

온톨로지 기반의 주소체계 의미 불일치 해결방법

김명희*, 김장원**, 이석훈**, 백두권**

*고려대학교 컴퓨터정보통신대학원 소프트웨어공학과

**고려대학교 컴퓨터전파통신공학과

e-mail : kenzo143@korea.ac.kr,

ikaros1223@gmail.com, brandonlee@korea.ac.kr, baikdk@korea.ac.kr

An Ontology-based Method for Solving Semantic Heterogeneity of Address System

Myeonghee Kim*, Jangwon Kim**, Sukhoon Lee**, Dookwon Baik**

*Dept. of Software Engineering, Computer Information and Communications, Korea University

**Dept. of Computer Science and Radio Communications Engineering, Korea University

요 약

우리나라 주소는 우편번호를 기준으로 한 주소체계, 법정동을 기준으로 한 주소체계, 행정동을 기준으로 한 주소체계, KT 에서 관리하는 KT 주소체계 등 주소관리가 통합되어 있지 않고, 각 기관, 기업별로 주소 관리를 별도로 하고 있다. 이에 따라 기관과 기업들간의 상이한 주소체계 관리문제로 인해 서로간의 주소체계의 의미가 일치하지 않는 문제가 발생하여 다양한 문제가 유발되고 있다. 본 논문에서는 이러한 문제점들을 해결하기 위해 주소 명칭에 대한 보다 자동화된 맵핑이 가능하도록 KT 주소, 우편번호주소, 행정동/법정동코드를 대상으로 온톨로지를 사용한 모델을 제안하였다. 제안된 모델은 주소정보간의 의미관계성을 유지하여 이기종 주소체계의 의미 불일치를 해결할 수 있으며, 보다 자동화된 데이터관리와 장기적인 비용절감이 가능하다.

1. 서론

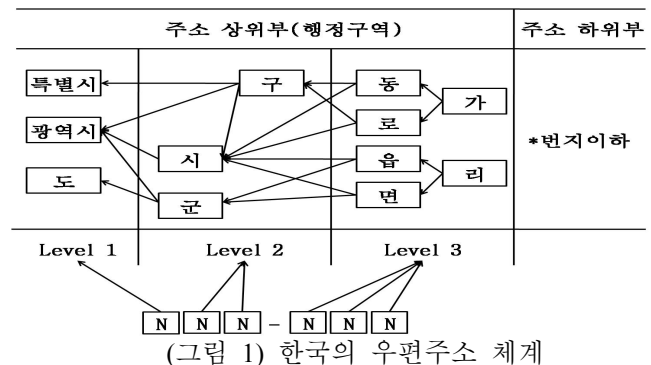
우리나라는 크게 우정사업본부에서 관리하는 우편번호 주소체계, 행정안전부에서 관리하는 토지 거래 및 등기를 위한 법 집행상에 필요한 법정동 주소체계, 행정업무 수행을 위한 행정동 주소체계, KT 에서 KT 자체의 동코드를 부여하여 관리하고 있는 KT 주소체계 등의 다양한 주소체계를 사용하고 있다. 각 기관, 기업별로 각기 다른 주소체계를 사용하고 있기 때문에, 동일한 지역을 가리키는 주소임에도 불구하고 서로 다른 명칭으로 사용하는 문제, 기존 주소와 새 주소체계와의 맵핑 문제, 기관간의 시스템통합이 이루어지는 경우 주소통합의 문제, 행정구역이 변경되는 경우 기존에 사용중인 주소명칭간의 매칭 문제들이 계속해서 발생하고 있다[1][2]. 기존의 접근방법인 주소단위의 접미사와 동의어 맵핑 방법은 주소정보의 확장성, 상호운용성, 활용성 측면에서 제한적이다. 현재 발생되고 있는 문제점들과 기존연구의 한계를 극복하기 위해 온톨로지를 기반으로 한 주소체계간의 의미의 불일치를 해결하는 방안을 제안한다.

2. 관련 연구

2.1 기존 연구 분석

현재 우리나라 주소는 지번을 기반으로 지번 주소

체계와 도로 명을 기반으로 한 새 주소체계를 병행하여 사용하고 있지만, 현재 사용중인 우편번호체계는 지번주소 체계를 기반으로 하고 있다[3]. 주소체계를 표준화하는 방안으로 기존 연구에서 지번 체계를 중심으로 우편물에 다양하게 표기된 주소들을 표준화하여 주소들을 매칭한 모델이 있다[4]. 지번중심으로 사용중인 우리나라 우편주소 체계는 그림 1 과 같다[5].



