

# 맞춤형 국토정보의 효과적 제공을 위한 빈발 패턴 탐사 기법을 활용한 오픈맵 API 기반 프로토타입

(An Open Map API based-Prototype  
Utilizing Frequent Pattern Mining  
Technique for Efficient Service of  
Customized Land Information)

이 동 규\* 이 경 민\*\* 신 동 문\*\*\*  
(Dong Gyu Lee) (Gyeong Min Yi) (Dong Mun Shin)  
김 재 철\*\*\*\* 류 근 호\*\*\*\*\*  
(Jae-Chul Kim) (Keun Ho Ryu)

**요약** 유비쿼터스 도시 환경에서 사용자에게 맞춤형 국토정보를 제공하기 위한 공간 정보 시스템 개발이 대두되고 있다. 공간 정보 시스템은 사용자들 위해 언제 어디서든 공간 정보를 탐사할 수 있다. 또한, 다른 사용자들에게도 데이터마이닝 기술을 통해 분석된 정보를 제공할 수 있다. 따라서, 본 논문에서는 빈발 패턴 기법을 활용한 오픈맵 API 기반 프로토타입을 제안한다. 제안하는 프로토타입은 지오포토의 위치 데이터를 기반으로 사용자에게 흥미로운 여행 코스와 우리가 알지 못했던 명소를 탐사할 수 있다. 또한, 제안된 프로토타입은 20가지의 오픈맵 API 기반 지도 중 사용자에 의해 선택된 지도에 분석된 공간 패턴을 표현할 수 있는 최초의 시도이다. 우리의 프로토타입은 스마트폰 등의 모바일 단말기에서도 응용될 수 있다.

**키워드** : 빈발패턴탐사, 지리정보시스템, 맞춤형 국토정보, 공간 데이터 마이닝, 오픈맵API

\* 본 연구는 국토해양부 첨단도시기술개발사업 - 지능형국토정보기술혁신 사업과제의 연구비지원(과제번호:07국토정보C05)과 2009년 정부(교육과학기술부)의 지원(지역거점연구단육성사업/충북BIT연구중심대학육성사업단)에 의해 수행되었습니다.

\* 충북대학교 전자계산학과 박사과정, dglee@dblab.chungbuk.ac.kr

\*\* 충북대학교 전자계산학과 박사과정, min9709@dblab.chungbuk.ac.kr

\*\*\* 충북대학교 전자계산학과 석사과정, mastershin@dblab.chungbuk.ac.kr

\*\*\*\* 한국전자통신연구원 선임연구원, kimjc@etri.re.kr

\*\*\*\*\* 충북대학교 전자정보대학 컴퓨터공학부 교수, khryu@dblab.chungbuk.ac.kr (교신저자)

**Abstract** Spatial information systems have developed in order to provide users with customized land information in u-City environments. The spatial information systems can detect spatial information for users anytime anywhere. Information which is analyzed by data mining techniques can be offered for other users. Therefore, we propose open map API-based prototype which utilizes frequent pattern mining technique. Proposed prototype can mine interesting trip routes and unknown attractions in location data of geophoto. Also, proposed prototype is the first attempt which analyzes spatial patterns can be represented on a map which is selected by users. Our prototype can be applied to the smart phone like mobile devices.

**Keywords** : Frequent pattern mining, GIS, Customized land information, Spatial data mining, Open map API

## 1. 서론

현재 유비쿼터스 도시(Ubiquitous City) 환경을 기반으로 하여 국토정보의 효과적인 활용을 위해서 다양한 원천자료와 소프트웨어를 통합해 플랫폼이 필요하며 구글(Google), 마이크로소프트(Microsoft), 야후(Yahoo) 등의 대형업체들이 플랫폼을 기반으로 전 세계 위성사진과 지도 서비스를 선보이고 있다. 이들은 지도 위에 사용자들이 갱신한 사진, 동영상, 지역 정보 등의 다양한 형태의 맞춤형 국토정보(Customized Land Information)을 서비스 하며 영상 센서 등을 통해 지역의 동적인 정보들을 흥미롭게 제공한다. 하지만 아직까지는 사용자들의 데이터를 웹을 통해 제공해주는 기능이 주류를 이루며 분석을 통해 우리가 알지 못한 지식, 정보, 및 패턴들을 제공하는 기능은 미흡하다. 그러한 모든 데이터가 사용자에게 의미 있는 맞춤형 국토정보로 제공되지 않을 수 있기 때문이다. 따라서 축적된 대용량의 데이터를 분석하고, 필요한 의미 있는 정보만을 저장 및 제공할 수 있는 데이터 관리 기술이 요구 된다 [1, 2, 3, 4]. 유비쿼터스 국토정보 환경의 축적된 대용량의 국토정보를 분석하여 필요한 의미 있는 정보를 제공해 줄 수 있는 기법이 공간 데이터마이닝 기법이다.

