

우리나라 도로명주소를 활용한 지오코딩 및 역 지오코딩 기법 개발

Development of Geocoding and Reverse Geocoding Method Implemented for Street-based Addresses in Korea

석상묵¹⁾ · 이지영²⁾

Seok, Sangmuk · Lee, Jiyeong

Abstract

In Korea, the address-point matching technique has been used to provide geocoding services. In fact, this technique brings the high positional accuracy. However, the quality of geocoding result can be limited, since it is significantly affected by data quality. Also, it cannot be used for the 3D address geocoding and the reverse geocoding. In order to alleviate issues, the paper has implemented proposed geocoding methods, based on street-based addresses matching technique developed by US census bureau, for street-based addresses in Korea. Those proposed geocoding methods are illustrated in two ways; (1) street address-matching method, which of being used for not only 2D addresses representing a single building but also 3D addresses representing indoor space or underground building, and (2) reverse geocoding method, which has converting a location point to a readable address. The result of street-based address geocoding shows 82.63% match rates, while the result of reverse geocoding shows 98.5% match rates within approximately 1.7(m) the average position error. According to the results, we could conclude that the proposed geocoding techniques enable to provide the LBS(Location Based Service). To develop the geocoding methods, the study has performed by ignoring the parsing algorithms for address standardization as well as the several areas with unusual addresses, such as sub-urban areas or subordinate areas to the roads, etc. In the future, we are planning the improved geocoding methods for considering these cases.

Keywords : 3D GIS, Street-based Address, Indoor Space, Reverse Geocoding, Address-matching

초 록

현재 국내 포털사이트 및 기타 공공기관에서 제공하는 지오코딩 서비스는 어드레스-포인트 지오코딩 기법을 활용한다. 이 기법은 매우 높은 위치 정확도를 가지는 반면, 데이터의 품질이 지오코딩 결과물에 큰 영향을 미치며, 3차원 주소에 대한 지오코딩 및 역 지오코딩에 활용될 수 없다는 한계점이 있다. 이에 본 연구에서는 미국 센서스 국에서 개발한 도로 기반 지오코딩 기법에 기반 하여, 우리나라 도로명주소를 활용한 지오코딩 기법을 제시한다. 이 때, 본 연구에서 제안하는 지오코딩 기법은 (1) 단일 건물을 표현하는 2차원 주소뿐만 아니라 지하 건축물 또는 실내 공간까지 포함하는 3차원 주소의 지오코딩을 지원하는 지오코딩 기법과, (2) 특정 지점을 주소로 반환하는 역 지오코딩 기법으로 구분된다. 실제 연구 대상지역을 대상으로 제안한 지오코딩 기법을 적용한 결과, 도로명주소의 지오코딩 시 82.63% 매칭률, 역 지오코딩 시 98.5% 매칭률을 가지는 것으로 나타났으며, 평균 위치 오차가 1.7미터로 나타남에 따라 제안한 지오코딩 기법을 활용한 위치 기반 서비스가 가능성을 보였다. 지오코딩 기법 개발에 있어, 본 연구에서는 주소 정규화를 위한 파싱 알고리즘 및 농촌지역, 도로종속 구간 등의 일부 지역을 고려하지 않고 수행하였다. 이에 향후 연구에서는 이와 같은 사항을 고려한 개선된 지오코딩 기법이 제시될 필요가 있다.

핵심어 : 3차원 GIS, 도로명주소, 실내 공간, 역 지오코딩, 주소 정합

Received 2015. 12. 18, Revised 2016. 01. 13, Accepted 2016. 02. 01

1) Member, Dept. of Geoinformatics, University of Seoul (E-mail: rltk00@uos.ac.kr)

2) Corresponding Author, Member, Dept. of Geoinformatics, University of Seoul (E-mail: jlee@uos.ac.kr)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. 서론

지오코딩(geocoding)은 위치 정보를 기술(記述)하는 데이터를 지오레퍼런싱하는 과정으로, 일반적으로 주소와 같은 준공간 데이터를 공간 데이터로 변환하는 주소 정합 기술로 활용된다. 이와 같은 주소 정합 기술은 문자열로 이루어진 주소를 GIS 데이터로 변환할 수 있기 때문에, 주소를 포함하는 여타 시스템 내에서 널리 활용될 수 있다(Lo and Yeung, 2006).

우리나라의 주소체계는 종전 지번주소 체계에서 2007년 국제적 보편성 확보 및 위치 식별의 편리성을 위하여 도로명주소 체계로 변환되어 이를 법정주소로서 활용하고 있다. 국내 도로명주소는 각 도로구간에 인접한 좌·우측 건물에 도로명과 함께 건물번호를 부여하는 주소체계로, Fig. 1과 같이 구성·표기된다. 이때, 건물 등이 지하에 있을 경우에는 건물번호 앞에 ‘지하’를 붙여 표기한다. 상세주소는 공동주택과 같은 복합 건축물 내부에 관한 주소로, 동(洞)번호 및 호(號)수, 층수를 말한다.

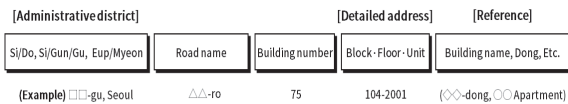


Fig. 1. Address format of street addressing system in Korea

건물번호는 건물 주 출입구의 위치를 기준으로, 각 도로구간마다 일정한 간격으로 부여된 기초번호에 기반 하여 할당된다. 기초번호는 일반적으로 대로로의 경우 20미터, 길의 경우 10미터 간격으로 도로구간의 시작 지점부터 순차적으로 부여되며, 좌측에는 홀수, 우측에는 짝수가 부여 된다. 이때, 건물번호는 해당 건물의 주 출입구에 인접한 도로구간의 기초번호를 사용한다. 이와 같이 우리나라 도로명주소의 건물번호는 일정 간격에 따라 부여된 기초번호에 근거하여 할당되기 때문에, 해당 기초구간 내에 여러 건물이 포함될 경우에 대비하여 가지번호(부번)를 사용하고 있다(예. △△로 96, △△로 96-1).

