



샤모아: 컴포넌트 기반의 지식공학 프레임워크

이화여자대학교 김 원 · 김 명** · 박승수 · 박현석 · 반효경 · 옹환승** · 이미정*
이민수 · 이상호** · 정안모 · 조동섭** · 채기준** · 최병주**

1. 서론

컴퓨터 분야에서 지식공학으로 불리는 기술은 산업계에서 지식경영 응용을 위한 기반 기술로 더 잘 알려져 있다. 지식경영을 위한 응용들로는 고객관리 시스템(CRM), 전자 상거래(e-commerce), 정보 개인화(information personalization) 등이 있다. 넓은 의미에서 지식공학은 부가가치를 지닌 데이터의 저장(value-added data storage), 네트워크를 통한 데이터 전송, 그리고 데이터로부터의 지식 추출 및 정보 시각화 등을 포함하기도 한다.

본 연구팀에서는 두 가지 목표를 가지고 샤모아 프로젝트를 시작했다. 우선 지식공학과 관련된 핵심 기술을 연구하고 또 하나는 그러한 결과를 한데 묶을 수 있는 지식공학 프레임워크의 프로토타입을 만드는 것이다. 현재 구현된 샤모아 프로토타입은 상용 소프트웨어 제품들과 프로젝트에서 개발 중인 소프트웨어 모듈들을 묶어주는 기능을 제공하고 있다. 샤모아란 유럽과 남부 러시아 지방에 서식하는 영양류의 일종으로서 짧은 걸음만으로도 높이 도약하는 능력을 가진 것으로 유명하다. 본 연구팀의 바람은 이 프로젝트를 통하여 참여 교수들과 학생들이 짧은 기간 안에 한 단계 높은 수준으로 기술적인 도약을 하는

것이다.

샤모아는 두 개의 주요 부분으로 구성되어 있다. 인프라 구조 부분은 지식공학의 핵심 기술을 제공하는 상용 소프트웨어와 프로토타입 모듈들로 구성된다. 응용 부분은 인프라 구조에서 제공하는 데이터와 지식을 활용한다.

본 연구에서는 지식공학 기술을 이용하는 다양한 컴포넌트들로 구성된 전자 상거래 시스템을 실험 및 검증에 위한 주요한 응용으로 설정하였다. 컴포넌트들이 제공하는 지식공학 기술들은 개인화(personalization), 보안, 멀티미디어 데이터의 고품질(QoS) 데이터 전송, 질의어 처리, XML 문서관리 등을 포함한다.

본 논문의 구성은 다음과 같다. 2절에서 샤모아의 구조를 설명하고 3절에서는 샤모아의 관리 기능을 다룬다. 4절에서는 샤모아의 테스트 기능에 대해 살펴봄으로써 5절에서는 샤모아의 컴포넌트들을 설명하고 6절에서 결론 및 추후 연구계획을 제시한다.

2. 샤모아의 구조

샤모아 프레임워크의 설계는 본 연구팀에서 개발한 소프트웨어 모듈들과 상용 소프트웨어들을 통합한 후 이들을 다시 프레임워크 컴포넌트들로 만드는 것이 해당된다. 그리하여 샤모아 프레임워크는 회사 소유의 상용 소프트웨어 패키지와 본 연구팀에서 개발한 소프트웨어 모듈들에 대한 공통의 응용 프로그래밍 인터페이스를 제공한다.

다행히도 기업 내외부의 업무 프로세스를 통합하기 위한 전사적 응용 통합 프레임워크(EAI: Enterprise Application Integration)는 이미 개발되어 있다 [1]. EAI는 통합된 정보에 대해서 실시간 접근을 제공하기 위하여 개발되었다. EAI는 인터넷처럼 새로

감사의 글 : 다음의 졸업생 및 학생들은 샤모아의 연구 개발에 기여하였음. 조태남, 최지영, 김호숙, 김현희, 광미라, 이정원, 이월영, 임지영, 박소영, 용승림, 윤혜성, 장경자, 김지심, 김수경, 김성하, 김예경, 임현숙, 임윤선, 오승희, 송지숙, 송영원, 서승현, 양자영, 이미자, 강나영, 강지선, 김은주, 배은주, 송은하, 옥지혜, 이병희, 최미라, 윤혜정, 박주현, 함화진, 김은희, 김명미, 최윤정, 이미연, 윤성희, 김애영, 정옥란, 현영순, 김은아, 박경희, 김마리

* 정회원, ** 중신회원

운 정보가 계속해서 발생하거나 혹은 무선 응용 프로토콜처럼 계속해서 새로운 프로토콜이 만들어지는 IT 시스템들이 더욱 효율적으로 구성될 수 있는 기능을 제공한다. 이 프레임워크는 상용 소프트웨어를 통합할 수 있도록 설계되어서 여러 회사들에 의해서 개발된 ERP, CRM 시스템 및 데이터웨어하우스, 새로운 응용 소프트웨어 및 데이터들을 하나로 통합할 수 있다.

그림 1은 샤모아의 구조를 보여주는데, 이는 EAI와 유사한 세 계층의 확장 가능한 구조로 구성되어 있으며, 다음과 같은 요소들을 포함하고 있다.

