

유비쿼터스 컴퓨팅 환경에서 데이터 마이닝을 이용한 서비스 조합

이 선 영[†] · 이 종 연^{††}

요 약

유비쿼터스 컴퓨팅 환경에서 사용자가 원하는 서비스를 다양한 상황에 맞게 적합한 서비스를 검색하고 조합하여 제공하는 것은 중요한 일이다. 그러나 기존 연구는 주로 서비스 발견에 관한 연구이며 사용자의 위치나 주변 환경, 선호도에 관한 고려가 부족하고, 사용자 정보를 이용하여 기본 서비스들로부터 새로운 조합 서비스를 찾아내는 것이 미흡하다. 또한 서비스 조합에 있어 기존 연구는 단순한 기본 서비스들의 나열에 불과하고, 사용 이력을 고려하기 위한 구체적인 방안을 제시하지 않는다. 따라서 본 논문에서는 사용자의 상황 정보 및 서비스 사용 이력을 고려한 데이터 마이닝 기반 서비스 조합 기법을 제안한다. 세부적 연구내용은 첫째, 최적의 서비스를 동적으로 생성하여 제공하는 COSEP 프레임워크를 설계하고 둘째, COSEP 프레임워크에서 데이터 마이닝 기능을 겸비한 온톨로지를 이용한 서비스 조합 기법을 제안한다. 본 연구는 사용자의 상황 정보에 능동적으로 반응하여 서비스를 발견하고 데이터 마이닝 기법을 가진 온톨로지를 이용하여 서비스를 조합함으로써 사용자에게 최적의 서비스를 제공하는 것이다.

키워드 : 유비쿼터스 컴퓨팅, 서비스 프로비저닝, 서비스 조합, 데이터 마이닝

Service Composition with Data Mining in Ubiquitous Computing Environment

Sun Young Lee[†] · Jong Yun Lee^{††}

ABSTRACT

Since users want to have services correctly in their own position and surrounding circumstance in ubiquitous computing environment, it is very important to search, compose basic services and provide suitable services according to various circumstances. However existing techniques have been studied on service discovery mainly and lack consideration for position or preference of users. Furthermore, on service composition, they lists basic services simply and do not propose concretely method of use service history data for service composition. Therefore we propose a framework for context based service provisioning middleware system, called COSEP, and Ontology engine with data ming. This research discovers services by reacting dynamically to circumstance information such as time and position of user, composites services using Ontology engine with data ming and offers newly created optimal services to users.

Key Words : Ubiquitous Computing, Service Provisioning, Service Composition, Data mining

1. 서 론

유비쿼터스 컴퓨팅 환경에서 이용 가능한 자원과 서비스들은 사용자와 장치들의 이동성에 따라 변한다. 그러나 사용자들은 자신들의 위치나 주변 환경에 상관없이 최적의 서비스 사용을 원하므로, 다양한 상황에 맞게 적합한 서비스를 검색하여 제공해 주는 것은 매우 중요한 일이다. 만약 기존의 등록된 서비스 집합에서 사용자의 요구에 맞는 서

스가 존재하지 않는다면 새로운 서비스를 생성하거나 기존의 기본 서비스를 조합하여 제공해주는 것이 필요하다. 서비스 조합은 합성 서비스를 고안하기 위해 논리, 애플리케이션 도메인, 제어 흐름에 의존하여 서비스들을 연결하는 것이며[1], 또한 작고 간단하며 쉽게 실행되는 서비스들 또는 요소들로 복잡한 서비스를 창조하는 기술로 유비쿼터스 컴퓨팅 환경에서 사용자들의 복잡한 질의를 해결하여 서비스들을 이용 가능하게 한다[2].

* 이 논문은 2006년도 교육인적자원부 지방연구중심대학 육성사업의 지원에 의하여 연구되었음.

† 준 회원 : 충북대학교 컴퓨터교육과 박사과정

†† 종신회원 : 충북대학교 컴퓨터교육과 교수(교신저자)

논문접수 : 2006년 4월 18일, 심사완료 : 2006년 6월 13일

1.1 연구 동기

서비스 발견에 관한 기존의 연구들은 Sun Microsystems

사가 사용자와 그 사용자들이 이용하려고 하는 자원의 유기적인 결합을 지원하기 위해 제안한 JINI[3], Microsoft사가 기존의 IP프로토콜과 HTTP 프로토콜을 사용하여 제안한 UPnP[4]가 있으며 그 밖에 [5-9] 등이 있다. 이 서비스 발견에 관한 연구들은 사용자의 위치나 주변 환경, 선호도에 대한 고려가 부족하며 사용자 정보를 이용하여 기본 서비스들로부터 새로운 서비스를 찾아내는 것이 미흡하다. 그러므로 사용자의 시간과 위치와 같은 상황정보에 동적으로 반응하여 서비스 발견, 조합하는 사용자 중심의 서비스 제공 메커니즘이 필요하다.

서비스 조합은 서로 연관성이 있는 서비스들끼리 그룹을 형성하는 특징을 가지고 있으나, 전혀 연관성이 없어 보이는 서비스들끼리 연관이 되는 경우도 발생한다. 이러한 경우 사용자의 서비스 사용에 대한 이력 정보를 활용하면 새로운 서비스 조합을 발견할 수 있다. 그러나 기존의 서비스 조합[1, 10-14]에 관한 연구들은 단순한 기본 서비스들의 나열에 불과하고 서비스 조합에 대한 구체적인 방안을 제시하지 못하며, 사용자의 서비스 사용 이력에 대한 관심이 부족하다. 그러므로 서비스 조합에 있어 특정 사용자의 서비스 사용이력을 이용하여 새로운 서비스 조합의 규칙을 제공하는 것이 필요하다.

1.2 기여도

따라서 본 연구에서는 사용자의 서비스 사용에 대한 이력 정보를 이용하여 데이터 마이닝을 통해 새로운 규칙을 발견하고, 그 규칙들을 서비스 조합에 활용한다. 본 논문에서는 첫째, 사용자의 상황정보(context information)를 고려하여 최적의 서비스를 동적으로 생성하고 제공하는 COSEP(Context-based Service Provisioning System) 프레임워크를 설계하고 둘째, 데이터 마이닝 기능을 이용한 서비스 조합 기법을 제안한다. 데이터 마이닝 기법은 사용자의 서비스 사용에 대한 이력 정보를 분석하여 새로운 규칙을 발견하고, 새로 발견된 조합 규칙들을 온톨로지 엔진에 추가한다. 데이터 마이닝을 통해 새로운 서비스 조합 규칙들을 발견하면 사용자에게 제공 될 수 있는 서비스 조합의 수가 증가하여, 조합 서비스의 제공 능력을 향상 시킬 수 있다. 즉, COSEP를 통해 사용자의 상황 정보에 능동적으로 반응하여 서비스를 발견하고, 데이터 마이닝 기법을 적용한 온톨로지 엔진을 이용하여 서비스를 조합함으로써, 사용자에게 최적의 서비스를 제공한다.

