

# 시맨틱 GIS 시스템을 위한 Geo-온톨로지 Population 모델 설계

송원용\*, 김장원\*\*, 정동원\*

\*군산대학교 정보통계학과

\*\*고려대학교 컴퓨터학과

e-mail: {swy0111,djeong}@kunsan.ac.kr

ikaros1223@gmail.com

## A design of Geo-Ontology Population Model for Semantic GIS

Wonyong Song\*, Jangwon Kim\*\*, Dongwon Jeong\*

\*Dept. of Informatics & Statistics, Kunsan National University

\*\*Dept. of Computer & Radio Communication Engineering, Korea University

### 요 약

인터넷이 등장하고 대중화되면서 웹 1.0의 개념에서 웹 2.0을 지나 웹 3.0을 바라보고 있다. 이러한 웹의 발전 과정에서 시맨틱 웹이 등장하게 되었고 시맨틱 웹이 GIS와 접목되어 시맨틱 GIS가 등장하게 되었다. 기존 GIS 시스템에서는 사용자들이 국한된 정보만을 제공받아 왔지만 웹이 발전하고 수많은 콘텐츠가 생성되면서 웹의 데이터를 GIS와 연결시켜서 사용자에게 풍부한 정보를 제공 할 수 있는 시스템이 연구, 개발되고 있다. 하지만 기존 연구들은 Learning 및 Population 기법 부재로 인하여 시맨틱 웹에서 지향하는 풍부한 서비스 제공에 어려움을 겪고 있다. 따라서 이 논문에서는 Learning 및 Population 기법을 통해 웹으로부터 비공간 정보를 추출하여 서비스 할 수 있는 시맨틱 GIS 시스템 모델을 설계한다. 이 구현 모델의 기대효과로 사용자는 구현 될 시스템을 사용하여 특정 공간에 대한 비공간 정보들을 웹으로부터 수집하여 사용자에게 풍부한 정보를 제공할 수 있다.

키워드 : 시맨틱 웹, 시맨틱 GIS, Learning 및 Population

### 1. 서론

인터넷이 등장하고 대중화되면서 웹 1.0의 개념에서 시작하여 현재에는 웹 2.0을 지나 웹 3.0을 바라보고 있다. 웹 1.0 환경은 기존의 오프라인 환경과는 달리 온라인 환경에서 등록된 정보를 사용자가 사용하는 콘텐츠 중심이었다. 이러한 콘텐츠 중심의 환경에서 사용자들은 참여와 공유, 개방이

라는 기본 정신을 바탕으로 사용자 스스로 정보를 생성하고 웹에 등록하는 웹 2.0 환경으로 발전 하였다[1]. 웹 2.0 환경에서는 사용자들의 인터랙션(Interaction)이 중요한 키워드로 부각되면서, 다양한 서비스를 제공하기 위해 근간이 되는 데이터가 가장 중요한 요소로 되었다. 또한 사용자들의 적극적인 참여를 기반으로 하여 계속적으로 진화하는 Wiki, Blog, Social Network Service와 같이 스스로 진화해 나가는 데이터를 제공할 수 있는 환경들이 웹 2.0 환경의 중요한 요소로 대두되었다. 이와 같은 환경 속에서 웹 1.0 기반의 기존 웹 서비스들도 진화하여, 데이터 중심의 시맨틱 웹 서비스

\* 2008년 정부(교육과학기술부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구(KRF-2008-314-D00485)이며 또한 이 연구에 참여한 연구자는 'BK 21 2단계 사업'의 지원을 받았음.

(Semantic Web Service)가 등장하였다.

시맨틱 웹이란 컴퓨터가 웹에 존재하는 모든 데이터들의 의미를 이해하고 조작하여 사용자에게 서비스 할 수 있으며, 그런 자원에 대해 사용자들도 이해할 수 있는 환경을 말한다.

이렇듯 웹에 존재하는 데이터 및 서비스가 방대해 지고 지능화 되자 웹을 기반으로 한 GIS(Geographic Information System) 분야도 점차 지능적으로 발전하고 있다. GIS의 발전 과정을 살펴보면 기존 지리정보시스템에서는 전문가가 비공간 정보들을 GIS 시스템에 저장하여 국한된 정보만을 사용자에게 제공해왔다. 하지만 최근 연구, 개발되고 있는 GIS는 일반 사용자 중심으로 시맨틱 웹과 융합되어 시맨틱 GIS라는 새로운 서비스가 개발되고 있다[2]. 시맨틱 GIS를 실현하기 위해 현재 시맨틱 웹 서비스의 선두 기업인 Google과 Microsoft의 경우 웹 기반의 GIS를 바탕으로 Wiki나 YouTube에 있는 콘텐츠를 제공함으로써 시맨틱 GIS를 구현하고 있다[2]. 또한 매쉬업(Mash-up)을 통하여 웹 콘텐츠의 활용을 강화 할 수 있는 환경을 제공하고 있다[2].

