의 충돌 해결 방안 연구(과제번호: 104573)의 일부로 수행된 결과임

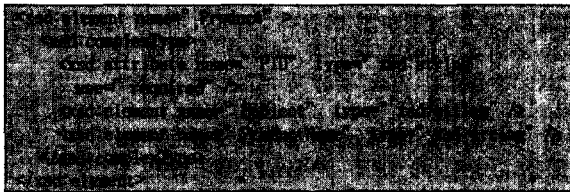
2.1 스키마 충돌

2.1.1 이름 충돌 (naming conflict)

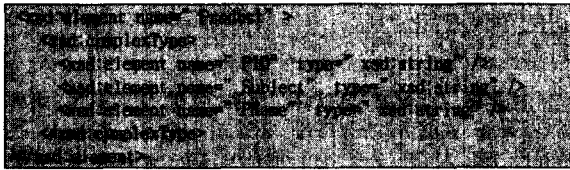
이름 충돌은 XML Schema 에서 쓰인 element, attribute 등의 이름에 관한 충돌로, 다시 두가지 경우로 나뉜다.

(1) 같은 이름이 다른 의미로 쓰인 경우 (synonym)  
[그림 1]과 [그림 2]의 <Subject> 를 보면 이름은 같지만 [그림 1]은 상품이 어느 제조사 소속인지를 [그림 2]는 상품이 어떤 분류에 속하는지를 나타낼 의도로 사용해서 같은 이름이 다른 의미로 쓰이고 있다.

(2) 같은 의미로 다른 이름이 쓰인 경우 (homonym)  
상품이름을 표현하기 위한 이름으로 [그림 1]에서는 <ProductName> 을 [그림 2]에서는 <PName> 을 사용했다.



[그림 1] XML Schema 1



[그림 2] XML Schema 2

2.1.2 구조적 충돌 (structural conflict)

데이터의 스키마 구조가 다른 경우에 발생하는 충돌로 한쪽에서는 attribute 로 쓰인 것이 다른 쪽에서는 element 로 쓰인 경우를 들 수 있겠다. [그림 1], [그림 2]를 보면 상품 일련번호를 나타내기 위한 PID 를 [그림 1]에서는 attribute 로 [그림 2]에서는 element 로 나타냈다.

2.1.3 타입 충돌 (type conflict)

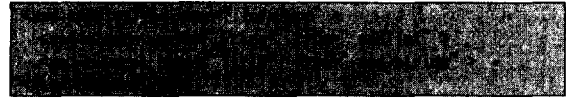
저장되어 있는 데이터의 타입이 서로 다른 경우에 발생하는 충돌로, DTD 에 비해 XML Schema 의 데이터 타입 부분이 더욱 방대해진 점을 고려할 때 충돌 문제의 가장 큰 비중을 차지한다.

XML Schema 에서 제공하는 데이터 타입은 크게 XML Schema 자체에서 제공하는 기본형(Built-in Type) 과 사용자가 기본형에서 유도해서 재정의할 수 있는 유도형(Derived Type) 의 2 가지로 나누어 볼 수 있는데, 이에 따라 발생할 수 있는 타입 충돌은 다음과 같다.

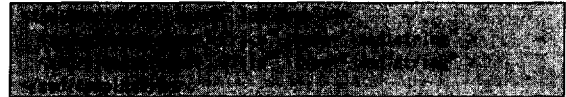
(1) 기본형과 기본형 사이의 충돌

XML Schema 기본형 사이의 데이터 타입이 달라서 발생하는 충돌이다.

위의 XML Schema 에서 <Price> 에 대한 데이터 타입으로써 [그림 3]은 int 를 [그림 4]는 string 을 사용했다.



[그림 3] XML Schema 3



[그림 4] XML Schema 4

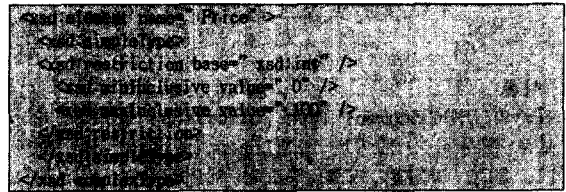
(2) 기본형과 유도형 사이의 충돌

XML Schema 기본형과 상속형 사이에서 발생하는 충돌이다.

[그림 5]와 [그림 6]은 두 경우 모두 <Price> 가 정수형을 표현하지만, [그림 6]은 기본형 int 에서 데이터 타입을 상속받아 값의 범위를 0 에서 100 사이로 제한한 상속형이 쓰인 경우이다.



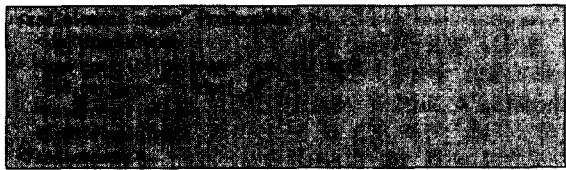
[그림 5] XML Schema 5



[그림 6] XML Schema 6

(3) 유도형과 유도형 사이의 충돌

XML Schema 유도형 사이에서 발생하는 충돌이다. [그림 7]과 [그림 8]은 두 경우 모두 상품의 수량을 나타내는 <ProductNum> 을 표현함에 있어서 각각 다른 기본형 타입으로부터 유도되어진 경우이다.