공간 데이터 마이닝(spatial data mining)은 공간 데이터베이스에서 명백히 저장되지 않은 지식추출, 공간관계 및 다양한 흥미 있는 객체의 패턴을 추출하는 방법이다 [1, 2, 5]. 공간 데이터에서는 실제로 공간 객체들이 자주 상호 관련되어 있거나 두 객체들이 더 가까이 놓여 있을수록 그들이 유사한 속성을 공유할 확률이 높다는 관점에서 이것을 지리학의 첫 번째 법칙으로 간주한다. 즉, “모든 것은 그 밖의 모든 다른 객체와 관계가 있다. 그러나 가까이 있는 객체는 먼 객체보다 더 관계가 있다.”라는 의미이다 [1].

본 논문에서는 빈발 패턴 탐사 기법을 활용한 오픈맵 API 기반 프로토타입을 제안한다. 제안하는 프로토타입은 지오포토(geophoto)의 GPS 데이터를 기반으로 사용자에게 흥미로운 여행 코스와 우리가 알지 못했던 명소를 탐사하여 분석된 결과를 사용자가 원하는 지도에 표

현할 수 있다. 전체 빈발패턴 중 특정 빈발 패턴을 지도 상에서 찾아 볼 수 있는 검색 기능을 제공하고 특정 빈발 패턴만이 지도에 표현될 수 있다. 또한, 제한한 프로토타입은 20가지의 오픈맵 API 기반 지도 중 사용자에게 의해 선택된 지도에 분석된 공간 패턴을 표현할 수 있다.

## 2. 관련 연구

GeoMiner는 관계형 데이터마이닝인 DBMiner를 개발한 경험을 바탕으로 설계되고 개발되어졌다 [6]. GeoMiner의 데이터 마이닝 능력은 특정 규칙, 비교 규칙, 연관 규칙과 같은 규칙을 공간 데이터베이스에서 탐사하는 것이고 또한 분류 규칙과 군집화 규칙을 포함한다.

SAND(Spatial And Nonspatial Data) 아키텍처는 공간데이터베이스 모델링에서 활용되고 반면에 GeoMiner는 공간 데이터 마이닝을 위한 공간 데이터 큐브 구성 모듈(Spatial data cube construction module), 공간 온라인 분석 처리 모듈(spatial on-line analytical processing module, OLAP), 공간 데이터 마이닝 모듈(spatial data mining module)을 포함한다. 공간 데이터 마이닝 언어(spatial data mining language)인 GMQL(Geo-Mining Query Language)은 Spatial SQL를 확장하여 설계하고 구현했다. 또한 친화적인 그래픽 사용자 인터페이스(Graphic User Interface)를 가지며, 발견된 공간 지식을 그래픽화한다.

SDAM(Spatial Data Analysis and Modeling)은 분산 네트워크 환경에서 각 지역마다 처리 시스템을 두고 지역별 데이터를 처리하고 모델을 생성하여 분석한다 [7]. 분석된 결과는 중앙 그래픽 사용자 인터페이스를 통해 사용자에게 제공되며 만약 전체 지역의 분석이 필요한 경우 한 곳의 시스템에 지역별 데이터 또는 모델을 전송하여 중앙에서 처리하는 구조를 보여준다. 또한, 하나의 코어 안에서 동작하는 소프트웨어를 보여준다. 공간 및 비공간 속성을 포함한 다양한 데이터베이스로부터 제공되는 데이터를 전처리하고 모델링하여 분석한 결과를 제공하며 전체 지역의 분석이 필요한 경우 모델을 통합하는 연산을 수행하여 결과를 그래픽 사용자 인터페이스를 통해 사용자에게 제공하는 과정이다.

스핀 프로젝트는 최신 지리정보시스템 기술과 데이터 마이닝 기능성을 통합하므로써 웹기반 공간 데이터마이닝 시스템을 개발하는 것이다 [8]. 엔터프라이즈 자바 빈즈(Enterprise Java Beans)를 사용하여 사용자 접속 권한, 다중 사용자 접속, 연결 및 캐싱 공유, 영속성 유지 및 트랜잭션 관리를 엔터프라이즈 자바 빈즈 컨테이너(EJB container)에서 수행하는 장점이 있다. 지역별로 속성 값을 색깔로 나타내며, C4.5 분류 모델을 이용하여 지식을 탐사한다.

