

산업입지정보시스템 공장정보 개선에 관한 연구

최유정¹, 임재덕², 김성건^{3*}

¹(주)지앤티솔루션 기술연구소 연구원, ²(주)지앤티솔루션 기술연구소 책임연구원, ³(주)지앤티솔루션 기술연구소 연구소장

Improvement of Factory Data in Industrial Land Information System

Yu-Jeong Choe¹, Jae-Deok Lim², Seong-Geon Kim^{3*}

¹Researcher, Technical Research Center, GnT Solution Inc.

²Senior Researcher, Technical Research Center, GnT Solution Inc.

³Research Director, Technical Research Center, GnT Solution Inc.

요약 산업입지정보시스템에서 제공하는 공장정보는 한국산업단지공단으로부터 원시데이터를 받아 필터링 과정을 거친 후 등록되기 때문에 새로운 공장 정보 갱신이 느린 특징이 있다. 본 연구에서는 산업입지정보시스템 공장 정보 갱신 문제를 해결하고자, 비교적 갱신 주기가 빠른 도로명 주소의 건물 데이터와 부동산의 건물 데이터를 이용하여 기존의 산업입지정보시스템 공장정보와 비교하고 새로운 공장정보를 추출하였다. 속성정보 매칭을 수행하고, 속성정보의 누락으로 매칭할 수 없는 공장의 경우에는 공간객체 매칭을 수행하였다. 위 과정에서 점 데이터와 다각형 데이터와 같이 형식이 다른 공간객체를 비교하는 방안을 제시하였다. 제안된 공간분석 방법의 정확도 평가 결과 약 79%의 정확도를 보였으며, 위 매칭 기법을 활용하여 도로명주소건물데이터, 부동산건물데이터와 산업입지정보시스템 공장정보의 융합 가능성을 확인하였다.

주제어 : 산업입지정보시스템, 공장정보, 부동산건물데이터, 도로명주소건물데이터, 속성추출, 공간분석

Abstract The factory information provided by the Industrial Location Information System (ILIS) is provided as raw data by the Korea Industrial Complex Corporation and registered after a filtering process, so the new factory information update is slow. In this study, to solve the problem of updating factory information of industrial location information system, using building data of road name address with relatively fast renewal cycle and building data of real estate, we compared the factory information of existing ILIS and extracted new factory information. In the process of comparison, a method was proposed to compare spatial objects of different types with point data and polygon data. Attribute information matching and object matching were performed, and attribute values of new factory information were extracted. The accuracy evaluation of the proposed spatial analysis method showed 79% accuracy, and the above matching technique was used to confirm the possibility of convergence of road name address data, real estate data and factory information of ILIS.

Key Words : Location of Industry, Industrial Complex, Smart Industrial Complex, Web GIS, Information Provider System

*This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 20AUDP-B119587-05)

*Corresponding Author : Seong-Geon Kim(sgkim@gntsolution.com)

Received September 18, 2020

Revised October 7, 2020

Accepted October 20, 2020

Published October 28, 2020

1. 서론

산업입지정보시스템(Industrial Land Information System, ILIS) 공장정보는 전국 산업입지 현황, 위치 및 직원 수와 같은 자세한 정보를 체계적으로 제공하고 있다[1,2]. 이러한 공장정보의 활발한 이용을 촉진시키기 위해서는 데이터 최신성과 정확성이 확보되어야 한다. 그러나 산업입지정보시스템 공장정보는 한국산업단지공단으로부터 제공 받아 필터링 과정을 거친 후 등록되기 때문에 갱신주기가 1년 이상으로 비교적 길고, 현재 기준 최신 데이터가 2018년인 단점이 있다. 또한 공장정보가 누락되거나 실제와 불부합하면 산업보상 업무에 문제가 생기므로 공장정보 최신화 및 갱신이 필요하다[3]. 한편 인터넷건축 행정시스템, 수치지도관리시스템, 한국토지정보시스템, 부동산시스템, 내비게이션시스템, 도로명주소시스템과 같은 건축물 정보가 입력되는 시스템 정보 갱신을 개선한 연구가 진행되었지만, 산업입지정보시스템과 관련한 데이터 갱신에 관한 연구는 부족한 실정이다[4-11].