반면에 미국은 도로구간을 교차지점마다 분할하고, 분할된 도로 세그먼트에 할당될 수 있는 주소 범위(address range)를 먼저 부여한 뒤, 실제 건물 또는 필지가 위치하는 지점까지의 비율을 활용하여 주소를 할당하고 있다(Drummond, 1995). 즉, 우리나라는 미국과 동일한 도로 기반 주소체계를 활용하고 있음에도 불구하고 주소를 할당하는 규칙이 서로 상이하게 나타난다. 국내 도로명주소의 건물번호의 기준이 되는 기초번호는 모든 도로구간을 일정한 간격으로 분할하여 할당

되기 때문에, 비율이 균일하지 않다. 예를 들어, 110미터 도로구간에 기초번호를 20미터단위로 부여하였을 경우 20미터간격의 기초번호 5개와 10미터간격의 기초번호 1개가 부여되는 것이다. 이러한 이유로, 주어진 주소의 범위를 동일한 비율로 분할하는 선형 내삽에 기초한 미국식 지오코딩 기법을 우리나라 도로명주소에 그대로 적용하기에는 한계점이 존재한다.

이에 본 연구는 우리나라 도로명주소체계에 유연하게 적용될 수 있는 지오코딩 기법을 제시하는 것을 목적으로 한다. 이때, 지오코딩 기법은 (1) 단일 건물을 표현하는 2차원 주소뿐만 아니라 지하 건축물 또는 실내 공간까지 포함하는 3차원 주소의 지오코딩을 지원하는 지오코딩 기법과, (2) 특정 지점을 주소로 반환하는 역 지오코딩(reverse geocoding) 기법으로 구분하여 제시한다.

2. 선행연구 분석

기존 연구에서는 각 국의 다양한 주소체계에 대한 지오코딩을 위하여, 1) 도로 기반 지오코딩 기법(US Census, 1992), 2) 필지 기반 지오코딩 기법(Lee, 2009), 3) 어드레스-포인트(address-point) 지오코딩 기법(Ordnance Survey, 2004)에 관한 연구가 수행되어 왔다.

도로 기반 주소의 지오코딩 기법 중 가장 잘 알려진 방법은 미국 센서스 국에서 개발한 방법으로, 각 도로 세그먼트 데이터에 부여된 주소 범위 속성을 기반으로 선형 내삽을 통해 지오코딩을 수행한다(Drummond, 1995; Lee and Kim, 2006). 도로 세그먼트 데이터는 단일 데이터가 많은 주소 정보를 포함할 수 있기 때문에, 가장 경량화 된 지오코딩 참조 데이터 중 하나로 활용될 수 있으며, 여러 분야에서 활용 가능한 기초 자료로서 활용성이 높다는 장점을 가진다.

필지 기반 지오코딩 기법은 지번과 같이 개별 필지 데이터가 가지고 있는 필지 식별자 정보를 참조 데이터로 활용하는 방법으로, 주소가 불규칙하게 부여된 지역이나 도로로부터 멀리 떨어진 지역의 경우 유용하게 활용 가능하다. 그러나 하나의 필지가 다수의 주소를 포함하는 경우 지오코딩의 매칭률(match rate) 측면에서 한계점이 존재한다. Lee(2009)는 이러한 한계점을 보완하고자 교차로를 기점으로 폐합된 구역(block)을 BlockObject model로 정의하고, 이를 참조 데이터로 활용하여 선형 내삽에 기초한 개선된 필지 기반 지오코딩 기법을 제시하였다.

영국 Ordnance Survey(2004)에서는 필지 식별자 정보에 기초한 지오코딩 기법의 한계점을 보완하기 위해 개별 건물 주소에 대한 위치 정보를 속성으로 가지는 어드레스-포인트 데

이터를 활용한 지오코딩 기법을 제안하였다. 이 기법은 어떠한 주소체계에 대해서도 유연하게 적용될 수 있으며 매우 높은 위치 정확도를 가지는 반면, 데이터 품질이 지오코딩 결과물에 큰 영향을 미친다는 한계점이 있다. 또한 3차원 주소에 대한 지오코딩에 활용될 수 없으며(Lee, 2009), 주소가 부여되지 않은 지점에 대한 위치정보를 가지지 않기 때문에 역 지오코딩(reverse geocoding)을 위한 참조 데이터로 활용될 수 없다는 한계점이 존재한다. 이처럼, 주소 정합을 위한 지오코딩 기법은 서로 다른 참조 데이터를 활용하기 때문에 참조 데이터의 특성에 따라 지오코딩 품질이 상이하게 나타난다. 선행연구 결과에 따르면, 지오코딩 매칭률은 도로 기반 지오코딩, 어드레스-포인트 지오코딩, 필지 기반 지오코딩 순서로 높은 것으로 나타났다(Zandbergen, 2008).

현재 국내에서 제공하고 있는 지오코딩 서비스는 어드레스-포인트 지오코딩 기법을 활용한다. 앞서 언급한 것처럼, 이 기법은 3차원 지오코딩 및 역 지오코딩이 불가능하며 참조 데이터의 품질이 지오코딩 결과물에 큰 영향을 미친다는 한계점이 존재한다. 반면에 도로망과 같은 네트워크 데이터는 3차원 지오코딩에 적용이 가능하며, 다수의 주소 정보를 포함하고 있기 때문에 역 지오코딩에 활용 가능하다는 점에서 강점이 있다. 이에 본 연구에서는 미국 센서스 국에서 제안한 도로망 데이터 기반의 지오코딩 기법에 기초하여, 우리나라 도로명주소체계에 보다 적합하게 적용될 수 있는 개선된 지오코딩 기법을 제시한다.

