

XML Registry를 이용한 이질 데이터베이스 통합을 위한

프레임워크 설계

손강석^o 이상훈

국방대학교

kafa42@hanmir.com^o, hoony@kndu.ac.kr

Design A Framework for Integration of Heterogeneous Database Using XML

Registry

Kang-Seok Sohn^o Sang-Hoon Lee

Korea National Defense University

요 약

최근 XML은 다양한 운영체제와 데이터베이스, 여러 응용체제를 포함하는 이질적인 컴퓨팅 환경에서 통합을 위한 표준으로 각광받고 있다. XML은 특히 데이터베이스 시스템, 파일 시스템 등을 포함한 이질적인 형태의 데이터 처리 시스템 사이에서 데이터를 교환하기 위한 각종 표준을 제공하고 있다. 본 논문의 목적은 여러 분야에서 데이터공유를 위한 많은 노력 끝에 제시한 각종 표준 및 지침을 연구하여 통합적인 정보체계의 기반을 구축할 수 있는 통합 프레임워크를 제안하는 것이다. 통합 프레임워크는 다양한 정보 시스템에서 운영, 유지하고 있는 이질적인 데이터를 통합하기 위한 방법으로 XML Registry를 이용하여 메타데이터를 생성하고 이를 참조하여 통합정보체계에서 필요로 하는 공유데이터를 보관, 관리하는 공유 데이터 Repository를 생성한 후 각종 정보를 제공할 수 있는 구조이다.

1. 서 론

현재 운영되고 있는 대부분의 정보체계는 설계단계부터 상호운용성을 고려하지 않았기 때문에 서로 다른 인터페이스와 운영체제, 다양한 데이터 소스로 구성된 연통형 시스템이 되었다.

최근에는 하드웨어나 DBMS 등의 이질성은 대부분 표준화 되어 해결되고 있으나 동시성이나 의미 이질성은 여전히 실정이다. 이를 해결하기 위해 분산 환경에서 시스템의 성능개선이나 유지보수에 많은 비용을 투자하고 있고 서로 다른 운영환경에서 이질적인 데이터 모델들로 구축된 데이터 통합에 대한 많은 연구가 활발히 진행되고 있다[1, 2].

그러나 이러한 연구의 결과들이 특정 부분의 기술에 대해서만 이루어져 있고 통합정보체계 구축을 위한 전체적인 구조는 제시하지 못하고 있다.

따라서 본 논문에서는 지금까지 연구되어 온 EAI[3], 웹서비스[4], XML[5] 등 각종 표준과 기술들을 바탕으로 통합적인 정보체계의 핵심인 데이터 공유를 위한 공유데이터 Repository와 이를 구축하기 위한 참조 데이터의 집합인 XML을 이용한 메타데이터 레지스트리[6], 그리고 구축된 공유데이터를 활용하기 위한 Web Application Server와 전체적인 관리를 수행할 수 있는 통합관리서버 등으로 구성된 통합 프레임워크를 제시하고자 한다.

2. 관련 연구

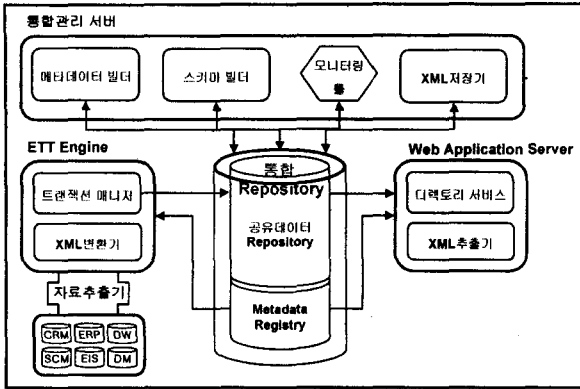
본 절에서는 통합정보체계 구축과 관련된 연구로 이질 데이터 통합을 위한 기반으로 각광받고 있는 EAI, 웹 서비스, XML과 데이터의 원활한 유통과 공유를 가능하게 하는 메타데이터 레지스트리에 대해서 살펴보고자 한다.

2.1 전사적 응용통합(EAI: Enterprise Application Integration)

EAI는 기업 내외부의 업무 프로세스를 통합하기 위한 통합 프레임워크로서 통합된 정보에 대해서 실시간 접근을 제공하기 위해 개발되었다[3]. EAI의 궁극적인 목적은 기존 애플리케이션 사이의 데이터 교환을 위해 표준화된 단일 허브를 제공하고 새로운 개발에 있어서 개방화된 기술을 제공하는 것이다. EAI는 기업들이 이미 보유하고 있는 레거시 업무와 ERP, SCM, CRM 등 시스템들을 통합하는데 아주 유용한 기반이다. 그러나 EAI는 표준기반의 기술을 사용하지 않기 때문에 각 프로젝트의 필요에 맞는 별도의 전용 EAI 솔루션을 도입해야 한다. 물론 업체에서 제공하는 이미 제작된 커넥터나 어댑터를 구입하는 것이 경제적인 수는 있으나 본 논문에서는 업체에서 제공하는 구조를 그대로 사용하기보다 EAI의 기술 중 데이터 추출 및 변환 등을 XML이라는 국제표준 기술을 통해서 구현하는 것을 제안한다.