GeoDMA는 이미지와 로컬 데이터베이스의 더 정확한 분석을 위해 그래픽화와 데이터 저장 기능을 테라뷰(TerraView)에 의해서 제공된다 [9]. C4.5 알고리즘[10]과 SOM(Self-organizing map) 알고리즘[11]으로 구현되

었고, 사용자에게 의해 선택된 입력 속성들을 고려하여 패턴 분류를 수행한다. 이 시스템은 입력은 이미지와 공간영역(region)으로 모든 공간 어플리케이션에서 사용될 수 있다. GeoDMA는 사진과 공간영역 데이터의 비교를 통해 데이터를 획득하고 획득된 데이터를 기반으로 분류 모델을 생성하여 새로운 사진의 변화된 위치의 형태(마을, 강, 도로, 밀립 등)를 예측할 수 있다. 사진과 공간영역 데이터 획득하여 마이닝을 적용하는 새로운 기술을 제시하였다.

## 3. 빈발패턴 탐사 기법

본 논문에서는 오픈맵 API 기반 프로토타입에 적용한 데이터마이닝 기법으로 빈발패턴 탐사 알고리즘을 적용한다. 빈발패턴 탐사 기법은 연관규칙 마이닝 분야의 가장 핵심적인 부분으로 상호관계(correlations), 인과관계(Causality), 순차패턴(Sequential patterns), 에피소드(episodes), 다차원패턴(multi-dimensional patterns), 최대패턴(max-patterns), 부분주기성(partial periodicity), 신흥패턴(emerging patterns) 등의 많은 연구가 활발히 진행되고 있다 [12, 13, 14]. 많은 빈발패턴 탐사 알고리즘들 중 가장 대표적인 알고리즘이 FP-Growth이며, FP-Growth의 자료구조인 FP-Tree의 성능을 개선한 CFP-Tree[15]를 빈발패턴 탐사 알고리즘에 적용한다. CFP-Tree는 밀집 및 희소 빈발 항목 집합 모두 효율적으로 탐색할 수 있는 자료구조이다. CFP-Tree는 빈발패턴 탐사를 위해 데이터 정제기(Data Refiner), 트리 구성기(Tree Builder), 그리고 패턴 탐사기(Pattern Miner)로 구성된다. 데이터 정제기 단계는 대용량 데이터베이스에서 데이터를 읽어드려 정제하는 과정을 나타낸다. 트리 구성기 단계는 노드 삽입, 분할, 평가를 수행하여 트리를 구성하는 과정이다. 패턴 탐사기 단계는 패턴들을 추출하는 과정이며, 위 3단계를 보여주는 것이 그림 1이다.

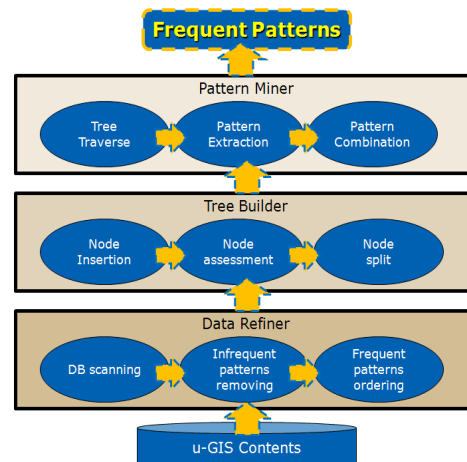


그림 1. CFP-Tree의 빈발패턴 탐사 과정

데이터 정제기 단계는 대용량 데이터베이스로부터 데이터를 읽어드려 정제하는 과정을 보여주며, 데이터베이스 스캔(DB Scanning), 빈발 패턴 제거(Infrequent Patterns removing), 빈발 패턴 정렬(Frequent Patterns Ordering)의 3단계로 구성된다. 데이터베이스 스캔 단계는 대용량 데이터베이스로부터 위치 데이터를 읽어드려 항목 단위로 빈발횟수를 계산한다. 빈발패턴 제거 단계는 데이터베이스 스캔 단계에서 최소지도를 만족하지 못하는 항목들을 제거한다. 빈발패턴 정렬 단계에서는 빈발패턴제거 단계에서 빈발하는 많은 항목들이 제거된 항목들을 빈발한 순으로 재 정렬하여 트리의 헤더 테이블(header-table)을 구성한다.

