

공간 시멘틱 웹을 위한 텍스트 공간정보의 위치 맵핑 기법

Location Mapping Techniques of Textual Spatial Information for Spatial Semantic Web⁺

하태석* · 이수욱** · 남광우*

Tae Seok Ha · Su Wook Ha · Kwang Woo Nam

*군산대학교 컴퓨터정보공학과, **한국전자통신연구원 u-공간연구팀

{hts00, kwnam}@kunsan.ac.kr, **suwook.ha@etri.re.kr

요 약

웹에서 다양한 웹 지리 지역 정보를 검색할 수 있는 시스템에 대한 요구가 증가하고 있다. 그러나 현재의 웹 검색 시스템은 사용자가 키워드로 지역 웹 문서를 검색하고 해당 웹 문서를 지도와 비교하여 공간정보를 취득하며, 다른 관련 정보를 얻기 위해서는 검색과 비교를 반복해야 하는 어려움이 있다. 따라서 본 논문에서는 비구조화 된 텍스트 웹 자원으로부터 지리정보 온톨로지(geo-ontology)를 확장할 수 있는 통합된 검색시스템을 제안한다. 이를 위해 문서의 정보에서 위치 정보를 추출하고 공간정보 위치 맵핑 기법을 적용하여 텍스트의 공간정보를 추출한다.

1. 서론

GeoWeb의 대표적인 예로, Google의 Google Maps나 Mircsort의 Virtual Earth 등이 있다. 하지만 이러한 서비스는 웹 문서를 지도와 검색 비교하여 공간정보를 취득해야 하는 단점이 있다.

웹에서 지리 지역정보를 검색하는 수단으로 GIS와 정보검색시스템을 동시에 이용하고 있다. GIS에서 사용되는 데이터는 지형공간에 관한 모든 정보, 즉, 지리정보가 정형화된 데이터인 반면, 정보검색시스템에서는 텍스트, 이미지, 동영상과 같은 다양한 종류의 데이터를 이용하고 있다.

웹을 위치 기반의 지리정보공간으로 바라보는 GeoWeb으로 접근하는 연구가 진행되고 있다. GeoWeb은 웹정보에서 질적인 속성을 추출하여 현실 공간의 특정한 위치로 정보를 축적하고, 사용자에게 위치를 기반으로 정보를 쉽고 편하게 활용

할 수 있도록 구조화 된 공간이다[1].

이종의 시스템을 이용한 정보 검색은 사용자에게 결과 비교, 추가적인 키워드 입력등 번거로움을 준다. 사용자가 웹에서 지역정보를 수집하고, GIS를 이용해 공간정보를 검색한 후에 이종의 정보를 비교 분석하고, 이를 통합하여 취득하는 문제가 있다.

본 논문에서는 이러한 이종의 시스템을 통한 검색 과정을 단순화하기 위해 웹 지리 지역정보에 포함된 지리공간단어를 활용하여 “공간 시멘틱 웹을 위한 텍스트 공간정보 위치 맵핑 기법”을 정의 하고, 이 모델을 기반으로 웹 정보와 공간정보가 유기적으로 통합된 검색 시스템을 제안한다.

본 논문의 구성은 다음과 같다. 제2장에서는 관련연구에 대해 기술하고, 제3장에서는 공간정보 위치 맵핑 기법의 구성과 특징을 제4장에서는 프로토타입 구현

⁺ 본 연구는 국토해양부 첨단도시기술개발사업-지능형국토정보기술혁신사업과제의 연구비 지원(07국토정보C05)과 2008년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2008-0061967)

및 평가에 대해 기술 한다. 제5장에서는 결론 및 향후 연구에 대하여 기술 한다.

2. 관련 연구

2.1 형태소 분석

웹의 발전과 함께 정보의 양은 점점 더 방대해지고 있다. 웹페이지를 대상으로 한 단순한 검색에서부터, 다양한 유형의 웹페이지를 대상으로 사회적 이슈의 판별, 이슈에 대한 사용자 의견 분석 등 웹에 존재하는 정보를 유용하게 사용하고자 하는 노력이 계속되고 있다. 이를 위해서는 어느 정도의 자연어 처리가 필요한데, 형태소 분석은 가장 기초적인 자연어 처리 과정 중 하나이다[2]. 한국어는 교착어의 특성으로 인하여 조사가 부착되거나 어미가 활용되면서 단어의 형태가 다양하게 변화한다. 이로 인해, 원래 의미를 가지는 단어를 추적하고, 원형을 복원하는 과정이 필요한데, 이는 영어에서의 단순한 스템밍(Stemming)이나 품사 부착(POS Tagging: Part Of Speech Tagging)에 비하여 훨씬 복잡하고, 전문적인 통사론이나 문법 지식을 요구하는 어려운 과정이다. 하지만 꼬꼬마 형태소 분석을 통한 보타 쉽게 텍스트 정보를 추출할 수 있었다. 꼬꼬마 형태소를 통해 분석된 단어는 다음과 같은 표로 작성 된다. 추출된 정보는 크게 체언, 용언, 관형사등으로 나누어지고 각 형태마다 품사로 단어를 선택할 수 있다.