2.2 웹 서비스(Web Service)

웹 서비스는 표준화된 XML 메시지를 통해 네트워크 액세스가 가능한 운영 기능들의 집합을 설명하는 인터페이스이다. 웹 서비스는 “서비스”들이 어떻게 구축되어야 하는지를 정의하는 것이 아니며, 새로이 작성되는 애플리케이션 또는 이미 사용되고 있는 기존 애플리케이션이 될 수 있다[4]. 웹 서비스는 표준의 세트로 만들어지는데 세부 구성요소로 UDDI(Universal Description, Discovery and Integration)는 요구되는 서비스를 찾는 데 사용되고, WSDL(Web Service Description Language)은 웹 서비스를 설명하는 데 사용되며, SOAP(Simple Object Access Protocol)은 원격으로 실행되는 웹 서비스를 위한 프로토



<그림 1> 통합 프레임워크의 구조

콜이다. 그러나 현 단계의 웹 서비스 기술은 인증, 트랜잭션 처리 등은 제공하지 않는다. 또한 웹서비스 기술을 이용하면 통합을 위한 별도의 전용 솔루션이 필요 없이는 반면 통합의 대상이 되는 시스템들이 웹 서비스 표준 프로토콜을 지원해야만 하는데 이를 위해서는 위에서 언급한 표준세트들을 모두 갖춰야 하는 번거로움이 있다. 따라서 본 논문에서는 웹 서비스를 정보를 공유할 수 있는 Public Interface의 개념으로만 사용하며 구축된 공유데이터의 정보를 UDDI 개념을 사용해서 서비스 할 수 있는 부분만 참조하고 통합관리 서버를 통해서 서비스를 제공하는 체계의 인증과 트랜잭션 매니저를 통한 트랜잭션 관리를 추가함으로써 웹서비스의 단점을 보완 하였다.

2.3 XML(Extensible Markup Language)

XML은 최근 웹 상에서 데이터 교환의 표준 형식으로 주목을 받고 있으며 사용자가 문서의 구조를 정의할 수 있기 때문에 다양한 형태의 데이터가 XML 문서로 표현 될 수 있다. 이러한 이유로 XML이 이기종 컴퓨터, 이질적인 데이터 소스 사이에서 데이터 교환의 좋은 매개체로 사용될 수 있다. 본 논문에서는 메타데이터의 관리와 각 응용체계에서 발생한 자료의 전송과 교환을 위해 XML을 이용한다.

2.4 메타데이터 레지스트리(MDR: Metadata Registry)

메타데이터 레지스트리에 대한 정의는 ISO 11179[6]를 참조하였다. 메타데이터 레지스트리는 메타데이터의 등록과 인증을 통하여 표준화된 메타데이터를 인증, 유지, 관리하며 데이터의 접근과 사용이 용이하도록 메타데이터의 명세와 의미의 공유를 목적으로 한다. 통합 프레임워크의 메타데이터 레지스트리에서 이 구조를 사용한다.

3. 통합 프레임워크의 구조

통합 프레임워크는 크게 네 부분으로 구성된다. 첫째, 통합 Repository는 메타 데이터와 공유 데이터를 수록하는 각종 데이터 보관 장소이고 둘째, 시스템 관리를 위한 각종 도구가 운영되는 통합관리 서버 셋째, 메타 데이터를 참조하여 각 조직에서 운용 중인 자료 중에서의 사결정에 필요한 자료를 추출 및 변형을 통해 전송을 하

는 기능을 가진 ETT Engine과 넷째, 웹 애플리케이션 서버는 공유 데이터를 요구하는 각 부서 및 응용체계에 XML 형태로 정보를 제공하는 서버이다. <그림 1>은 통합 프레임워크의 구조로서 세부기능은 다음과 같다.