3. 도로명주소를 활용한 지오코딩 및 역 지오코딩 기법

3.1 도로명주소 기반 지오코딩 기법

우리나라 도로명주소는 단독 건물을 표현하는 '단일주소(single address)'와 건축물 실내 정보를 포함하는 '상세주소(detailed address)'로 구분될 수 있으며, 상세주소는 단독 건물 내부 주소, 대학, 아파트 등의 복합 건축물 내부 주소, 지하상가 등의 여러 가지 유형으로 존재할 수 있다. 이에 본 연구에서 제안하는 지오코딩 기법은 단일주소 지오코딩을 위한 2차원 지오코딩 기법과 상세주소 지오코딩을 위한 3차원 지오코딩 기법으로 구분하여 제시한다. 이때, 상세주소는 복합 건축물 내부 주소로 한정한다. 이를 위한 선행 단계로 각 기법에 따른 참조 데이터를 설계하고, 일반적인 지오코딩 수행 절차에 따라 주소 정규화, 참조 데이터 매칭, 위치 결정 단계로 구분하여 제시한다(Lee, 2009). Fig. 2는 본 연구의 도로명주소 지오코딩 수행 절차를 나타낸 것이다.

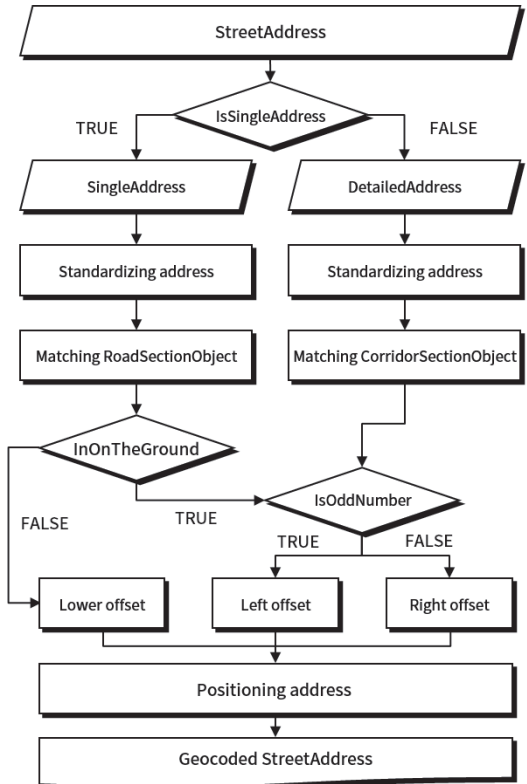


Fig. 2. Geocoding process

3.1.1 단일주소 지오코딩

(1) 참조 데이터 설계

참조 데이터는 지오코딩 시 입력받은 주소와 실질적으로 매칭을 수행하는 데이터로 해당 주소를 대표할 수 있는 데이터여야 한다. 이에 본 연구에서는 단일주소(실외 공간)를 대표하는 참조 데이터로 외부 공간의 연결성을 대표하는 도로구간 데이터(RoadSectionObject)를 활용한다.

도로구간 데이터(RoadSectionObject)는 외부 공간의 지상-지하 단일주소의 매칭을 위한 참조 데이터로, 도로구간 일련번호(RDS_ID) 및 양 측면 기초번호 범위(FR_BN_L, TO_BN_L, FR_BN_R, TO_BN_R)를 기본 속성으로 가진다. 도로구간 데이터는 단일주소(SingleAddressObject)와 1:1 Association 관계를 통해 행정구역명(metro, ward), 도로명(roadName), 도로 유형(roadType) 속성을 참조하여 활용한다. 이때, 도로구간 데이터의 주소 범위 속성은 각 도로구간에 할당된 기초번호를 기준으로 정의한다. 단, 기초번호가 할당되지 않은 구역은 주소 범위를 0으로 정의한다(Fig. 3).

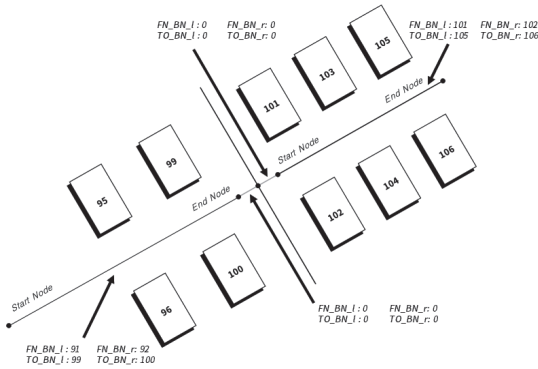


Fig. 3. Example of RoadSectionObject model

(2) 주소 정규화

우리나라 주소체계는 큰 객체에서부터 점차 작은 객체로 서술하는 계층적 구조를 가진다(Lee and Kim, 2006). 이에 본 연구에서는 지오코딩을 위한 도로명주소를 시/도, 시/군/구, 도로명, 도로 유형, 지상/지하, 건물 본번, 건물 부번의 객체별로 분할하여 정규화 한다. 이때, 부번을 가지지 않는 주소의 경우 건물 부번을 0으로 정의한다.

(3) 참조 데이터 매칭

주소 정규화 단계를 거친 단일주소는 도로구간 데이터와 매칭을 수행한다. 이때, 건물 본번이 홀수일 경우에는 FR_BN_L, TO_BN_L 속성에, 짝수일 경우에는 FR_BN_R, TO_BN_R 속성에 포함되는지 판단하여 일치하는 도로구간 데이터를 선택한다.

(4) 위치 결정

도로명주소의 부여규칙에 따라, 단일주소가 표현하는 개별 건물의 출입구는 도로구간에 할당된 기초구간 내에 반드시 포함된다. 이에 착안하여 본 연구에서는 각 도로구간에 할당된 기초구간 중심점을 각 도로명주소의 지오코딩을 위한 기본 좌표 값으로 정의하며, 기초구간 내 포함되는 건물이 2개 이상일 경우에는 건물의 개수만큼 해당 기초구간을 분할한 뒤 분할된 세그먼트의 중심점을 지오코딩 기본 좌표 값으로 활용한다. 단일주소 지오코딩 수행 절차는 다음과 같이 요약된다.

Step 1 : 주소의 도로 유형에 따라 도로구간 시점(Start Node)에서 종점(End Node)까지 기초간격(d_i)으로 분할하는 좌표 값($pt_i(x_i, y_i, z_i)$)을 산출한다. 이때 d_i 는 대로로일 경우

20미터, 길에서는 10미터로 정의한다.