산업입지정보시스템 공장정보의 최신화를 위하여 공장의 자세한 속성을 제공하는 다른 공공데이터를 조사 분석 및 검토해 본 결과, 갱신 주기가 빠르고 정확한 부동산건물데이터와 도로명주소건물데이터가 산업입지정보시스템 공장정보를 갱신하는데 적합하였다. 위 두 데이터를 이용하여 공장정보를 속성정보로 우선 매칭하여 신규 공장정보를 찾았으며, 공간객체 매칭 방법을 이용하여 나머지 누락된 공장정보를 추가로 추출하였다. 신규로 찾아낸 공장정보가 산업입지정보시스템에서 누락된 공장정보가 맞는지 정확도 분석을 통하여 일치율이 의미있는지도 확인하였다.

따라서 본 연구에서는 산업입지정보시스템 공장정보를 한국산업단지공단에서 제공하는 정보뿐만 아니라, 정부에서 제공하는 공공데이터인 부동산건물데이터와 도로명주소건물데이터를 이용하면 누락된 공장정보를 찾고 최신화가 가능한 것을 확인하였다.

2. 현황 분석

현재 산업입지정보시스템에서 제공하는 공장정보는 전국의 공장이름, 전화번호, 주소뿐만 아니라 산업코드, 공장면적 및 직원 수를 점 데이터 표시하여 제공하고 있다. 이렇게 제공되는 공장정보는 한국산업단지공단으로

부터 제공받고 있으며, 일일이 검토 과정을 거친 후 등록되기 때문에 과다한 인력 비용 발생 및 갱신 주기도 느린 단점이 있다. 또한 갱신이 필요한 공장정보를 현행화하는데 시간적 소모뿐만 아니라 참조하는 데이터가 한국산업단지공단에만 의지하고 있기 때문에 공장정보를 비교 검토하기에도 어려움이 있는 실정이다. 따라서 공장정보를 다른 공공데이터에서 제공하는지 조사 및 분석하였고, 이러한 참조데이터에서 산업입지정보시스템 공장정보를 추가 및 갱신할 수 있는지 검토하여 보았다. 또한 정확한 공장정보를 비교하기 위해서는 적합한 참조 데이터 선택과 비교 방법도 필수적이기 때문에 참조할 수 있는 데이터 현황과 비교 방법 현황도 분석하여 보았다.

2.1 참조 데이터 현황

공장정보를 제공하는 데이터를 조사한 결과, 공공데이터로 제공하는 정보 중 산업입지정보시스템 공장정보를 갱신 및 최신화 가능한 것으로 보이는 참조데이터는 부동산건물데이터, 도로명주소건물데이터, 연속수치지도데이터와 연속수치지형도데이터가 있는 것으로 조사 및 분석되었다.

2.1.1 부동산건물데이터

부동산건물데이터는 국토교통부 국가공간정보포털 오픈마켓에서 제공하고 있다. 다각형 형태의 건물객체를 각 시도별로 제공하며, 속성정보로는 건물이름, 건물지번주소, 건물용도, 건물면적을 포함한 30개 정도 정보를 제공하고 있다. 또한 갱신 주기는 일년에 2회 데이터를 공개하며, 현재 2020년 4월이 최신데이터이다. Fig. 1은 부동산건물데이터의 예시이다.

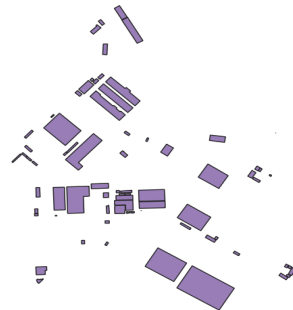


Fig. 1. Example of real estate data

2.1.2 도로명주소건물데이터

도로명주소건물데이터는 행정안전부 주소정책과에서 제공한다. 부동산건물데이터와 마찬가지로 다각형 형태

의 건물을 시도별로 제공하며, 속성정보로는 건물 시군구 코드, 도로명코드, 건물본번, 건물부번을 포함한 총 20개 정보를 제공하고 있다. 각 지방 자치단체에서 매 달 정보가 등록되기 때문에 가장 최신데이터인 장점이 있다. Fig. 2는 도로명주소건물데이터 일부분이다.