따라서 이 논문에서 제안하는 모델은 기존의 GIS에서 제공하던 제한된 비공간 정보를 웹과 연동하여 Wiki나 Blog등에서 풍부한 비공간 정보를 추출하여 제공할 수 있는 시맨틱 GIS 모델을 설계한다.

## II. 관련 연구

이 장에서는 Geo-온톨로지 시스템을 위해 필요한 기본 개념에 대해 기술한다.

### 1. 시맨틱 GIS

시맨틱 GIS는 시맨틱 웹과 GIS가 접목된 지능적인 GIS 시스템을 말한다. 웹 2.0 환경에서 제공하는 수많은 정보들을 이용하여 지도상에 정보를 표현 한다[3-5]. 예를 들어, 특정 건물에 대한 상가나 음식점에 대한 정보를 건물의 위치와 이름을 통해 Wiki, Blog, Cafe, Social Network에서 비공간 정보들을 검색해 사용자에게 보여줌으로써 사용자는 특정 건물에 대한 위치와 특성, 속성 등을 쉽게 이해할 수 있다.

- KML : 구글어스(Google Earth), 구글맵(Google Maps), 모바일 구글맵(Google Maps for mobile) 등 어스 브라우저에 지리 정보를 표현 하는데 사용되는 파일 포맷이다. KML은 웹 브라우저가 HTML(혹은 XML)을 처리하는 방식과 유사한 방식으로 처리된다. KML은 화면표시 목적에 맞는 이름과 속성이 있는 태그 기반의 구조를 가지고 있다. 즉, 구글어스와 구글맵은 KML 파일에 대한 브라우저 역할을 하게 된다[9].
- GML:GML (Geography MarkupLanguage)은 XML 특성을 이용하여 지리정보를 교환하거나 공유하기 위해 개발된 언어이다. GML은 단순한 교환이나

공유의 목적 외에도 지리정보의 모델링 및 저장을 위한 XML 스키마로 작성된 XML 문법이다[10].

### 2. 웹 크롤러(Web Crawler)

웹 크롤러는 URL을 지정하여 그 URL의 인덱스 페이지에 접속한다. 접속한 페이지를 분석하여 그 페이지의 하이퍼링크를 검색하여 링크 URL에 접속한다[4]. 방문했던 URL을 판단하여 중복된 URL일 경우 다음 URL로 접속한다. 이러한 방법으로 계속 웹 페이지에 접속한다[7].

### 3. RSS

RSS는 뉴스, 블로그, 실시간 검색어와 같은 수시로 업데이트 되는 성격의 사이트를 위한 XML 기반의 포맷으로 자료 교환을 하기 위한 규격이다[5]. 이것은 'RDF Site Summary' 또는 'Really Simple Syndication'의 약자로 콘텐츠 수집과 신디케이트를 구성하기 위해 사용하는 도구이다. 이것을 이용하여 자신의 홈페이지나, 블로그 등에 원하는 정보가 있는 블로그 사이트의 파일을 계속적으로 수집하여 그 사이트의 업데이트된 콘텐츠를 자동화된 과정에 의해 자신의 사이트에서 사용할 수 있다.

### 4. 블로그(Blog)

블로그는 최근 네티즌들에게 자기만의 공간을 만들기 위해 사용하는 1인 미디어로 각광받고 있다. 블로그는 웹(Web)과 일지(Lob)의 합성어로 인터넷상에서 개인이 운영하는 일지 또는 일지 형태로 기록되는 형식을 말한다[11]. 블로그에 관한 용어에는 블로거(B.oger)와 포스트(Post)가 있다. 블로거는 블로그를 이용하는 모든 유저들을 뜻한다. 하지만 일반적으로 블로그를 운영하는 사람을 말한다. 포스트는 엔트리라고도 불리우며 블로그에 작성된 글, 문서, 게시물, 사진, 자료 등 모두를 의미한다. 블로그의 특징은 최근 정보를 가장 빠르고 정확하게 제공한다는 것이다. 그리고 HTML 형식으로 작성하여 FTP를 통해 올려야 하는 번거로움을 거치지 않고 인터넷에 바로 업데이트 할 수 있고 작성된 포스트는 페이지에 접속하는 것만으로 포스트의 내용을 확인할 수 있다는 특징 때문에 가장 빠른 미디어로 각광받고 있다.