Step 2 : 건물/방 번호(k)를 기반으로 k 가 위치한 기초구간 위치 인덱스(i)를 산출하여 기초구간(S_i)을 선택한다. 이때, i 는 Eq. (1)과 같이 연산한다. 여기서, FN 은 k 가 홀수일 경우 시작지점의 좌측 기초번호, 짝수일 경우에는 시작지점의 우측 기초번호를 활용한다.

$$i = (k - FN + 2) / 2 \tag{1}$$

where, k : Building main-number/room number of target addresses, FN : Address range attribute for starting point of the reference data.

Step 3 : S_i 의 시/종점 좌표 값($pt_i(x_i, y_i, z_i), pt_{i+1}(x_{i+1}, y_{i+1}, z_{i+1})$)을 이용해 세그먼트 중심점($ptm_{0j}(x_{0j}, y_{0j}, z_{0j})$)을 도출한다. 여기서, S_i 가 2개 이상의 건물을 가지는 경우를 대비하여 건물의 개수(j)*2만큼 S_i 를 분할하는 좌표 값($pts(x_s, y_s, z_s)$)을 도출하여, 이에 따른 시/종점으로부터 중심점을 도출한다.

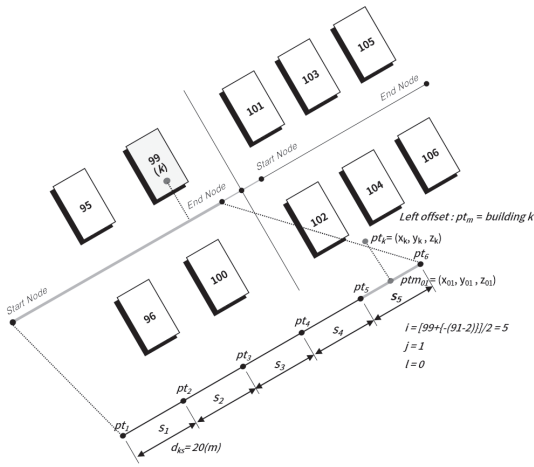
Step 4 : S_i 를 기준으로 도로 유형 및 지상/지하, 건물번호 본번의 홀/짝수 여부에 따라 오프셋 거리(offset distance)와 방향을 정의한다.

Step 5 : 부번을 가지는 단일주소의 최종 지오코딩을 위한 기초구간 중심점($ptm_{0j}(x_{0j}, y_{0j}, z_{0j})$)은 Step 4에서 연산된 pts 를 기반으로 주소의 부번(l)을 활용하여 Eq. (2)를 통해 연산된다. 이후 도출된 $ptm_{0j}(x_{0j}, y_{0j}, z_{0j})$ 을 기준으로, 정의된 오프셋 방향/거리에 따라 해당 기초구간에 직교하는 벡터의 좌표 값($pt_l(x_l, y_l, z_l)$)을 산출하여 지오코딩을 완료한다. 상세주소의 경우 정의된 오프셋 방향에 따라 기초구간의 중심점으로부터 오프셋 하여 지오코딩을 완료한다.

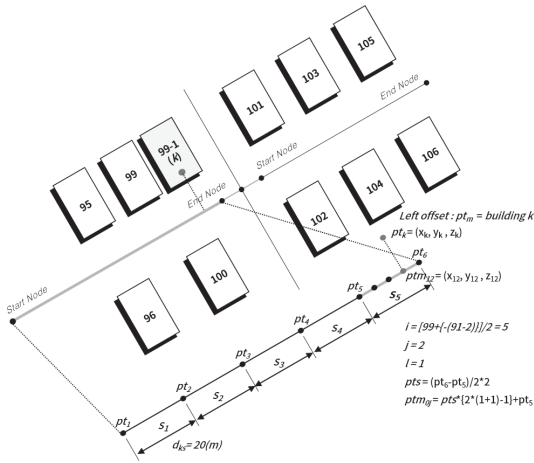
$$ptm_{0j} = pts * \{2 * (l + 1) - 1\} + pt_i \tag{2}$$

where, ptm_{0j} : midpoint of the line segment(S_i) numbering building/room k , j : the number of addresses allocated in the line segment(S_i), pts : geographic coordinates divide into line segment by $2j$, l : Building sub-number of a matched address, and pt_i : geographic coordinates of the starting point of the line segment(S_i)

Fig. 4는 이와 같은 절차에 따라 도로구간을 기반으로 건물 부번 유무에 따른 건물(k)의 위치를 도출하는 예시를 나타낸 것이다.



(a) Single addresses without building sub-number



(b) Single addresses with building sub-number

Fig. 4. Locating single address

위치결정 단계의 마지막 단계로 수행되는 오프셋은 매칭된 주소의 좌표 값을 실제 건물에 가깝게 표현하기 위하여 수행되는 절차로, Step 3, Step 5에서 도출된 기초구간 중심점(pt_m)을 Step 4에서 정의된 오프셋 방향과 거리에 따라 위치시키는 것이다. 이때, 도로구간의 좌·우측 방향은 도로구간의 진행방향에 따라 변하므로, 본 연구에서는 이를 판단하기 위하여 방위각(azimuth angle)을 활용하였으며, 기초구간 중심점($pt_{m0}(x_{0j}, y_{0j})$)으로부터 직교하는 선상에서 정의된 오프셋 거리만큼 떨어진 지점을 최종 지오코딩 좌표 값으로 결정하였다(Hill, 1994; Zwillinger, 1995).

3.1.2 상세주소 지오코딩

(1) 참조 데이터 설계

외부 공간에서 개별 건물과 인접한 객체로서 공간의 연결성을 나타낼 수 있는 데이터가 도로라면, 실내 공간에서 방(호)과 인접하고 연결성을 나타낼 수 있는 객체는 복도가 될 수 있다. 근래에는 실내 위치기반 서비스의 구현을 위하여 방과 방 또는 층과 층 등의 실내 공간 간 위상관계를 복도와 같은 전이(transition) 공간과의 연결성을 통해 그래프 형태의 네트워크 데이터로 구축·활용하는 연구가 활발하게 진행되고 있다(Lee, 2004a; Lee, 2004b; Li and Lee, 2013).