우편물에 표기된 우편주소는 표 1 과 같은 특징들이 있어 우편물의 순로 구분 작업을 자동화하기 위한 주소해석 과정이 필요하며, 다양한 표현 형태를 수용하여 데이터를 저장하여 검색해야 한다. 기존 연구에서는 계층적 행정 구역의 일관성 있는 표현, 번지가 포함된 주 배달 점과 번지 내에서 세분화된 배달 점의 표현, 주소 표기의 변형 흡수를 위한 키워드와 동

이 연구에 참여한 연구자는 ' 2 단계 BK21 사업' 의 지원을 받았음.

의어 표현의 개념을 가지고 총 5 개의 테이블을 구성하며, 각각의 역할은 표 2와 같다.

<표 1> 우편번호 주소의 다양한 특징과 예

특징	예(표준형)	예(변형)
① 계층적 행정구역 체계	서울특별시 종로구 을지로 3가	서울 종로 을지 3가
② 법정지명, 행정지명의 세분화 및 대표지역 명칭 사용	인후동 1가, 인후동 2가	인후 1동, 인후 2동, 인후 3동, 인후동
③ 택지 개발 또는 공단 지역의 가 지번 형태의 주소 존재	남동공단 37블록 49로트	남동 공단 37B49L
④ 사서함의 이용	대전유성우체국 사서함 106호	대전 유성 PO Box 106
⑤ 건물, 아파트 등의 표기 형태가 다양	목동 7단지, 한빛타운아파트	목동@, 한빛 Apt.
⑥ 건물, 아파트 등의 동, 호, 수 등의 세부 주소 형태가 다양	710동 1234호	710-1234
⑦ 번지 표기 형태가 다양	256번지 1호, 256-1	256-1번지, 256의 1, 산 4-1
⑧ 동일 지점에 대한 유사 명칭 사용	한국과학기술원	과기원

<표 2> 우편주소파일을 구성하는 테이블 설명

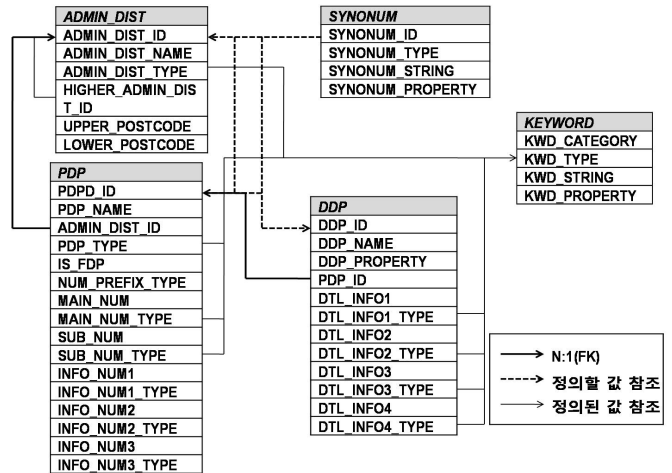
테이블명	설명 및 기능
ADMIN_DIST	계층적인 행정구역(Administrative District)을 표현하기 위한 테이블
PDP	번지 단위로 부여된 주 배달점(Primary Delivery Point)을 표현하기 위한 테이블
DDP	번지만으로 표현되지 않는 상세 배달점(Detailed Delivery Point)을 표현하기 위한 테이블
KEYWORD	주소에서 사용되는 키워드(Keyword)를 정의한 테이블
SYNONYM	주소에서 사용된 동의어(Synonym) 또는 유사어를 정의한 테이블

행정구역(ADMIN_DIST), 주 배달 점(PDP), 상세 배달 점(DDP), 키워드(KEYWORD), 동의어(SYNONYM)의 5 개로 구성된 우편주소파일의 테이블간의 관계를 나타내면 그림 2와 같다. 이 테이블들간의 관계를 기반으로 주소검색을 할 경우, 먼저 우편번호 인식을 시도하고, 그 결과에 따라 주소의 상위부, 즉 우편번호에 해당하는 행정구역명까지 인식한 다음에는 번지와 건물명 등의 주소의 하위 부를 해석하고 최종 결과를 출력한다. 기존 연구방법은 우편번호를 기준으로 한 빠른 검색과 정확한 맵핑 결과를 얻을 수 있는 장점이 있으나, 정부 및 기업들의 주소검색 및 우편번호검색 서비스에 다양한 검색방법을 제공할 수 있다고 제안한 점에 대해서는 한계점이 있다. 실제로 기업, 공공기관은 우편번호와 표 3과 같이 별도의 서로 다른 주소코드관리시스템을 사용하고 있다[6]. 그러므로, 명칭을 이용한 기존 연구인 맵핑 방법으로는 주소코드간의 의미 불일치를 해결할 수 없다.

