

B2B상에서 메타 데이터의 의미값을 이용한 다른 기종간

데이터 베이스 통합 구조

이진수⁰ 노희영

강원대학교

{binary, rohhy}@pllab.kangwon.ac.kr

Using Semantic Values of Meta Data for Heterogeneous Database Integration

Architecture on B2B

Jin-Soo Lee⁰ Hi-Young Roh

Dept. of Computer Science, Kangwon National University

요 약

본 논문에서는 기존에 개발된 데이터베이스내의 데이터 소스를 공유하기 위해 통합 라이브러리를 구축하고 이를 통해 데이터베이스내의 소스 데이터에 대한 의미적인 통합을 위한 구조를 제안한다. 데이터의 의미적인 통합을 위해 데이터 스키마가 가지고 있는 실질적인 의미값을 기존의 각 데이터베이스 시스템에서 참조하고 데이터 공유에 대한 동의를 갖도록 메타 데이터를 이용하였다. 따라서 먼저 통합 라이브러리를 구축하여 메타 데이터에 대한 정보와 메타 데이터간에 존재하는 함수적인 관계를 변경 함수로 함께 구현하였으며, 데이터 통합을 위한 참고 자료는 각 시스템마다 온톨로지(Ontology)를 작성하여 활용하였다. 본 논문에서 제안하는 방법은 기존의 방법에 비해 메타데이터의 수정과 추가가 간단하며, 새로운 시스템을 통합 라이브러리로 합병하는 비용이 현저히 감소되었다.

1. 서 론

최근 기업은 급격한 기업환경의 변화에 발빠르게 적응하기 위해 기존의 데이터베이스 시스템내의 데이터를 공유하고자 많은 노력을 기울이고 있다. 즉, 정보 공유를 통해 기업간 합병이나 다른 기종간의 데이터베이스의 통합을 꾀하고 있다. 이러한 데이터베이스 통합을 바탕으로 기업간 전자상거래 형태인 B2B가 일반 개인을 상대로 하는 B2C보다는 투자 가치가 높은 사업으로 부각되고 있다. B2B는 기업간 정보를 공유함으로써 부가적인 인력이나 자원의 낭비를 막을 수 있고 협력 기업간의 데이터를 공유함으로써 새로운 정보를 창출하는데 드는 비용을 최소화 할 수 있다.

기업간 업무 통합 및 신뢰성과 안정성을 보장하여 협력체제 구축 비용을 절감할 뿐만 아니라 제품 개발 기간 단축과 판매촉진 비용을 절감하기 위해서 여러 가지 방법들이 제시되고 있다.

본 논문에서는 B2B를 추진하고 있는 여러 기존의 시스템에 최소한의 비용을 통해 데이터의 의미적인 통합을 꾀하고자, 기존 시스템에서 의미값(Semantic value)을 포함한 메타 데이터(meta data)를 이용하여 데이터 소스에 대한 공유를 유도하고자 한다. 이러한 메타 데이터를 이용해 다른 기종간의 데이터베이스에서 의미상으로는 같지만 서로 다른 객체 이름과 특성을 가진 데이터를 공유할 수 있는 방법을 제시하고 이에 적합한 모델로 항공예약 시스템을 구현하여 기존의 시스템과 제안한 통합 시스템간의 성능을 비교 분석하였다.

2. 관련 연구

기업간 업무 통합 및 신뢰성과 안정성을 보장하여 협력체제를 구축하는 비용을 절감할 뿐만 아니라 제품을 개발하는 기간을 단축하고 판매를 촉진하기 위해 여러 가지 방법들이 제시되고 있다. 과거 70년대부터 발전되어온 EDI(Electronic Data Interchange)기법은 제한된 분야에 대한 서비스와 표준 프로토콜을 사용해야 하는 부담으로 인해 그 한계를 드러내고 있다. 최근에는 유수의 기업들이 앞다투어 B2B 솔루션으로 제시하고 있는 XML(eXtensible Markup Language)이 부각되고 있다. 이 XML은 기업간 비즈니스 문서 형식과 메시지를 형태에 관계없이 교환이 가능하고, 기존에 사용중인 어플리케이션과 쉽게 연동할 수 있다는 장점을 가지고 있다. 뿐만 아니라, 고급 검색 기능을 제공하며, 현재 웹상의 데이터를 통합할수 있는 기술로 각광을 받고 있다. 또한,

