

## 공공데이터를 활용한 주소 기반 데이터 분석과 제언

박 하 람<sup>1</sup> · 김 학 래<sup>2\*</sup><sup>1</sup>중앙대학교 문헌정보학과 정보학 박사과정<sup>2</sup>중앙대학교 문헌정보학과 교수

## Address-based Data Analysis and Recommendations Utilizing Open Data

Haram Park<sup>1</sup> · Haklae Kim<sup>2\*</sup><sup>1</sup>Ph.D's Course, Department of Library and Information Science, Chung-Ang University, Seoul 06974, Korea<sup>2</sup>Professor, Department of Library and Information Science, Chung-Ang University, Seoul 06974, Korea

### [요 약]

공공데이터에서 주소는 국민의 생활 전반에 활용되는 핵심 데이터이지만, 품질 문제가 지속적으로 제기되고 있다. 주소의 활용성 제고를 위해 주소가 활용되는 규모와 양상을 파악하는 것이 필요하지만, 광범위한 공공데이터에서 주소가 포함된 데이터를 찾는 것이 어려운 실정이다. 본 연구는 공공데이터포털을 포함한 총 13개의 국내 데이터 포털에서 주소가 포함된 데이터셋을 식별하는 방안을 제시하고, 이를 기반으로 주소 데이터의 활용 양상을 분석한다. 모든 데이터셋의 컬럼명에서 알고리즘과 전문가 리뷰로 주소와 관련된 컬럼명을 선별하고, 이 컬럼명이 포함된 데이터셋을 주소가 포함된 데이터셋으로 선정한다. 분석 결과, 공공데이터의 약 37%가 주소를 포함하고 있었다. 특히, 공공데이터에서 도로명주소와 함께 지번주소도 광범위하게 활용되는 양상을 확인할 수 있었다. 연구 결과에 따르면, 주소와 관련된 컬럼명의 표준화가 필요하며 도로명주소와 지번주소의 체계적인 연계 방안이 마련되어야 한다.

### [Abstract]

Addresses are a critical component of open government data, used across many aspects of citizens' lives. However, the quality of address data remains an issue. Understanding the volume and scope of address data usage is essential to enhancing its value, yet identifying address-related information within vast public datasets is challenging. This study presents a method for identifying datasets containing addresses from 13 domestic data portals, including data.go.kr, and analyzes address data usage patterns based on this approach. We identify address-related datasets using an algorithm and expert review to select relevant column names from the datasets. The analysis reveals that about 37% of open data includes addresses. Notably, land lot addresses are often used alongside street number addresses. Based on these findings, there is a need to standardize column names related to addresses and establish a systematic linkage between road name addresses and land lot number addresses.

**색인어** : 공공데이터, 주소, 컬럼명, 도로명주소, 지번주소**Keyword** : Open Data, Address, Column Names, Road Name Address, Land Lot Number Address<http://dx.doi.org/10.9728/dcs.2024.25.9.2617>

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.

Received 06 July 2024; Revised 05 September 2024

Accepted 11 September 2024

\*Corresponding Author; Haklae Kim

Tel: +82-2-820-5561

E-mail: haklaekim@cau.ac.kr

## I. 서론

공공데이터는 공공기관이 업무를 수행하는 과정에서 생성 또는 취득한 다양한 유형의 모든 자료[1]로, 국민의 이용권을 보장을 위해 개방과 활용이 강조되고 있다[2]. 이 중 주소는 행정이나 교통, 물류, 안전 등 국민 생활 전반에 걸쳐 활용되는 핵심 데이터다. 주소는 국가의 7대 등록 공부(주민등록, 가족관계등록, 외국인등록, 사업자등록, 법인등록, 건축물등록)에 등록되어 행정 업무의 기반이 되고, 소방과 경찰, 재난 등의 분야에서 실시간으로 갱신되어 국민의 안전 보장에 핵심적인 데이터로서 기능하고 있다[3]. 또한, 주소는 특정 객체의 위치를 식별할 수 있는 정보를 담고 있어 다양한 분야의 데이터를 연계 또는 융합할 수 있는 핵심 식별자로 주목받고 있다[4]-[7].

공공데이터에서 주소는 광범위하게 활용되고 있지만, 주소 데이터의 품질 이슈가 지속적으로 논의된다. 주소 데이터의 오류를 식별하고 정제하기 위한 방법론이 제안되고 있고 [8],[9], 실제 공공데이터에서 행정구역 누락(‘가좌로7가길 35’) 또는 다양한 행정구역의 표기(‘전라남도’ 대신 축약어 ‘전남’ 사용), 띄어쓰기 오류(‘123길11’이 아닌 ‘123길 11’의 잘못된 도로명 띄어쓰기), 도로명주소 대신 지번주소 기입(또는 그 반대), 부정확한 값 입력(‘산마루놀이터 내’), 공백 등과 같은 주소 오류가 빈번하다[10],[11]. 한편, 대규모 언어모델(LLM; Large Language Model)의 등장으로 최근 인공지능의 발전에 고품질 학습 데이터의 확보가 중요해졌다. 이 맥락에서 공공데이터는 LLM의 학습과 검증에 활용될 수 있는 데이터세트의 원천으로 활용되고 있다. 특히, 공공데이터는 주소의 다양한 오류를 포괄하는 광범위한 데이터를 갖고 있고, 이를 기반으로 LLM을 위한 한국어 주소 데이터세트를 구축할 수 있다.

그러나, 대규모의 공공데이터에서 주소를 포함하고 있는 데이터세트를 찾는 것은 쉽지 않다. 주소를 정제 또는 수집하기 위해 주소 값에 접근할 수 있어야 하지만, 개별 데이터세트에서 주소를 표현하는 컬럼명은 매우 다양하다. 더불어 도로명주소가 전면 시행되었음에도 여전히 데이터 상에서 지번 주소를 사용하고 있어, 실제로 주소 데이터가 활용되는 양상을 파악해볼 필요가 있다. 본 연구는 중앙부처와 지방자치단체가 운영하는 국내 데이터 포털을 대상으로 주소가 포함된 데이터세트를 식별할 수 있는 방법을 제안하고, 선별된 공공 데이터를 기반으로 주소의 활용 양상을 분석한다.

본 논문의 구성은 다음과 같다. 2장은 주소와 관련된 국내 선행연구를 정리하고, 주소가 포함된 데이터세트를 찾는 전반적인 연구 방법은 3장에서 설명한다. 4장은 연구 방법을 실제로 수행한 결과와 검증을 설명하고, 5장은 주소가 포함된 데이터세트의 특성을 분석한다. 6장은 연구결과의 요약과 시사점을 기술한다.