Fig. 2. Example of road name address data

2.1.3 연속지적도데이터

연속지적도데이터는 국토교통부 국가공간정보포털 오픈 마켓에서 각 시도별로 제공한다. 연속지적도의 경우 전산화된 지적도 및 임야도의 도면상 경계점을 연결하여 연속된 형태로 작성하여 정확한 도면을 보여주지만, Fig. 3과 같이 건물 객체가 따로 존재하지 않으며 건물 일련 번호만 부여하고 이름과 같이 공장과 관련한 자료를 포함하고 있지 않아 산업입지정보시스템 공장정보 참조데이터로는 부적합하다.

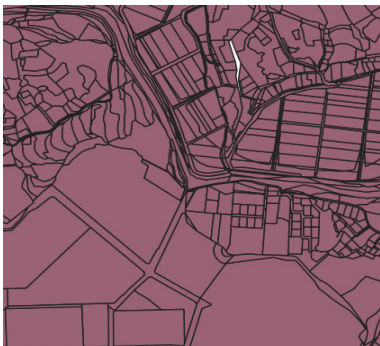


Fig. 3. Example of continuous cadastral map data

2.1.4 연속수치지형도데이터

연속수치지형도데이터에서 제공하는 건물정보는 수치지형도 2.0 건물 레이어를 통합한 정보로 다각형 형태 건물을 시도별로 제공한다. 속성정보로는 법정동코드(bjcd),

건물종류(kind), 건물용도(serv), 건물이름(anno), 건물층수(nmly), 제작정보(jj_code)로 6가지를 제공하고 있다. 산업입지정보시스템 공장정보 최신화를 위한 필수 속성 정보인 상세 건물 주소를 제공하지 않아 산업입지정보시스템 공장정보 참조데이터로 이용하기는 어려운 것으로 분석되었다. Table 1은 연속수치지형도 속성테이블 일부이다.

Table 1. Example attributes of digital map

bjcd	kind	serv	anno	nmly	jj_code
4420035000	BDK004	BDS015	T.O.P. Co., Ltd.	1	3404034010008
4155038000	BDK004	BDS015	Inopra Co., Ltd.	1	3122038020002
4155034000	BDK000	BDS015	creative industries co., ltd.	1	3122034020003
4159033000	BDK004	BDS015	Urae Iron &Steel Co., Ltd.	1	3124035010006

위와 같이 공장정보를 포함한 건물데이터를 분석한 결과 산업입지정보시스템 공장정보 갱신 및 최신화를 위한 참조데이터로 부동산건물데이터와 도로명주소건물데이터가 적합하다고 판단하였다.

2.2 데이터 비교 방법 현황

서로 다른 공간정보 데이터의 비교와 갱신 방법 현황을 검토한 결과, 크게 속성정보 매칭과 공간객체 매칭 방법이 있었다. 속성정보 매칭은 같은 속성 값을 지닌 열을 추출하여, 각 열에 있는 값의 일치율을 계산하여 비교한다. 박슬아 등은 서울시 관악구를 대상으로 최신 정보에 근거한 갱신 작업을 수행하였다. 도로명주소건물데이터와 연속수치지형도 데이터를 이용하여 같은 속성 값인 건물의 용도 값을 이용하여 건물 객체를 갱신하였다[12].

공간객체 매칭은 지도 평면에 띄웠을 때 겹치는 지의 여부를 판단하여 비교하는 방법이다. 김성삼 등은 수치지형도 공간정보를 갱신하고 개선하기 위하여 LiDAR자료를 이용하여 건물 객체를 갱신하였다[13]. 김효중 등은 비교적 갱신주기가 빠른 건축행정정보시스템에 등록되어 있는 건물정보와 연속수치지형도를 비교하여 항공사진과 현지측량의 대안으로 인터넷건축행정시스템의 가능성을 제시하였다[14]. 염준호 등은 서로 다른 건물 객체를 매칭하기 위하여 도로명주소지도와 수치지형도를 비교하고 다대다 매칭을 수행하여 95% 이상의 정확도를 확보하였다[15]. 위 연구들은 모두 다각형 대 다각형을 비교한 연

구로 같은 공간 형태를 지닌 데이터를 비교하였다.