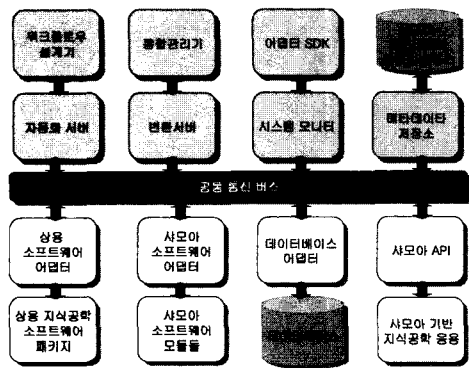


그림 1 샤모아 프레임워크의 구조

공통 통신 버스 : 이 버스는 여러 상용 소프트웨어와 샤모아 연구팀이 개발한 소프트웨어 모듈들을 연결한다. 또한 샤모아를 이용하여 개발한 여러 지식공학 응용들도 이 버스를 통해서 연결된다. 샤모아 구조는 인터넷 환경을 위해서 마이크로소프트사의 COM(Component Object Model)을 사용하며, 인터넷 환경을 위해서는 웹 서비스(Web services)[2]를 사용한다.

이를 위해서 COM+를 선택하였다. 왜냐하면, COM+가 현재 많이 쓰이고 있고 또 그 위에서 작동하는 지식공학 API가 안정적이고 효율적이라는 사실이 증명되어 있기 때문이다. 이 API는 OLE DB, OLE DB for OLAP, 그리고 OLE DB for data mining을 포함한다. 또한, 웹 서비스를 선택하였는데, 그 이유는 Microsoft, IBM, Ariba, Bea, SAP, IONA, SUN 및 TIBCO를 비롯한 130개가 넘는 IT 회사들이 설계와 검토에 참여하여 다음 세대의 EAI 연결 버스로 채택될 가능성이 높기 때문이다. 웹 서비스에 관련된 기

술은 UDDI(Universal Description, Discovery and Integration), WSDL(Web Service Description Language), SOAP(Simple Object Access Protocol)을 포함한다[3,4].

상용 소프트웨어와 어댑터 : 샤모아 프레임워크는 상용의 비즈니스 인텔리전스 인프라를 구성하는 소프트웨어(예를 들어, Informatica의 PowerMart ETL 도구, Accrue/Pilot 소프트웨어의 MOLAP 서버, Microstrategy의 ROLAP 서버, IBM DB2, Microsoft SQL 서버, Oracle 관계형 데이터베이스 시스템, IBM의 Intelligent Miner 텍스트 마이닝 툴킷, CRM 응용을 위한 MaxScan 데이터 서버 및 DB miner 데이터 마이닝 툴)에 의해 제공되는 여러 API들에 쉽게 접근할 수 있는 공통적인 방법이 필요하다. 상용 소프트웨어 어댑터들을 사용하여 상용 소프트웨어 API를 샤모아의 공통 API로 동적으로 바꾸어준다.

샤모아 소프트웨어 모듈과 어댑터 : 개발한 소프트웨어 모듈들은 개별적인 API들을 갖고 있는데 이들 모듈들은 Java, Corba, Call Level Interface 및 COM+에 기반한 API를 갖고 있다. 샤모아 소프트웨어 어댑터는 이러한 API들을 동적으로 샤모아 공통 API로 변환해 주는 실시간 라이브러리들이다.

워크플로우 설계기 및 자동화 서버 : 워크플로우 설계기를 사용하면 응용 프로그램 및 샤모아 프레임워크의 컴포넌트들 사이의 자동화된 프로세스를 설계할 수 있다. 자동화 서버는 샤모아 프레임워크의 모든 요소들 사이의 메시지 전송과 사건의 호출 및 알림을 제어하는 워크플로우 프로세스를 실행한다.

통합 관리기와 변형 서버 : 샤모아 프레임워크는 응용 프로그램 개발자에게 두 가지 지식공학 서비스를 제공한다. 첫째는 워크플로우 설계기와 주요 샤모아 컴포넌트들로 구성되며 다음의 기능들을 제공한다.

- 데이터를 데이터웨어하우스로 추출, 변형, 적재하여 지식베이스를 구성하는 기능
- 관계형 데이터베이스에서의 XML 문서를 저장하는 기능
- 빠른 질의어 처리를 위한 OLAP 요약 테이블을 생성하는 기능
- 데이터 마이닝 모델을 훈련시키는 기능

둘째는 샤모아 기반의 응용 소프트웨어로부터 지

식베이스에 대한 질의를 가능하게 하고, 그 결과를 제시하고 가시화한다.

통합 관리기는 응용 소프트웨어 개발자들이 고차원적이고 추상적으로 업무 지능 프로세스를 정의할 수 있도록 여러 가지 사모아 컴포넌트 기능들을 조합해준다. 개발자들은 먼저 통합 관리기를 이용하여 지식베이스를 구성한 후에 주기적으로 수정을 한다. 이러한 방법을 사용하면 응용 소프트웨어 개발자들은 사모아 컴포넌트의 구조나 데이터를 모르는 상태에서 질의어 처리 서비스를 사용할 수 있게 된다.

통합 관리기는 응용 소프트웨어 개발자들에게 편리한 그래픽 사용자 인터페이스를 제공하여 사모아 프레임워크 서비스를 단계별로 응용 소프트웨어와 통합할 수 있도록 해준다. 통합 관리기는 각 단계들 사이의 호출 매개 변수와 정보들을 메타 데이터 저장소에 등록한다. 어댑터는 소프트웨어 모듈 사이의 인터페이스 조정을 담당하나 데이터 무결성을 보장하기 위해서 변형 서버가 관여하여 매개 변수들을 동적으로 변화시킨다.