이 논문의 구성은 다음과 같다. 2장에서는 서비스 조합에 관한 기존 연구를 살펴보고 기존 연구들의 문제점을 제시한다. 3장에서는 기존의 서비스 조합의 단점을 보완하기 위해 상황 기반 서비스 프로비저닝 시스템의 프레임워크를 설계하고, 4장에서 데이터 마이닝 기법을 적용한 온톨로지 엔진을 제안한다. 5장에서는 제안한 서비스 조합 기법의 성능을 실험을 통하여 분석 한다. 마지막으로 6장에서는 논문의 결론을 간략히 요약한다.

2. 관련 연구

이 장에서는 유비쿼터스 컴퓨팅 환경의 프레임워크 설계

와 기존의 서비스 조합 기법에 관해 기존 연구 내용을 검토한다.

2.1 프레임워크

Jini[3]는 선 마이크로시스템에서 개발한 미들웨어로 Java를 기반으로 JVM(Java Virtual Machine)에서 동작한다. 네트워크의 모든 멤버를 서비스로 알리고, 그 서비스들은 룩업서비스에 등록된다. 서비스 제공자들은 서비스 프록시 객체를 업로드하고 클라이언트는 프록시 객체를 다운로드하기 위해 룩업서비스에 접촉한다. 클라이언트는 서비스들이 어떻게 구현되고 어디에 있는지에 관계없이 이용할 수 있는 서비스들을 간단하고 균일하게 제공 받을 수 있다. SLP[6]는 IETF에 의해 발전된 것으로, 엔터프라이즈 네트워크에서 공유 자원을 찾아내기 위한 프로토콜이며 매우 규모가 크다. UPnP[4]는 마이크로소프트가 제안한 미들웨어 솔루션으로 사용하기 쉽고 유연하며, 어디서든 상용 네트워크에 표준 기반 연결을 지원하는 초창기 기술이다. 작은 크기부터 중간 크기의 IP 네트워크에서 장치를 발견하고 조정을 위한 프로토콜들과 시스템 서비스들로 구성된다. 모든 종류의 기기들을 연결하는 네트워크 구조로서 사용자의 작업 없이 기기를 네트워크에 연결하며, 서비스 묘사에 XML을 사용하여 사용이 용이하고 유연하다. UPnP의 가장 좋은 점은 운영체제와 프로그래밍 언어에 독립적인 구조와 메시지를 제공하는 것이다. 그 밖의 서비스 발견 프로토콜에는 Bluetooth SDP[8], Salutation[7], HAVi[5] 등이 있다.

사용자의 상황 정보를 고려한 기존 연구에는 유비쿼터스 컴퓨팅 환경에서 보안 서비스의 설계와 전개에 대한 상황-인지 접근 제어 프레임 워크인 UbiCOSM[15]와 SUN의 Jini에 상황-인지를 첨가한 reggie[16]가 있다. 또한 다양한 사용자 인증단계로 통합된 네트워크나 서비스 단계의 여러 클러스들이 동작하는 네트워크에서 사용자 레벨에 맞게 인증된 서비스를 발견하고 사용하는 COPS-SD[17]가 제안된 바 있다.

2.2 기존 연구의 서비스 조합

UBIDEV[18]는 사용자 환경의 상황 중심 관리를 제안하고, 애플리케이션이 상황의 변화에 따라 자동적으로 재구성하도록 한다. 동차 공간(Homogeneous space)의 애플리케이션 레벨에서 자원, 서비스, 상황 정보에 대한 통합된 관리 모델을 나타낸다. (그림 1)의 Ubiquitous Access 계층의 Coordination Manger는 애플리케이션 층으로부터 입력된 복잡한 질의를 분해하여 기본 서비스를 합성한다. 여기서 온톨로지는 상황 정보의 분류에 사용된다.

CB-SeC[1]는 서비스 실행에 있어 사용자의 위치와 자원에 상관없이 사용자의 요구를 만족시키기 위해 상황정보를 고려하여 서비스 발견과 조합의 기능을 증진시키기 위해 제안되었다. 조합 모듈에서 조합 서비스를 만들기 위해 서비스 검색을 관리하는 작업을 수행하고 클라이언트가 서비스를 요청할 때 저장된 기본 서비스로부터 복잡한 서비스를

조합한다. 검색 단계 후 서비스 조합 모듈은 발견된 기본 서비스들의 집합을 갱신한다. 사용자 선호도를 만족하는 서비스가 선택되고 대응되는 캡슐들은 서비스 실행 모듈로 보내진다. 단, UBIDEV와 CB-SeC는 상황 정보를 고려하여 기본 서비스들의 집합을 제공하는 역할만을 하고 사용자의 선호도나 서비스 사용 이력에 대한 고려가 없다.

USON[10, 11]에서는 서비스 엔터티(Service Entity)와 서비스 템플릿(Service Template)을 기초로 한다. (그림 2)는 USON네트워크와 USON 엔진을 가지는 사용자 사이에서의 서비스 요청(service request), 발견(discovery), 응답(response) 등의 과정을 보여준다. 서비스 조합 단계에서 USON은 무엇을 원하는지, 어디에 위치했는지 등의 사용자 정보에 기초하여 서비스 템플릿을 검색하고 발견한 후 획득한 서비스 템플릿 중에 하나의 후보를 선택하고 이에 적당한 서비스 엔터티의 발견을 요청한다. 서비스 출현 단계에서 새로운 상황에 만족하는 새로운 서비스 템플릿은 서비스 엔터티와 서비스 템플릿의 사용이력을 기초로 얻어진다. 아무것도 없는 곳에서 서비스 템플릿이 생성되는 것이 아니라 현재의 서비스 템플릿이나 그 일부에서 새로운 서비스 템플릿이 생성된다. 그러나 서비스 엔터티와 서비스 템플릿의 사용이력을 어떻

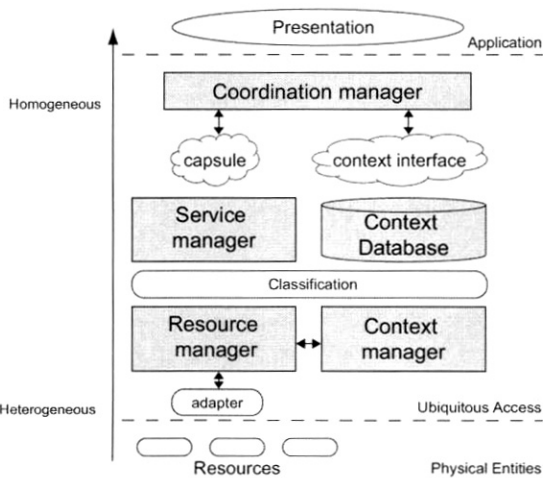
[2]에서는 pervasive/ad-hoc 네트워크에서 중개자 기반 분산 서비스 조합 프로토콜을 제안한다. GSD 발견 프로토콜의 사용으로 서비스 탐사부분에 유연성의 증가를 가져오고 서비스 요청의 전달 제어에 효율적인 광대역 사용으로 서비스 탐사의 효율성이 향상되었다. 중개자 선출 매커니즘을 사용하여 중개자가 결정되었을 때 조합서비스의 분산된 실행이 허락된다. 조합 능력, 중개자 중재 능력, 조합 환경에서 Fixed-Source based Composition과 비교하여 우수성을 보인다.