트리구성기 단계는 트리를 구성하는 과정으로써 노드 삽입(node insertion), 노드 평가(node assessment), 노드 분할(node split)로 구성된다. 노드 삽입 단계는 대용량 데이터베이스에서 데이터를 읽어드려 데이터정제기 단계에서 구성한 헤더테이블을 이용하여 빈발한 순으로 재 정렬된 항목들을 노드로 구성하여 트리에 삽입한다. 노드 평가 단계는 트리를 구성하는 과정 중에 노드의 분할 발생 유무를 판별하고 노드 분할 단계에서는 노드를 분할하고 링크들을 재설정하게 된다.

패턴 탐사기는 패턴들을 추출하는 과정으로써 트리 운행(Tree Traverse), 패턴 추출(Pattern Extraction), 그리고 패턴조합(Pattern Combination)의 3단계로 구성된다. 트리 운행 단계는 트리구성기로부터 구성된 트리를 순회하여 패턴 경로를 추출한다. 이렇게 추출된 패턴 경로가 단일 경로라면 패턴을 추출하고 단일 경로가 아니라면 재귀적으로 재 탐사과정을 수행하는데 이러한 과정은 패턴 추출 단계에서 이루어진다. 또한, 패턴 조합 단계에서는 재귀적으로 탐색한 단일경로들의 패턴을 조합하여 다중경로의 패턴을 추출한다. 탐사된 빈발 패턴들은 최소지도를 만족하는 모든 패턴들로 구성되며, 하나의 패턴부터 N개의 패턴까지 탐색이 가능하다.

#### 4. 오픈맵 API 기반 프로토타입

본 장에서는 3장에서 설명한 CFPF-tree 구조를 적용한 오픈맵 API 기반 프로토타입의 특징을 설명하고 지오태깅(geotagging)된 지오포토(geophoto)의 위치 정보를 통해 미등록 명소를 찾아주는 시나리오를 설명한다.

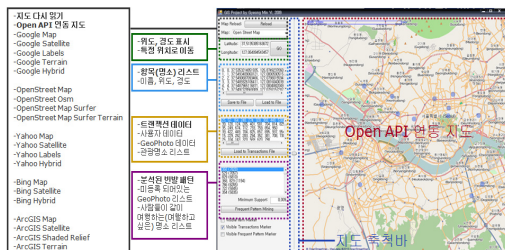


그림 2. 오픈맵 API 기반 프로토타입 인터페이스

개발된 프로토타입은 오픈소스인 GMap.NET을 이용하여 구현되었다 [16]. 기본적인 실행화면 구성은 그림 2이고 좌측은 빈발패턴 탐사 메뉴들로 구성되어 있으며 우측은 오픈맵 API와 연동된 지도 인터페이스로 구성된다. 메뉴와 지도 인터페이스 사이의 축척 바를 통하여 지도의 축척을 사용자가 보기 편한 축척으로 조정할 수 있다. 빈발패턴 탐사 관련 메뉴들은 개발된 프로토타입의 여러 기능들을 편리하게 이용할 수 있도록 구성되어 있다. 먼저 지도를 다시 로드하는 기능은 프로토타입이 오픈맵 API를 이용하기 때문에 네트워크의 사정에 따라 지도가 원활하게 표현되지 않는 경우 지도를 다시 로드(load)할 수 있게 해준다. 지도를 선택하는 기능은 프로토타입과 연동된 총 20가지의 오픈맵 API를 사용자 요구에 맞게 지도를 선택하여 사용한다. 입력 데이터로는 지도 상의 지오포토의 위치 좌표와 기존 관광 명소의 위치 좌표이며, 위도 및 경도의 값을 표시해준다. 지오포토는 최근 모바일 장치에 GPS 수신기가 장착되어 사진을 찍은 위치의 GPS 좌표 값을 저장하고 있는 사진이다.