메타 데이터 저장소 : 사모아 메타 데이터 저장소는 응용 소프트웨어 개발자에게 데이터의 원천, 목표 데이터베이스 혹은 데이터 매트, 변환 규칙과 매핑, 데이터 정제 및 응용의 개발에 필요한 규칙 등에 대한 정보를 제공한다. 또한 저장소는 사모아 프레임워크의 모든 요소들을 병합하기 위한 메타 데이터도 관리한다.

데이터베이스 어댑터 : 이 어댑터는 API를 충분히 제공하지 못하는 상용 소프트웨어를 사모아 프레임워크에 통합하기 위해서 데이터베이스를 사용한다. 다른 서버나 응용 소프트웨어들이 데이터베이스 어댑터를 사용하여 사모아 공통 API를 통해서 특정 상용 소프트웨어들에 접근할 수 있다.

어댑터 소프트웨어 개발 키트 : 설계자들은 이 SDK를 사용해서 다양한 어댑터를 개발할 수 있다. 이것은 통합된 개발 환경, 사모아 공통 API 명세서 및 샘플 코드로 구성되어 있다. 메타 데이터 저장소는 새로 개발된 어댑터에 관한 정보를 자동적으로 등록하여 나중에 통합 관리기에 의해 사용될 수 있도록 한다.

사모아 API와 응용 소프트웨어 SDK : 사모아 API는 두 개의 인터페이스로 구성되어 있다. 하나는 지식공학 응용 소프트웨어 개발자로 하여금 사모아 프레임워크에서 제공하는 서비스를 접근할 수 있도록

해준다. 다른 하나는 사모아 자동화 서버가 응용 소프트웨어에게 정보를 알려주는 것이다. 각 응용 소프트웨어 SDK는 API 브라우저, 헤더 파일, 라이브러리 및 응용 소프트웨어 생성을 위한 통합 개발 환경 등으로 이루어져 있다.

시스템 모니터 : 이 도구는 시스템 및 분산되어 있는 지식공학 응용 소프트웨어 등을 감독하고 관리한다. 이를 이용하면 시스템 관리자는 응용 소프트웨어의 매개 변수, 작동 상태, 처리 작업, 네트워크 등을 감독할 수 있고, 또한 사전에 정의된 조건이 발생하는 경우 이에 대한 예방 조치를 취할 수 있게 해준다.

3. 사모아의 관리 기능

사모아 플랫폼의 관리 기능을 실행시킬 수 있도록 다양한 상용 제품들의 공통 인터페이스를 정의하였다. 사모아 컴포넌트들의 구성 및 입출력, 그리고 실행을 위한 공통 인터페이스와 식별 기능도 정의하였다. 공통의 통신 버스로는 COM+를 채택하였고 어댑터들은 COM 오브젝트로 구현하였다.

프로젝트 생성 : 사모아는 컴포넌트 풀더라는 것을 자동으로 생성하여 이를 통해서 기본적인 관리 기능들을 식별, 통합, 제어한다. 기본 관리 기능들로는 워크플로우 프로세스, 데이터 원천, 데이터 매트, OLAP 큐브, 데이터 마이닝 모델들의 관리를 들 수 있다.

지식공학 컴포넌트들의 식별과 통합 : 사모아에서는 다양한 기능과 API를 제공하는 사모아 프로토타입 컴포넌트들을 식별, 통합, 제어하기 위해서 확장성 있는 표준 인터페이스를 지원한다. 사모아 컴포넌트들을 위해서 설계, 구현된 어댑터들은 이러한 인터페이스를 모두 지원한다.

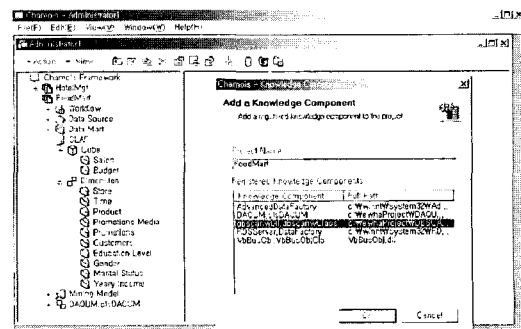


그림 2 사모아의 통합 관리기

샤모아 인터페이스에 맞추어 개발된 응용들도 하나의 컴포넌트로써 샤모아 플랫폼에 통합될 수 있다. 그림 2는 데이터 품질 측정 프로토타입과 공간 데이터 마이닝 알고리즘인 DBSCAN - W가 하나의 컴포넌트로서 샤모아에 통합된 모습을 보여준다.

워크플로우 프로세스 설계 : 샤모아는 워크플로우 프로세스 설계를 위해 편리한 그래픽 사용자 인터페이스를 제공한다. 이 사용자 인터페이스는 Microsoft의 Visio를 사용하여 구현하였으며 컴포넌트의 실행과 컴포넌트들 사이의 데이터 전송을 정의하는 부분들로 구성된다. 워크플로우 프로세스는 스케줄러에 의해서 주기적으로 실행되며, 클라이언트 질의를 신속하게 처리하기 위해서 지식 베이스를 자동적으로 관리한다. 그림 3은 데이터 품질 측정 프로토타입을 사용하여 데이터 마트의 데이터 품질을 측정하여 응용들에게 제공하는 워크플로우 프로세스를 나타낸다.