기존의 EDI 시스템에 XML의 개념을 도입하여 두 가지의 장점을 모두 수렴하는 방법이 대두되고 있다. 그러나 XML은 현재까지 완전한 시스템을 구축한 사례가 드물고, 기존의 웹상의 데이터들이 대부분 HTML 형식을 취하고 있기 때문에 이를 XML로 변환해주는 과정이 필요하다. 또한 XML 문서의 경우에 의미적인 통합을 유도하기 위해서는 부가적인 태그간의 관계유지가 필요하는 등 기존의 데이터베이스 시스템을 통합하기 위해서는 많은 선행 과제들이 존재한다.

기존에 개발된 데이터베이스 시스템의 데이터간에 통합을 유도하기 위한 또 다른 방법으로 각 데이터베이스 시스템이 가지고 있는 데이터베이스 스키마에 대한 의미값을 포함하고 있는 메타 데이터를 이용해 통합을 유도하는 방법이 있다. 이 방법은 메타 데이터에 의미값을 부여하여 이 메타 데이터를 기반으로 각 시스템간에 소스 데이터의 의미값을 통합하여 동일한 객체에 대해서 동일한 의미로 해석할 수 있도록 해준다. 이 방법은 현재 데이터베이스 구조에 대한 명확한 이해를 토대로 메타 데이터를 구축하고 이 메타 데이터를 각 시스템에서 유지함으로써 소스 데이터에 대한 의미적 통합을 유도한다. 그러나 이 방법은 기존의 데이터 베이스 시스템에서 의미값을 가진 메타 데이터를 바탕으로 새로운 데이터 스키마를 동일하게 구축해야 하는 문제가 있다. 이런 경우에는 각 데이터베이스 시스템에서 공유할 수 있는 통일된 표준안이 필요하게 된다.

본 논문에서는 이러한 메타 데이터를 이용해 기존의 데이터베이스 시스템의 소스 데이터를 통합하고자 한다. 그러기 위해 본 논문은 각 데이터베이스 시스템의 데이터 스키마를 통일된 표준안으로 변경하지 않고, 통합된 라이브러리를 모든 데이터베이스에서 참조하여 데이터 스키마에 대한 정보만 통합 시스템에 함으로써 기존의 메타 데이터를 이용한 통합 방법보다 효율적인 데이터 통합 구조를 제안하였다.

3. 데이터베이스 통합 구조

3.1 메타 데이터(Meta data)

일반적으로 메타 데이터는 웹기반 응용 프로그램에서 주로 비 구조적인 데이터 소스에 대해서 직접적으로 그 데이터에 대한 의미적인 특성이나 데이터 구성에 대한 정보를 얻을 수 없는 경우에 사용되었다. 그리고 다른 기종간 데이터 소스의 통합을 위해 멀티 데이터베이스 시스템이나 정보 소스 관리 시스템에서 소스 데이터 간에 존재하는 관계

나 데이터 구성요소 등을 부가적인 레벨의 메타 데이터로 정의하고 이를 기초로 하여 통합화의 과정에 사용된다. 메타 데이터는 다른 기종간의 데이터베이스 시스템내에 저장되어 있는 데이터 소스를 공유하기 위해 이용되는 데이터 소스에 대한 정보인 뿐만 아니라, 소스 데이터에 대한 스카마로서 각 데이터 소스가 가지고 있는 의미와 특성 및 구조에 대한 정보를 포함하고 있다.

메타 데이터 값은 엄밀히 두 종류로 구분할 수 있다. 즉, 구조적인 부분에 대한 정보를 포함하고 있는 구조적인 메타 데이터 값과 의미적인 부분에 대해 정의되어 있는 의미적인 메타 데이터 값으로 구별할 수 있다. 메타 데이터는 실세계의 동일한 객체나 현상에 대해 다양하게 표현이 가능하다. 그러므로 응용 프로그램에 필요한 정보를 모두 부가적으로 데이터 통합 라이브러리에 정의 해두고 이를 참조하여 데이터 소스를 원하는 클라이언트나 서버의 데이터 포맷으로 메타 데이터의 내용을 변형하여 전달해 줄 수 있다.

3.2 변형 함수(Conversion Function)

데이터 소스를 의미적으로 동일한 수신측의 데이터 포맷으로 변형하기 위해서 데이터 변형 함수가 필요하다.