## III. 제안 모델

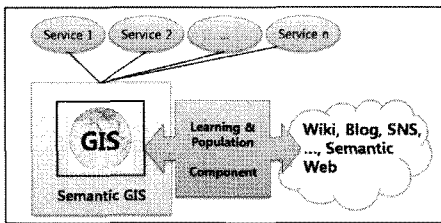
이 장에서는 논문에서 제안하는 Geo-온톨로지 시스템의 전체적인 프레임워크 구조와 각각의 컴포넌트와 역할에 대해 기술한다.

### 1. 개념 모델

〈그림 1〉은 시맨틱 GIS를 위한 Geo-온톨로지 Population 제안 모델의 개념도이다. 제안 모델은 사용자에게 다양한 비

공간 정보를 웹으로부터 추출하여 지리정보시스템과 연동함으로써 기존의 지리정보 시스템보다 많은 비공간 정보를 제공할 수 있고 추출된 비공간 정보를 기반으로 다양한 서비스를 제공한다. 예를 들어, 특정 공간에 대한 비공간 정보로 맛집이나 숙박시설 등과 같은 사용자가 방문하지 않았지만 맛집이나 숙박업소를 다녀온 손님들로부터 작성된 웹 게시판이나 블로그에 있는 정보를 비공간 정보로 제공함으로써 사용자는 맛있는 맛집, 깨끗한 숙박업소를 찾을 수 있을 것이다.

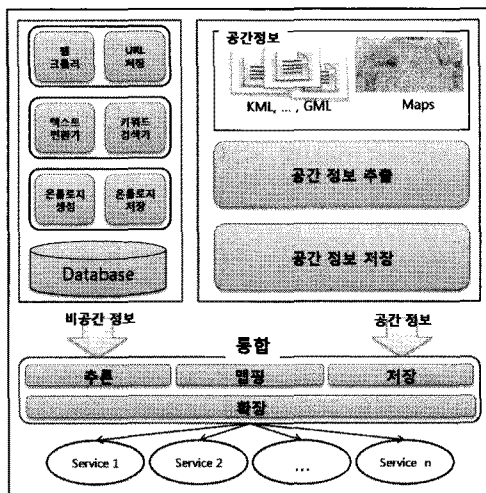
제안 모델의 핵심요소로 웹 자원의 비공간 정보를 추출하기 위해 Learning 및 Population 기술이 사용되는데 이 기술은 계속적으로 데이터를 수집하는 Learning 기술과 수집된 데이터를 온톨로지(Ontology)에 추가하는 Population 기술로 나뉜다.



〈그림 1〉 제안 모델의 개념도

## 2. 제안 모델 구조

〈그림 2〉는 제안 모델을 위한 시스템 구조를 보여준다.



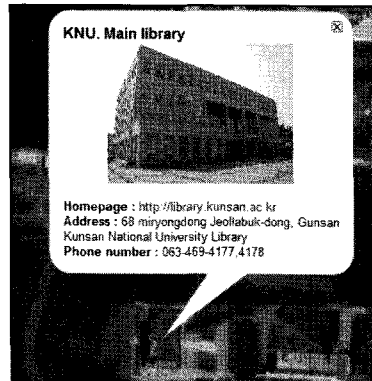
〈그림 2〉 제안 모델의 구조도

비공간 정보를 추출하기 위해서는 웹 크롤러, URL 저장, 텍스트 변환 컴포넌트, 키워드 검색 컴포넌트, 온톨로지 생성 컴포넌트, 온톨로지 저장 컴포넌트 그리고 데이터베이스로 구성된다. 먼저, 웹 크롤러는 웹에 있는 자원을 활용하기 위해 웹 페이지의 URL을 추출한다. 그리고 추출된 페이지들의

하이퍼링크 URL에 접속하며 URL을 추출한다. 추출된 URL을 계속 적재해나가며 수집된 웹 페이지들을 텍스트 변환 컴포넌트로 변환하고 변환된 텍스트를 검색 컴포넌트를 통해 찾고자하는 정보가 있는지 검색한다. 검색된 정보를 기반으로 각 컨셉(Concept)를 정의하고 Geo-온톨로지를 생성한다. Geo-온톨로지 생성 후 각각의 인스턴스를 데이터베이스에 저장한다.

공간 정보를 추출하기 위해서는 공간정보를 표현하는 KML, GML 등과 같은 언어와 지도, 공간정보 추출 컴포넌트, 공간정보 저장 컴포넌트로 구성된다.