## II. 선행 연구

주소는 전세계적으로 우편과 배송, 비즈니스, 공공서비스, 재난 관리와 긴급상황 등에서 핵심적인 역할을 수행하고 있다[12]. 그러나, 개별 국가는 고유한 주소 체계를 갖고 있고, 이는 국제적인 수준에서 주소 데이터의 통합을 어렵게 한다. 이런 맥락에서 주소 데이터의 상호운용을 위해 다양한 국제 표준 논의가 진행되었고, 가장 대표적인 사례가 ISO(International Organization for Standardization)의 19160이다[13]. ISO 19160은 다양한 국가의 주소 체계를 포괄하는 개념 모델을 제시하고, 주소의 부여와 유지/관리, 품질 등에 대한 프레임워크를 제공하고 있다[14]. 한편, 주소를 중심으로 데이터를 연계하기 위한 연구가 다양하게 진행되고 있다. 대표적으로 UPRN(Unique Property Reference Number)은 영국의 공공표준이자 주소 할당이 가능한 위치에 부여한 식별자로, 주소 데이터를 중심으로 민간과 공공의 데이터를 연계하는 데 활용된다[15]. UPRN은 공공서비스 [16], 보건과 의료분야[17],[18] 등의 분야에서 활발히 사용되고 있다. 이 외에도 링크드 데이터 기술을 활용해 주소를 중심으로 다양한 공공데이터를 연계한 연구가 진행되고 있다 [19],[20].

2014년에 도로명주소가 전면 시행된 이후로, 주소와 관련된 국내 연구는 크게 2가지로 구분할 수 있다. 정책적인 측면에서 도로명주소 중심으로 한국의 주소체계에 대해 논의하는 연구와 데이터 측면에서 주소를 정제·검증·활용하려는 연구다.

정책적 관점에서 한국은 일본의 지번주소를 벗어나 도로명 주소를 정착시키려는 다각도의 노력을 수행해왔다. 한국은 2011년부터 주소와 관련된 국제 표준 ISO 19160 표준 제정에 참여하고, 한국의 주소체계를 국제 표준에 반영하려는 노력을 진행했다[21]-[30]. 노력의 결과로 ISO 19160-1의 개념모델에 대한 프로파일의 사례에 한국의 도로명주소 예시가 추가되었고[13], 2023년에 제정된 ISO 19160-2(주소 부여와 유지관리)에 사물주소 부여 방식과 입체 이동경로 등이 우수사례로 포함되었다[31]. 한편, 도로명주소가 2014년부터 본격적으로 시행되면서 관련 법률과 시행령, 시행규칙, 각종 규정 등이 마련되었음에도 불구하고, 실질적인 이행 단계에서 여전히 어려움을 겪고 있다. 건물군 내의 건물별 위치 찾기[32]와 상세주소 부여[33], 주소정보기반대상(예: 도로 구간)의 등록[34] 등은 도로명주소에서 보완되어야 할 사항으로 논의되고 있다. 더불어 도로명주소가 부여되기 어려운 객체나 지역 등에 주소를 부여하기 위해 국가지점번호, 사물주소와 같은 새로운 주소체계도 도입되었다. 도로명주소는 도로가 없는 지역(예: 산악지역)이나 시설물(예: 상하수도)에 접근하는 데 한계가 있고[35], 이를 보완하기 위해 비거주지역의 위치를 표시할 수 있는 국가지점번호[36], 시설물에 주소를 부여하는 사물주소[37],[38] 등이 도입되었다.

데이터 관점에서 주소는 데이터의 활용성에 초점을 두어

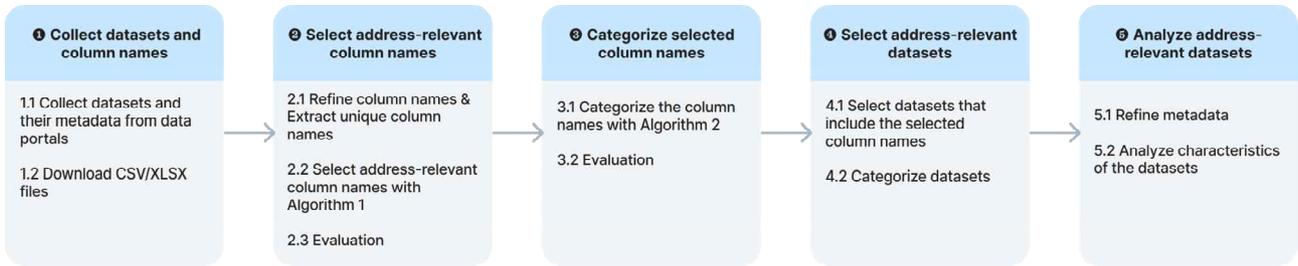


그림 1. 연구 방법  
Fig. 1. Research method

연구가 진행되어왔다. 주소는 다양한 정보를 연결할 수 있는 기본 키(primary key)로 활용될 수 있지만[6],[7], 공공데이터와 연계하기 위해 주소 데이터의 정제는 필수적이다[10]. 특히, 주소는 행정구역과 도로명, 건물명 등의 조합으로 구성되기 때문에 개별 단위의 정제를 위해 딥러닝 기반의 주소 파싱 모델을 구축하고 오류를 정제하는 방안이 제안되고 있다 [8],[9]. 한편, 공공데이터는 주소 기반으로 시설물이나 건물 등의 다양한 속성정보를 연계할 수 있는 데이터를 갖고 있고, 이를 융합·연계하는 시도가 지속되고 있다[7],[39]. 행정구역과 같이 주소에 포함된 구성요소도 우편번호, 행정기관 등의 공공데이터와 연계해 재사용과 확장성이 높은 데이터로 구축될 수 있다[20]. 이와 같이 공공데이터에 있는 주소 데이터를 활용하기 위한 연구가 다양하지만, 광범위한 공공데이터에서 주소 데이터에 접근하기 위한 방안에 대한 연구는 미흡하다. 기존 연구는 공공데이터포털의 일부 데이터만 선정하고 있어[33],[34] 광범위한 공공데이터를 대상으로 연구 결과를 일반화하기 어렵다. 본 연구는 공공데이터포털을 포함한 17개의 국내 데이터 포털을 대상으로 주소가 포함된 공공데이터를 선별하는 자동화 방안을 제안하고, 주소의 활용 양상을 분석한다.

### III. 연구 방법

연구 방법은 그림 1과 같이 크게 5가지 단계로 진행된다. 모든 절차는 파이썬(버전 3.9)으로 자동화된다. 1단계는 국내 데이터 포털에서 개방된 데이터셋을 수집하고, 컬럼명을 추출하는 과정이다. 데이터셋의 수집이 용이한 공공데이터로 범위를 한정하고, 중앙부처와 지방자치단체에서 운영하는 데이터 포털을 중심으로 데이터셋과 메타데이터(예: 분류체계, 키워드)를 크롤링한다(1.1). 이 중 컬럼명의 추출이 가능한 CSV 또는 XLSX 형태의 파일을 다운로드한다(1.2).