3.1 통합 Repository

3.1.1 Metadata Registry(MDR) : 통합 Repository의 일부분으로 레거시 시스템에서 추출할 데이터의 원천, 포맷변환 형태, 데이터의 구조 및 변환대상, 변환규칙, 데이터의 정제 및 응용의 생성에 필요한 규칙 등에 관한 정보를 저장하고 제공한다. 또한 추출된 정보가 정확한 XML 형태로 변경되었는지 검증하기 위한 문서형태정의(DTD, Document Type Definition)[2]도 같이 제공한다. 이것은 데이터 공유를 위한 기본 틀로서 공유를 위한 기본요소로 데이터요소를 사용하고, 데이터 요소를 그 내포적인 의미를 충분히 나타내어 줄 수 있는 구조 내에 등록하여 각 응용에 활용도록 제공한다. 추가적으로 표준화된 XML 태그의 사용을 위해서 각 분야별 구분이 명확한 이름공간(Namespace) 개념을 도입하여 사용한다.

3.1.2 공유데이터 Repository : 통합 Repository의 일부분으로 Metadata Registry의 정보를 참조하여 레거시 자료에서 추출되어 전송된 XML 자료가 관계형 데이터베이스의 형태로 저장되는 곳이다. 이 구조는 다양한 응용 및 조직에 의해 작성되는 데이터를 의사결정을 위한 글로벌 뷰로 통합하고, 이와 같이 통합된 데이터를 필요에 따라 사용자에게 제공하는 기능을 갖는다.

3.2 통합관리 서버

3.2.1 메타데이터 빌더 : 이 도구는 메타데이터 관리를 책임진 메타데이터 관리자가 사용할 부분으로, 의사결정에 필요한 메타데이터에 수록할 자료를 식별하고 표준으로 등록할 자료들을 수록하는 기능을 수행한다. 먼저 각 조직에서 표준으로 사용하기를 원하는 정보를 제출하면 해당 정보가 기존 메타데이터에 있는지 확인하는 기능(매칭)과 표준 데이터에 없으면 데이터 변환규칙, 정제규칙, 생성규칙 등(매핑)을 MDR에 기록한 후 표준으로 등록하고 DTD를 자동 생성한다.

3.2.2 스키마 매니저 : 이 도구는 메타데이터 빌더에 의해 제안되어 표준이 될 후보가 된 데이터 및 표준으로 승인된 메타데이터의 물리적인 스키마를 생성해 주는 것으로 관련정보는 메타데이터 레지스트리에서 획득한다.

3.2.3 모니터링 툴 : 이 도구는 통합 프레임워크에 접근하여 공유데이터 Repository에 수록되어 있는 자료를 참조하려는 사용자뿐만 아니라 ETT Engine이 설치된 각 서버에 대한 인증을 해주는 것으로, 정확한 데이터 생성정로를 위해 필수적인 기능이다. 또한 시스템 자체의 이상을 진단하고 관련 응용 체계에 대한 운영을 감독하며 관리하는 역할도 수행한다. 이것을 이용해서 시스템관리자는 각 기능들의 상태를 실시간으로 확인할 수 있고 네트워크 상태 점검 및 백업 등의 주기적인 업무를 자동화할 수 있을 뿐 아니라 사전에 등록된 이상의 경우 예방조치도 할 수 있다.

3.2.4 XML 저장기 : 이 도구는 레거시 시스템에서 XML 형태로 전달된 정보를 공유데이터 Repository에 저장하는 역할을 수행한다. 문서 저장을 위해 전달된 XML 문서를 먼저 모니터링 틀에서 제공한 인증서를 사용해서 복호화한다. 변환방법에는 XML 데이터 형식을 기존의 관계형 데이터베이스 시스템이나 객체지향 데이터베이스 시스템의 데이터 형식으로 변환[7]하거나 전용 데이터베이스 시스템을 개발[8]하기 위한 방법이 있는데 통합 프레임워크에서는 XML 데이터를 관계형 데이터베이스 시스템의 데이터 형식으로 변환하여 저장한다. XML에서 필요한 자료를 추출할 때 DTD를 입력받아 유효성을 검증한 뒤 관계형 데이터베이스 스키마 추출모듈을 통해 SQL Script를 자동으로 생성하고, 생성된 SQL문을 이용하여 데이터베이스에 자료를 입력하도록 하는 것이다. 본 논문에서 사용하는 자료 변환방법은 XMLDBMS의 Java Package[9]의 소스코드를 수정하여 구현한다. [9]는 가장 일반적인 스키마 추출방법의 하나로 XML 데이터와 관계형 데이터베이스 시스템 사이에서 데이터 교환을 하는 미들웨어로서 Java Packages 형태로 배포 중이다.

3.3 ETT Engine

3.3.1 자료추출기 : 이 도구는 레거시 시스템에 파일 또는 관계형 데이터베이스 형태로 저장되어 있는 자료를 추출하는 도구이다. 관계형 데이터베이스의 경우 Stored Procedure 형태로 저장되어 특정 데이터베이스 자료구조에 자료변경에 대한 이벤트가 발생할 경우 변경자료를 실시간으로 추출하여 XML변환기에 통보하고, 파일 시스템의 경우 일일 일괄처리를 수행할 때 일일 작업 종료 후 파일내용을 XML변환기에 전달하는 역할을 수행한다.