시맨틱 정보를 이용한 동적 서비스 조합[13]은 미래 애플리케이션의 유연성과 적응성을 가져오며, 요소(component)들을 선택하고 조합함으로써 애플리케이션은 각각 개인의 선호도에 적응하고 이용 가능한 상황 정보를 고려한다. 현재 조합 서비스 시스템이 데이터 타입, 서비스 템플릿, 논리식과 같은 정확한 구문 형식을 요구하는 것에 대한 방안을 제시한다.

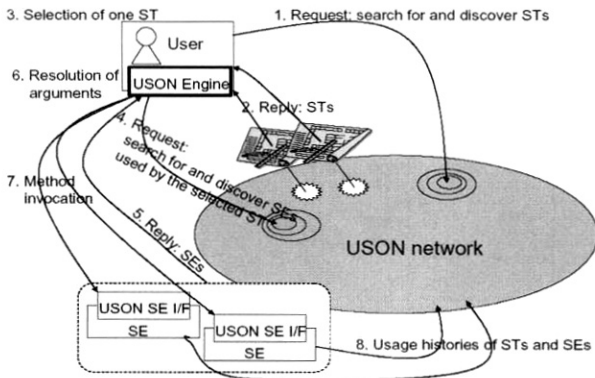
3. COSEP: 서비스 프로비저닝 프레임워크의 설계

이 장에서는 사용자의 위치, 서비스 사용이력 등과 같은 상황 정보를 이용하여 서비스 사용자에게 제공할 수 있는 적합한 서비스를 발견하고, 조합하는 서비스 프로비저닝 프레임워크를 제안한다. (그림 3)은 네 개의 층으로 구성된 상황 기반 서비스 프로비저닝 미들웨어 시스템(COSEP: Context-based Service Provisioning Middleware System)을 나타낸다.

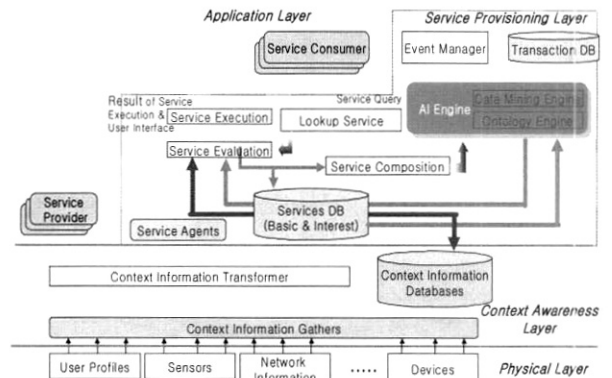
첫 번째 애플리케이션 계층(Application Layer)은 COSEP에게 서비스를 질의하고 최적의 서비스를 제공받는 서비스 소비자(Service Consumer)와 자신의 서비스를 COSEP에 등록하고 소비자에게 서비스를 제공하는 서비스 제공자(Service Provider)가 위치하고 있다. 두 번째 서비스 프로비저닝 계층(Service Provisioning Layer)은 서비스 소비자로부터 서비스 질의를 받아 최적의 서비스를 발견하고 새로운 서비스를 생성하는 곳으로 데이터 마이닝 엔진과 온톨로지 엔진을 기반으로 기본 서비스들 간의 조합을 통한 조합 서비스를 생성한다. (그림 4)는 (그림 3)의 서비스 프로비저닝 계층의 질의처리 과정을 상세히 나타낸다. (그림 4)에서 서



(그림 1) UBIDEV의 구조



(그림 2) USON의 서비스 조합에 이용하는데 대해 구체적으로 제시하지 못한다.



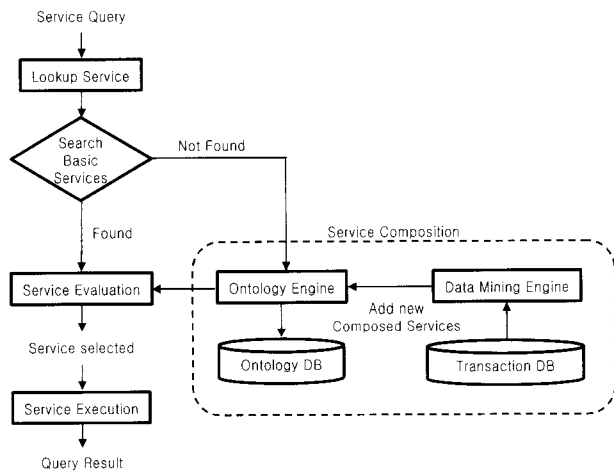
(그림 3) 상황 기반 서비스 제공 시스템의 프레임 워크

비스 질의를 받은 룩업 서비스(Lookup Service)부는 서비스 데이터베이스(Services DB)에서 서비스 소비자의 상황 정보에 알맞은 서비스를 검색하고 서비스 평가부(Service Evaluation)로 검색한 서비스들을 보낸다. 그러나 알맞은 서비스가 없는 경우는 서비스 조합부(Service Composition)에 서비스 조합을 요청한다. 조합 서비스는 상황 정보와 사용자 선호도를 고려하여 기본 서비스들로부터 생성되고, 온톨로지 엔진에 정의된 규칙을 기반으로 한다. 이때 온톨로지 엔진은 데이터 마이닝에서 사용자의 서비스 사용 이력을 통해 발견된 새로운 지식도 이용한다. 발견된 서비스는 평가부에서 사용자에게 평가받고, 그 중 선택된 서비스는 서비스 실행부(Service Execution)를 통해 소비자에게 제공된다. 그리고 평가 받은 서비스들의 정보는 상황 데이터베이스와 서비스 데이터베이스에 저장되어 다음 서비스 요청 시 이용된다. 서비스 소비자의 서비스 사용 이력은 트랜잭션 데이터에 저장되어 데이터 마이닝 엔진에서 서비스 소비자의 서비스 사용 패턴에 대한 새로운 규칙의 발견에 이용한다. 데이터 마이닝 엔진에서 발견된 새로운 조합 서비스들은 온톨로지 엔진을 재구축 한다.

세 번째 상황 인지 계층(Context-Aware Layer)은 최적의 서비스 제공을 위한 상황 정보를 인식하는 곳으로 상황 정보 수집기(Context Information Gatherers), 상황 정보 변환기

(Context Information Transformation), 상황 정보 데이터베이스로 구성된다. 상황 정보 수집기는 물리 계층(Physical Layer)으로부터 생성된 상황 정보를 수집하고, 상황 정보 변환기에서 수집된 상황정보를 서비스 프로비저닝 층에서 사용할 수 있게 계산 가능한 정규화된 형태로 변환하여 데이터베이스에 저장한다. 마지막 물리 계층은 서비스 제공에 영향을 미칠 수 있는 상황 정보를 발생하는 계층이다. 발생된 상황 정보는 상황 인지 계층(Context Awareness Layer)의 상황 정보 수집기에 의해 수집된다.