3. 공간정보 위치 매핑 기법

3.1 지리정보 온톨로지 정의

변화하는 웹 환경하에서, 현재 제공되고 있는 국내의 지리정보 시스템의 문제점은 비공간 정보들 간의 관계가 정의되어 있지 않아 컴퓨터가 이해할 수 있는 환경이 제공되지 않는다는 점이다. 이는 보다 다양하고 정확한 서비스 제공을 어렵게 한다. 이 논문에서는 이러한 문제를

해결하기 위하여 지리정보 온톨로지를 정의함으로써 정확한 정보를 제공할 수 있다.

3.2 지리정보 온톨로지 확장

웹 자원 즉, 블로그나 게시판 Wikipedia 등과 같은 시맨틱 웹 서비스에서 비공간 정보를 추출하여 온톨로지의 인스턴스를 확장함으로써 풍부한 비공간 정보를 제공할 수 있다. 지리정보 온톨로지를 생성하고 키워드를 입력하여 웹으로부터 자원을 추출한다. 추출된 자원은 형태소 분석 처리 과정을 통해 최종적으로 지리정보 온톨로지의 인스턴스로 추가하게 된다.

3.3 위치 매핑 구조

형태소 분석기를 통해 얻어진 품사로 현재 사용 중인 우편번호 구조로 행정명, 우편번호 데이터를 가진다. 용언의 처리 속도 향상을 위해 메모리 trie 구조를 사용한다. 프로토타입에서는 행정명, 우편번호 데이터만을 지역정보로 간주하고 있어 향후 개선이 필요하다. 발전 모델은 다음과 같은 구조를 가진다[3].

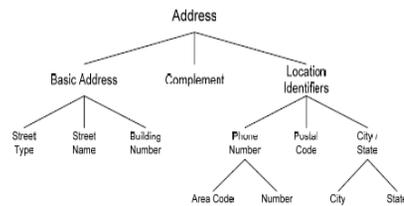


그림 1. 지리정보 온톨로지 구조

확장된 지리정보 온톨로지는 유선전화, 지역 코드, POI, 도시 정보를 보완해줄 수 있는 관광명소 구조를 가진다.

4. 구현 및 평가

이 장에서는 맛집을 대상으로 한 공간정보 위치 매핑위한 프로토타입을 구현한다.

4.1 구현

이 절에서는 공간정보 위치 매핑에 대한 프로토타입을 기술한다. 구현에 앞서, 지도는 구글의 Google Maps를 사용하고, 비공간 정보 검색 API는 네이버의 Open API를 사용한다.

pp.31-36.

[4] <http://maps.google.co.kr/>

[5] <http://dev.naver.com/openapi/>



그림 2. 구현 화면

그림 2는 구현된 시스템의 화면으로 비공간 정보인 맛집 뉴스의 비공간 정보를 확장한 것을 보여준다.

5. 결론 및 향후 연구

이 논문에서는 현재 웹에서 제공하는 지리정보 시스템의 문제점을 정의하고, 비 구조화된 웹 자원을 통한 지리정보 온톨로지의 매핑기법을 제안하였다. 또한 제안한 방법을 통해 지리정보 온톨로지를 매핑한 예를 보였다.

향후 연구로는 주소구조의 확장과 보다 정확한 형태소의 정보 제공을 위한 연구가 요구된다.

참고문헌

- [1] 양단희·김연수, 2008, “IS의 진화 : Geospatial Web & u-GIS” 인터넷 정보학회지 제9권 제1호, pp.44-55.
- [2] <http://kkma.snu.ac.kr/>
- [3] Karla A.V. Borges, Alberto H.F. Laender, Claudia B. Medeiros, Clodoveu A. Davis Jr, 2007, “iscovering Geographic Location in Web Pages Using Urban Addresses” Proceedings of the 2007 workshop on Geographic information retrieval,