데이터 변형 함수는 분명한 데이터 소스의 의미를 전달하기 위해 필요한 함수이다. 이 함수를 통해 서로 다른 시스템간의 데이터 소스를 수신측에 부합되는 데이터 형태로 변형하게 되는데, 그 과정은 그림[1]의 메타 데이터 연결자에서 수행하게 된다. 변형함수의 호출이 이루어지면, 이 변형함수는 데이터 인스턴스 뿐만 아니라, 관계형 데이터베이스의 어트리뷰트가 가지고 있는 의미적인 내용을 참조하여 동일한 의미를 갖는 다른 이름의 어트리뷰트 또한 변경해 준다. 그러나 변형함수를 구현하는 과정에는 많은 어려움이 있다. 모든 데이터에 대해 변형 함수를 구현하기 위해서는 명확한 의미적인 통합이 요구되기 때문이다. 따라서 변형함수는 일반적으로 일대일로의 변형을 피하거나 변형 대상이 변형하고자 하는 대상의 부분집합일 때만 가능하다.

3.3 데이터 통합 구조 설계

그림[1]은 본 논문에서 제안하는 데이터 통합 구조를 보여주고 있다. 이 구조상에서 가장 중요한 역할을 수행하는 것은 메타 데이터 연결자이다. 메타 데이터 연결자는 로컬 시스템으로부터 질의를 받으면 이 질의를 다른 로컬 시스템에 부합되는 질의로 변형하기 위해 데이터 통합 라이브러리를 참조하여 변형된 질의를 다른 시스템으로 전달하는 역할을 한다.

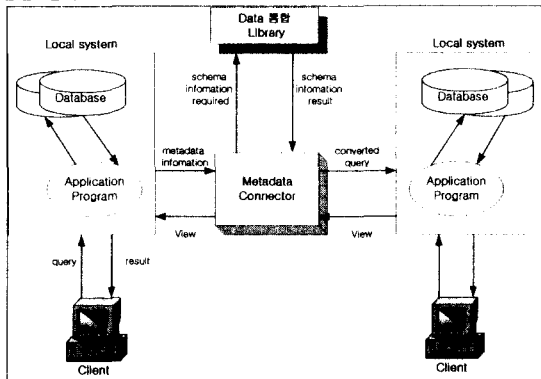


그림 [1] 데이터 통합 시스템 구조

이 과정에서 관계형 데이터베이스의 어트리뷰트와 인스턴스를 변형하기 위해 변형 함수가 필요하며, 이 변형함수는 데이터통합 라이브러리에 구현되어 있다. 메타 데이터 연결자는 또한 변경된 질의에 대한 응답을 수령하여 질의를 요청한 측에 부합되는 스키마형태로 모든 질의의 결과를 변형하여 돌려주는 역할을 한다. 통합 라이브러리내에는 변형함수뿐만 아니라 관련된 모든 시스템에 대한 메타 데이터를 가지고 있으며, 이 메타 데이터의 사전인 온톨로지를 포함하고 있다. 본 문항에서는 통합 라이브러리의 구성자인 메타 데이터와 변경함수, 온톨로지 테이블형태로 설계하였다. 그림[2]는 메타 테이블과 변경함수 테이블, 온톨로지 테이블의 구성을 설명하고 있으며, 메타 테이블의 경우는 각 어트리뷰트 이름과, 타입정보, 위치정보, 의미값, 속성값을 가지며, 변경함수 테이블의 경우는 두 개의 속성값에 대해 일대일로

변경될 함수와 이 변경함수의 의미값을 포함하고 있다. 온톨로지 테이블은 메타 데이터의 사전과 같은 역할로 객체의 의미값에 대해서 각 로컬 시스템별로의 동일한 객체의 이름정보를 가지고 있다.