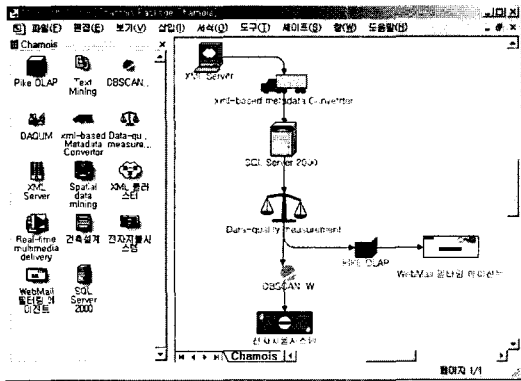


그림 3 샤모아를 이용한 워크플로우 프로세스의 설계

클라이언트 질의 처리를 위한 API : 샤모아는 OLAP 큐브 또는 훈련된 데이터 마이닝 모델과 같은 지식 베이스에 대하여 'OLE DB for OLAP'이나 'OLE DB for data mining'을 지원하느 질의들을 위한 API를 제공한다.

데이터 원본 관리 : 이 관리 기능은 외부로부터 가져온 데이터베이스나 데이터 테이블들을 식별하고 연결하는 역할을 한다. 현재 샤모아는 OLE DB나 ODBC API를 지원하는 DBMS를 연결할 수 있다.

데이터 마트 스키마 설계 : 샤모아는 데이터의 다차원적 분석을 할 수 있도록 하기 위해 별모양 스키

마나 눈송이 스키마를 설계할 수 있는 GUI를 제공한다.

OLAP 큐브의 설계 : 샤모아는 사용자들이 OLAP 컴포넌트들을 제어, 접근, 통합하는데 사용할 수 있는 GUI 를 제공한다. 샤모아는 또한 메타 데이터 저장소를 제공하고, 데이터 마트의 차원과 레벨을 정의하는 마법사와 OLAP 큐브 구조를 설계하는 마법사도 제공한다.

데이터 마이닝 모델 설계 : 데이터 마이닝 기술들과 이들을 실행시키는데 쓰이는 수많은 API들이 이미 사용되고 있기 때문에 공통의 인터페이스를 정의하는 것은 쉽지 않다. 샤모아는 사용자가 다차원 또는 관계형 데이터베이스 원본으로부터 데이터 마이닝 모델을 생성할 수 있도록 해 준다. 이 때 마이크로소프트사의 SQL 서버의 'analysis services'의 데이터 마이닝 모델 마법사를 사용한다.

4. 프레임워크의 테스트와 검증

컴포넌트 기반의 구조를 이용하면 네트워크의 컴퓨터에 분산된 컴포넌트들을 서로 연계시킬 수 있을 뿐 아니라, 각 컴포넌트를 독립적으로 변화시킬 수 있다. 즉, 각 컴포넌트를 후속 버전으로 쉽게 대처할 수 있다. 샤모아의 컴포넌트 기반의 구조에서는 컴포넌트들 간에 메타 데이터를 공유할 수 있도록 공통의 API를 통해서 다량의 데이터와 서비스들을 접근할 수 있게 해 준다. 또한 통합 관리기, 워크플로우 설계기, 어댑터 SDK는 시스템 전체를 더 향상시키고 관리할 수 있도록 해준다.

샤모아 프레임워크를 설계, 개발, 테스트 하는 데에는 컴포넌트 기반의 소프트웨어 개발 기술(CBSD: Component-Based Software Design)을 사용하였는데 이러한 기술은 이미 개발된 소프트웨어들을 재사용함으로써 소프트웨어의 설계 및 개발 비용이 절감될 수 있도록 한다. CBSD에는 맞춤(customization), 조립(composition) 등과 같은 필수적인 기술이 포함된다. 그러나 타인에 의해 설계, 개발된 소프트웨어를 사용하는 경우 위험 부담이 따른다. 일반적인 CBSD 연구뿐만 아니라 샤모아 프로토타입의 개발에 있어서도 한가지 주요한 연구문제는 이미 구현된 컴포넌트들을 사용하여 만든 복잡한 소프트웨어를 테스트하는 방법이다.

상품화된 소프트웨어들과 개별적으로 만든 모듈

들이 CBSD 환경에서 혼합되어 있기 때문에 샤모아를 테스트한다는 것은 매우 복잡하고 어려운 일이다. CBSD 문제들을 컴포넌트 제공자 및 컴포넌트 사용자의 두가지 측면에서 바라볼 수 있다. 실제로 컴포넌트 제공자만이 소스코드를 접근할 수 있으므로 소스코드를 접근할 수 없는 컴포넌트 사용자들은 매우 제한적인 테스트를 할 수 밖에 없다.

이 프로젝트에서는 컴포넌트 사용자들의 관점을 고려한 테스트 기술을 개발하였다. 이 기술을 쓰면, 사용자들은 컴포넌트 제공자들의 제품을 현재 개발 중인 소프트웨어에 맞게 반복적으로 계속 맞추어 가면서 컴포넌트 기반의 소프트웨어를 개발할 수 있다.

샤모아 팀에서는 일반적인 컴포넌트들을 개발하였고 이를 전사적 자바빈즈 (Enterprise JavaBeans) 로 확장하였다. 이 기술로 EJB 사용자들은 컴포넌트들의 인터페이스에 관한 정보를 사용하여 각 부분들이 서로 상호작용하는 가운데 발생할지 모르는 오류를 테스트한다. 여기서 결함 주입 표적 (FIT: fault injection target)을 정의하였는데, 이는 맞춤형 컴포넌트의 인터페이스에서의 특정 부분이다. 그런 후에 FIT에만 결함을 주입함으로써 이 컴포넌트를 위한 테스트 케이스들을 생성한다. 이 기법은 테스트 케이스들에 효과적으로 검증되었고 비용 절감 효과를 가져다주었다[5].