한편, 다운로드 받은 모든 데이터셋에서 주소 값이 포함된 데이터셋을 일일이 찾기는 어렵다. 이 데이터셋을 찾는 기본적인 방법은 주소 값을 포함하는 컬럼명을 선정하고, 이 컬럼명이 포함된 데이터셋을 찾는 것이다. 이를 위해 2 단계는 주소가 값으로 포함될 가능성이 있는 컬럼명을 선정

한다. 이 때, 주소가 데이터셋 값으로 포함된 컬럼명은 ‘주소와 관련된 컬럼명’으로, 주소 값을 포함하는 데이터셋은 ‘주소 데이터셋’로 정의한다. 모든 데이터셋에서 컬럼명을 추출하고, 전체 컬럼명을 정제한 다음 고유한 컬럼명만 남긴다(2.1). 추출된 모든 컬럼명은 표 1의 함수로 정제해 공백, 특수문자나 영문 등을 제거한 한글만 남긴다. 고유한 컬럼명  $(c_1, c_2, \dots \in UC)$ 은 정제된 컬럼명에서 중복을 제거한 컬럼명을 선정한다. 고유한 컬럼명에서 주소와 관련된 컬럼명의 후보군(AC)은 표 2의 알고리즘 1로 추출한다(2.2). 알고리즘 1의 tokenizer는 soynlp의 형태소 분석기를 사용하여 컬럼명을 형태소 단위의 토큰 리스트로 생성한다. 토큰 리스트에서 ‘주소’ 또는 ‘소재지’와 일치하는 토큰이 포함된 컬럼명은 주소와 관련된 컬럼명으로 추출한다.

2.2단계에서 선정된 주소와 관련된 컬럼명이 주소 값을 갖고 있는 데이터셋을 효과적으로 선별할 수 있는지 검증할 수 행한다(2.3). 검증 방법은 주소와 관련된 컬럼명으로 예측된 데이터셋의 F1 score를 측정하고, 이를 기반으로 전문가 리뷰를 진행하여 주소와 관련된 컬럼명을 최종 선정한다. F1 score는 무작위로 100개의 데이터셋을 추출해 주소 값이

표 1. 컬럼명 정제에 사용한 함수

Table 1. A function used to refine column names

<b>Input:</b> A original column name
<b>Output:</b> A refined column name
<pre>def refine_col_name(col):     col = re.sub("[^가-힣]", "", str(text).strip())     return col</pre>

표 2. 알고리즘 1: 주소와 관련된 컬럼명의 후보군 추출

Table 2. Algorithm 1: Extracting candidate column names related to addresses

<b>Input:</b> A Unique Column Set UC
<b>Output:</b> Get the address-related column names
<pre>Let: AC ← ∅ for all c ∈ UC do     token_list = tokenizer(c)     if '주소' in token_list or '소재지' in token_list then         AC ← AC ∪ {c}     end if end for</pre>

\*The Korean terms “juso(주소)” and “sojae(소재지)” are used to identify column names that contain the Korean words.

포함된 정답 데이터셋을 선정한 다음, 예측한 데이터셋의 결과와 비교하는 데 활용한다. 주소와 관련된 컬럼명은 F1 score의 결과를 검토하여 전문가 리뷰를 통해 확정된다. 주소에 사전지식이 있는 저자 2명과 주소에 사전지식이 없지만 공공데이터에 전문지식이 있는 1명이 주소와 관련된 컬럼명에 대해 체계적인 검토를 수행한다. 주소와 관련된 컬럼명은 전문가의 의견을 종합적으로 고려하여 최종적으로 선정된다.

3단계는 상세한 분석을 위해 주소와 관련된 컬럼명을 주소의 유형에 따라 분류한다. 주소의 유형은 광범위하게 사용되는 도로명주소와 지번주소로 구분하고, 이 유형으로 구분되지 않는 컬럼명은 기타에 포함한다. 예를 들어, ‘소재지도로명주소’와 같은 컬럼명은 도로명주소 유형에, ‘영업장소재지번’과 같은 컬럼명은 지번주소 유형에 분류한다. 기타는 ‘소재지주소’ 또는 ‘사업장주소’와 같이 주소의 유형을 알 수 없는 컬럼명을 포함한다. 주소와 관련된 컬럼명(FAC)의 유형 구분은 표 3의 알고리즘 2로 수행한다(3.1). 컬럼명이 문자열인 ‘도로’와 ‘지번’을 모두 포함하거나 모두 포함하지 않는 경우는 기타(EA)로 분류하고, ‘도로’만 포함하는 컬럼명은 도로명주소(RA)로, ‘지번’만 포함하는 컬럼명은 지번주소(LA)로 분류한다. 분류한 결과는 전문가 리뷰에 참여한 3명이 분류한 컬럼명과 코헨의 카파 계수(Cohen’s kappa efficient)로 일치도를 검증하고, 최종적으로 컬럼명의 유형을 확정한다(3.2).

표 3. 알고리즘 2: 주소의 유형에 따라 주소와 관련된 컬럼명 분류

Table 3. Algorithm 2: Categorizing column names associated with an address based on the type of address

<b>Input:</b> A Column Set FAC
<b>Output:</b> Get the address-related column names by category
<b>for all</b> $c \in \text{FAC}$ <b>do</b>
contains_road = ‘도로’ in $c$
contains_land = ‘지번’ in $c$
<b>if</b> contains_road and contains_land <b>then</b>
$EA \leftarrow EA \cup \{c\}$
<b>elif</b> contains_road and not contains_land <b>then</b>
$RA \leftarrow RA \cup \{c\}$
<b>elif</b> contains_land and not contains_road <b>then</b>
$LA \leftarrow LA \cup \{c\}$
<b>else</b>
$EA \leftarrow EA \cup \{c\}$
<b>end if</b>
<b>end for</b>

\*The Korean terms “*doro*(도로)” and “*jibun*(지번)” are used to identify column names that contain the Korean words.

네 번째 단계는 주소 데이터셋을 선정하는 과정이다. 2 단계에서 선정한 주소와 관련된 컬럼명을 포함하는 데이터셋이 주소 데이터셋으로 선정되고(4.1), 컬럼명의 분류에 맞게 주소의 유형별 데이터셋이 분류된다(4.2). 이 때, 개별 데이터셋은 도로명주소와 지번주소를 모두 포함할 수 있기 때문에 주소 유형별 데이터셋의 개수는 중복을 포함한다.

마지막 단계는 주소 데이터셋의 특징을 분석한다. 제목, 키워드, 분류체계와 같은 메타데이터가 정제되고, 이를 활용해 주소를 포함하는 데이터셋의 특성을 분석한다.

