

# 이기종 데이터베이스 환경의 정보 통합을 위한 I2System(:Information Integration System) 설계

권도훈\*, 박성공, 이정옥, 백두권  
고려대학교 컴퓨터학과 소프트웨어 시스템 연구실  
e-mail:{dhkwon,skpark,ljo,baik}@software.korea.ac.kr

## I2System(:Information Integration System) Design for an Information Integration of Heterogeneous Databases

Do-Hoon Kwon\*, Sung-Kong Park, Jeong-Oog Lee, Doo-Kwon Baik  
Software System Lab, Dept. of Computer Science & Engineering, Korea University  
요 약

인터넷과 통신기술의 발전으로 온라인 정보소스와 정보사용자의 수가 증가되었다. 이것은 정보시스템간 특히 데이터베이스간의 통합을 요구한다. 데이터베이스간 통합에 대한 연구는 오래 전부터 시작되었고, 많은 결과들이 발표되고 있다. 기존의 데이터베이스간 통합 방법은 일반적으로 전역데이터모델(Global Data Model)을 제공해야 하며, 전역스키마(Global Schema) 작성, 전역스키마와 지역스키마간 번역과 같은 복잡한 작업들을 필요로 한다. 이는 최종 사용자에게 다양한 질의(정보)와 이질적인 지역 데이터베이스들에 대한 단일접근방법을 제공하기 위함이다. 하지만 데이터베이스간 통합이 요구되는 영역에서 필요한 정보만 제공해 준다면 다양한 질의를 제공하기 위한 기존 정보통합시스템들의 요구사항은 오버헤드로 작용한다. 본 논문은 데이터베이스간 통합 시 필요한 정보를 기반으로 데이터베이스간 통합을 하는 I2System(Information Integration System)을 제안한다. 이는 기존의 정보통합방법의 오버헤드를 줄이며, 단순하고 확장성 있는 데이터베이스간 통합을 제공한다.

### 1. 서 론

인터넷과 통신기술의 확산으로 온라인 정보소스와 정보사용자의 수가 증가되었다. 이것은 정보시스템간 특히 데이터베이스간의 통합을 요구하게 되었다. 지금까지 데이터베이스간 통합을 위한 많은 방법들이 제안되었다. 이 방법들은 유사한 유형의 모델을 바탕으로 하고 있다. 기존의 데이터베이스간 통합을 위한 정보통합방법은 데이터모델과 스키마의 이질성을 극복하기 위해 GDM(Global Data Model)과 전역스키마를 제공해야 하며, 사용자의 정보서비스요청(질의)에 대한 부질의(subquery) 집합으로 변환, 그리고 전역스키마와 지역스키마간 번역 등 복잡한 구조를 기반으로 하고 있다. 이는 최종 사용자에게 다양한 질의(정보)와 이질적인 지역 데이터베이스들에 대한 단일의 접근방법을 제공하기 위함이다.

정보통합이 필요한 문제영역에서 필요한 질의(정보)의 수는 구축된 정보통합시스템에서 제공하는 사용자인터페이스의 기능수만큼 유한하다. 결국, 문제영역에서 필요로 하는 정보만 제공할 수 있다면 다양한 질의를 제공하기 위한 기존 정보통합시스템의 요구사항들은 오버헤드가 된다. 본 논문에서는 데이터베이스간 통합이 요구되는 영역에서 필요로 하는 정보를 이용하는 정보통합시스템(I2System :Information Integration System)을 제안한다. I2System은 정보관련 API(Application Program Interface)를 제공한다. 기존의 API가 클라이언트에게 필요한 기능을 제공한 것과 달리, I2System에서의 API는 클라이언트에게 필요한 미리 식별된 정보를 함수의 형태로서 제공한다. 클라이언트 프로그램은 이 정보관련 API 내의 함수를 호출함으로써 통합된 정보를 얻을 수가 있다. I2System은 클라이언트 프로그램에서 사용하게되는 정보관련 API를 제공하는 APIFI2, 각 지역 데이터베이스로의 접근을 위한 IC, APIFI2로 IC 접근에 필요한 정보를 제공하는 IRDelegator를 컴포넌트로서 제공하고, 실제 IC에 대한 정보를 가지고 있는 KBFIC와 각 지역 데이터베이스로부터의 지역 결과를 임시로 보관하는 TR을 데이터베이스 형태로 제공하며, TR에 저장된 정보를 바탕으로 통합된 정보를 작성하는 RDelegator를 컴포넌트로서 제공한다. I2System의 구성 요소 중 APIFI2, IC를 제외한 다른 구성요소

들은 임의의 문제영역에서 재사용이 가능하다.