[그림 7] XML Schema 7



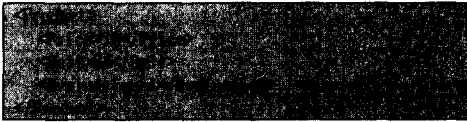
[그림 8] XML Schema 8

2.2 의미 충돌

2.2.1 단위 충돌 (measurement conflict)

데이터의 단위가 다를 경우에 발생하는 충돌이다.

상품가격을 나타내는 <Price>에 대해 [그림 3]에서는 미화(달러)로 표현하였고, [그림 4]에서는 한화(원)로 표현하였다.



[그림 9] XML 문서1



[그림 10] XML 문서2

2.2.2 포함성 단위 충돌 (granularity conflict)

단위 충돌과 같이 단위가 다른 경우이나, 여기서의 단위들은 서로 포함 관계가 있는 경우이다.

상품의 판매개수를 나타내는 <Sell>에 대해 [그림 3]에서는 주간 판매량을 의미하지만, [그림 4]에서는 월간 판매량을 의미한다.

2.2.3 표현 충돌 (representation conflict)

동일한 의미의 데이터를 달리 표현하는 경우에 발생하는 충돌이다.

상품 평가를 나타내는 <Evaluation>에 대해 [그림 3]에서는 각 등급을 A,B,C 로 나타내는 것을 기준으로 표현하였고, [그림 4]에서는 Good, Normal, Bad로 나타내는 것을 기준으로 표현하였다.

위에서 나타낸 각 충돌의 경우를 XML Schema를 구성하는 요소를 기준으로 분류해 보면 [표 1]과 같다.

[표 1] 스키마 통합시 발생하는 충돌의 분류

<ul style="list-style-type: none"> <li>■ Element-to-Element                     <ul style="list-style-type: none"> <li>■ One-to-One Element                             <ul style="list-style-type: none"> <li>▶ Element Name                                     <ul style="list-style-type: none"> <li>- different names for equivalent elements</li> <li>- same name for different elements</li> </ul> </li> <li>▶ Element Constraints                                     <ul style="list-style-type: none"> <li>- Element occurrence</li> <li>- Element's content model(=Data Type)</li> <li>- Element default value</li> <li>- Element Nullable</li> </ul> </li> <li>▶ Element structure                                     <ul style="list-style-type: none"> <li>- missing attributes</li> <li>- missing but implicit attributes</li> </ul> </li> </ul> </li> <li>■ Many-to-Many Element                             <ul style="list-style-type: none"> <li>▶ Element Name</li> <li>▶ Element Constraints                                     <ul style="list-style-type: none"> <li>- Occurrence</li> </ul> </li> </ul> </li> </ul> </li> </ul>
---

<ul style="list-style-type: none"> <li>▶ Element Structure(=ComplexType)                     <ul style="list-style-type: none"> <li>- Children order</li> <li>- Children occurrence</li> <li>- Children types</li> <li>- Children node name</li> <li>- Children node structure</li> </ul> </li> <li>■ Attribute-to-Attribute                     <ul style="list-style-type: none"> <li>■ One-to-One Attribute                             <ul style="list-style-type: none"> <li>▶ Attribute Name                                     <ul style="list-style-type: none"> <li>- different names for equivalent attributes</li> <li>- same name for different attributes</li> </ul> </li> <li>▶ Attribute Constraints                                     <ul style="list-style-type: none"> <li>- occurrence</li> <li>- default values</li> <li>- data type</li> <li>- facets</li> </ul> </li> <li>▶ Default values</li> </ul> </li> <li>■ Many-to-Many Attribute</li> </ul> </li> <li>■ Element-to-Attribute</li> <li>■ Different Representation for Equivalent Data                     <ul style="list-style-type: none"> <li>■ Different Expression denoting same information</li> <li>■ Different Units</li> <li>■ Different Levels of Precision</li> </ul> </li> </ul>
--

3. 결론 및 향후 연구 과제

본 논문에서는 기존 DTD 를 이용했을 때의 단점으로 지적되어 오던 점들을 반영해서 새롭게 선보인 XML Schema 라는 표준을 통해 스키마 통합시 발생할 수 있는 충돌 문제에 대해 분류해 보았다. 특히 타입 충돌 부분은 많은 기본형과 다양한 형태로 만들어질 수 있는 유도형으로 인해 많은 복잡한 충돌 문제가 발생하게 된다.

본 논문에서 분류한 충돌 문제는 향후에 있을 XML Schema 를 이용한 스키마 통합 시스템의 연구에 있어 충돌문제라는 근원적인 고려사항을 제시하고 있고, 이에 따라 XML Schema 를 이용한 스키마 통합시 충돌 문제의 해결이라는 향후 연구가 진행중에 있다.

4. 참고 문헌

- [1] 이강찬, 이경하, 이규철, "XML 기반의 인터넷 정보 자원 통합", 데이터베이스 연구, 한국데이터베이스연구회, 제16권 2호, pp. 5-21, 2000.12.
- [2] C. Baru, A Gupta, B. Ludascher, R. Marciano, Y.Papakonstantinou, P. Velikhov, V. Chu, "XML-Based Information Mediation with MIX", exhibition program, ACM Conf. on SIGMOD' 99, Philadelphia, USA
- [3] W3C, Recommendation, " XML Schema Part 0: Primer" , <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/>, 2001
- [4] W3C, Recommendation, " XML Schema Part 1: Structures" , <http://www.w3.org/TR/2001/REC-xmlschema-1-20010502/>, 2001
- [5] W3C, Recommendation, " XML Schema Part 2: Datatypes" , <http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/>, 2001