## IV. 주소 데이터셋의 선정

### 4-1 데이터셋의 수집

데이터셋은 중앙부처와 지방자치단체에서 운영하는 데이터 포털에서 수집한다(2022년 9월 기준, 표 4 참고). 17개의 지방자치단체 중 개별 데이터 포털이 없는 대전광역시와 세종특별자치시, 외부 링크(예: 공공데이터포털)로 데이터셋이 연결된 울산광역시와 경상북도, 데이터셋의 다운로드가 불가능한 광주광역시는 수집대상에서 제외한다. 따라서, 공공데이터포털을 포함한 총 13개의 데이터 포털에서 데이터셋이 수집된다. CSV 또는 XLSX 형식의 데이터셋 중에서 컬럼명의 추출이 가능한 총 123,498개의 데이터셋이 분석 범위이다. 수집한 데이터셋의 규모는 약 13만개의 행과 약 178만개의 컬럼을 갖는다.

표 4. 데이터 포털별로 수집한 데이터셋과 행, 컬럼의 개수  
Table 4. The number of collected datasets and rows, columns per data portal

data portals	# of datasets	total amount of rows	total amount of columns
Gangwon	469	961,952	15,847
Gyeonggi	1,310	11,587,993	19,656
Gyeongnam	2,682	16,923,436	28,728
public data portal	file data	69,143	897,976,558
	standard dataset	8,126	3,964,313
Daegu	8,739	31,423,581	290,088
Busan	2,990	10,314,614	24,713
Seoul	5,356	47,562,896	164,383
Incheon	4,093	13,679,538	50,285
Jeonnam	136	412,929	1,237
Jeonbuk	1,951	5,763,793	16,945
Jeju	650	143,461,117	5,281
Chungnam	80	162,396	2,152
Chungbuk	17,773	125,162,173	223,208
total amount	123,498	1,309,357,289	1,784,722

### 4-2 주소와 관련된 컬럼명의 선정

총 123,498개의 전체 데이터셋 중에서 추출된 모든 컬럼명은 총 1,784,722개다. 이 중에서 표 1의 함수로 정제되어 중복이 제거된 고유한 컬럼명은 총 92,058개다. 한편, 그림 2는 모든 고유한 컬럼명에 대한 문자열 길이의 분포를 나

타낸다. 컬럼명의 길이는 1~30자까지 분포되어 있고, 그 이상은 희박하다. 컬럼명의 길이가 1자이거나 30자 이상인 경우는 이상치로 간주하여 분석의 범위에서 제외한다. 최종적으로 선별된 고유한 컬럼명은 총 91,844개다. 주소와 관련된 컬럼명은 고유한 컬럼명 중에서 표 2의 알고리즘 1로 추출하고, 총 1,018개가 선정된다.

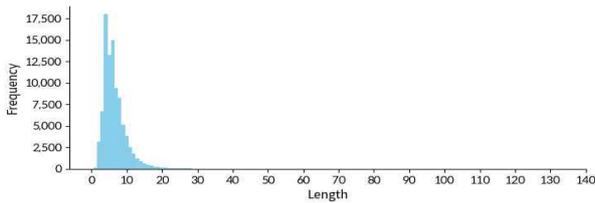


그림 2. 고유한 컬럼명 길이의 분포  
Fig. 2. Distribution of unique column names' lengths

### 4-3 주소와 관련된 컬럼명의 검증

총 1,018개의 주소와 관련된 컬럼명이 주소 값을 갖고 있는 데이터셋을 실제로 추출할 수 있는지 검증을 수행한다. 주소와 관련된 컬럼명으로 5번의 교차 검증을 수행한 결과는 표 5와 같다. 표 5의 개별 지표의 값은 소수점 3자리까지 표기한다. 5번 무작위 추출한 100개의 샘플 데이터에서 약 30~40%의 데이터셋이 주소 값을 포함한다. F1 score는 평균적으로 약 0.885로, 주소 값을 갖고있는 데이터셋의 예측에 높은 성능을 보인다고 평가할 수 있다. 그러나, 주소와 관련된 컬럼명의 추출은 데이터셋이 실제로 주소 값을 포함하는지에 대한 정밀한 검토가 필요하다. 평균적인 정밀도(precision)는 약 0.913으로 높은 예측율을 보이지만, 주소와 관련된 컬럼명에 관련 없는 컬럼명이 포함되어 있어 보완이 필요하다. 반면, 평균적인 재현율(recall)은 약 0.860으로 정밀도보다 비교적 낮은 수준이다. 이는 주소와 관련된 컬럼명이 주소 값을 갖는 데이터셋의 예측에 충분하지 않다는 것을 의미하므로, 주소와 관련된 컬럼명으로 선정되지 않은 컬럼명을 재검토하는 것이 필요하다.

검증 결과를 토대로 주소와 관련된 컬럼명에 대해 전문가 리뷰를 수행한다. 3인의 전문가가 주소와 관련된 컬럼명을 최

표 5. 주소와 관련된 컬럼명으로 예측한 데이터셋의 검증  
Table 5. Validation of datasets predicted by address-related column names

sample	% of address datasets	F1 score	precision	recall
Sample 1	31%	0.909	0.967	0.857
Sample 2	37%	0.846	0.804	0.891
Sample 3	39%	0.876	0.941	0.820
Sample 4	37%	0.931	0.944	0.918
Sample 5	38%	0.861	0.911	0.815

종적으로 선정하기 위해 검토한 기준은 다음과 같다. 첫째, 선정된 주소와 관련된 컬럼명은 다음의 기준으로 제거한다. ‘링크주소’와 같이 동음이의어가 포함된 컬럼명은 제외하고, ‘업소소재지전화번호’와 같이 ‘주소’ 또는 ‘소재지’가 포함되어 있지만, 주소 값이 아닌 다른 값을 포함하는 컬럼명은 제거한다. ‘임시주소’와 같이 모호한 경우는 해당 컬럼명이 포함된 데이터셋의 값을 확인하고 포함 여부를 결정한다. 둘째, 주소와 관련된 컬럼명으로 선정되지 않은 고유한 컬럼명은 휴리스틱한 방식으로 주소와 관련된 컬럼명을 선별한다. 알고리즘으로 선별된 1,018개의 컬럼명 중에서 274개 제외, 53개 추가되어 총 797개의 컬럼명이 최종선정된다.

### 4-4 주소와 관련된 컬럼명의 분류

주소와 관련된 컬럼명은 표 3의 알고리즘2를 사용해 도로명주소와 지번주소, 기타로 분류한다. 총 797개의 컬럼명은 197개의 도로명주소, 132개의 지번주소, 468개의 기타 유형으로 구분된다. 검증은 코헨의 카파 계수를 사용해 전문가 리뷰에 참여한 3명의 결과와 알고리즘의 결과의 일치도를 확인한다. 저자 2명이 분류한 결과에 대한 카파 계수는 약 0.936, 0.958이며, 공공데이터 전문가의 분류 결과에 대한 카파 계수는 0.916이다. 평균적인 카파 계수는 약 0.936으로 상당히 높은 일치율을 보이지만, 정확도를 높이기 위해 불일치한 203건의 컬럼명에 대해 검토를 수행한다. 알고리즘과 전문가 3명의 분류가 일치하지 않은 컬럼명은 의견을 종합한 다음, 최종적인 분류를 선택한다. 대표적으로 기타로 분류되었던 ‘신주소’ 또는 ‘새주소’는 도로명주소로, ‘구주소’는 지번주소로 분류된다. 컬럼명이 분류된 최종 결과는 표 6과 같다.