샤모아의 기능들과 컴포넌트들간의 상호작용을 검증하기 위해서 전자 상거래 프레임워크인 이화숍 (EwhaShop)을 설계하고 개발하였다. 이 전자 상거래 프레임워크는 다음과 같은 기능들을 갖는 컴포넌트들로 구성된다.

- (1) 상품 데이터베이스로부터의 카탈로그 생성
- (2) 고객 정보 관리
- (3) 재고 관리
- (4) 구매, 결제 정보 관리
- (5) 점포 소유권 관리
- (6) 매입/매출 일지 생성
- (7) 판매촉진을 위한 세일 관리
- (8) 경매 관리
- (9) 공급체인 관리
- (10) 웹 로그 서버를 통한 웹 로그 추출

초기에는 B2C 상거래에 초점을 맞추고 마이크로소프트사의 SQL 서버를 사용하여 전자 상거래 서버

를 구현하였다. 서버측의 스크립트로는 ASP, Perl, PHP 등을 사용하였다. 이 전자 상거래 프레임워크는 마이크로소프트사의 Siteserver와 intershop과 같은 일반 제품들의 모델을 따랐다. 핵심 컴포넌트 모듈들은 마이크로소프트사의 윈도우 2000 advanced 서버 상에서 ASP와 비주얼 C++로 코딩하였다. 이와 같이 구축된 전자 상거래 프레임워크는 가상적으로 합성하여 만든 고객들의 요구들, 예를 들면 제품 검색; 제품 목록의 커스터마이징; 구매 물품 쇼핑카트에 넣기; 물품의 진열, 수정, 제거; 주문 처리; 주문 물량 계산; 지불 처리 등을 사용하여 검증하였다.

웹상에서 생성되는 대용량 데이터 세트를 처리하려면 성능저하의 문제가 발생할 수 있는데 이를 해결하기 위해서 이 프로젝트에서는 지능적 로그 서버 (ILS: Intelligent Log Server)를 구현하였다. 독립적인 웹 서버인 ILS는 전자 상거래 시스템의 고객들에 대한 성향 정보를 원격으로 수집한다. 이 서버는 마이크로소프트사의 ADO 기술에 기반한 고성능 파일 포맷을 사용한다. 그리고 일반 로그 포맷으로는 수집할 수 없는 사용자들의 브라우징 정보까지 수집할 수 있다.

5. 샤모아에 통합된 소프트웨어 모듈

샤모아 팀은 통합 지식공학 구조를 구성하는 핵심 기술 전반에 대한 연구를 수행하였다. 지금까지 설계되고 구현된 샤모아 컴포넌트들이 그 결과를 보여주고 있으며 전자 상거래 시스템 프레임워크, 전자지불 시스템 테스트베드, OLAP 서버, 공간 데이터 마이닝 알고리즘 외 몇몇 간단한 데이터 마이닝 응용들, 몇 개의 암호화 알고리즘, XML 기반 메타 데이터 변환기, XML 문서 서버, XML 마이닝 전처리 방법, QoS 멀티캐스트 시뮬레이터, 가상사설망(VPN) 성능 테스트베드 등이 있다. 지면관계와 이 논문의 컴포넌트 구조에 초점을 맞추기 위해서 지금까지 샤모아 프레임워크에 포함된 모듈에 대해서만 간략히 설명한다. 나머지 모듈들과 개발 중인 새 모듈들은 이 프레임워크에 향후 1년 내에 포함될 것이다. 대부분의 소프트웨어 모듈들은 지속적으로 수정되고 기능과 성능이 평가될 것이며 전체 샤모아 프레임워크의 기능과 성능 또한 계속 타당성을 검증받을 것이다.

전자 상거래 시스템 테스트베드 컴포넌트: 전자 상거래 프레임워크로 전자 화폐와 새로운 전자 지불

시스템의 특성을 잘 이해하기 위한 도구로서 전자지불 시스템 테스트베드를 개발하였다.

이 테스트베드는 발행자(issuer), 고객(customer), 상점(merchant), 인증 기관(certificate authority)과 은행(bank)이라는 5개의 주요 컴포넌트를 가진다. 발행자는 은행이 아닐 수 있으며 전자 화폐를 발행하고 사용 화폐의 데이터베이스를 유지하여 이중 사용 여부를 체크한다. 고객은 전자 화폐의 사용자이다. 은행은 발행자가 발행한 화폐에 서명을 한다. 상점은 전자 화폐를 받는 가게이며, 인증 기관은 은행의 공용 인증서를 발행한다.

이 테스트베드는 온라인 주화-기반 전자 화폐 시스템이다. 온라인 전자 화폐는 발행자가 보관하는 디지털 화폐의 잔액을 추적하며 오직 네트워크에서만 접근 가능하다. 전자 화폐를 나타내는 한 방법은 전자주화의 집합을 사용하는 것이다. 각 주화는 유일한 일련번호와 은행의 서명을 가진다. 일반적인 주화와 같이 각 전자주화는 고정된 액면금액을 가진다. 주화들의 액면금액의 합계를 사용하여 임의의 금액도 지불할 수 있다. 전자 거래가 이루어지기 위해서는 철회, 지불, 예금 프로토콜이라는 최소한 세가지 프로토콜이 실행되어야 한다. 본 연구팀에서는 이 세가지 프로토콜을 모두 구현하였다.