(그림 5)는 유비쿼터스 컴퓨팅 환경에서 서비스 제공자(Service Provider), 서비스 소비자(Service Consumer), COSEP의 관계를 보여준다. 서비스 제공자는 자신의 서비스를 COSEP에 등록하고 자신의 서비스를 이용하고자 하는 서비스 요청의 정보를 받고, 서비스 소비자는 COSEP에게 서비스를 질의하고 서비스 제공자의 서비스를 사용한다. COSEP은 서비스 소비자에게 이용 가능한 서비스 목록을 제공해주며, 서비스 제공자에게는 서비스 사용자의 정보를 제공해준다. 따라서 서비스 사용자는 서비스 제공자를 확인하고 최적의 서비스 제공자를 선택한 다음 서비스 제공자와 P2P 통신을 통해 서비스를 이용할 수 있다.



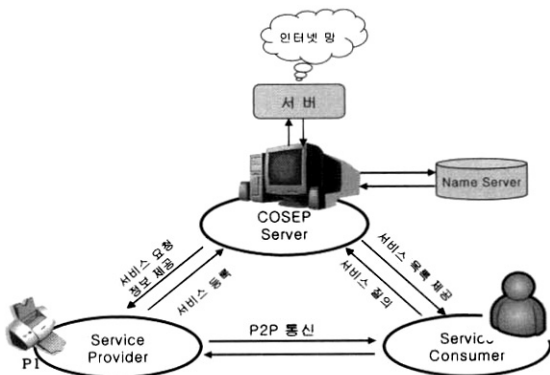
(그림 4) 서비스 프로비저닝 계층의 질의처리 과정

4. 데이터 마이닝 기법을 적용한 온톨로지 엔진

이 장에서는 위에서 설명한 COSEP의 서비스 조합에 대해 기술한다. 이미 정의된 표현들과 규칙을 통해 온톨로지 엔진에서 동적으로 서비스 조합이 이루어지고, 새로운 서비스 조합은 데이터 마이닝을 통해 생성된다.

4.1 온톨로지 엔진

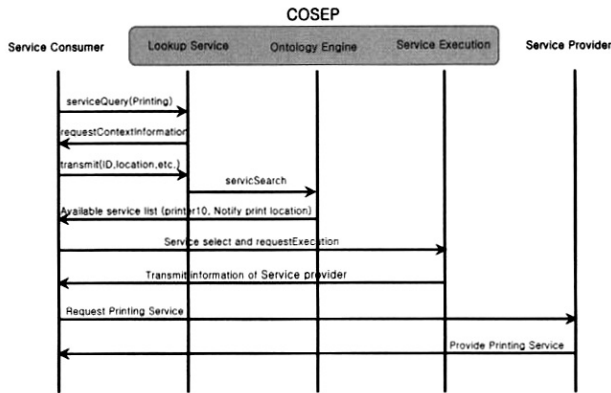
온톨로지는 공유된 개념화(shared conceptualization)에 대한 정형화되고 명시적인 명세(formal and explicit specification)를 뜻한다. 또한 단어와 관계들로 구성된 일종의 사전으로서 생각할 수 있으며, 그 속에는 특정 도메인에 관련된 단어들 이 계층적으로 표현되어 있고, 추가적으로 이를 확장할 수 있는 추론 규칙이 포함되어 있어, 웹 기반의 지식 처리나 응용 프로그램 사이의 지식 공유, 재사용 등이 가능토록 되어 있다. 온톨로지는 시맨틱 웹 응용의 가장 중심적 개념으로서, 이를 표현하기 위해 스키마와 구문 구조 등을



(그림 5) 서비스 제공 플랫폼

Time	Location	UserID	UserLevel	Service 1	Service 2	...
x	Room A-1	x	x	Printer 1 printing	Search location of print	...
x	Room A-1	x	Adm	Printer 10 printing	Search location of print	...
x	Room A-2	UserA	Adm	Printer 2 printing	Search location of print	...
x	Room A-3	UserB	x	Ramp Turn On	Computer Power On	...
⋮						

(그림 6) COSEP의 온톨로지 엔진에서 서비스 조합의 예



(그림 7) 서비스 제공 흐름도

정의한 언어가 온톨로지 언어(ontology language)이며, 현재 DSML+OIL, OWL, Ontolingun 등이 있다. 이와 같은 온톨로지를 이용하여 온톨로지 엔진은 이미 정의된 관계나 규칙을 통하여 서비스를 조합한다.

(그림 6)은 COSEP의 온톨로지 엔진에 구축되어 있는 몇 가지 서비스의 예를 보여준다. 서비스 소비자의 위치, ID 또는 등급에 따라 이용 가능한 서비스와 서비스 조합이 달라진다. 예로 만일 서비스 소비자의 위치가 방 A-1구역이라고 해도 소비자의 등급이 관리자리아만 프린터 10을 이용하여 프린터를 할 수 있고 다른 등급의 소비자들은 프린터 1만을 이용할 수 있다. 이렇게 미리 정의된 서비스 조합의 정보를 이용하여 서비스 소비자가 서비스를 요청할 때의 간단한 서비스 조합 예를 살펴본다(그림 7). 서비스 소비자 A는 현재 방 A-1구역에 위치하고 있고 등급은 관리자이다. 이 사용자가 COSEP에게 프린팅 서비스를 요청하면 COSEP은 사용자의 상황 정보를 이용하여 사용자에게 적당한 서비스를 제공하기 위해 서비스들을 탐사한다. 온톨로지 엔진의 서비스 조합을 이용하여 소비자 A가 이용할 수 있는 프린터 10과 프린트의 위치를 검색해주는 서비스를 제공한다. 이후 사용자는 프린터 10의 위치를 찾고, 그 프린터를 사용하기 위한 관련 프로그램을 설치한 후 프린터 10을 통해 프린팅 서비스를 수행한다.

이렇듯 온톨로지 엔진은 이미 정의된 관계나 규칙을 통하여 서비스 조합이 이루어지므로 미리 정의된 서비스 조합에 관한 정보가 없으면 사용자에게 적당한 서비스를 제공할 수 없는 단점이 있다. 그래서 향후 서비스의 추가나 새로운 서비스 조합을 위해서 새로운 관계와 규칙을 도출하는 데이터 마이닝을 적용한다.