이에 본 연구에서는 복도구간 데이터(CorridorSection Object)를 실내 공간을 표현하는 상세주소의 지오코딩을 위한 참조 데이터로 활용한다. 이는 양 측면 방 번호 범위(FR_RN_L, TO_RN_L, FR_RN_R, TO_RN_R)와 최소 구성단위(minimalUnit)를 기본 속성으로 가지며, 상세주소(DetailedAddressObject)와 Association 관계를 이루어 해당 건물의 도로구간 일련번호(RDS_ID)와 건물번호 본번(buildMainNum), 건물명(buildName), 층수(floor) 속성을 참조하여 활용한다.

호수(방 번호)는 도로구간을 따라 부여된 건물번호와 같이 진행방향의 좌·우측에 따라 홀수, 짝수 등으로 구분되는 표준화된 형태를 가지지 않기 때문에, 복도구간 데이터의 방 번호 범위 속성은 개별 방 번호와 대조될 수 있는 대체번호(alternative number)를 기반으로 부여된다. 이때 대체번호의 부여 간격은 부번의 생성을 미연에 방지하기 위하여 해당 복도구간의 최소 구성단위를 통해 결정한다. Fig. 5는 복도구간 데이터 구축 예시를 나타낸 것이다.

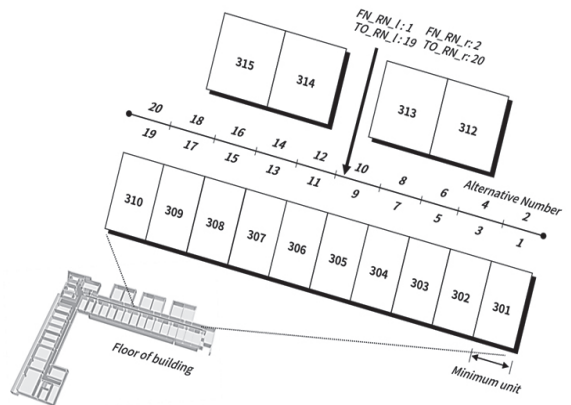


Fig. 5. Example of CorridorSectionObject model

(2) 주소 정규화

상세주소의 정규화는 단일주소와 마찬가지로 주소를 구성하는 객체별로 분할하여 수행한다. 단, 해당 방 번호가 존재하는 복도를 식별하기 위하여 복도의 층 정보를 추가적으로 참조하여 활용한다.

(3) 참조 데이터 매칭

상세주소의 참조 데이터 매칭은 주소의 건물명(동)을 포함한 층, 호수(방 번호) 속성과 복도구간 데이터의 주소 범위속성과의 비교를 통해 최종 복도 세그먼트를 선택한다. 이때, 입력받은 주소의 방 번호는 이와 대조되는 대체 번호가 작성된 LOOKUP TABLE을 추가적으로 참조하며, 동일한 이름의 건물이 다른 지역에 존재할 수 있으므로 건물 주소에 해당하는 도로구간의 일련번호와 건물 본-부번 속성을 추가로 참조한다.

(4) 위치 결정

상세주소 지오코딩의 위치 결정 단계는 단일주소와 마찬가지로 개별 복도에 할당된 대체번호 구간의 중심점을 지오코딩을 위한 기본 좌표 값으로 활용한다. 단, 복도구간 시점에서부터 중점까지 d_k 으로 분할하는 좌표 값($P(x_n, y_n, z_n)$)의 산출시, d_k 는 최소 구성단위를 기준으로 정의한다. 이후 오프셋 단계에서는 호수의 홀·짝수 여부에 따른 좌·우측 오프셋 방향만을 결정하여 지오코딩을 수행한다.

3.2 역 지오코딩 기법

역 지오코딩(Reverse geocoding)은 특정 지점에 관한 좌표 값을 주소, 또는 지명 등으로 반환하는 기법을 말한다(Stenneth et al., 2011). 이를 활용하면 주소가 부여되지 않은 지점 또는 주변 구조물 등의 위치를 주소로 반환할 수 있기 때문에, 지오코딩 기법과 병행하여 활용할 시 측위 기술의 일종으로 활용할 수 있다.

이에 본 연구에서는 도로명주소 기반의 지오코딩의 활용성을 증진시키기 위하여, 특정 좌표 값($P(x_p, y_p, z_p)$)이 주어졌을 경우 도로명주소를 반환하는 역 지오코딩 기법을 제시한다. 이때, 역 지오코딩의 범위는 단일주소로 한정한다. 역 지오코딩 수행 절차는 다음과 같이 요약된다.

Step 1 : 주어진 좌표 값($P(x_p, y_p, z_p)$)로부터 가장 인접한 도로구간에 P 를 투영(projection)한 좌표 값($P_{projected}(x_{pr}, y_{pr}, z_{pr})$)을 연산한다(Lee, 2004a).

Step 2 : $P_{projected}$ 이 도출되면, 인접한 도로구간 데이터의 시

점에서부터 $P_{projected}$ 까지의 거리(d_p)에 따른 비율을 연산하여, 이에 해당하는 건물 번호(k)를 반환한다. 이때, 반환되는 건물 번호의 홀·짝수 여부는 선택된 도로구간 시작점(ptb)과 끝점(pte), P 와 $P_{projected}$ 가 이루는 방위각을 통해 결정한다.

이와 같이 역 지오코딩을 위한 건물 번호(k)는 도로구간 시작지점에서부터 P_p 까지의 거리(d_p)와 기초간격(dks)를 활용하여 연산된 인덱스($index_k$)를 통해 도출된다. Fig. 6은 역 지오코딩 예시를 나타낸 것이다.