표 1. 오픈맵 API 지도

제공자	지도 종류	설명
Goggle	Map	거리지도
	Satellite	위성사진지도
	Labels	명칭지도
	Terrain	지형도
	Hybrid	거리지도+위성사진지도
Open Street Map	Map	거리지도
	OSM	사용자 참여거리지도
	Map Surfer	거리지도
	Map Surfer Terrain	지형도
Yahoo	Map	거리지도
	Satellite	위성사진지도
	Labels	명칭지도
	Hybrid	거리지도+위성사진지도
Bing	Map	거리지도
	Satellite	위성사진지도
	Hybrid	거리지도+위성사진지도
ArcGIS	Map	거리지도
	Satellite	위성사진지도
	Shaded Relief	지형도
	Terrain	지형도

표 1은 제안한 프로토타입에 적용된 5개 기업의 지도들을 보여주고 있다. 거리 지도 서비스는 실제 거리에서의 주변 정보 및 건물 모형을 확인할 수 있는 기능이 제공된다. 위성 사진 서비스는 60cm급 해상도를 가진 위성 사진을 기반으로 주요 시설 및 거리를 표시하였으며, 사용자가 가보지 않고도 실제 거리의 모습과 각 종 시설의 위치를 확인할 수 있는 장점이 있다. 명칭지도는 거리 및

행정구역에 대한 명칭을 표시하며, 지형지도는 색상 및 명암, 등고선을 이용하여 지형의 형태를 표현한다. 하이브리드 지도 서비스는 배경을 위성 사진을 사용하며, 위성 사진 위에 거리 지도를 포개어서 행정구역 및 주요 시설의 명칭을 표시하여 실제 거리와 건물들의 모습을 동시에 볼 수 있다.

빈발패턴 마이닝을 위해 지오포토의 위치 좌표를 로딩 (loading)하여 최소 지지도를 만족하는 빈발 패턴을 탐사하게 된다. 그림 3은 최소지지도 0.9%로 설정하며, 100,000개의 트랜잭션 데이터로부터 빈발한 순서대로 패턴을 나열해준다. 패턴은 공간식별자(빈발횟수) 형식으로 표현한다.

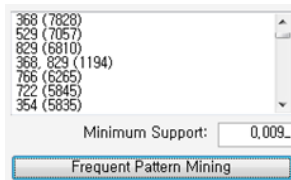


그림 3. 빈발패턴 표현

공간식별자는 지오포토의 위치 값을 가리키며, 빈발횟수는 데이터베이스로부터 해당 공간식별자의 빈발한 횟수를 계산하여 높은 값을 가지는 공간식별자를 상위에 나타낸다. 예를 들면, 1-패턴인 368 지역에 사용자의 지오포토가 7828번 빈발했다는 의미는 사용자들 368 지역에서 적은 지오포토를 많이 등록했고 사용자들에게는 의미 있는 장소라고 분석할 수 있다. 또한 2-패턴인 368, 829 패턴의 경우는 사용자들이 두 지역을 함께 여행하며 적은 지오포토를 등록한 패턴이 1194번 빈발하여 두 지역을 동시에 여행하는 것이 빈발하다는 의미이다. 이와 같이 구현한 프로토타입에서는 어떤 지역에 많은 지오포토들이 등록되어 있다면 우리는 그 지역이 빈발하고 사용자들이 흥미로워하는 여행지역이라고 정의한다.

탐색된 빈발 패턴은 그림 4와 같이 지도상에 붉은색 마커로 표시된다. 우리는 사용자의 입력 데이터인 지오포토의 위치를 하늘색 마커로 표시하고, 1,000개의 임의의 장소를 등록된 명소로써 녹색 마커로 표시한다. 녹색과 붉은색 마커가 가까운 거리에 위치하면 알려진 명소에 여행자들이 빈번하게 방문하여 사진을 찍고 흥미로운 명소로 정의할 수 있으며, 반대로 붉은색 마커 주변에 녹색 마커가 존재하지 않은 지역은 기존에 알려지지 않은 명소로써 여행객이 자주 방문하고 사진을 찍는 지역으로 정의할 수 있다.

이러한 패턴들은 기존의 빈발패턴을 탐사하는 데이터 마이닝 도구로는 의미 해석이 어려우며 오픈맵 API를 연동하여 탐색된 빈발 패턴을 직관적으로 식별할 수 있어 가능하다. 빈발패턴은 각각의 빈발횟수에 따라 적색의 농도를 다르게 설정한다. 즉, 빈발도가 높을수록 적색의 농도가 진하고 낮을수록 농도가 희리다. 또한, 빈발패턴 탐사 기법을 적용하여 단순 통계적 수치 계산으로 찾을 수

없는 N개의 지역이 하나의 여행 코스로써 탐사되거나 미등록 명소를 찾을 수 있다는데 의미가 있다.