• meta table

Attirb	A_type	Emb_table	DB_Name	Semantic	Property
--------	--------	-----------	---------	----------	----------

• Coversin table

Property1	Property2	Conversion	Semantic
-----------	-----------	------------	----------

• Ontology

Semantic	Local system	Local system
----------	--------------	--------------	-------

그림[2] 통합 라이브러리 구성

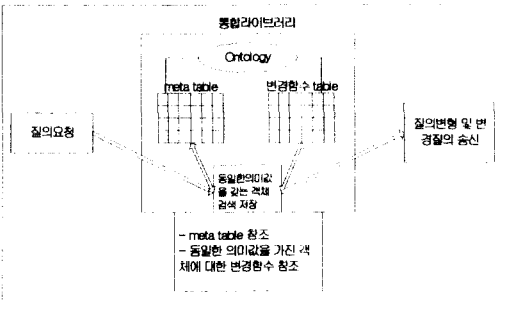
변경함수는 일반적으로 동일한 의미값을 가지는 데이터 소스에 따라서 비례적으로 증가하게 된다. 동일한 의미값을 가지는 데이터 소스가 n 개 일 때, 각 n 개에 대해 하나씩의 변경함수가 존재하므로 변경함수의 개수는 n! 개가 된다. 변경함수의 개수가 급속히 증가하게 되지만, 사실상 동일한 의미값을 갖는 데이터 소스 자체의 개수는 한정되어 있으며, 또한 실제로 기존에 사용하던 데이터 소스를 기초로 하여 새롭게 추가 변경되는 데이터 소스에만 변경함수를 추가하여 주면 모든 동일한 데이터 소스에 대한 변경함수를 정의할 수 있게 된다.

$con_semantic(p_1, p_2) = conversion_2 * p_1(data\ source)$
$con_semantic(p_1, p_3) = conversion_3 * p_1(data\ source)$
$con_semantic(p_1, p_n) = conversion_n * p_1(data\ source)$
$con_semantic(p_n, p_{n-1}) = conversion_{n-1} * p_n(data\ source)$

그림[3] 변경함수의 정의

위 그림[3]의 변경함수의 정의를 보여주고 있다. 변경함수의 간단한 예로 의미값이 '거리(semantic = 'distance')를 나타내는 단위원 'km'로 표기된 값을 'mile'로 환산하는 함수를 구하고자 할 때, 'km'를 'mile'로 변경하는 함수 $con_distance(p_2, p_1)$ 을 적용하면 다음과 같은 결과를 얻을 수 있다. 이러한 함수는 변경 함수 테이블에 저장되어 있는 인스턴스 값을 참조하여 추출하게 된다.

4. 통합 시스템의 기능



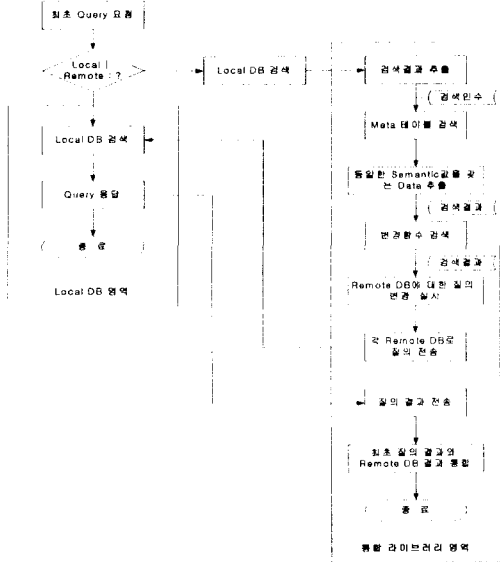
그림[4] 통합 라이브러리

본 논문에서 제안한 통합 시스템에서는 통합 시스템내 연결되어 있는 모든 데이터베이스에 대한 소스 데이터를 검색할 수 있으며, 로컬 데이터베이스내에서 데이터 스키마가 추가되거나 수정될 경우에도 곧바로 통합 라이브러리로 이 내용을 반영해 줄 수 있다. 뿐만 아니라 새로운 데이터베이스 시스템을 추가할 경우에도 통합 라이브러리에 있는 온톨로지를 표준으로 추가하고자 하는 데이터베이스의 스키마를 등록하여 줌으로써, 연결된 모든 데이터베이스에 이를 감지하고 통합된 결과를 추출할 수 있다. 여기서, 통합 라이브러리의 기능이 시스템 성능을 좌우하게 된다. 통합 라이브러리의 주된 역할은 그림[4]에서 보는 바와 같이 메타 테이블과 변경함수 테이블을 참조하여 질의를 요청한 측의 스키마에 부합되는 형태로 원격 데이터베이스 시스템의 내용을 검색 변환하여 통합된 결과를 얻는 것이다. 제안한 통합시스템에서는

항공 예약 시스템을 적용하여 통합된 결과를 구현하였다.