표 6. 주소 유형별 컬럼명과 데이터셋의 개수  
Table 6. The number of column names and datasets by address type

address type	# of column names	# of datasets
road name address	202	27,559
land lot number address	140	14,932
others	455	27,158
total amount	797	46,207

### 4-5 주소 데이터셋의 선정 결과

총 123,498개의 데이터셋 중에서 총 797개의 주소와 관련된 컬럼명을 포함한 데이터셋이 주소 데이터셋으로 선정된다. 최종적으로 총 123,498개의 데이터셋 중에서 선정된 주소 데이터셋은 총 46,207개로, 전체 공공데이터의 약 37%를 차지한다. 주소 유형별 데이터셋은 도로명주소가 27,559개(22%), 지번주소가 14,932개(12%), 기타가 27,158개(21%)다.

## V. 주소 데이터세트의 분석

### 5-1 분석을 위한 데이터 정제

주소 데이터세트는 분류체계나 키워드와 같이 개별 데이터 세트가 갖고 있는 메타데이터와 함께 분석한다. 이 중 분류체계는 데이터 분석을 위해 추가적인 정제가 필요하다. 데이터 포털마다 서로 다른 분류체계를 갖고 있어 통합적인 분석을 위해 포털의 분류체계를 매핑하는 것이 필수적이다. 강원, 경남, 대구, 부산, 전북, 충청은 공공데이터포털과 유사한 분류체계를 사용하고 있고, 경기와 서울, 인천, 전남, 제주는 개별적으로 정의한 분류체계를 사용한다. 분류체계 매핑의 기준은 다수의 포털이 사용하는 공공데이터포털의 분류체계로 정의한다. 공공데이터포털의 분류체계와 매핑하기 어려운 기타, 인구는 새로운 분류체계로 추가한다. 분류체계를 제공하지 않는 충남의 모든 데이터세트는 기타로 분류한다. 총 13개의 포털에서 매핑된 18개의 통합 분류체계는 다음과 같다:

공공행정(ADMI), 인구(POP), 과학기술(TECH), 교육(EDUC), 교통물류(TRAN), 국토관리(LAND), 농축수산(AGRI), 문화관광(CULT), 법률(LAW), 보건의료(HEAL), 사회복지(WELF), 산업융합(INDU), 식품건강(FOOD), 재난안전(DISA), 재정금융(FINA), 통일외교안보(DIPL), 환경기상(ENVI), 기타(ETC)

### 5-2 데이터세트 단위의 분석 결과

그림 3은 분류체계별 전체 데이터세트 대비 주소 데이터 세트의 비율을 나타낸 결과다. 대부분의 분류체계에서 주소 데이터세트가 차지하는 비율은 0.2 이상으로, 주소는 광범위하게 공공데이터에서 활용되고 있다. 특히, 절반 이상의 데이터세트가 주소를 포함하는 분류체계는 기타(0.9), 식품건강(0.62), 문화관광(0.61), 사회복지(0.58), 보건의료(0.56)이며, 개인의 실생활과 밀접한 분야라는 공통점을 갖고 있다. 주소는 개인의 거소나 특정 객체(예: 건물, 시설물)의 위치를 식별할 수 있는 기본정보이며, 생활 전반과 밀접하게 연관된 행정과 교통, 경제활동, 배송 등의 다양한 영역에서 활용되고 있다. 이런 맥락에서 주소는 다방면의 공공데이터에서 사용되고 있는 핵심 데이터라고 평가할 수 있다.

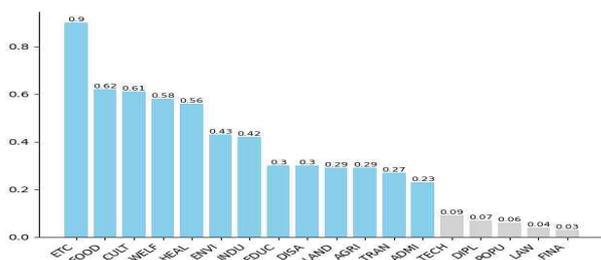


그림 3. 분류체계별 주소 데이터세트의 비율  
Fig. 3. Address dataset ratios by category

한국에서 광범위하게 사용되고 있는 주소 유형은 도로명주소와 지번주소다. 도로명주소는 공공데이터의 약 22%, 지번주소는 공공데이터의 약 12%를 차지할 정도로, 공공데이터에서 이원적인 주소체계가 폭넓게 활용된다(표 6 참고). 도로명주소가 2014년부터 전면 시행된 이후로 약 10년이 지난 현재는 도로명주소의 정착과 더불어 지번주소도 함께 활용되는 추세를 보인다.

도로명주소와 지번주소의 활용 분야를 파악하기 위해 분류체계별 주소 데이터세트의 비율을 확인한 결과는 그림 4와 같다. 대부분의 분류체계에서 도로명주소와 지번주소가 활용되고 있지만, 주목해야 할 점은 도로명주소와 지번주소의 비율이 크게 차이나지 않는 분류체계다. 교통물류(0.02p%), 재난안전(0.03p%), 공공행정(0.03p%), 국토관리(0.04p%)는 0.05p% 이하의 극소한 차이를 보인다. 이 분류체계의 지번주소 데이터세트의 상위 키워드 빈도를 분석하면, ‘도시광역철도(37건)’, ‘버스(20건)’, ‘주차장(12건)’(교통물류), ‘민방위(73건)’, ‘급수(38건)’, ‘지진(31건)’, ‘화재(28건)’(재난안전), ‘사회서비스(10건)’, ‘주민편의(8건)’, ‘관공서(6건)’(공공행정), ‘하수(24건)’, ‘아파트(23건)’, ‘공동주택(18건)’, ‘공개공지(5건)’(국토관리)이 두드러진다. 이와 같은 결과는 도로명주소와 지번주소가 갖고 있는 내재적 특성과 연관지어 해석할 수 있다. 건물과 도로 기반으로 부여되는 도로명주소는 건물 찾기에 유리하지만, 시설물(‘하수’, ‘민방위’, ‘급수’)이나 부동산거래(‘아파트’, ‘공동주택’), 건물이 없는 토지(‘주차장’, ‘공개공지’, 산악지역(‘지진’, ‘화재’) 등은 도로 또는 건물이 없기 때문에 주소 부여에 한계가 있다. 반면, 토지를 기반으로 부여되는 지번주소는 도로명주소의 부여가 어려운 지역에도 주소를 부여하는 것이 가능하다. 이런 특성 때문에 도로명주소가 부여되기 어려운 분야에서 지번주소가 광범위하게 이용되는 것으로 보인다. 도로명주소의 한계를 보완하기 위해 사물주소나 국가지점번호 등을 도입하고 있지만, 지번주소가 대안으로 폭넓게 활용되고 있다고 추측할 수 있다.