공간형태가 다른 데이터를 비교할 경우, Fig. 4와 같이 동일한 건물 객체를 표현하더라도 매칭이 되지 않는 문제가 있다. 따라서 정확한 비교를 위해서는 같은 공간 객체 형태로 맞춰야 한다. 대표적인 방법으로는 주로 다각형을 점으로 바꾸는 centroid 방법을 이용한다[16,17]. 위 방법은 데이터의 용량을 줄이고 다각형 객체의 무게 중심을 표현하는 장점이 있지만 2차원의 다각형 정보를 점으로 줄이게 되므로 점의 대표성 문제와 위치적 오차를 포함하는 한계가 있다. 이러한 centroid 방법을 보완하고자, Juan 등은 다각형 형태의 공간 정보의 꼭짓점과 테두리를 추출하여 데이터를 비교하였지만 이 방법 또한 다각형 객체를 점으로 추출하기 때문에 오차를 포함하는 단점이 있다[18].

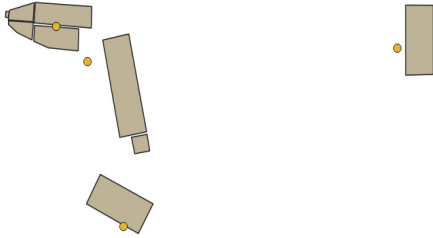


Fig. 4. Example of inconsistency due to object type differences.

위와 같이 속성정보 매칭과 공간객체 매칭 방법을 검토한 결과, 부동산건물데이터와 도로명주소건물데이터에서 제공하는 공장주소와 이름을 이용하여 산업입지정보시스템 공장정보와 속성정보 매칭이 일부 가능한 것으로 분석되었다.

그러나 공간객체 매칭의 경우, 산업입지정보시스템 공장정보는 점으로 표현되지만, 부동산건물데이터와 도로명주소건물데이터는 다각형으로 표현되어 형태가 다른 문제가 있다. 위 언급한 centroid 방법이나 point 추출 방법으로 수행하게 되면 공장건물을 점으로 변환하는 과정에서 위치정보가 손실되기 때문에 다각형을 점으로 전환하는 방법보다는 점 데이터를 다각형 데이터로 변환하는 것이 필요한 것으로 분석되었다.

3. 연구 방법 및 내용

기존 연구를 바탕으로 세 종류의 공장정보를 비교하여

누락된 신규 공장을 추출하는 연구를 수행하였다. 최신화가 필요한 산업입지정보시스템 공장정보(Target data)와 참조할 수 있는 부동산건물데이터(Reference data 1)와 도로명주소건물데이터(Reference data 2)를 각각 전처리한 뒤 두 가지 단계로 매칭하였다. 우선 1단계로 Fig. 5와 같이 각 공장정보 주소와 공장명으로 매칭하였다. 위 과정으로 매칭되지 않는 공장을 맞추기 위하여 다음 단계로 Fig. 6과 같이 공간 분석으로 매칭하고 최종적으로 신규 공장 정보를 추출하여 속성 값을 확인하였다.