2.2 온톨로지 모델

주소코드간의 의미 불일치를 해결하기 위해 의미 기술, 즉 온톨로지를 이용하여 서로 다른 주소체계에

포함되어 있는 의미를 정의하고, 정의된 관계성을 이용하여 보다 자동화된 맵핑이 가능하도록 한다. 온톨로지(Ontology)는 사물이 “존재한다”(그리스어의 On)라는 것의 의미를 다루는 존재론적 철학을 의미하는 낱말이다. 인공지능 및 지식표현 분야에서는 “대상세계에 존재하는 것들의 체계적인 분류와 그들 사이의 관계를 명시적/형식적으로 기술한 것”을 온톨로지라고 부르고 있다[7]. 즉 해당 영역의 개념들과 이들 개념 간의 상호 관계를 정의하는 것을 온톨로지라고 한다[8]. 온톨로지는 클래스와 속성의 성질, 관계를 보다 상세하게 표현할 뿐만 아니라, 다양한 커뮤니티가 독자적으로 정의한 어휘를 상호 관련시키거나 어플리케이션이 모순을 발견해내는 “발전성”에 관한 기능을 제공한다[9].



(그림 2) 우편주소파일 테이블간의 관계도

<표 3> KT 동코드와 행정안전부코드 비교 예

지역	동명	행정안전부		
		KT 동코드	법정동코드	행정동코드
용인시 기흥구	신갈동	364860	41463101	41463510
	하갈동	364863	41463104	41463515
	영덕동	364870	41463111	

2.3 Geonames 모델

Geonames 프로젝트는 지명, 기관 명 등 명칭에 대한 온톨로지를 구축하여 의미의 이질성을 해결하고자 하고 지도 서비스와 연계된 서비스를 제공하기 위한 프로젝트이다[10]. 동일한 의미를 가지는 명칭에 대해 다양한 언어, 다양하게 불리는 명칭들을 사용자들이 직접 수정하거나, 추가할 수 있고, WGS84(World Geodetic System 1984)에 있는 좌표 값들을 이용하여 지도 서비스를 제공한다[11]. Geonames 에서 제공하는 여러 가지 웹 서비스 중에서 findNearestAddress 서비스는 US 의 우편번호와 주소 값을 매치하여 데이터베이스화하여 우편번호나 지명의 주소를 입력하면 해당 위치를 지도에 표시해준다. Geonames 의 클래스는 크게 공간적인 것에 관한 클래스(SpatialThing), 주소코드에 관한 클래스(Code), 명칭의 개념에 관한 클래스(ConceptScheme), 웹문서(Document)에 관한 클래스로 나뉜다[12]. SpatialThing 클래스는 국가코드, 인구수, 우편번호, 명칭, 변경된 명칭들의 정보에 관한 클래스

그림 5 는 표 4 에서 표현된 온톨로지를 Jena 를 통해 트리플(Triple) 구조 형태로 데이터베이스에 저장한 결과를 보여준다.

5. 평가

제안된 온톨로지를 구현한 후, 다음 표 5 의 항목들을 기반으로 제안모델을 평가하였다.

먼저, 제안모델은 정보간의 의미 관계성을 정의할 수 있다. 사업자 별로 주소코드체계가 다르기 때문에 기존 모델은 주소코드간의 의미를 정의할 수 없으나, 제안된 모델은 주소코드간의 관계성을 정의하여 주소체계 의미의 불일치를 해결 할 수 있다.

두 번째로 제안된 모델의 장점은 정보의 획득 범위이다. 기존 모델에서는 한 사업자의 데이터가 변경될 경우, 시스템 관리자가 사업자로 부터 변경 데이터를 받거나 웹에서 찾아서 검증하여 별도로 관리하는 독자적인 시스템에 의존한다. 따라서 각 사업자들은 서로의 데이터를 공유할 수 없다. 제안 모델은 사용자 누구나도 변경된 데이터를 등록할 수 있어 사업자들이 서로의 정보를 공유할 수 있는 장점이 있다[15].