I2System은 데이터베이스간 통합에 필요한 정보를 기반으로 GDM, 전역 스키마 작성, 부질의 작성, 지역스키마와 전역스키마간 번역 등을 배제하고, 문제 영역에서 필요로 하는 정보들 기반으로도 보다 간단하고 확장이 용이한 데이터베이스간 통합을 제공한다.

### 2. 관련 연구

#### 2.1 다양한 정보 시스템 통합 방법

지금까지 여러 연구에서 분산되며, 이질적인 데이터베이스간 통합방법이 연구되어 오고있다. 이들은 일반적으로 GDM과 부질의 작성, 지역스키마와 전역스키마간 번역 등의 복잡한 기능을 요구한다. 다양한 정보 시스템 통합 방법을 5가지 유형으로 살펴본다.

첫째, 고전적 접근 방식이 있다. 이 방식은 다양한 지역 소스(local source) 스키마의 차이를 해결하기 위한 하나의 전역 스키마 작성이 핵심이 된다. 즉, 이질적인 지역소스의 스키마를 바탕으로 전역 스키마를 작성하고, 이를 통해 시스템 사용자는 다양한 지역 소스에 대해 일정한 형태(전역 스키마)로의 접근이 가능하며, 다양한 지역 소스가 통합되는 효과를 준다. 하지만 지역소스가 새로이 추가되거나 수정되면 지역 소스를 바탕으로 한 전역 스키마의 갱신이 요구되는 단점이 있다[1,2]. 둘째, 고전적 접근방식이 유발했던 데이터베이스 사용자 이질성의 문제를 해결한 연합된 접근 방법(federated approach)이 있다. 하지만 여전히 전역 스키마의 방식을 사용하고 있고, 전역 스키마 또한 여전히 정적이기 때문에 지역 소스의 추가나 수정 시 전역 스키마의 갱신이 요구되어진다[1,3]. 셋째, 분산 객체 관리방식(distribute object management approach)이 있다. 이는 분산되고 이질적인 데이터베이스를 분산 객체 공간의 객체 집합 모델 기반으로, 연합된 접근 방법을 일반화 시켰다. 이것은 공통 객체 모델과 공통 객체 질의 언어를 바탕으로 하고 있다[1,4]. 넷째, 지능적 정보 통합 방식(intelligent information integration(I3))이 있다. 이것은 3계층 구조를 바탕으로 한 중계자(mediator) 아키텍처를 사용한다 [1,5]. 중계자는 일반적으로 지식을 기반으로 데이터베이스 스키마 정보를 가지고 있으며,

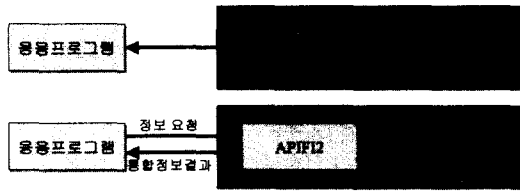
룰(rule)을 통하여 서로 다른 데이터베이스의 내용을 통합할 수 있다[6]. 다섯째, 온톨로지(ontology) 기반의 정보 통합 방식이 있다. 온톨로지는 일반적으로 특정 도메인의 개념과 그들의 관계를 정의해 놓은 지식 베이스이다. 이들은 실제 소스들과 독립적으로 개발되기 때문에 지역 소스의 추가, 삭제, 수정 등에 유연하다[7]. 일반적으로 온톨로지 기반의 정보통합방법은 자율성(autonomy), 지능성(intelligence), 이동성(mobility), 사교성(social ability) 등의 특성을 가진 에이전트와 함께 사용된다[8].

**3. I2System을 이용한 데이터베이스간 통합**

웹기반 응용프로그램의 최종 사용자는 사용자 인터페이스를 통하여 작업을 하게 된다. 이런 시스템들의 질의 유형은 보통 사용자 인터페이스에서 제공하는 서비스의 수와 거의 유사하게 제한적이다. 일반적인 도메인의 경우도 유사하다. 동일 도메인에서 제작된 응용프로그램들은 보통 비슷한 일을 처리하고 있다. 결국, 유사한 사용자 인터페이스를 가진다. 이것은 응용프로그램 수준에서 유사한 정보를 필요로 한다는 것을 의미한다. 이런 특성을 바탕으로 우리는 임의의 도메인에서 한정된 질의만을 필요로 한다는 것을 알 수 있고, 그것을 일반화시킬 수 있다. 본 논문에서는 제안하는 I2System은 이러한 일반적이고 제한된 질의들을 바탕으로 구현된다.