4.2 데이터 마이닝 기법을 적용한 온톨로지 엔진

데이터 마이닝이란 일반적으로 “대량의 데이터로부터 새롭고 의미 있는 정보를 추출하여 의사결정에 활용하는 작업”이다[19]. 데이터 마이닝을 통해 도출되는 정보의 종류는 다양하나 본 논문에서는 연관규칙을 이용하였다. 연관 규칙은 마케팅에서 손님의 장바구니에 들어있는 품목간의 관계를 알아본다는 의미에서 장바구니 분석(market basket

analysis)이라고도 한다. 한 항목을 알 경우 다른 항목을 예측할 수 있다면, 이들의 관계는 종속관계가 존재한다. 데이터 안에 존재하는 항목간의 종속관계를 찾아내는 것이 연관 규칙을 발견하는 작업이다. 즉, 연관 규칙 마이닝은 거대한 데이터 아이템 집합 사이에 유용한 연관성이나 상관관계를 발견하는 것이다. 예를 들어 서비스 소비자 A가 저녁에 집에 오면 전등을 켜고 TV를 켜는 서비스를 계속하여 실행한다면 트랜잭션 DB에는 사용자 A의 사용이력이 저장된다. 이 서비스 소비자의 서비스 사용이력이 어느 정도 저장되면, 데이터 마이닝 엔진은 이 사용자의 서비스 사용 이력에 대한 분석을 한다. 전혀 연관이 없어 보이는 두 서비스가 연관 규칙을 가지고 있음을 발견하여, 그 결과로 전등을 켜는 서비스와 TV를 켜는 서비스를 조합하여 새로운 조합 서비스를 생성한다. 그러나 데이터 마이닝은 다량의 데이터에서 발생된 트랜잭션이 없으면 서비스를 조합 할 수 없는 단점이 있으나, 시간이 흘러 많은 수의 트랜잭션이 발생되면 더욱 정확하고 다양한 서비스 조합을 만들어낼 수 있다.

온톨로지 엔진과 데이터 마이닝 엔진이 독립적으로 작동했을 경우 장점과 단점이 있다. 온톨로지 엔진만을 사용하면 미리 정의된 관계와 규칙을 이용하여 서비스 조합을 함으로써 서비스 소비자에게 동적으로 서비스를 제공해 줄 수는 있지만 새롭게 요청되는 조합 서비스는 제공해 줄 수 없다. 반면 데이터 마이닝 엔진은 서비스 사용에 대한 트랜잭션들이 데이터 마이닝을 할 수 있는 정도의 크기가 되어야 새로운 조합 서비스에 대한 지식을 발견해 낼 수 있다. 두 엔진의 장단점을 결합하여 데이터 마이닝 기법을 적용한 온톨로지 엔진을 구축하면 동적으로 서비스를 제공할 뿐 아니라 사용자의 서비스 사용 이력에 대한 정보를 통해 새로운 조합 서비스도 얻어내어 사용자가 요청하는 서비스를 제공해 줄 수 있다. (그림 8)은 데이터 마이닝에서 Apriori 알고리즘[20]을 이용하여 서비스를 조합하는 알고리즘이다. (그림 8)에서 입력은 사용자의 사용이력을 저장한 트랜잭션 데이터베이스(TDB)와 최소 지지도(MSC: Min Support Count)이고, 출력은 조합된 서비스들(ComposedServices)이다. 서비스들은 트랜잭션 데이터베이스에 저장된 사용이력의 수가 최소 지지도 이상일 때 Apriori 알고리즘에 이용된다. 단계 2에서 트랜잭션 데이터베이스를 가져와 각 트랜잭션에 대해

```

Procedure Service Composition
Input: TDB, Transactional Database; MSC
Output: ComposedServices
1 initialize ComposedServices;
2 For each transaction t ∈ TDB
3   ComposedServices = apriori_gen(t, MSC);
   //generate composition service using Apriori Algorithm
   If ComposedServices is not in the OntologyDB then
   // check if they are existing composition services
5     add ComposedServices to OntologyDB;
6   End if
7 End for
    
```

(그림 8) Apriori 알고리즘을 이용한 서비스 조합

새로운 서비스 조합 규칙을 찾는 것을 시작한다. 단계 3에서는 Apriori 알고리즘을 이용하여 서비스들 간의 연관 규칙을 발견하여 새로운 조합 서비스를 생성하고, 단계 4~6에서는 조합된 서비스가 온톨로지 엔진에 존재하는지 아닌지를 확인한다. 새롭게 조합된 서비스가 온톨로지 엔진에 존재하지 않으면 온톨로지 엔진에 추가(단계 5)한다. 이 과정을 거쳐 새로운 조합 서비스가 추가되어 온톨로지 엔진이 재구축된다.

서비스 조합에 이용된 Apriori 알고리즘은 이진 연관 규칙을 위한 빈발 항목(frequent itemsets)을 마이닝하기 위해 자주 사용하는 알고리즘으로 다음 두 단계로 구성된다.

- 1) 조인 단계(The join step)
빈발 항목 집합 L_k 를 찾기 위해 후보 C_k 의 집합은 L_{k-1} 과 L_{k-1} 의 조인으로 생성된다.
- 2) 가지치기 단계(The prune step)
후보 C_k 의 $k-1$ 항목 부분집합이 L_{k-1} 에 속하지 않을 때 이를 모두 제거한다. L_k 는 C_k 에서 최소 지지도를 만족하지 못하는 항목들을 제거하여 L_k 를 생성한다.

Apriori 알고리즘의 간단한 예를 살펴보기 위해 다음 <표 1>과 같은 서비스 트랜잭션을 고려해 보자. 각 번호들은 서비스에 주어지는 특정 번호이다. 이때 전체 항목 집합 $I = \{101, 102, 103, 104, 105\}$ 이며, 지지도(support count)는 2로 가정한다.

<표 1> 트랜잭션 데이터

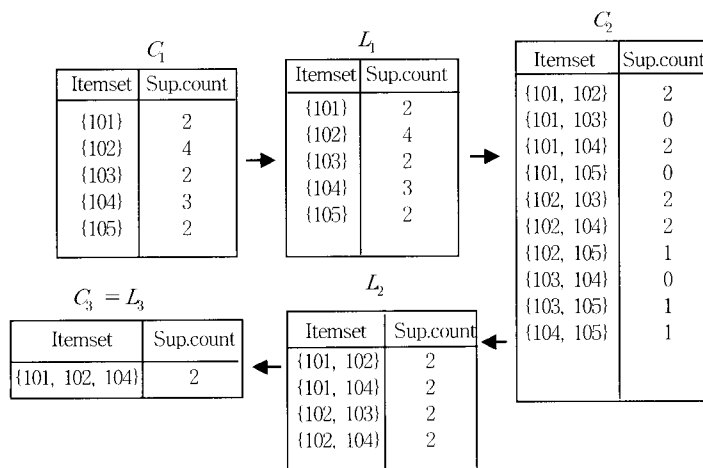
Transaction ID	Items
1	102, 103
2	101, 102, 104
3	102, 103, 105
4	101, 102, 104
5	104, 105

트랜잭션 데이터를 바탕으로 빈발 항목집합을 생성하면 (그림 9)와 같다. (그림 9)에서 C_1 은 1개의 아이템들로 생성이 되고, 모두 최소 지지도를 만족하므로 L_1 을 구성한다. 다음으로 C_2 가 L_1 의 조인으로 생성이 되고 이중 최소 지지도를 만족하지 못하는 항목집합들을 제거하여 L_2 를 만든다. C_3 도 L_2 의 조인으로 생성이 되는데 $\{101, 102, 103\}$ 의 부분집합 $\{101, 103\}$ 이 L_2 에 속하지 않으므로 C_3 에서 제거하였다. C_3 의 집합이 최소 지지도를 만족하므로 L_3 와 C_3 는 동일하다. 이렇게 Apriori 알고리즘을 통해 생성된 L_2 와 L_3 에서의 항목집합들이 새로운 서비스 조합 정보이며, (그림 10)의 네모안의 서비스 조합처럼 추가된다.