4.1 통합 데이터 검색

통합 라이브러리로 최초 질의가 요청되면 로컬 영역의 데이터베이스를 검색한 후 그 결과값을 통합 라이브러리로 보내게 된다. 이렇게 전달된 결과값은 그 결과 인수를 매개로 메타 테이블을 다시 검색하게 된다. 메타 테이블은 동일한 의미를 가진 데이터와 그 데이터의 위치 정보가 함께 포함되어 있기 때문에 메타 테이블을 검색함으로써 실질적으로 원하는 데이터에 대한 검색만을 통해 최초 질의에 대한 충분한 응답을 얻을 수 있다.



그림[5] 통합 검색 과정

메타 테이블을 검색한 결과를 토대로 최초 질의에 부합되는 데이터 타입을 추출하기 위해서 변경함수 테이블을 참조하게 된다. 변경함수 테이블은 동일한 의미를 갖더라도 그 타입이 틀린 경우에 변경함수 테이블의 데이터를 이용해 두 타입에 대한 일치를 유도할 수 있다. 이 과정은 모든 메타 테이블 검색결과에 적용되는 것은 아니다. 대부분의 기업 시스템 통합은 동일분야나 유사 분야의 데이터 통합을 목적으로 하기 때문에 대부분 특정 수치데이터나 지역명과 같은 위치 정보, 날짜와 시간을 나타내는 데이트 정보에 대해서만 변경함수를 적용하게 된다. 따라서 메타테이블 검색 결과에 이러한 인수가 포함되어 있는 경우에만 변경 테이블 검색 요청을 하게 된다. 그렇지 않은 경우에는 메타 테이블의 검색 결과에 대해서 곧바로 리모트 데이터베이스에 적용할 수 있는 질의로 최초 질의 결과값을 변형해 주면 된다. 이렇게 변경된 질의는 메타 테이블의 검색 결과에서 도출된 관련 리모트 데이터베이스로만 그 질의를 전송하게 된다. 이렇게 전송된 질의는 리모트 데이터베이스 서버에서 로컬 검색과 같이 질의에 대한 결과값을 도출하게 된다. 도출된 결과값은 다시 통합 라이브러리로 그 전송되게 되며, 그 결과값을 다시 최초 질의에 부합되는 결과로 변형해서, 최초 결과값과 통합하여 통합된 검색 결과값을 추출하게 된다. 그 과정을 그림[5]와 같이 설명할 수 있다.

4.2 통합 데이터 수정

전통적인 데이터베이스 시스템에서 자료의 수정은 간단하다. 여기서 자료는 소스 데이터를 변경하는 것을 의미한다. 물론, 소스 데이터 자체뿐만 아니라 현재 사용하고 있는 어트리뷰트의 이름과 타입을 변경할 수도 있다. 그러나 본 논문에서 제안하는 자료의 수정이란 메타 테이블의 내용 자체가 변경되는 것을 의미한다. 데이터의 의미적 통합을 유지하기 위해서는 로컬 시스템에서 변경된 데이터 스키마의 내용이 통합 라이브러리에도 반영이 되어야 한다. 이러한 일련의 과정은 다음과 같다. 먼저 로컬 시스템에서 데이터 스키마의 수정이 있을 경우, 그 수정된 어트리뷰트 정보를 메타 테이블에서 검색한다. 메타 테이블에

서 변경 이전의 정보를 새롭게 업데이트하고, 변경된 메타 데이터가 변경함수와 관련이 있는 데이터일 경우에는 변경함수에도 이와 동일한 수정을 해준다. 반대로 리모트 시스템의 소스 자료를 변경할 때에는 먼저, 통합 라이브러리를 변경하고, 그 후에 리모트 시스템의 스키마를 변경해 준다.