**3.1 I2System의 개요**

I2System(Information Integration System)은 데이터베이스간 통합이 이루어지는 영역에서 필요한 정보를 API형태로 제공하는 표준인터페이스[9]를 구현한 IC 컴포넌트를 기반으로 작성된 정보통합시스템을 의미한다. I2System에서는 OS나 응용프로그램에서 API를 통해 자신이 서비스 할 수 있는 기능을 제공하는 것과 마찬가지로, 응용프로그램에서 필요로 하는 정보를 APIF12컴포넌트를 통해 제공하게 된다<그림1>.



<그림1> I2System의 개요

**3.2 I2System의 아키텍처**

기존의 단일 데이터베이스 시스템은 최근 분산 환경(클라이언트 서버환경)의 보급과 함께 응용프로그램(비즈니스 로직과 사용자 인터페이스로 구성), 미들웨어, 소스 층으로 구성된 3계층 아키텍처로 구성되어있다. I2System을 이용하는 데이터베이스간 통합방법은 이 3계층 아키텍처를 응용했다.

<그림2>와 같이 I2System을 이용한 정보통합은 클라이언트 층, I2System 층, 지역소스 층으로 구성되어있다.

**3.2.1 클라이언트 층**

**• Client**

기존의 3-Tier의 클라이언트와 동일한 개념이다. 사용자로부터 서비스의 요청을 받는 사용자 인터페이스로서 동작한다. 또한 APIF12로부터의 최종 결과를 사용자에게 비주얼하게 제공한다. 그리고 사용자의 요청에 따라 APIF12로 정보요청을 전달한다.

**3.2.2 I2System 층**

APIF12, IRDelegator, RDelegator, 웹서버, TR, KBFIC, IC컴포넌트로 구성되어 있다. IRDelegator, RDelegator, TR, KBFIC

는 I2Framework을 구성한다. I2Framework는 데이터베이스간 통합에 있어서 요구되는 정보에 상관없이 재사용 된다.

**• APIF12(API For Information Integration)**

I2System에서 제공하는 정보에 관한 API로서 클라이언트는 이 컴포넌트에서 제공하는 API를 이용해서 원하는 통합정보 서비스를 호출하게 된다. 이것은 표준인터페이스[9]를 기반으로 한다.

**• IRDelegator(Information Request Delegator)**

APIF12로부터의 각각의 IC 컴포넌트에 대한 정보 요청을 받고, KBFIC의 정보를 바탕으로 각각의 IC 컴포넌트의 정보를 APIF12로 전달한다.

**• RDelegator(Results Delegator)**

TR로부터의 원시정보를 취합해서 최종결과를 작성하고 그 결과를 APIF12 컴포넌트로 전달한다.

**• KBFIC(Knowledge Base For Information Component)**

통합된 로컬소스와 연결된 각 IC 컴포넌트들에 관련된 정보를 저장한다. KBFIC는 데이터베이스로 구현된다. 코바 기반 구현 시 ORB 참조 변수, POA 이름, ServantID-Visibroker 4.0 기반-로 구성된다.

**• TR(Temporary Results)**

각 IC 컴포넌트는 해당 로컬소스로부터의 결과를 TR에 임시적으로 저장하게 된다. 즉, 각 IC 컴포넌트를 통해 모든 통합된 로컬 소스로부터의 정보가 TR에 일시적으로 저장되게 된다. 이 저장된 정보는 RDelegator 컴포넌트가 APIF12 컴포넌트에 전달하게되는 정보 요청의 결과를 작성하는 기본 데이터로 쓰인다. TR은 데이터베이스로 구현된다. TR의 내용은 키워드 문자열, 결과, 결과의 타입으로 구성된다.

**• IC(Information Component)**

미리 식별된 정보를 기반으로 표준인터페이스[9]를 구현한다. 이것은 APIF12 컴포넌트로부터의 정보 요청에 따라, 해당 데이터베이스를 기반해서 정보 서비스를 수행하고 결과를 TR로 저장한다.