서비스 소비자 A와 B는 저녁 시간에 집에 들어와 이용하는 서비스가 다르다. 사용자 A는 집에 들어오면 항상 전등을 켜고 TV를 본다. 그러나 소비자 B는 전등을 켜 후 오디오를 켜고 음악을 듣는다. 그러면 데이터 마이닝 엔진은 각각의 사용자에 대한 정보와 서비스 사용에 대한 이력을 분석하여 사용자 A와 B에 맞는 서로 다른 새로운 조합 서비스를 찾아 온톨로지를 재구축한다. 즉, 서비스 사용이력을 이용한 새로운 서비스 조합뿐만 아니라 개인화된 서비스 사용에 대한 분석을 통하여 각각의 서비스 소비자에게 최적화된 서비스를 제공해 줄 수 있다. 또한 서비스 인스턴스 단

Time	Location	UserID	UserLevel	Service 1	Service 2	...
x	Room A - 1	x	x	Print1 printing	Search location of print	...
x	Room A - 1	x	Adm	Print10 printing	Search location of print	...
x	Room A - 2	UserA	Adm	Print2 printing	Search location of print	...
x	Room A - 3	UserB	x	Ramp Turn On	Computer Power On	...
pm 7:00	Room B	UserA	x	Ramp Turn On	TV Turn On	...
pm 7:00	Room B	UserB	x	Ramp Turn On	Radio Turn On	...
⋮						

(그림 10) 데이터 마이닝을 통해 추가된 서비스 조합



(그림 9) Apriori 알고리즘 예

위로 서비스 사용에 대한 분석이 가능하다. 만일 서비스 소비자가 이용 가능한 프린터들 중 하나의 특정 프린터 P1만을 이용한다면 데이터 마이닝 엔진은 이 사용 이력을 통하여 인스턴스 단위로 서비스를 조합한다.

데이터 마이닝 기법을 적용한 온톨로지 엔진은 서비스 사용자의 질의에 대해 실시간 서비스 조합은 온톨로지를 통하여 처리하고, 서비스 사용에 대한 트랜잭션들이 데이터 마이닝을 할 수 있는 만큼의 수가 되면 COSEP의 유희 시간에 데이터 마이닝을 통해 새로운 서비스 조합을 만들어 온톨로지 엔진을 재구축한다. 온톨로지 엔진에 데이터 마이닝 기법을 적용함으로써 동적 서비스 조합뿐만 아니라 다양한 서비스 조합을 통해 사용자에게 더 알맞은 서비스를 제공할 수 있다.

5. 실험 평가

이 장에서는 본 논문에서 제시한 데이터 마이닝 기법을 이용한 온톨로지 엔진의 성능평가를 위한 실험 환경을 기술하고 이론적인 실험 모델과 실험 결과를 통해 향상된 서비스 조합 성능을 분석한다. 본 실험은 Intel Pentium4 2.8GHz CPU, 1GB DDR RAM, 160GB HDD의 Windows XP sp2 환경에서 구현하여 실험하였다. 프로그램 개발은 J2SE Development Kit 1.4를 사용하였고, 개발 툴은 Eclipse 3.1을 사용하였다. 각 실험을 위해 생성한 데이터들과 중간 단계의 데이터를 저장하기 위하여 데이터베이스 Oracle9i를 이용하였으며, 서버는 Sun Blade 2000을 사용하였다. 데이터 마이닝을 이용한 서비스 조합에 대한 실험을 위하여 연관 규칙 데이터 마이닝 기법으로 많이 사용되는 Apriori 알고리즘 [20]을 구현하였다. Apriori 알고리즘은 생성된 연관 규칙에 대해 지지도와 신뢰도를 이용하여 이를 만족하는 연관규칙을 찾아낸다.

5.1 실험 데이터

5.1.1 BasicService DB

BasicService DB는 기본 서비스를 저장하는 데이터베이스로, 본 실험을 위해 생성한 기본 서비스의 수는 30개이고 101부터 130까지의 서비스 식별자를 가진다.

5.1.2 Ontology DB

Ontology DB는 고정적으로 서비스 조합을 할 수 있는 경우를 저장하는 데이터베이스이다. CreateOntology 클래스는 조합되는 서비스의 수를 2~5개로 선택하고 선택된 개수만큼 BasicService DB에서 데이터를 랜덤하게 가져와 하나의 서비스 조합을 만든다. 본 실험에서는 온톨로지에 의해 생성되는 조합 서비스의 수를 100개로 제한하였다. <표 2>은 CreateOntology 의해 생성된 일부 서비스 조합들을 나타내고 있다.

<표 2> 온톨로지의 서비스 조합

Ontology ID	BasicServices
1	106 113 127
2	117 119 129
3	108 112 114 115
4	113 120 128 129
5	101 129
6	101 121
7	116 121
8	129 130
9	103 105 113 117

5.1.3 Transaction DB

Transaction DB는 서비스 사용에 대한 트랜잭션을 저장하는 데이터베이스이다. 트랜잭션을 발생시키기 위하여 우선 동시적으로 발생하는 트랜잭션의 수를 2에서 5사이로 결정하고 선택된 수만큼 BasicService DB에 있는 기본 서비스들을 무작위로 가져온다. 이 때 기본 서비스 아이디 101~110까지의 발생 확률을 가장 높게 하고, 111~120까지를 중간 확률로, 121~130까지를 낮은 확률로 설정하여 트랜잭션을 발생시킨다. 특정 서비스들끼리 함께 나타나는 확률을 높게 하여 데이터 마이닝에 의하여 서비스 조합으로 생성될 수 있도록 하기 위해 발생 확률을 다르게 주었다. 본 실험을 위하여 생성된 트랜잭션의 수는 10000개이고 2000개 단위로 트랜잭션에 타임스탬프를 부여하였다. 즉 트랜잭션 아이디에 따른 타임스탬프는 다음 <표 3>과 같고 타임스탬프 1에서의 일부 트랜잭션들은 <표 4>와 같다.