데이터 품질 측정 컴포넌트 : 데이터 품질 측정(DAQUM: The Data Quality Measurements) 컴포넌트는 본 연구팀에서 개발한 오염 데이터(dirty data) 분류를 기준으로 데이터 집합의 품질을 분석하여 품질에 대한 정량적 척도를 사용자가 얻도록 하여 준다.

지식공학 프레임워크는 다양한 운영 데이터 원천으로부터 데이터를 추출하여 변환한 후에 데이터 웨어하우스로 적재하여 다양한 응용들이 접근하여 사용하도록 한다. 데이터 웨어하우스의 데이터 품질은 이 데이터로부터 유도되는 응용이나 사용자들이 보게 될 지식의 품질을 상당 부분 결정하게 된다. DAQUM 컴포넌트는 지식공학 시스템에서 데이터의 품질을 측정하기 위해 개발된 방법론을 구현한다. 본 연구팀은 ISO/IEC 9126에서 정의한 소프트웨어 품질 정의를 확장한 방법론을 개발하였다.

이 방법론의 일부를 구성하는 것은 표준 연속 계층형 정제 방법론[6]을 사용하여 개발된 오염 데이터에 대한 포괄적인 분류이다. 이 분류는 오염 데이터는 누락 데이터, 누락되지 않고 틀린 데이터, 누락이

나 틀리지 않고 불필요한 데이터라는 세 가지임을 기반으로 한다. 본 연구팀에서 개발된 분류는 세가지 기본 오염 데이터를 계층적으로 분해하여 33가지의 최종 단계의 오염 데이터 타입으로 구성된다. 본 방법론의 또다른 부분을 구성하는 것은 최종 사용자의 시각에서 ISO/IEC 9126과 오염 데이터 분류를 평가하여 얻어진 데이터 품질의 특성과 부특성을 나타내는 데이터 품질 척도이다.

공간 데이터 마이닝 컴포넌트 : 공간 데이터 마이닝은 공간 데이터베이스에 묵시적으로 존재하는 흥미있는 관계성과 특성을 탐사하는 과정이다. 여러가지 공간 클러스터링 알고리즘들이 논문으로 발표되었다[7].

본 연구팀은 DBSCAN-W라고 하는 새로운 클러스터링 알고리즘을 개발하였고 Informix Spatial Datablade 모듈을 이용하여 샤모아 컴포넌트로 구현하였다. DBSCAN-W는 기존의 밀도 기반 클러스터링 알고리즘인 DBSCAN[8]을 확장한 것이다. DBSCAN은 객체들을 클러스터링하기 위하여 객체의 위치만을 고려하는데 DBSCAN-W는 각 객체의 위치뿐만 아니라 주어진 응용과 관련된 일부 특성을 고려한다. DBSCAN-W에서 각 객체는 영역이라고 불리는 원으로 표현되며 원의 반경은 객체의 특성(예를 들면 각 고객에 대한 판매금액)에 의해 결정된다. 그래서 DBSCAN-W는 동일한 클러스터에 있는 중첩된 객체들도 처리할 수가 있다.

실시간 멀티미디어 전송 컴포넌트 : 오디오-비디오 정보를 적시에 전송하는 것은 네트워크 서비스 분야의 새로운 과제이다. 샤모아 컴포넌트들 간에 멀티미디어 트래픽의 효과적인 실시간 전송과 네트워크 서비스를 제공하기 위해서는 두 가지 핵심 연구 문제로 QoS(Quality of Service) 라우팅과 멀티캐스트 전송이 제기된다.

QoS 라우팅은 요청된 QoS 매개 변수들을 만족할 수 있는 충분한 자원을 갖는 네트워크 라우트를 선택한다. QoS 라우팅이 잘 되면 QoS를 보장하면서 네트워크 성능을 현저히 증대시킬 수 있다. 그러나 이러한 장점은 추가적인 라우팅 비용들을 수반하게 되는데 더욱 복잡하고 빈번한 라우팅 연산과 함께 이들 연산에 관련된 네트워크 자원 상태에 대한 분산 갱신을 위한 프로토콜 부하와 같은 것들이다.

최근의 연구는 이 비용을 감소하는데 초점이 맞추어져 있다. 발표된 대부분의 논문들은 네트워크의

QoS 상태를 분산 갱신한다. 그러나 본 연구팀에서는 이러한 부하를 완전히 제거하는 집중형 서버 기반 QoS 라우팅 기법을 설계하고 구현하였다[9]. 이 방식에서는 라우트 서버가 라우팅 도메인내의 모든 라우터를 대신하여 QoS 경로 결정을 담당한다. QoS 경로 계산에 필요한 동적 링크 QoS 상태 정보는 라우트 서버가 QoS 경로를 할당하고 회수할 때마다 목시적으로 유지된다. 네트워크 상태 정보를 유지함으로써 네트워크 갱신 메시지 교환 부하를 없앨 뿐만 아니라 경로 계산에서 보다 정확한 네트워크 상태 정보를 사용함으로써 높은 라우팅 성능을 얻게 된다. 시뮬레이션 결과 제안한 방식이 현저하게 라우팅 성능을 증대시키며 라우팅 서버에 가해지는 부하는 최신 기술이 수용할 수 있는 범위 내에 있음을 보여준다.