4.3 통합 데이터 추가

데이터 베이스에 자료를 추가하는 것은 두 가지로 생각할 수 있다. 현재 존재하는 데이터베이스 시스템 내의 테이블에서 어트리뷰트를 추가하거나 혹은 새로운 테이블을 추가하는 경우가 그 한가지이고, 두 번째는 새로운 데이터베이스를 추가하는 경우가 그 두 번째 경우가 될 것이다. 전자의 경우는 추가되는 어트리뷰트의 의미값을 온톨로지 상에서 대조하여 객체의 의미값이 이와 동일한 경우를 찾아서 메타 테이블내에 그 내용을 추가시켜주고, 동일한 의미값이 없는 경우에는 온톨로지상에 이를 추가시키고, 다시 메타 테이블에 이러한 내용을 추가시켜주어야 한다. 하지만 이러한 과정은 데이터베이스 관리자가 직접 수작업을 통해서 이루어져야 한다. 일반적으로 데이터베이스내의 스키마에 대해서 잘 알고 있는 데이터베이스 관리자가 각 스키마가 가지고 있는 의미값을 온톨로지상에서 찾아서 추가시켜주는 방법을 이 논문에서는 사용하였다.

후자의 경우에는 새로운 데이터베이스 시스템을 현재 통합된 데이터베이스 시스템에 추가하여 데이터 통합의 과정을 의미한다. 새롭게 추가될 데이터베이스 시스템에서는 자신의 시스템에서 가지고 있는 모든 자료에 대해서 온톨로지 상에 존재하는 데이터의 의미값을 기초로 하여 메타 테이블에 새로운 스키마를 추가하여 주어야 한다. 이 과정을 통해서 새롭게 추가되는 데이터베이스 시스템을 통합할 수 있다.

4.4 성능 평가

기존의 메타 데이터를 이용한 통합 시스템과 비교할 때, 제안한 방법은 자료를 수정하거나 추가하는데 있어서 많은 이점을 가지고 있다. 먼저 자료를 수정할 경우에는 수정된 스키마를 통합 라이브러리에서 변경만 시켜줌으로써, 간단히 스키마를 수정할 수 있다. 하지만 기존의 메타 데이터를 이용해 데이터를 통합하는 방법은 각 로컬 시스템에서 일된 데이터 스키마를 구축하고 있기 때문에 하나의 로컬 시스템에서 자료의 수정이 발생하면, 모든 연결된 시스템에서 이와 동일한 작업을 해주어야 한다. 또한 새로운 시스템을 추가할 경우에도 새롭게 추가되는 데이터베이스에서 기존의 표준을 따라 데이터 스키마를 변경해 주어야 한다. 그러나 제안된 시스템은 스키마를 수정하거나 새로운 시스템을 추가할 경우에 모든 작업이 통합 라이브러리 내에서 이루어지므로 신뢰성 있는 데이터의 의미적 통합을 유도할 수 있다.

5. 결론 및 향후 연구과제

본 논문에서는 기존에 개발된 관계형 데이터 시스템들간에 소스 데이터의 변형없이 데이터 스키마가 가지고 있는 의미있는 데이터인 메타 데이터를 이용해 각 시스템간의 통합을 유도할 수 있는 통합 구조를 제안하였다. 통합된 라이브러리를 구축함으로써, 관련된 시스템에서 자료의 수정이나 추가가 발생할 경우에 일괄적으로 통합 라이브러리만을 수정함으로써, 의미적 통합에 대한 신뢰성을 가질 수 있다. 제안된 시스템은 예약시스템뿐만 아니라 B2B 분야에 분산되어 있는 데이터들을 통합하는데 유용하게 적용될 수 있다. 또한 기존의 방법에 비해 통합 비용이 줄어들고, 시스템을 유지하는데 큰 장점이 있다. 그러나 동일한 객체에 대해서 동일한 타입 정보를 갖도록 소스 데이터를 변경하는 함수에 대한 적용범위의 확대와 관계형 데이터베이스 스키마로 제한되어 있는 현재의 시스템을 객체지향 데이터 베이스나 객체 관계 데이터베이스와 같이 다른 종류의 데이터베이스에도 적용할 수 있도록 하는 연구가 필요할 것이다. 또한 현재 웹 상의 수많은 소스 데이터를 통합하기 위해 통합 라이브러리를 XML로 구축하여 고급 검색기능과 대용량의 데이터를 처리할 수 있는 연구가 필요할 것이다.

참고문헌

[1] Edward Sciore, Michael Siegel, "Using Semantic values to Facilitate Interoperability Among Heterogeneous Information Systems" ACM Tran. on Database Systems, Vol. 19, No2, June 1994, p.254-290
 [2] Christof Bornhvd, "Semantic Metadata for the Integration of Web-based data for Electronic Commerce",