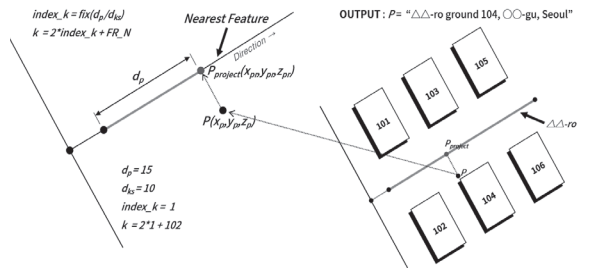


Fig. 6. Example of the reverse geocoding

4. 실험 및 결과 분석

4.1 참조 데이터 구축

국가공간정보유통시스템에서는 도로명주소의 보급·확산을 위하여 우리나라 전역을 대상으로 도로구간, 실폭도로 등의 교통 관련 데이터뿐만 아니라 행정경계, 건물 등 30개 레이어로 구성된 도로명주소 전자지도 데이터를 제공하고 있다. 이 데이터는 GRS80 타원체, UTM-K 좌표체계에 구축되어 있으며, 도로구간과 기초구간 레이어를 분리하여 제공한다. 두 레이어는 동일하게 도로중심선을 표현하고 있으나, 도로구간 레이어는 각 도로구간마다 하나의 선형 데이터로 이루어진 반면 기초구간 레이어는 정의된 기초간격에 따라 분할된 세그먼트로 구축되어 있다. 또한 도로구간 레이어는 도로구간에서 분기되는 골목길의 분기 지점을 표현하고 있으나, 기초구간 레이어는 기초번호 시작 지점부터 구축되어 기초번호가 부여되지 않은 일부 교차로 지역은 표현하지 않는다.

이에 본 연구에서는 2015년도 8월 11일자 동대문구 도로명주소 전자지도의 도로구간 데이터를 기반으로, 중속구간을 제외한 도로구간을 대상으로 좌·우측 기초번호 범위를 속성으로 추가하여 단일주소 지오코딩을 위한 도로구간 데이터를 구축하였다. 여기서, 기초구간 레이어에서 포함하지 않는

분기 지점의 주소 범위 속성은 0으로 할당하였으며, 각 도로 구간 Z값은 모두 0으로 가정하였다.

상세주소 지오코딩을 위한 복도구간 데이터는 서울시립대학교 21세기관 준공도면을 기반으로 실내 공간을 표현하는 벽, 복도, 방 객체를 추출하여 구축하였다. 이때, 각 층의 높이는 2.5미터로 가정하였으며, 좌표체계는 별도의 지오레퍼런싱 단계를 거치지 않은 상대좌표계를 활용하였다.

4.2 결과 분석

4.2.1 도로명주소 기반 지오코딩 기법 적용

본 연구의 지오코딩 기법 적용을 위한 도로명주소는 종속 구간을 제외한 동대문구 전농동 일대 단일주소 2,447개와 서울시립대학교 21세기관 5, 6, 7층에 해당하는 상세주소 89개로 한다. 이를 위한 전처리 단계로, 도로명주소의 형식을 정규화 된 형태로 변환하여 입력 데이터로 활용하였다.

(1) 단일주소 지오코딩

정규화 된 단일주소와 도로구간 데이터와의 매칭을 수행하였다. 이때, 각 기초구간에 포함되는 주소의 개수는 도로명주소 안내시스템(<http://www.juso.go.kr>)에서 제공하는 도로명주소 DB를 참조하였다. 지오코딩 수행 시 좌·우측 오프셋 값은 도로의 유형에 따라 대로 30미터, 로 18.5미터, 길 6.5미터로 정의하였다. Fig. 7은 단일주소 2,447개에 대한 지오코딩 결과를 매핑한 것이다.



Fig. 7. Mapping geocoded single addresses

건물 부번을 가지는 주소 261개 또한 개별적으로 지오코딩되어 각기 다른 좌표 값을 가지는 것으로 나타났다. Fig. 8은 두 개의 부번을 가지는 주소(195, 195-1, 195-2)에 대한 지오코딩 결과를 나타낸 것이다.

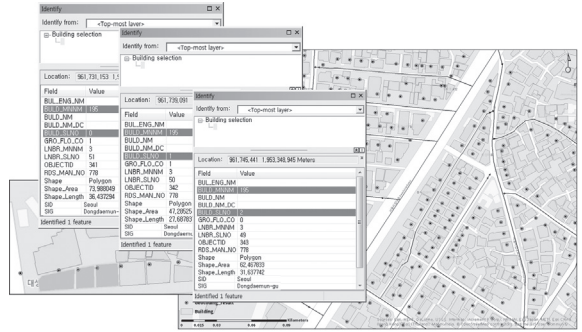


Fig. 8. Geocoded single addresses with building sub-number

지하 주소는 지상 주소와 다르게 지상 하부 층으로 지오코딩 되는 것을 확인할 수 있었다. Fig. 9는 지하철역에 대한 지오코딩 결과를 나타낸 것이다.

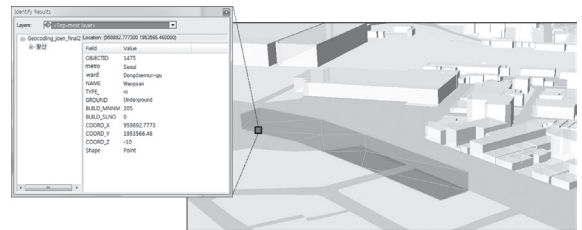


Fig. 9. Geocoded underground addresses

단일주소 지오코딩 결과 전체 2,447개 결과물 중 1,651개가 건물 폴리곤 내에 위치하는 것으로 나타났으며, 796개는 건물 폴리곤 바깥에 위치하는 것으로 나타났다. 여기서, 건물 폴리곤 내 위치한 1,651개 지오코딩 결과물 중 건물 본번만이 일치하는 경우는 1,507개로 나타났으며, 건물 본·부번이 모두 만족하는 경우는 1,448개로 나타났다. 그 외 144개는 지오코딩 결과물이 해당 건물의 위치에 정확하게 표현되지 않은 것으로 나타났다.