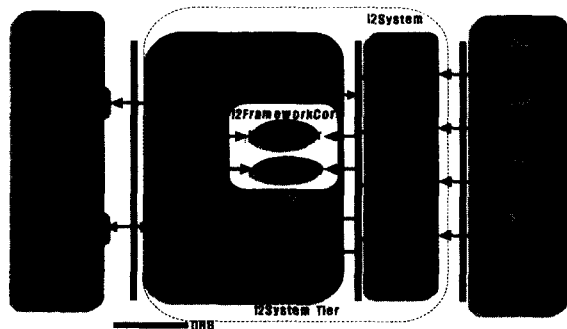
**• 웹 서버**

I2System의 APIF12 컴포넌트에 접근하기 위한 방법으로 클라이언트는 인터넷을 이용할 수도 있다. 이때, 웹서버를 통해서 서비스를 제공한다.

**3.2.3 지역소스 층**

**• 지역소스**

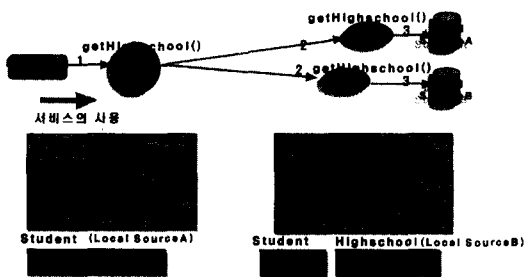
통합되어 지는 각 지역 데이터베이스를 의미한다.



<그림 2> I2System 아키텍처

3.3 I2System 기본 아이디어

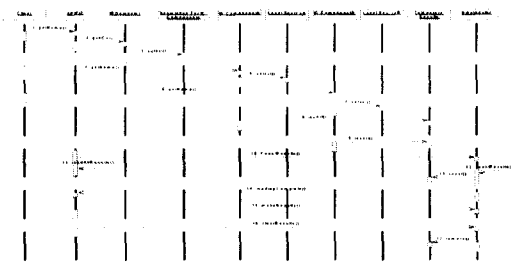
I2System을 이용한 데이터베이스간 통합 동작원리는 미들웨어를 사용하는 3계층 아키텍처를 응용한다. <그림3>에서 독립된 두 지역 소스에는 스키마의 이질성이 발견된다. 즉, 학생정보에 대해 Local SourceA는 학생이름과 출신고교를 Student라는 테이블 하나로서, Local SourceB는 Student와 Highschool 테이블 각각에 저장하고 있다. 이때, 두 로컬소스를 기반으로 학생마다의 출신고교가 필요하다고 하자. 이때, I2System은 각 학생의 출신고교에 대한 함수 String getHighschool(String StudentName)를 미리 식별, 이를 선언하고 이를 APIF12, IC 컴포넌트 각각 구현하게 된다. 이때, 각 IC 컴포넌트(IC\_a, IC\_b)는 <그림3>에서와 같이 해당 지역 소스 스키마와 데이터 모델을 바탕으로 선언된 함수 getHighschool()을 구현한다. 이것은 각 지역 소스의 이질성을 해결하게된다. 이제 필요한 정보에대해서 각 IC 컴포넌트의 구현된 해당 함수만 호출함으로써 통합된 지역 소스로부터 필요한 정보를 얻을 수가 있다.



<그림3> I2System 동작 원리

3.4 I2System의 프로세스 흐름

I2System의 전체 프로세스의 흐름은 <그림4>와 같이 정의된다. 클라이언트는 원하는 정보에 대응하는 함수를 APIF12에서 제공하는 API로부터 호출한다. APIF12는 요청된 정보에 대한 처리를 위해 IRDelegator로부터 호출될 각 IC 컴포넌트의 정보를 요청한다. IRDelegator는 KBFIC로부터 각 IC 컴포넌트의 정보를 얻어서 APIF12에 전달하게 된다. APIF12는 전달받은 각 IC 컴포넌트에 대한 정보를 바탕으로 클라이언트로부터의 정보 요청을 각 IC 컴포넌트에 전달한다. 각 IC 컴포넌트는 해당 지역 데이터베이스에 접근해 요청된 정보를 가지고 온다. 각 IC 컴포넌트가 얻은 정보는 TR에 임시로 저장되게 된다.



<그림4> I2System의 통합정보 순차 다이어그램

이때, APIF12는 RDelegator에 각 IC 컴포넌트의 작업종료에 대한 식별요청을 보낸다. RDelegator는 TR의 데이터를 기반으로 APIF12에 각 IC 컴포넌트의 작업종료 여부를 전달한다. APIF12는 RDelegator에 TR내의 정보결과의 통합을 요청한다. RDelegator는 최종결과를 작성한 후, 결과를 APIF12에 전달하

고 TR의 내용을 지운다. APIF12는 통합정보결과를 클라이언트에게 전달한다.