대부분의 주요 멀티미디어 응용들, 예를 들면 애드혹 네트워크(Ad-hoc network)는, 멀티포인트와 멀티포인트 통신을 요구한다. 애드혹 네트워크(Ad-hoc network)는 베이스 스테이션과 같이 고정된 네트워크 기반 구조를 갖지 않는 다중도약(multi-hop) 무선 네트워크를 말한다. 이들은 보통 통상적인 행동에 참여하는 이동 호스트들로 구성되며 특정한 지리적으로 제한된 영역을 갖는다. 애드혹 네트워크는 매우 심한 자원에 대한 제약조건을 가지며 무제한적인 노드 이동성을 갖는다는 특성이 있다. 본 연구팀에서는 애드혹 네트워크를 위한 멀티캐스트 전송 기법을 연구하였으며 PatchODMRP와 HMMRP라고 하는 두 개의 애드혹 멀티캐스트 전송 프로토콜을 개발하였다. PatchODMRP는 프로토콜 부하를 감소하면서 효과적인 이동 적응성을 보여준다[10]. HMMRP는 높은 이동성을 가지는 대규모 애드혹 네트워크에 적합하다. 멀티미디어 트래픽은 큰 대역폭을 요구하므로 멀티캐스트 통신이 필수적으로 요구되며 대역폭이 상대적으로 제한된 무선 네트워크에서 더욱 그렇다.

XML 기반 메타 데이터 변환기 컴포넌트 : XML은 다양한 응용에서 요구되는 구조 및 반구조 문서를 표현하는 표준 형식으로 급속히 자리잡고 있다[11]. 특히 XML은 데이터 웨어하우징, OLAP, 데이터 마이닝, 데이터베이스 등 서로 다른 도구들간의 메타 데이터 교환을 포함하는 데이터 교환을 위한 표준 규격으로 채택되어 다른 도구에서 생성된 메타 데이터를 활용 가능하게 하고 있다. 우리는 두 개의 관계형 데이터베이스 시스템간의 메타 데이터를 저장하고 적재하는 기능을 가지는 XML 기반 변환기를 개발하

였다. 이 변환기는 소스 메타 데이터를 대응하는 XML로 변환한 후에 다시 이 데이터를 목표 데이터베이스 시스템의 메타 데이터로 변환한다. 이 시스템은 모든 종류의 관계 데이터베이스 시스템이 가지는 테이블, 항목들의 특성, 항목에 대한 제약조건, 기본키, 외래키, 색인 등의 메타 데이터를 처리한다. 이 변환기는 메타 데이터를 갖는 샤모아의 모든 컴포넌트들 간의 메타 데이터 교환에 사용될 수도 있다.

이 변환기 설계를 위해 먼저 XML 기반 메타 데이터 교환 규격(X-MDIS: MetaData Interchange Specification by XML)을 정의하였다. 이 규격은 Meta Data Coalition 에서 제안한 MDIS(MetaData Interchange Specification)를 충족시킨다.

XML 서버 : 임의의 구조를 잘 표현하게 해주는 XML의 장점인 유연성은 오히려 XML 문서를 저장한 데이터베이스에 대한 간단한 형태의 질의를 구성하기가 어렵게 한다. 관계 데이터베이스가 현재 다양한 데이터 타입을 지원하고 저장하므로 XML 문서 역시 이러한 관계 데이터베이스에 저장하여 고급의 질의를 지원하도록 하는 것이 요구되고 있다. 이를 지원하기 위해서는 XML 문서들로부터 데이터베이스 스키마를 추출해야 한다.

본 연구팀에서 설계하고 개발한 관계 데이터베이스를 기반으로 한 XML 서버는 기존의 XML 문서를 저장하고 질의하기 위한 데이터 서버보다 더 좋은 성능을 가진다. XML 데이터 서버는 XML 문서를 파싱할 때에 XML DTD와 함께 개개의 문서들도 모두 고려한 관계형 스키마를 추출해 낸다. 그런 후 데이터 서버는 DOM(Document Object Model) 트리에서 엘리먼트의 순서, 부모-자식 관계, 그리고 깊이를 포함하는 엘리먼트 경로 식별자들(EPID: Element Path Identifiers)을 생성한다. EPID는 XML 문서의 저장과 질의를 지원한다.

XML 문서 마이닝 : XML 문서를 다루는 많은 웹 응용들은 앞으로 이들 문서들에 대한 클러스터링과 분류를 해야 할 필요가 생길 것이다. 직관적으로 보면 XML의 풍부한 의미 표현 능력은 다양한 XML 응용들에서 클러스터링과 분류를 하는데 매우 강력한 도구가 될 것이다. 그리하여 문서들간에 유사성을 측정할 수 있는 XML 문서를 마이닝할 수 있는 컴포넌트를 설계하고 구현하였다. 이 컴포넌트에서 사용되는 방법은 XML 문서의 엘리먼트와 내포된 구조들의 의미를 분석한다.

본 연구팀이 제안한 방법은 XML 문서 내에서 동의어, 복합어, 축약어를 나열한 확장 요소 벡터와 이들 벡터들을 비교한 유사도 행렬을 생성한다[12].