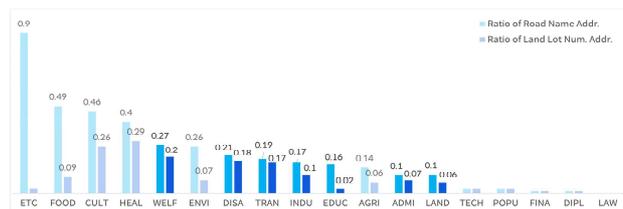


그림 4. 분류체계별 도로명주소와 지번주소를 포함하는 데이터세트의 비율  
Fig. 4. Ratios of datasets containing road name address and land lot number address

그림 5는 주소 데이터세트에서 주소가 어떻게 활용되는지 조건별로 분석한 결과다. 데이터세트의 분류 기준은 도로명주소만 포함하는 경우, 지번주소만 포함하는 경우, 도로명주소와 지번주소를 함께 포함하는 경우, 주소 유형을 판별할 수 없는 경우로 구분한다. 주소 유형을 알 수 없는 데이터세트

(39%)를 제외하고, 도로명주소만 포함된 데이터세트는 약 28%(13,133건), 지번주소만 포함된 데이터세트는 약 1%(506건), 도로명주소와 지번주소가 함께 포함된 데이터세트는 약 31%(14,426건)다. 주목할 점은 도로명주소만 존재하는 데이터세트만큼 도로명주소와 지번주소를 모두 포함하는 데이터세트가 많다는 것이다. 이러한 결과는 도로명주소가 공공데이터에서 지배적으로 사용되고 있지만, 지번주소는 도로명주소를 보완하는 방향으로 함께 사용되고 있다고 해석할 수 있다.

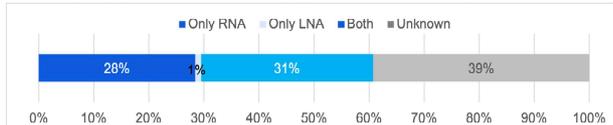


그림 5. 조건별 주소 데이터세트의 비율  
Fig. 5. Ratios of address datasets by conditions

### 5-3 컬럼명 단위의 분석 결과

주소 데이터세트에서 주소와 관련된 컬럼명이 어떻게 사용되고 있는지 파악하기 위해 총 979개의 주소와 관련된 컬럼명을 분석한다. 표 7은 유형별로 개별 데이터세트에서 출현 빈도가 높은 상위 10개의 컬럼명을 나열한 결과다. 공공데이터에서 빈번하게 등장하는 컬럼명은 ‘소재지도로명주소’, ‘소재지지번주소’, ‘소재지전체주소’, ‘소재지도로명’, ‘소재지지번’ 등과 같이 소재지라는 용어를 포함하는데, 이는 특정 기관이나 업체 등의 위치를 파악하는 데 필요한 기본정보로 활용되고 있다고 판단할 수 있다. 한편, 주소 또는 도로명주소, 지번주소를 표현하는 컬럼명은 매우 다양하다. 현재 공공데이터에서 가장 많이 활용되는 ‘소재지도로명주소’와 ‘소재지지번

표 7. 상위 출현빈도의 주소와 관련된 컬럼명  
Table 7. Top-frequency address-related column names

Type	Top-frequency column names
All types	소재지도로명주소(10,553), 소재지지번주소(9,190), 도로명전체주소(8,149), 소재지전체주소(8,127), 주소(7,682), 도로명주소(6,464), 소재지(6,378), 지번주소(4,490), 영문상호주소(1,433), 소재지주소(773) 등
Road name address	소재지도로명주소(10,553), 도로명전체주소(8,149), 도로명주소(6,464), 소재지도로명(415), 교육장도로명주소(388), 사업장도로명주소(379), 업소도로명주소(339), 실험실도로명특수주소(242), 실험실도로명주소코드(237), 시작지점도로명주소(199) 등
Land lot number address	소재지지번주소(9,190), 지번주소(4,490), 소재지지번(302), 업소지번주소(230), 시작지점소재지지번주소(199), 종료지점소재지지번주소(199), 사업장지번주소(195), 차고지지번주소(178), 영업소지번소재지(46), 사업장지번상세주소(28) 등
Others	소재지전체주소(8,127), 주소(7,692), 소재지(6,378), 영문상호주소(1,433), 소재지주소(773), 상세주소(291), 공장대표주소(289), 실험실특수주소(242), 실험실특수주소호(237), 사무소주소(214) 등

\*In order to accurately represent the column names used in public data, the names were expressed in Korean.

주소’는 공공데이터 개방 표준[40]에 정의된 컬럼명으로, 공공데이터의 컬럼명 표준화를 위한 결과로 평가할 수 있다. 그러나, 이외에도 상당 수의 주소와 관련된 컬럼명이 다양하게 활용되는 것이 현실이다. 다양한 주소와 관련된 컬럼명의 사용은 공공데이터의 주소 연계와 통합을 복잡하게 만들고, 주소 데이터의 활용을 제한하는 요인이 될 수 있다.

## VI. 결 론

본 논문은 국내 공공데이터에서 주소가 포함된 데이터세트를 식별하고, 공공데이터에서 주소가 활용되는 현황을 분석했다. 총 13개의 국내 데이터 포털에서 123,507개의 데이터세트와 91,844개의 고유한 컬럼명을 수집하고, 알고리즘과 전문가 리뷰를 통해 총 797개의 주소와 관련된 컬럼명을 추출했다. 이 컬럼명을 포함하는 총 46,207개의 데이터세트가 주소를 포함하는 것으로 파악됐다. 결과적으로 전체 공공데이터의 약 37%가 주소를 포함하고 있으며, 특히 개인의 실생활과 밀접한 분야에서 주소 데이터가 핵심 정보로 활용되고 있었다. 한편, 도로명주소가 공공데이터에서 광범위하게 사용되지만, 지번주소도 도로명주소를 보완하는 방향으로 함께 활용되고 있었다.