<표 3> 트랜잭션 ID

트랜잭션 ID	Timestamp
1 ~ 2000	1
2001 ~ 4000	2
4001 ~ 6000	3
6001 ~ 8000	4
8001 ~ 10000	5

<표 4> 서비스 사용의 트랜잭션 데이터

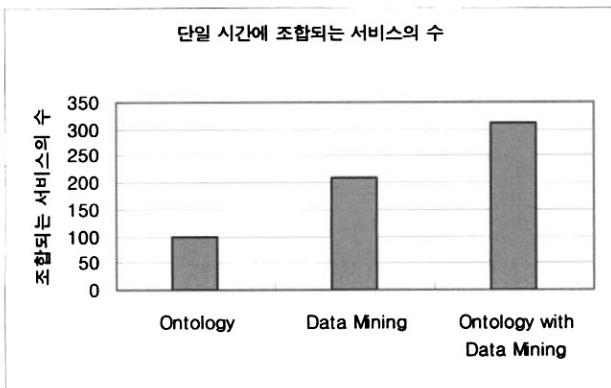
TransactionID	BasicServices	Timestamp
1	107 109 117	1
2	101 105	1
3	105 109	1
4	107 110 116 120 125	1
5	106 107 120	1
6	101 106 115	1
7	102 105 120	1
8	109 117 118	1
9	101 105	1

5.2 실험 결과

본 절에서는 온톨로지 엔진과 데이터 마이닝 엔진을 통해 조합되는 서비스의 수를 두 가지 경우에 대해서 비교해 본다. 첫 번째는 단일 시간에 조합되는 서비스의 수이다. 일정 시간 동안 온톨로지 엔진과 데이터 마이닝 엔진을 통해 조합되는 서비스의 수를 확인한다. 두 번째는 시간이 지남에 따라 트랜잭션의 수를 증가시켜 생성되는 서비스 조합의 수를 알아본다. 데이터 마이닝은 서비스 사용에 대한 이력 정보를 트랜잭션 형태로 보관하고 이를 분석하여 새로운 서비스 조합을 생성할 수 있으므로, 이 때 발생하는 서비스 조합에서 온톨로지에 의해 발생할 수 있는 서비스 조합은 제거하고 순수하게 데이터 마이닝을 통한 서비스 조합만을 추출하여 데이터 마이닝 엔진을 통해 조합된 서비스로 저장한다.

5.2.1 단일 시간에 조합되는 서비스의 수

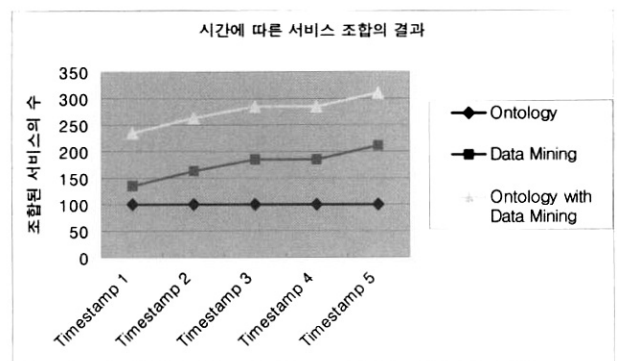
이 실험에서는 전체 트랜잭션(1000개)에서 발생될 수 있는 서비스 조합의 수를 알아보았다. 데이터 마이닝을 통해 발생된 서비스 조합 집합에서 각각의 서비스 조합이 Ontology DB에 있으면 삭제하여 순수하게 데이터 마이닝을 통해 발생한 서비스 조합만을 추출하였다. (그림 12)에서 온톨로지에 의해 추출될 수 있는 서비스 조합의 수는 100개이다. 이 결과는 초기에 Ontology DB에 미리 만들어둔 서비스 조합의 수와 동일하다. 데이터 마이닝을 통해 발생될 수 있는 서비스 조합의 수는 온톨로지에서 제공하는 조합 서비스를 제외하면 210개이다. 따라서 데이터 마이닝을 적용한 온톨로지를 통해 서비스 사용자에게 제공할 수 있는 서비스 조합의 수는 모두 310개가 된다. 이러한 실험 결과에서 온톨로지서 정의된 서비스 조합이 많다면 온톨로지의 서비스 조합의 수가 늘어날 수 있으므로 결과에서 보이는 두 기법 사이의 서비스 조합에 대한 차이는 중요하지 않다. 하지만 데이터 마이닝을 온톨로지에 적용함으로써 서비스 소비자에게 제공할 수 있는 서비스의 수가 증가하므로 두 기법을 혼합하여 사용하는 것이 서비스 조합에 좋은 결과를 보이는 것을 알 수 있다.



(그림 12) 단일 시간에 조합되는 서비스의 수

5.2.2 시간에 따라 조합되는 서비스의 수

이 실험은 각 타임스탬프 마다 트랜잭션의 수를 다르게 하여 데이터 마이닝을 적용해보았다. 각 타임스탬프마다 트랜잭션의 수가 2000개씩 추가로 발생하였다. (그림 13)의 실험 결과를 보면, 데이터 마이닝 기법은 시간이 지남에 따라 조합되는 서비스의 수가 증가하는 추세를 가진다. 실험 1과 마찬가지로 조합되는 서비스 집합의 원소들이 온톨로지에 의해 추출되는 서비스 집합에 존재하는 경우 삭제하여 순수하게 데이터 마이닝에 의해 추출되는 서비스 조합만을 나타내었다. 이러한 실험 결과를 보이는 것은 트랜잭션을 발생시킬 때 각 기본 서비스 식별자에 대한 발생 빈도를 다르게 하여 트랜잭션의 수가 많을수록 발생 빈도가 높은 서비스들의 출현수가 많아지기 때문이다. 즉 트랜잭션의 수가 많을수록 발생 빈도가 높은 기본 서비스를 포함하는 조합 서비스가 고정 변수 값으로 정해놓은 Min Support 값을 만족할 확률이 높아진다.



(그림 13) 시간에 따른 서비스 조합의 수

6. 결 론

유비쿼터스 컴퓨팅 환경에서 사용자가 원하는 서비스를 다양한 상황에 맞게 적합한 서비스 검색하고 조합하여 제공하는 것은 중요한 일이다. 그러나 기존 연구는 주로 서비스 발견에 관한 연구이며 사용자의 위치나 주변 환경, 선호도에 관한 고려가 부족하고, 사용자 정보를 이용하여 기본 서비스들로부터 새로운 조합 서비스를 찾아내는 것이 미흡하였다. 또한 서비스 조합에 있어 기존 연구는 단순한 기본 서비스들의 나열에 불과하였고, 사용 이력 고려에 대한 구체적인 방안을 제시하지 않았다. 따라서 본 논문에서는 사용자의 상황 정보 및 서비스 사용 이력을 고려한 데이터 마이닝 기반 서비스 조합 기법을 제안하였다. 세부적 연구내용은 첫째, 최적의 서비스를 동적으로 생성하여 제공하는 COSEP 프레임워크를 설계하고 둘째, COSEP 프레임워크에서 데이터 마이닝 기능을 겸비한 온톨로지를 이용한 서비스 조합 기법을 제안하였다. 실험을 통하여 데이터 마이닝 기법을 가진 온톨로지를 적용하여 서비스를 조합하는 것이 서

비스 소비자에게 더 많은 조합 서비스를 제공 할 수 있음을 확인하였다.

향후 우리는 본 연구를 바탕으로 조합된 서비스의 사용자 평가와 데이터마이닝 기법의 향상에 관하여 연구할 것이다.