건물 폴리곤 내 포함되지 않은 지오코딩 결과물 796개를 대상으로 가장 인접한 건물과의 거리를 측정한 결과 최소 0미터에서부터 최대 약 64.3미터까지 분포하는 것으로 나타났으며, 평균 2.0미터, 표준 편차는 약 3.7미터로 나타났다. 이와 같이 최소값과 최대값의 편차가 크게 나타나는 이유는 도로명주소의 부여 규칙 상 복합 건축물에 대한 건물번호는 개별 건물을 대상으로 부여되는 것이 아니라 이를 대표하는 출입구를 대상으로 부여되고 있기 때문인 것으로 나타났다. 이에 복합 건축물에 해당하는 단일주소 21개를 제외한 775개

Table 1. Distance from geocoded single addresses that not included in building polygons to nearest building polygons

Category	Distance		
	Max(m)	Mean(m)	Std(m)
Single Address (n = 775)	15.5	1.7	2.0

Table 2. Results of the geocoded single addresses

Category	Geocoding (n = 2,447)		
	Partly matched	Perfectly matched	Unmatched
Single Address 1 (n = 1,651)	1,507 (61.59)	1,448 (59.17)	144 (5.88)
Single Address 2 (n = 796)	618 (25.26)	574 (23.46)	178 (7.27)
Total	2,125 (86.84)	2,022 (82.63)	322 (13.16)

를 대상으로 인접한 건물과의 거리를 재 측정된 결과, Table 1과 같이 최소-최대값의 편차가 확연하게 줄어드는 것을 확인할 수 있었다.

Table 2는 앞서 도출된 모든 사항을 종합한 단일주소 지오코딩 최종 결과를 나타낸 것이다. 이와 같이, 전체 2,447개 단일주소 중 건물 본-부번이 모두 일치하는 형태로 지오코딩된 주소는 2,022개로 나타났으며, 해당 주소에서 벗어난 위치에 지오코딩된 주소는 322개로 나타났다.

(2) 상세주소 지오코딩

단일주소와 마찬가지로 전처리 단계를 거쳐 정규화 된 상세주소를 기반으로 지오코딩을 수행하였다. 그 결과 복도구간의 최소 구성단위에 따라 부여된 대체번호를 기반으로 각 호수별 지오코딩이 수행됨에 따라, 공간 질의 또한 가능한 것으로 나타났다(Fig. 10).

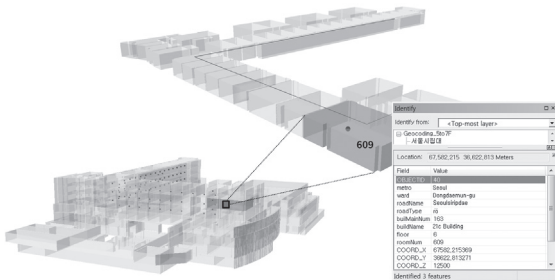


Fig. 10. Spatial query example of the geocoded detailed addresses

Table 3은 상세주소 지오코딩 결과를 나타낸 것이다. 이와 같이, 전체 89개 상세주소에 대한 지오코딩 결과는 89개 주소 모두 해당 호수 내에 위치하는 것으로 나타났으며, 모두 해당하는 호수와 일치하는 것으로 나타났다.

Table 3. Results of the geocoded detailed addresses

Category	Geocoding (n = 89)		
	Matched room number	Matched alternative number	Unmatched
Detailed Address	89 (100)	89 (100)	0 (0)

상세주소의 경우 단일주소보다 상대적으로 지오코딩 결과물의 정확도가 높게 나타나는 것을 알 수 있었다. 이는 일률적인 규칙에 따라 정의된 속성을 포함하는 도로구간 데이터와 달리, 비교적 협소하고 정형화된 공간을 대상으로 사용자가 정의한 속성을 포함하는 복도구간 데이터를 지오코딩 참조 데이터로서 활용함에 따라 상대적으로 높은 정확도를 보이는 것으로 나타났다.

4.2.2 역 지오코딩 기법 적용

본 연구에서는 역 지오코딩을 위한 샘플 데이터 생성을 위하여, 앞서 구축된 도로구간 데이터의 30미터이내 지역을 대

상으로 1,000개의 점(point) 데이터를 무작위로 생성하였다. 이후 샘플 데이터로부터 가장 인접한 객체를 식별하고 그에 따른 투영지점을 도출하기 위해 ArcMap에서 제공하는 인접(Near) 툴을 활용하였다. 이에 따라 1,000개의 점 데이터와 구축된 도로구간 데이터 간의 인접 관계를 연산하였으며, 연산된 결과물을 역 지오코딩 적용을 위한 입력 데이터로 활용하였다. 이때, 기초번호가 존재하지 않는 지역에 투영 지점이 발생하는 것을 방지하기 위하여 도로구간 데이터 중 주소 범위 속성이 0으로 부여된 분기 지점은 인접 관계 연산 시 제외하였다. 역 지오코딩 결과에 대한 정확도 검증은 도로명주소 전자지도에서 제공하는 기초구간 레이어와 역 지오코딩 결과물 간의 비교를 통해 수행하였다. Table 4는 본 연구의 역 지오코딩 기법 적용 결과를 나타낸 것이다. 이와 같이, 역 지오코딩 결과 전체 1,000개 지점 데이터 중 985개(98.5%)의 데이터가 가장 인접한 도로구간의 기초번호와 일치하는 주소를 반환하는 것으로 나타났다.

Table 4. Results of the reverse geocoding

Category	Reverse geocoding (n = 1,000)	
	Matched	Unmatched
Single address	985 (98.5)	15 (1.5)

기초번호와 불일치한 것으로 나타난 15건의 경우, 모두 역 지오코딩 결과로 도출된 건물번호와 연속적으로 인접한 기초번호로 반환된 것으로 나타났다. 이러한 현상은 건물번호 연산 시 도로구간 시작 지점에서부터 인접 객체에 투영된 지점까지의 비율을 연산하는 과정에서 발생하였다. 즉, 본 연구에서는 기초간격을 '대로:로'의 경우 20미터, '길'의 경우 10미터로 정의한 반면, 도로명주소 전자지도 기초구간 데이터에서의 기초간격은 19.9미터, 9.9미터 등과 같이 일정하지 않음에 따라, 세그먼트 상에서 기초번호가 전환되는 가장자리 지점에서 이와 같은 불일치가 발생하는 것으로 나타났다. 이에 따라, 도로명주소 전자지도 기초구간 레이어에서 제공하는 세그먼트는 해당 도로구간에 정의된 기초간격에 따라 20미터, 10미터 등으로 정확하게 할당될 필요가 있을 것으로 나타났다.