세 번째로 제안모델은 시스템 유지 비용을 낮출 수 있다. 데이터를 추가하여 매치할 경우, 간단한 수식으로 비용을 평가하면, 표 6 에서 보는 바와 같이 기존 모델의 비용보다 제안모델의 시스템 유지 비용이 낮다. 표 6 에서는 두 개의 사업자의 주소코드를 대상으로 적용하였지만, 사업자가 늘어날수록 더욱 높은 비용을 예측할 수 있다.

마지막으로 제안모델의 장점으로 데이터의 자동화된 이력관리를 할 수 있다. 기존 모델은 데이터가 변경될 경우, 별도의 이력데이터를 만들어야 하는 수동적인 이력관리를 요구하지만, 제안모델은 데이터가 변경될 때마 자동으로 이력관리가 되어 추후에 주소변경내역을 쉽게 알 수 있다.

<표 5> 제안 모델의 정성적 평가표

평가항목	기존 모델	제안된 모델
정보간 의미 관계성	관계성 없음	관계성 유지
정보의 획득 범위	제한적	제한 없음
시스템 유지 비용	고비용	저비용
데이터의 이력관리	수동적	자동화

<표 6> 비용 평가표

비용항목	기존모델의 비용	제안모델의 비용
데이터 추가비용	T	T
데이터 매치비용	m * n * T	T
데이터 저장비용	T	T
기타 비용	C	C
비용 합계	=m * n * T + 2T + C	=3T + C

6. 결론

이 논문에서는 현재 기업, 정부기관간의 서로 다른 주소체계로 인한 문제점을 정의하고, Geonames 의 온

톨로지 스키마를 활용하여 우리나라의 주소체계 의미 불일치를 해결하기 위한 모델을 제안하였다. 또한, 제안한 모델의 온톨로지를 구현하고, 구축된 온톨로지의 장점을 중심으로 한 평가에 대하여 기술하였다. 제안 모델의 평가의 결과를 통하여 제안모델은 주소정보간의 의미관계성을 유지하여 이 기종간의 주소체계의 의미 불일치를 해결하고, 주소정보의 이력관리를 할 수 있다. 비용측면에서 온톨로지 시스템 구축은 초기비용이 높으나, 주소정보는 변경사항이 자주 있지 않기 때문에, 평가의 결과를 통해 시스템 유지 비용을 줄일 수 있어 장기적인 관점에서 비용절감효과를 볼 수 있다.

참고문헌

- [1] 장태우, “주소정보와 위치정보의 통합 및 활용 방안 연구”, 한국전자거래학회지, 제 15 권 제 2 호, 2010.5
- [2] 최동훈, “온톨로지 기반의 이질적 정보의 의미적 통합”, 한국정보과학회 데이터베이스 소사이어티, 데이터베이스연구 제 20 권 제 4 호, 2004.12
- [3] 박정현, 김호연, 박종홍 “세계 주소 및 우편번호 체계와 행정구역”, ETRI, 2009.9.
- [4] 장태우, 김호연, 임길택 “주소해석 및 검색을 위한 우편주소과일 설계”, 한국산업정보과학회논문지, 제 12 권, 제 4 호, 2007. 12.
- [5] Jiyeong Lee, Hag-Yeol Kim, “A Geocoding Method Implemented for Hierarchical Areal Addressing System in Korea”, The Journal of GIS Association of Korea, Vol.14, No.4, pp.403-419, December 2006
- [6] 행정안전부, <http://www.mopas.go.kr>
- [7] Kanzaki Masahide, “시맨틱 웹을 위한 RDF/OWL 입문”, 홍릉과학출판사, 2008
- [8] 이재호, “시맨틱 웹의 온톨로지 언어”, 한국정보과학회지, 제 21 권 제 3 호, 2003.4
- [9] 최호섭, 옥철영, “정보검색 시스템과 온톨로지”, 정보과학회지, 제 22 권 제 4 호(통권 제 179 호), 2004.4
- [10] Geonames, <http://www.geonames.org>
- [11] C.Becker and C.Bizer, “Exploring the Geospatial Semantic Web with DBpedia Mobile”, Web Semantics: Science, Services and Agents on the World Wide Web, 2009
- [12] Geospatial Semantic Web Research. <http://www.geospatialmeaning.eu>
- [13] 탁문희, 김경화, 심준호, “상품 온톨로지 모델링 도구의 설계” 한국정보과학회 한국컴퓨터종합학술대회 논문집(B), 2005.7
- [14] Protégé, <http://protege.stanford.edu>
- [15] 송원용, 백두권, 정동원 “웹 자원을 이용한 지리 정보 온톨로지 확장”, 정보과학회논문지, 컴퓨팅의 실제 및 레터 제 15 권 제 10 호, 2009.10