본 논문의 결과는 두 가지의 주요 시사점을 제공한다. 첫째, 주소와 관련된 컬럼명은 매우 다양한 형태로 사용되고 있어 주소 데이터의 활용을 제약할 수 있다. 공공데이터에서 주소는 특정 객체의 소재를 파악할 수 있는 기초 정보로서 재사용성이 높은 유용한 데이터다. 그러나, 연구 결과에서 확인할 수 있듯이, 공공데이터에서 주소가 포함된 데이터세트를 식별하는 과정은 쉽지 않고 주소와 관련된 컬럼명은 매우 다양하다. 따라서 주소와 관련된 컬럼명에 대한 명확한 표준이 수립되고, 공공기관이 이를 준수하는 정책이 뒷받침되어야 한다. 둘째, 도로명주소가 전면적으로 도입된 지 10년이 지났음에도 지번주소는 여전히 다방면에서 사용되고 있다. 연구 결과와 같이 토지 또는 시설물과 관련된 분야는 여전히 지번주소가 도로명주소를 보완하는 방향으로 함께 활용된다. 이런 관점에서 지번주소를 폐기해야 할 대상으로 보기 보다, 도로명주소와 지번주소가 함께 사용될 수 있는 방향을 고려해야 한다. 특히, 데이터 활용 관점에서 두 가지의 주소체계를 연계하고 활용할 수 있는 방안이 마련될 필요가 있다. 정부 부처 사이의 경계를 넘어 도로명주소와 지번주소의 연계 지침을 마련하고, 범정부 차원에서 주소 데이터의 지속적인 품질 향상과 갱신을 보장하는 제도가 필요하다.

본 논문에서 도출한 797개의 주소와 관련된 컬럼명은 주소 값을 자동으로 추출하기 위한 기반이 될 수 있다. 이 컬럼명 목록을 기반으로 대규모의 공공데이터에서 주소 값을 포함하는 데이터세트를 빠르게 식별하고, 주소 데이터 정제가 적용될 컬럼을 효과적으로 선별할 수 있다. 더불어 광범위한 공공데이터에서 주소 값에 자동으로 접근할 수 있어 다양한

유형과 형식의 한국어 주소에 대한 학습 데이터셋 구축이 가능하다.

그러나, 본 논문은 주소와 관련된 컬럼명을 선정하는 과정에서 전문가의 전문지식과 판단에 의존했다는 한계가 있다. 이를 보완하기 위해 향후 연구는 자연어 처리 기술을 활용한 자동화 알고리즘을 개발하고, 대규모 데이터셋에서 확장성 있게 적용가능한지 검토한다. 이를 기반으로 주소 데이터 중심의 공공데이터 연계과 융합방안을 검토한다.

## 감사의 글

이 논문은 2023학년도 중앙대학교 CAU GRS 지원에 의하여 작성되었음

## 참고문헌

- [1] H. Kim, "Quality Evaluation of the Open Standard Data," *Journal of the Korea Contents Association*, Vol. 20, No. 9, pp. 439-447, September 2020. <https://doi.org/10.5392/JKCA.2020.20.09.439>
- [2] Ministry of the Interior and Safety, Act on Promotion of the Provision and Use of Public Data, Author, Sejong, Amended by Act No. 19408, May 2023.
- [3] Ministry of the Interior and Safety, The 1st Master Plan for the Utilization of Address Information (2022~2026), Author, Sejong, 2022.
- [4] G. Kim, K. Kim, D. Kim, and S. Park, "A Study on Construction of Building Inventory Map Using Road Name Address Map and Registered Building Data," in *Proceedings of the Korean Society of Civil Engineers Convention*, Pyeongchang, pp. 133-134, October 2019.
- [5] D.-W. Kim, S.-M. Oh, H.-Y. Yang, and S.-E. Lee, "Improvements on Public Buildings Energy Information System," *Journal of the Architectural Institute of Korea*, Vol. 36, No. 9, pp. 109-118, September 2020. <https://doi.org/10.5659/jaik.2020.36.9.109>
- [6] J.-W. Lee, S.-H. Kim, S.-E. Lee, and D.-W. Kim, "Address-Based Integration of Building Open Data Using OpenAPI -A Case Study of the Korean Public Institution Building List-," *Journal of the Architectural Institute of Korea*, Vol. 37, No. 4, pp. 11-22, April 2021. <https://doi.org/10.5659/JAIK.2021.37.4.11>
- [7] J. Y. Kim, "Exploring Method to Integrate and Utilize Open Data and Address of Things," *Journal of Korean Society for Geospatial Information Science*, Vol. 29, No. 1, pp. 43-54, March 2021. <https://doi.org/10.7319/kogsis.2021.29.1.043>
- [8] S. Saravit, J.-H. Bae, K.-H. Lee, and W.-S. Cho, "Global Address Data Quality Verification and Improvement Techniques using Deep Learning," *The Journal of Korean Institute of Information Technology*, Vol. 20, No. 12, pp. 15-24, December 2022. <https://doi.org/10.14801/jkiit.2022.20.12.15>
- [9] J. Y. Kim, H. J. Kim, and J. W. Lee, "Street Name Address Parsing Model based on Bidirectional Gate Recurrent Unit through Automatic Construction of Training Data," *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*, Vol. 41, No. 5, pp. 301-310, October 2023. <https://doi.org/10.7848/ksgpc.2023.41.5.301>
- [10] J.-W. Lee, S.-H. Kim, D.-W. Kim, and S.-E. Lee, "Errors in Address Data for Address Based Integration of Building Open Data Using Open API -A Case Study Pertaining to the Address List of Korean Public Facilities-," *Journal of the Architectural Institute of Korea*, Vol. 38, No. 4, pp. 25-34, April 2022. <https://doi.org/10.5659/jaik.2022.38.4.25>
- [11] H. Kim, J. Yang, H. Kim, H. Park, E. Seon, and C. Song, A Study on Address Sharing System and Address-Based Knowledge Graph, Ministry of the Interior and Safety, Sejong, December 2022.
- [12] S. Coetzee, A. K. Cooper, M. Lind, M. M. Wells, S. W. Yurman, E. Wells, ... and M. J. Nicholson, "Towards an International Address Standard," in *Proceedings of the 10th International Conference for Spatial Data Infrastructure (GSDI 10)*, St. Augustine, Trinidad and Tobago, February 2008.
- [13] ISO (International Organization for Standardization), Addressing - Part 1: Conceptual Model, Author, Geneva, Switzerland, ISO 19160-1:2015, 2015.
- [14] C. Song, H. Park, and H. Kim, "Analysis and Understanding of the ISO 19160-1:2015 Conceptual Model," *Journal of Digital Contents Society*, Vol. 25, No. 5, pp. 1359-1372, May 2024. <https://doi.org/10.9728/dcs.2024.25.5.1359>
- [15] GOV.UK. Identifying Property and Street Information [Internet]. Available: <https://www.gov.uk/government/publications/open-standards-for-government/identifying-property-and-street-information>.
- [16] G. Gander, "Making Votes Count," *Land Journal*, pp. 18-19, May/June 2015.
- [17] F. Santos, S. Conti, and A. Wolters, "A Novel Method for Identifying Care Home Residents in England: A Validation Study," *International Journal of Population Data Science*, Vol. 5, No. 4, 09, September 2021. <https://doi.org/10.23889/ijpds.v5i4.1666>