5. 결론

우리나라 도로명주소는 크게 단독주택을 표현하는 단일주

소와 공동주택을 표현하는 상세주소로 구분될 수 있다. 일반적으로 단일주소는 도로에 인접한 건물의 위치를 표현하는 2차원 주소인 반면, 상세주소는 건물 내부의 층과 호수까지 포함하는 3차원 주소이다. 국민의 50% 이상이 공동주택에 거주하고 있는 우리나라의 거주환경 특성을 감안할 때(Jeon *et al.*, 2010), 도로명주소 기반 지오코딩의 활용성을 높이기 위해서는 단일주소뿐만 아니라 상세주소 부문까지 지원할 수 있는 3차원 지오코딩 기법이 요구된다. 이에 본 연구에서는 우리나라 주소체계에 기반 한 지오코딩 및 역 지오코딩 기법을 제시함에 따라 주소를 활용한 위치 기반 서비스의 활용성을 증진시키고자 하였다.

본 연구에서 제안한 지오코딩 기법 적용 결과, 서울특별시 동대문구 전농동의 단일주소 중 82.63%의 결과물이 해당 건물의 위치에 정확하게 표현된 것을 알 수 있었다. 반면에 상세주소의 경우, 89개 주소 전체가 해당하는 호수 내에 매핑된 것으로 나타났다. 이는 법률적인 규칙에 근거하여 모든 도로를 대상으로 일률적으로 할당된 도로구간 기초번호와 달리, 상대적으로 정형화된 공간 내에서 사용자 정의에 따라 할당된 북도구간 기초번호를 지오코딩에 활용하였기 때문인 것으로 나타났다. 역 지오코딩 기법 적용 결과, 1,000개의 샘플 데이터 중 98.5%가 정확한 주소를 반환하는 것으로 나타나 매우 높은 정확도를 가지는 것으로 나타났다. 이와 같은 역 지오코딩 기법은 주소가 할당되지 않은 지점에 대한 거주소(假住所)를 부여하는 기술로 활용될 수 있다. 즉, 낯선 장소에서 화재 등을 목격하였을 경우, 역 지오코딩에 따라 반환되는 도로명주소를 활용하여 보다 신속하게 해당 지점의 위치정보를 취득할 수 있을 것이다.

우리나라 도로명주소 기반의 지오코딩 기법 개발에 있어, 본 연구에서는 주소 정규화를 위한 파싱 알고리즘 및 농촌지역, 도로중속 구간, 복합 건축물 내 개별 건물 주소 등의 일부 지역을 고려하지 않고 수행하였다. 지오코딩 수행 시, 입력받은 주소에 대한 정제 결과는 지오코딩의 품질에 큰 영향을 미칠 수 있다. 즉, 다양한 형식의 주소에 대해 유연한 처리가 가능한 파싱 알고리즘은 지오코딩 서비스 제공 시 필수적인 사항이라고 할 수 있다. 이에 실질적인 지오코딩 서비스 제공을 위해서는 향후 이와 같은 사항을 총괄적으로 고려한 연구를 통해 개선된 지오코딩 알고리즘이 개발될 필요가 있다.

알림의 글

이 논문은 저자의 석사학위 논문에 기반 하여 작성된 논문입니다.

References

- Drummond, W.J. (1995), Address matching GIS technology for mapping human activity patterns, *Journal of the American Planning Association*, Vol. 61, No. 2, pp. 240-251.
- Hill, F.S. Jr. (1994), The pleasures of “perp dot” products, In: Heckbert, Paul S. (ed.), *Graphics Gems IV*, Morgan Kaufmann, San Francisco, CA, pp. 138-148.
- Jeon, W., Kim, J., and Kim, J. (2010), A study of standardization model for detailed address writing of apartment house, *Journal of the Korean Cadastre Information Association*, Vol. 12, No. 2, pp. 249-264. (in Korean with English abstract)
- Lee, J. (2004a), A spatial access-oriented implementation of a 3-D GIS topological data model for urban entities, *GeoInformatica*, Vol. 8, No. 3, pp. 237-264.
- Lee, J. (2004b), 3-D GIS for geo-coding human activity in micro-scale urban environments, *Geographic Information Sciences: Springer's Lecture Notes in Computer Science Computers*, Vol. 3234, Springer, New York, NY, pp. 162-178.
- Lee, J. (2009), GIS-based geocoding methods for area-based addresses and 3D addresses in urban areas, *Environment and Planning B: Planning and Design*, Vol. 36, No. 1, pp. 86-106.
- Lee, J. and Kim, H. (2006), A geocoding method implementation for hierarchical areal addressing system in Korea, *The Journal of GIS Association of Korea*, Vol. 14, No. 4, pp. 403-419. (in Korean with English abstract)
- Li, K. and Lee, J. (2013), Basic concepts of indoor spatial information candidate standard IndoorGML and its applications, *Journal of Korea Spatial Information Society*, Vol. 21, No. 3, pp. 1-19. (in Korean with English abstract)
- Lo, C.P. and Yeung, A.K.W. (2006), *Concepts and Techniques of Geographic Information Systems 2nd Edition*, Prentice Hall, Upper Saddle River, NJ.
- Ordnance Survey (2004), *OS MasterMap User Guide: Product Specification*, Ordnance Survey, Southampton.
- Roh, H. (2009), Assessment and improvement method of position accuracy of geocoding application for new address system in Korea, *Journal of the Korean Cartographic Association*, Vol. 9, No. 1, pp. 63-72. (in Korean with English abstract)
- Stenneth, L., Wolfson, O., Yu P.S., and Xu B. (2011), Transportation mode detection using mobile phones and GIS information, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 01-04 November, Chicago, IL, pp. 54-63.
- U.S. Census Bureau (1992) *TIGER/Line Census Files, 1992: Technical Documentation*, Department of Commerce, Washington, DC.
- Zandbergen, P.A. (2008), A comparison of address point, parcel and street geocoding techniques, *Computers, Environment and Urban Systems*, Vol. 32, pp. 214-232.
- Zwillinger, D. (1995), Spherical coordinates in space, In: Zwillinger D. (ed.), *CRC Standard Mathematical Tables and Formulae*, CRC Press, Boca Raton, FL, pp. 297-298.