Geo-온톨로지 확장을 위해 비공간 정보와 공간정보를 추출하여 추론과 맵핑, 저장을 통해 확장한다. 추론은 여러 가지 비공간 정보 간의 관계를 파악하여 사용자에게 원하는 정보를 사상시킨다. 예를 들어, 〈그림 3〉와 같이 비공간 정보 공간정보 위에 사상시켜 군산대학교 도서관의 정보를 표현할 수도 있다. 앞의 예와 같이 통합된 정보들을 기반으로 사용자에게 질 높은 서비스를 제공할 수 있다.



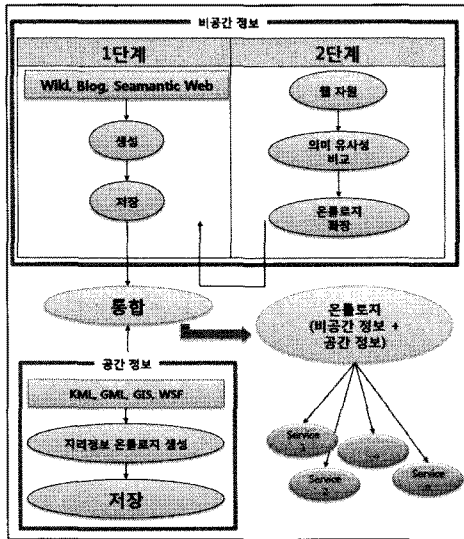
〈그림 3〉 공간 정보와 비공간 정보 사상 예제

## 3. 제안 모델 프로세스

〈그림 4〉는 제안 모델의 주요 프로세스를 보여준다. 비공간 정보를 위한 프로세스는 크게 2단계로 구성된다. 1단계에서는 Wiki, Blog와 같은 웹으로부터 비공간 정보들을 RSS나 웹 크롤러를 이용하여 생성하고 최초 Geo-온톨로지에 저장한다. 2단계에서는 기존의 Geo-온톨로지의 정보와 추출된 웹 자원들과의 의미적 유사성을 비교하여 점차 온톨로지를 확장한다.

공간정보는 KML, GML, GIS, WSF등에서 DOM이나 SAX 파서를 이용하여 추출하고 추출된 데이터를 기반으로 온톨로지를 생성하고 RDB에 저장한다.

생성된 비공간 정보의 온톨로지와 공간 정보의 온톨로지를 통합하여 풍부해진 Geo-온톨로지를 생성한다.



〈그림 4〉 제안 모델의 프로세스

#### IV. 결론 및 향후 연구

이 논문에서는 웹에 존재하는 데이터 및 서비스를 활용하여 지리정보시스템에 비공간 정보로 활용할 수 있는 Geo-온톨로지 확장 모델에 대한 시스템을 제안하였다.

향후 연구 과제로는 제안한 모델을 구현하여 실험 하고 웹에서 검색한 비공간 정보들의 신뢰성 검증을 통하여 정확한 데이터를 제공해야 할 것이다.

#### 참고문헌

- [1] T. O'Reilly, "What Is Web 2.0: DesignPatterns and Business Models for the Next Generation of Software", Sep. 2005.
- [2] 김은형, 김경옥, "맞춤형 국토정보 제공 서비스 모델 연구", 인터넷정보학회지 Vol.9, No.1, 2008.3.
- [3] A. Mikroyannidis, "Toward a Social Semantic Web," IEEE Computer, Vol. 40, No. 11, pp. 113-115, Nov. 2007.
- [4] Leonid Stoimenov, Slobodanka Djordjevic-Kajan "Framework for Semantic GIS Interoperability", Ser. Math. Inform. 2002.
- [5] Tanasescu, V. and Domingue, J., "Toward User Oriented Semantic Geographical Information Systems", 2nd AKT Doctoral Symposium, 2006.

- [6] Erik Hatcher, Otis Gospodnetic, "Lucene in Action", Manning Publications Co. Greenwich, CT, USA, 2004.
- [7] 김원, 김희철, 진용옥, "URL 분석을 위한 웹 로봇 구현 및 성능분석", 한국통신학회논문지 Vol.27 No.3C, 2002.3.
- [8] 윤현주, 김영민 외 2명, "규칙기반 온톨로지 자동생성 및 검색", 한국정보과학회 추계 학술발표논문집 Vol.31, No.2, 2004.
- [9] OGC, Keyhole Markup Language, <http://www.opengeospatial.org/standards/kml>.
- [10] 정동원, 김장원 외 2명, "GML 문서의 유효성 및 독립성을 고려한 지리공간 데이터 가시화 시스템 설계 및 구현", 한국IT서비스학회지, 2008.
- [11] 이영주, "블로그에 관한 연구", 학술발표대회논문집, 2004. 5.