- [18] H. Zhang, A. Casey, I. Guellil, V. Suárez-Paniagua, C. MacRae, C. Marwick, ... and B. Alex, "FLAP: A Framework for Linking Free-Text Addresses to the Ordnance Survey Unique Property Reference Number Database," *Frontiers in Digital Health*, Vol. 5, 1186208, November 2023. <https://doi.org/10.3389/fgdh.2023.1186208>
- [19] C. Stadler, J. Lehmann, K. Höffner, and S. Auer, "LinkedGeoData: A Core for a Web of Spatial Open Data," *Semantic Web*, Vol. 3, No. 4, pp. 333-354, 2012. <https://doi.org/10.3233/sw-2011-0052>
- [20] H. Kim, "Interlinking Open Government Data in Korea using Administrative District Knowledge Graph," *Journal of Information Science Theory and Practice*, Vol. 6, No. 1, pp. 18-30, March 2018. <https://doi.org/10.1633/JISTAP.2018.6.1.2>
- [21] D. Jeong, S. Hwangbo, Y. Lee, B. Lee, and H. Lee, A Study on the Response Plan for Address Standardization of ISO T211, Ministry of Public Administration and Security, Seoul, August 2011.
- [22] D. Cha, B. Lee, Y. Shin, B. Kwak, J. Yoo, and I. Na, A Study on the Response Plan for International Standardization of Location Finding Measures, Ministry of Public Administration and Security, Seoul, December 2012.
- [23] D. Cha, B. Lee, J. Kim, W. Kim, S. Park, and I. Na, A Study on the Response Plan for International Standardization of Location Finding Measures (2013), Ministry of Public Administration and Security, Seoul, December 2013.
- [24] M. Lee, B. Lee, B. Kwak, Y. Jung, and J. Choi, A Study on the Response Plan for International Standardization of Location Finding Measures (2014), Ministry of the Interior, Seoul, 11-17440000-000028-10, December 2014.
- [25] J. Kim, B. Lee, B. Kwak, Y. Jeong, J. Lee, and J. Choi, A Study on the Response Plan for International Standardization of Location Finding Measures (2015), Ministry of the Interior, Seoul, 11-1740000-000199-01, December 2015.
- [26] B. Lee, B. Kwak, D. Hwang, J. Kim, J. Lee, and J. Yoon, A Study on the Response Plan for Address Standardization (2017), Ministry of the Interior, Seoul, 11-1741000-000057-11, December 2017.
- [27] J. Kim, J. Choi, A. Kang, B. Lee, C. Kang, Y. Na, and T. Kim, A Study on the Response Plan for Address Standardization (2019), Ministry of the Interior and Safety, Sejong, 11-1741000-000229-01, December 2019.
- [28] J. Lee, J. Choi, H. Kim, S. Bae, Y. Na, and T. Kim, A Study on the Response Plan for Address Standardization (2020), Ministry of the Interior and Safety, Sejong, 11-1741000-000338-10, December 2020.
- [29] S. Won, J. Choi, S. Kim, B. Kim, S. Lee, and Y. Lee, A Study on the Response Plan for Address Standardization (2021), Ministry of the Interior and Safety, Sejong, 11-1741000-000416-01, February 2022.
- [30] S. Won, J. Choi, S. Kim, J. Park, B. Kim, M. Kim, and Y. Lee, A Study on the Response Plan for Address Standardization (2022), Ministry of the Interior and Safety, Sejong, 11-1741000-000338-10, December 2022.
- [31] S.-H. Kim, B.-E. Kim, and S.-H. Won, "The Trend of International Address Standardization and Implications -With a Focus on ISO 19160-2-," *Journal of Cadastre & Land InformatiX*, Vol. 52, No. 1, pp. 57-68, June 2022. <https://doi.org/10.22640/lxsiri.2022.52.1.57>
- [32] S. C. Yang, "A Study on the Road Name Address Assignment Method for Each Building in a Building Group," *Journal of the Korean Society of Cadastre*, Vol. 39, No. 1, pp. 69-80, April 2023. <https://doi.org/10.22988/ksc.2023.39.1.006>
- [33] S. W. Hwang Bo, "A Study on the Development Method of the Detailed Road Name Address," *Journal of the Korean Society of Cadastre*, Vol. 33, No. 3, pp. 65-74, December 2017. <http://doi.org/10.22988/ksc.2017.33.3.005>
- [34] S. Yang, "A Study on Improvement of the Registration System for Address Information Reference Object of Addressing Road Name Address," *Journal of Cadastre & Land InformatiX*, Vol. 51, No. 2, pp. 21-34, December 2021. <https://doi.org/10.22640/LXSIRI.2021.51.2.21>
- [35] J. Go, Y. Lee, and B. Lee, "Study on the Strategy for Corresponding the International Address Standard Trend," *Journal of Cadastre & Land InformatiX*, Vol. 42, No. 1, pp. 39-58, June 2012.
- [36] S. W. Hwang Bo and Y. J. Kim, "A Study on a Scheme of the Position Indication in the Street Addressing of Non-Assign Area," *Journal of the Korean Society of Cadastre*, Vol. 26, No. 2, pp. 149-169, December 2010. <http://doi.org/10.22988/ksc.2010.26.2.011>
- [37] J. Y. Kim and S. C. Yang, "Definition of Address of Things Based on Standards in Preparation for Revision of Road Name Address Act," *Journal of Korean Society for Geospatial Information Science*, Vol. 28, No. 2, pp. 49-57, June 2020. <http://dx.doi.org/10.7319/kogsis.2020.28.2.049>
- [38] J. Y. Kim and J. W. Lee, "Analysis of Address of Things(AOT) Assignment and its Policy Implications," *Journal of Korean Society for Geospatial Information Science*, Vol. 31, No. 2, pp. 13-24, June 2023.

<https://doi.org/10.7319/kogsis.2023.31.2.013>

- [39] S. C. Yang, "A Study on Construction of Hyper-Connected Building Database Using Public Data," in *Proceedings of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography Convention*, Online, pp. 82-84, July 2020.
- [40] MOSPA (Ministry of Security and Public Administration), *Open Data Standard*, Author, Seoul, MOSPA Notice No. 2014-46, October 2014.



**박하람(Haram Park)**

2023년 : 중앙대학교 문헌정보학과  
정보학 석사

2017년~2021년: 중앙대학교 사회학과 (문학사)  
2021년~2023년: 중앙대학교 문헌정보학과 정보학 석사  
2023년~현 재: 중앙대학교 문헌정보학과 정보학 박사과정  
※ 관심분야 : 지식그래프, 메타데이터, 공공데이터 등



**김학래(Haklae Kim)**

2010년 : 아일랜드 국립대학교  
(공학박사)

2004년~2009년: Digital Enterprise Research Institute,  
Ireland  
2009년~2016년: 삼성전자  
2017년~2019년: 한국과학기술정보연구원  
2019년~현 재: 중앙대학교 문헌정보학과 교수  
※ 관심분야 : 지식그래프, 인공지능, 데이터 사이언스 등