XML 문서의 구조를 이론적으로 표현하기 위해서 비결정적 유한 오토마타를 사용하여 엘리먼트와 그 속에 내포된 엘리먼트 간의 내포관계를 상태변이로 표현하였다. 반복된 상태 변이들을 제거하면 비결정적 유한 오토마타를 결정적 유한 오토마타로 변화시킬 수 있고 이 결과에다가 상태 축소 알고리즘을 적용하면 상태의 수가 감소한다.

다음으로 순차 패턴 마이닝 알고리즘을 사용해서 문서들의 계층 구조들간에 유사하거나 공통적인 경로들의 수를 결정함으로써 XML 문서들의 유사성을 계산할 수 있다. XML 문서를 이용한 실험에서는 기존의 벡터 기반의 모델과 비교할 때 유사성을 측정하는데 50에서 100 퍼센트 정도의 성능 향상을 보였다.

6. 결론 및 향후 계획

앞으로 1년 정도 본 연구팀에서는 샤모아 프레임워크의 소프트웨어 모듈들을 개선하고 새로운 모듈들을 추가할 예정이다. 각 모듈별로 기능성과 성능에 대한 검증 및 개선을 통하여 전체적인 프레임워크의 성능도 개선시키고자 한다.

프로젝트의 다음 단계에서는 프레임워크를 구성하는 소프트웨어 모듈들 중에서 기능성을 확장할 예정인 것들로 OLAP 서버, 공간 데이터 마이닝 알고리즘, XML 데이터 마이닝 알고리즘, 데이터 품질 관리, B2B 중심의 전자 상거래 등이다. 또 추후 각종 바이오 정보나 CRM을 위한 데이터 마이닝 알고리즘, OLAP 활용 모듈, 보안 알고리즘, 전자 지불 시스템, 무선 통신을 위한 모듈 등의 추가를 고려하고 있다.

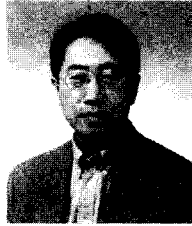
참고문헌

[1] W.A. Ruh et al, Enterprise Application Integration, John Wiley & Sons, New York, 2001, pp.12-38
 [2] M. Kirtland, Web Services Essentials, Microsoft Developer Network, Redmond, Wash., 2001.
 [3] E. Christensen et al., "Web Services Description Language(WSDL) 1.1," W3C Note, 15

Mar, 2001.

- [4] D. Box et al., "Simple Object Access Protocol (SOAP) 1.1," W3C Note, 8 May, 2000.
 [5] H. Yoon and B. Choi, "Component Customization Testing Technique Using Fault Injection Technique and Mutation Test Criteria," Proc. Mutation in the Twentieth and Twenty-First Centuries (Mutation 2000), Kluwer Academic Publishers, Boston, pp.71- 78
 [6] Won Kim, Byoungju Choi, Eui-Kyeong Hong, Soo-Kyung Kim, Doheon Lee, "A Taxonomy of Dirty Data," Data Mining and Knowledge Discovery, Vol.7 no.1, pp.81-99, 2003. 1
 [7] J. Han and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, San Francisco, 2001.
 [8] M. Ester et al., "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proc. 2nd Int'l Conf. on Knowledge Discovery and Data Mining (KDD 96), AAAI Press, Menlo Park, Calif., 1996, pp.226-231
 [9] S. Kim, and M. Lee, "Server-Based QoS Routing," Proc. Int'l Conf. on Parallel and Distributed Processing Techniques and Applications (PDPTA 2001) CSREA Press, 2001, pp.219-225
 [10] Y. Kim and M. Lee, "PatchODMRP: An Ad hoc Multicast Routing Protocol," Proc. 15th Int'l Conf. on Information Networking (ICOIN 2001), IEEE CS Press, Los Alamitos, Calif., 2001, pp.537-543
 [11] S. Abiteboul, P. Buneman, and D. Suciu, Data on the Web: From Relations to Semistructured Data and XML, Morgan Kaufmann, San Francisco, 2000.
 [12] J. Lee, K. Lee, and W. Kim. "Preparations for Semantics-Based XML Mining," Proc. 1st IEEE Int'l Conf. Data Mining (ICDM 2001), IEEE CS Press, Los Alamitos, Calif., 2001.

김 원



1970 Massachusetts Institute of Technology 물리학(학사)
 1971 Massachusetts Institute of Technology 물리학(석사)
 1980 University of Illinois, Urbana-Champaign 전산학(박사)
 1999~현재 이화여자대학교 과학기술대학원장
 현재 Cyber Database CEO, SIGKDD Chair, TOIT editor-in-chief
 관심분야 : 객체지향/객체관계 데이터베이스, 비즈니스 인텔리전스 시스템
 E-mail : won.kim@cyberdb.com

정 안 모



1981 서울대학교 물리교육학과(학사)
 1984~1992 삼성전자 선임연구원
 1992~현재 (주)아리스 대표이사
 2000~현재 이화여자대학교 컴퓨터학과 겸임교수
 관심분야 : 지식기반 시스템, 컴포넌트 기반 소프트웨어 시스템
 E-mail : jam06@ibis.co.kr

김 명

박 승 수

박 현 석

반 호 경

옹 환 승

이 미 정

이 민 수

이 상 호

조 동 섭

채 기 준

최 병 주

● The 9th International Conference on Database Systems for Advanced Applications ●

- 일 자 : 2004년 3월 17~19일
- 장 소 : 제주도
- 주 최 : 데이터베이스연구회
- 상세안내 : <http://aitrc.kaist.ac.kr/~dasfaa04>