참 고 문 헌

- [1] Soraya Kouadri Mostefaoui and Beat Hirsbrunner, "Context Aware Service Provisioning," Proceedings of The IEEE/ACS International Conference on Pervasive Services, ICPS 2004, pp.71~80, 19-23 July, 2004.
- [2] Dipanjan Chakraborty, Yelena Yesha and Anupam Joshi, "A Distributed Service Composition Protocol for Pervasive Environments," Wireless Communication and Networking Conference 2004. IEEE, Vol.4, pp.2575~2580, 2004.
- [3] SUN Microsystems, "Jini: Architectural Overview", Technical White Paper, 1999.
- [4] Microsoft Corporation, "Understanding Universal Plug and Play: a White Paper", Microsoft, June, 2000.
- [5] HAVi Consortium, "HAVi Specification V1.0," 2000.
- [6] E. Guttman, C. Perkins, J. Veizades, M. Day, "Service Location Protocol, Version 2," RFC2608, 1999.
- [7] Salutation Consortium, "Salutation Architecture Specification Version 2.1," Salutation, June, 1999.
- [8] Bluetooth Specification Part E. Service Discovery Protocol (SDP), <http://www.bluetooth.com>, 1999.
- [9] V. Sundramoorthy, Hans Scholten, Pierre Jansen and Piter Hartel, "Service Discovery at Home", Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia ICICS-PCM 2003, Vol.3, pp.1929~1933, 15-18 Dec. 2003.
- [10] Michiharu Takemoto, Tetsuya Oh-ishi, Tetsuya Iwata, Yoji Yamato and Yohei Tanaka, "A service-Composition and Service-Emergence Framework for Ubiquitous-Computing Environments," Proceedings of the 2004 International Symposium on Applications and the Internet Workshops (SAINTW'04), pp.313~318, 26-30, Jan., 2004.
- [11] Michiharu Takemoto, Hirosshi Sunaga, Kenichiro Tanaka, Hiroaki Matsumura and Eiji Shinohara, "The Ubiquitous Service-Oriented Network(USON): An Approach for a Ubiquitous World based on P2P Technology," Proceeding of the second International Conference on Peer-to-Peer Computing, IEEE, 2002.
- [12] Qun Ni, "Service Composition in Ontology enabled Service Oriented Architecture for Pervasive Computing," Workshop on Ubiquitous Computing and e-Research, May, 2005.
- [13] Keita Fujii and Tatsuya Suda, "Daynamic Service Composition Using Semantic Information," ICSOC'04, pp.15-19, November, 2004.
- [14] Keita Fujii and Tatsuya Suda, "Component Service Model With Semantics (CoSMoS): A New Component Model for Dynamic Service Composition," Proceedings of the 2004 International Symposium on Applications and the Onternet Workshops (SAINTW'04), 2004.
- [15] Antonio Corradi, Rebecca Montanari and Daniela Tibaldi, "Context-Based Access Control for Ubiquitous Service provisioning," Computer Software and Applications Conference, 2004, COMPSAC 2004, Vol.1, pp.444~451, 2004.
- [16] Choonhwa Lee and Sumi Helal, "Context Attributes: An Approach to Enable Context-Awareness for Service Discovery," Symposium on Applications and the Internet, 2003, pp.22~30, 27-31, Jan., 2003.
- [17] Samir Chamri-Doudane and Nazim Agoulmine, "Hierarchical Policy Based Management Architecture to Support the Deployment and the Discovery of Services in Ubiquitous Networks," 29th Annual IEEE International Conference on Local Computer Networks, 2004, LCN'04, pp.126~133, 16-18, Nov., 2004.
- [18] S. Maffioletti, S. kouadri Mostefaoui and B. Hirsbrunner, "Automatic Resource and Service Management for Ubiquitous Computing Environments," Proceedings of the Second IEEE Annual Conference on Pervasive Computing and Communications Workshops (PERCOMW'04), pp.219~223, 14-17, March, 2004.
- [19] 장남식의 2명, 데이터 마이닝(p19-47), 대청미디어, 1999.
- [20] Jiawei Han and Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2001.
- [21] 이선영, 신병철, 이종연, 배정숙, 신경철, "유비쿼터스 컴퓨팅 환경에서 서비스 프로비저닝 프레임워크", 한국정보과학회 제32회 추계학술발표회, Vol.32, No.2(1), pp.604~606, 2005년 11월 11일.
- [22] K. El-khatib, Zhen E. Zhang, N. Hadibi and G. v. Bochmann, "Personal and Service mobility in Ubiquitous computing environment," Wirel. Commun. Mob. Comput. Vol.4, pp.595~607, 2004.
- [23] Swaroop Kalasapur, Mohan Kumar and Behrooz Shirazi, "Personalized Service Composition for Ubiquitous Multimedia Delivery," World of Wireless Mobile and Multimedia Networks (WoWMoM'05), Sixth IEEE International Symposium, pp.258~263, 13-16, June, 2005.
- [24] Nikola Milanovic and Miroslaw Malek, "Current Solutions for Web Service Composition," IEEE Internet Computing, 2004, Vol.8, Issue 6, pp.51~59, Nov.-Dec., 2004.
- [25] Jose Andre Moura, Jose Manuel Oliveira, Eurico Caarapato and Renato Roque, "Service Provision & Resource Discovery in the VESPER VHE," IEEE International Conference on Communications, ICC 2002, Vol.4, pp.1991~1995, 28, April-2, May, 2002.
- [26] S. Kouadri Mostefaoui, Amine Trafat-Bouazid and B. Hirsbrunner, "Using Context Information for Service Discovery and Composition," Proceedings of the Fifth International Conference on Information Integration and Web-based Applications and Services, Jakarta, Indonesia, pp.129~138, 15-17, September, 2003.
- [27] Nishkam Ravi, Peter Sterm, Niket Desai, and Liviu Iftode,

“Accessing Ubiquitous Service Using Smart phones,” Third IEEE International Conference on Pervasive Computing and Communications, PerCom 2005, pp.383~393, 8-12, March, 2005.



이 선 영

e-mail : elesun@nate.com

2001년 충북대학교 전기공학과(학사)
2005년 충북대학교 전기공학과(공학석사)
2005년~현재 충북대학교 컴퓨터교육과
박사과정

관심분야: 데이터베이스, 유비쿼터스
컴퓨팅, Bioinformatics



이 종 연

e-mail : jongyun@chungbuk.ac.kr

1985년 충북대학교 전자계산기공학과
(공학사)

1987년 충북대학교 전자계산기공학과
(공학석사)

1999년 충북대학교 전자계산학과(이학박사)

1989년 비트컴퓨터(주) 개발부

1990년~1994년 현대전자산업(주) 소프트웨어연구소 주임연구원

1994년~1996년 현대정보기술(주) CIM사업부 책임연구원

1999년~2003년 삼척대학교 정보통신공학과 조교수

2003년~현재 충북대학교 컴퓨터교육과 부교수

관심분야: 질의 최적화, 시공간 데이터베이스, 데이터 마이닝,

Bioinformatics, Ubiquitous Computing, u-learning, GIS.