3.5 데이터베이스 통합 모델의 구축 시나리오

I2System을 통해 데이터베이스간 통합 시 다음과 같은 절차를 통한다.

- A. 먼저 적용될 도메인의 분석하여 요구되는 정보를 식별한다.
- B. 식별된 정보를 기반으로 각각에 대한 함수를 선언하고 각각을 API로써 묶는다.
- C. 작성된 API를 기반으로 APIF12와 각각 로컬소스에 대응하는 IC 컴포넌트를 구현한다.
- D. APIF12와 각 IC 컴포넌트를 I2Framework에 통합해서 I2System을 구현한다.
- E. I2System의 APIF12에서 제공하는 함수를 바탕으로 클라이언트 프로그램을 작성한다.

4. 기존 데이터베이스통합모델과의 비교

본 논문에서 제안한 모델의 장단점은 다음과 같다. 첫 번째 장점으로서는 시스템의 구조가 간단하다. I2System은 전역데이터 모델이 없고, 전역 스키마의 구축이 필요 없으며, 해당 질의물 지역 소스로 번역하는 래퍼 대신 데이터베이스 접근 관련 함수들의 집합인 IC 컴포넌트를 사용한다. 이는 시스템을 훨씬 단순화 시켰으며 구축하기도 많이 쉬워진다는 장점을 가지고 있다. 둘째, 데이터베이스의 확장성이 높다. 미리 식별된 API를 각 지역 데이터베이스에서 구현만 한다면 어떤 데이터 모델과 스키마 구조에 상관없이 확장이 가능하다.

단점은 다음과 같다. 우선 통합되어 질 지역 데이터베이스는 항상 제공한 IC 컴포넌트를 구현해야 한다. 즉, 통합 시 지역 데이터베이스 쪽의 오버헤드가 크다. 둘째, 질의어를 제공하지 않기 때문에 제공되는 정보가 한정된다. 보통 제공하는 함수의 수를 늘리면 정보 통합 모델을 좀더 일반화 할 수 있지만, 지역 소스의 부담이 더해진다.

5. 결론 및 향후 연구 과제

본 논문에서는 데이터베이스간 통합이 필요한 도메인에서 보다 쉽게 데이터베이스를 통합하는 I2System을 제안했다. 이것은 OS나 응용프로그램이 제공하는 API와 유사하게 특정 도메인에서 필요한 정보에 대한 API를 작성하고 그것을 바탕으로 스키마와 의미 레벨의 데이터베이스 통합을 한다. 이는 단일 도메인에서 기존의 데이터베이스통합 모델에 비해 간단하고 확장성 있게 데이터베이스 통합을 수행한다.

추후에는 데이터베이스간 통합을 수행하는 도메인에서 요구되는 정보가 변경 시 처리에 대한 연구가 필요하다.

< 참고 문헌 >

- [1] Maurizio Panti, Luca Spalazzi, and Alberto Giretti, "A Case-Based Approach to Information Integration", Proceedings of the 26th VLDM Conference, Cairo, Egypt, 2000.
- [2] S.Ram, "Special issue on heterogeneous distributed database systems", IEEE Computer Magazine, 24(12), December 1991.
- [3] A. Sheth and J. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases", ACM Transaction on Database Systems, 22(3), 1990.
- [4] T.Ozsu, U.Dayal, and P.Valduriez, "Distributed Object Management", Morgan Kaufmann, San Mateo, CA, 1993.
- [5] G. Wiederhold, "Mediators in the Architecture of Future Information Systems", IEEE Computer Magazine, 25:38-49, March 1992.
- [6] 이강찬, 김덕현, 이규철, "차세대 데이터베이스 통합모델에 관한 연구", 98춘계 학술발표논문집, 정보과학회, 1998.
- [7] Jerry Fowler, Brad Perry, Marian Nodine, and Bruce Bargmeyer, "Agent-Based Semantic Interoperability in InfoSleuth", SIGMOD Record 28(1): 60-67, 1999.
- [8] K. Decker and K.P. Sycara, "Intelligent adaptive information agents", Journal of Intelligent Information Systems, 9(3):239-260, 1997.
- [9] 권도훈, 박성공, 이정옥, 백두권, "표준 인터페이스를 이용한 데이터베이스 통합", 2001 춘계 학술발표논문집, 정보과학회, 2001