3.3.2 XML 변환기 : XML은 문서의 구조를 사용자가 원하는 대로 정의하여 웹에서 사용되는 데이터가 동일한 형태로 저장, 처리될 수 있는 구조적 유동성을 제공하는데 바로 통합 프레임워크의 핵심 개념이라고 할 수 있다. 이 도구는 이러한 XML의 기능을 이용하여 레거시 시스템에서 추출된 정보를 Metadata Registry에 있는 변환규칙과 DTD 등을 이용하여 XML문서 형태로 변경시키는 기능을 제공한다.

3.3.3 트랜잭션 매니저 : 변환된 XML문서는 이 도구에 의하여 XML저장기로 전달된다. XML문서의 전송에 있어서 가장 중요한 사항 중 하나가 문서의 암호화인데 이 도구는 모니터링 틀에서 제공하는 인증서를 바탕으로 XML문서의 주요 요소를 암호화한다. 암호화된 XML문서는 Java 기반의 JMS 애플리케이션 아키텍처를 사용하여 전달된다[10].

3.4 Web Application Server

3.4.1 디렉토리 서비스 : 공유데이터 Repository에 수록된 내용은 이 도구를 통하여 사용자에게 서비스의 내용을 알려준다. 사용자가 특정정보를 선택하면 XML검색기에 해당 내용을 통보한다. 현재 설계된 내용은 사용자에게 고정적인 내용의 정보를 보여주는 형식이나 향후에는 사용자가 직접 질의입력을 통해 다양한 형태로 자료를 검색할 수 있는 질의처리 관리기도 추가할 예정이다.

3.4.2 XML 추출기 : 사용자가 디렉토리 서비스에서 조회를 원하는 특정 정보를 선택하면 이 도구는 요구사항을 토대로 공유데이터 Repository에 있는 내용을 검색한 후 검색결과를 XML 형태로 추출한다. 이 때 MDR을 검색하여 DTD를 참조한 후 구조에 맞도록 XML 문서를 생성하여 사용자의 웹 브라우저에 전달한다.

4. 결론 및 향후 연구과제

컴퓨터 사용의 보편화와 네트워크의 성능 향상에 따라 하드웨어뿐만 아니라 소프트웨어와 자료의 이용을 서로 독립적인 형태가 아니라 통합적으로 운영하는 요구가 지속적으로 증가하고 있다. 급변하는 정보환경 하에서 상호운용성에 바탕을 둔 통합데이터의 활용은 신속한 의사결정을 위해 반드시 필요하다. 본 연구에서는 이러한 사용자의 요구를 실제적으로 구축할 수 있는 통합프레임워크와 세부 기능들을 소개하였는데 이는 이질적인 시스템과 데이터베이스들이 가지고 있는 기존의 기능과 특성을 최대한 활용하면서 동시에 통합의 장점도 발휘할 수 있다.

향후 연구과제로는 이러한 통합 프레임워크에 관련된 관련 표준과 기술들을 좀 더 심층적으로 연구하여 서로 배타적이 아닌 상호 보완적인 관계 속에서 이질적인 자료를 쉽게 통합할 수 있는 체계를 실제로 구현하는데 있다.

참고문헌

- [1] A. P. Sheth and J. A. Larson, "Federated Database Systems for managing Distributed Heterogeneous Database," ACM Computing Surveys, Vol.22, No. 3, 1990.
- [2] 이정욱, "멀티데이터베이스 시스템의 정보공유를 위한 개념 기반 의미망", 고려대학교 대학원 컴퓨터학과 박사학위논문, 2001
- [3] W.A. Ruh et al, Enterprise Application Integration, John Wiley & Sons, New York, 2001.
- [4] M.Kirtland, Web Services Essentials, Microsoft Developer Network, Redmond, Wash., 2001.
- [5] T. Bray, J. Paoli, and C. M. Sperberg-McQueen, "Extensible Markup Language (XML) 1.0(Second Edition)", <http://www.w3c.org/TR/REC-xml>, W3C Recommendation 6, October, 2000.
- [6] ISO, Metadata Registry, ISO-11179, 2000.
- [7] D. Florescu and D. Kossmann, "Storing and Querying XML Data Using an RDBMS," Proc. of Int. Conf. on Data Eng., 1999
- [8] C. Kanne and G. Moerkotte, "Efficient Storage of XML Data," Proc. of Int. Conf. on Data Eng., 1998.
- [9] <http://www.rpbouret.com/xmldbms/index.htm>
- [10] <http://java.sun.com/products/jms/nonlicensed-vendors.html>