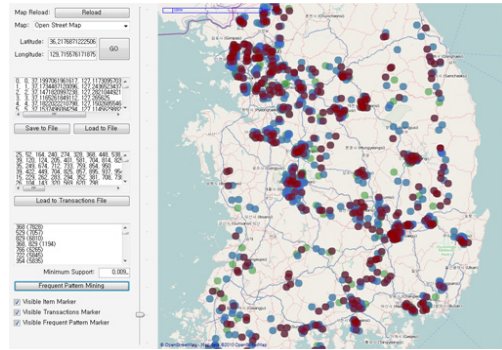


그림 4. 미등록 명소 찾기

### 5. 결론

본 논문에서는 빈발패턴 탐사 기법을 이용하여 지오포토의 위치 데이터를 기반으로 사용자에게 흥미로운 여행 코스와 우리가 기존에 알지 못했던 여행 장소를 탐사할 수 있는 웹 기반 프로토타입을 개발하였다. 또한, 데이터 마이닝 분야에서 마이닝 결과의 표현 방식은 그래픽적으로 직관적일수록 좋다. 방대한 양의 결과물 또한 전문가의 해석이 요구되므로 시간과 노력이 상당히 많이 소요되기 때문이다. 우리의 프로토타입이 20가지의 지도상에 사용자 요구에 맞게 표현할 수 있는 것은 분석 결과의 단순 서술 또는 도표를 뛰어넘는 효과적인 표현이라고 할 수 있다.

### 참고 문헌

- [1] J. Han, M. Kamber and J. Pei, "Data Mining: Concepts and Techniques Second Edition", Morgan Kaufmann, 2005.
- [2] V. Kumar, M. Steinbach and P. Tan, "Introduction to Data Mining," Addison- Wesley, 2005.
- [3] 박성승, 손호선, 이동규, 지은미, 김희석, 류근호, "u-City응용에서의 시간 패턴을 이용한 단기 전력 부하 예측", 한국공간정보시스템학회 논문지, 제11권 제2호, 2009, pp. 177-181.
- [4] 서성보, 이용미, 이준욱, 남광우, "RFID 데이터 스트림에서 이동패턴 패턴의 탐사", 한국공간정보시스템학회 논문지, 제11권, 제1호, 2009, pp. 127-136.
- [5] 배덕호, 백지행, 오현교, 송주원, 김상욱, 최명희, 조현주, "공간 데이터 마이닝 시스템의 설계 및 구현", 한국공간정보시스템학회 논문지, 제11권 제2호, 2009, pp. 119-132.
- [6] J. Han, K. Koperski and N. Stefanovic, "GeoMiner: a system prototype for spatial data

- mining”, In Proc. ACM SIGMOD Int’l conference on Management of data, 1997, pp. 553-556, USA.
- [7] A. Lazarevic, T. Fiez and Z. Obradovic, “A Software System for Spatial Data Analysis and Modeling, In Proc. Hawaii Int’l Conference on System Sciences, 2000, USA.
- [8] M. May and A. Savinov, “An architecture for the SPIN! spatial data mining platform”, In Proc. the New Techniques and Technologies for Statistics, 2001, pp. 467-472, UK.
- [9] T. S. Korting, L. Fonseca, M. Escada, F. Silva and M. Silva, “GeoDMA - A novel system for spatial data mining”, In Proc. IEEE Int’l Conference on Data Mining Workshops, 2008, pp. 975-978, Italy.
- [10] J. R. Quinlan, “C4.5: Programs for Machine Learning”, Morgan-Kaufmann Publishers, 1993.
- [11] T. Kohonen, “Self-Organizing Maps Third Edition,” Springer Series in Information Sciences, Vol. 30, 2001.
- [12] R. Agrawal, T. Imielinski, and A. Swami, “Mining association rules between sets of items in large databases,” In Proc. ACM-SIGMOD Int’l Conference Management of Data, 1993, pp. 207-216, USA.
- [13] R. Agrawal, and R. Srikant, “Fast algorithm for mining association rules in large databases,” In Research Report RJ 9839, 1994, USA.
- [14] J. Han, H. Cheng, D. Xin and X. Yan, “Frequent pattern mining: current status and future directions”, In Proc. Data Mining and Knowledge Discovery, 2007, pp. 55-86, Netherlands.
- [15] 이경민, 정석호, 신동문, 이브라힘, 이동규, 손교용, “효율적인 밀집 및 희소 빈발 항목 집합 탐색 방법”, 제31회 한국정보처리학회 춘계학술대회 논문집, 2009, pp 375-376.
- [16] The MIT License (MIT), “GMap.NET”, <http://greatmaps.codeplex.com/>