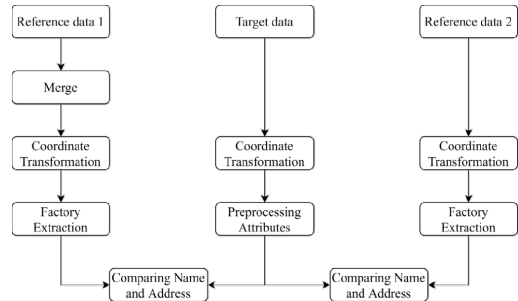


Fig. 5. Flow chart-1

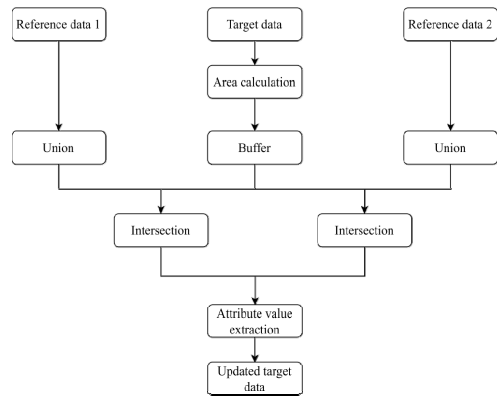


Fig. 6. Flow chart-2

3.1 부동산건물데이터 전처리

부동산건물데이터는 국가공간정보포털 오픈 API로부터 획득하였다. 부동산건물데이터는 건물의 소유주를 바탕으로 제작되기 때문에 건물의 소유주가 개인인 일반 GIS와 아파트와 같이 건물의 소유주가 여러명인 집합 GIS로 각각 내려 받았다. 일반 GIS와 집합 GIS의 건물을 Merge 기법을 이용하여 합친 후 좌표계를 확인하였다. 제공하는 좌표계는 EPSG: 5174로 Bessel 타원체를 기반으로 하고 있다. 그러나 Bessel 타원체는 이격이 존재하므로 투영체계 계수를 추가하여 이용하였다. 이용한

투영체계 계수는 아래와 같다.

<투영체계 계수>

towgs84=-115.80,474.99,674.11,1.16,-2.31,-1.63
.643

데이터의 공간 객체를 비교하기 위해서는 좌표계가 통일되어야 하므로 연구에서 진행되는 모든 자료의 좌표계를 EPSG: 5181로 맞추었다. 부동산건물데이터 중에서 공장만 추출하기 위하여 속성 값 쿼리를 이용하였다.

건물의 용도를 나타내는 속성 컬럼인 A30 = '공장'을 이용하여 공장을 추출하였다. Fig. 7과 같이 추출된 공장 객체 256,090개에서 같은 공장이지만 건물이 여러 개인 경우를 찾아 한 공장으로 식별할 수 있도록 공간 처리를 수행하였고, 최종 추출된 공장 개수는 144,111개로 확인되었다. Fig. 8은 여러 공장 건물을 지니지만, 한 공장으로 대표할 경우 한 개의 객체로 인식할 수 있도록 멀티폴리곤을 적용한 모습이다.

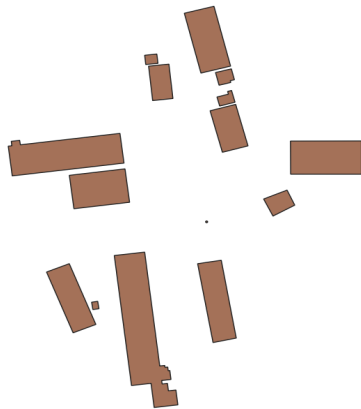


Fig. 7. Raw real estate data

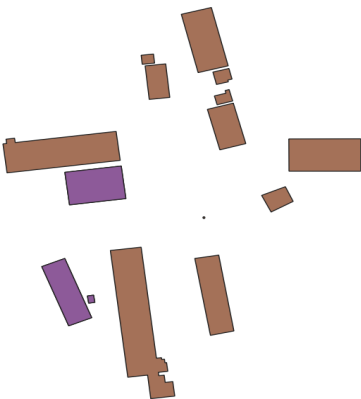


Fig. 8. Preprocessed real estate data

전처리가 완료된 114,111 개의 데이터에 대하여 속성 정보 매칭을 수행하였다. 위 과정에서 공장의 주소와 이름을 이용하였다. 정확한 속성정보를 비교하기 위해서는 입력 값들의 형식이 일치되어야 한다. 부동산건물데이터의 경우, 공장주소의 시군구와 읍면동이 한 컬럼에 입력되고 번지는 다른 컬럼에 따로 입력되어 있다. 주소를 한 컬럼에 입력하기 위하여 두 컬럼을 병합하여 이용하였다. 또한 주소의 비교를 위하여 모든 주소 값에 공백을 없애고 '시군구, 읍면동, 건물 본번, 건물 부번' 형식을 갖도록 수정하였다. Table 2는 속성 값 전처리 이전 공장의 주소와 번지 컬럼을 타나내고, Table 3은 전처리를 수행하여 주소의 형식을 맞춘 모습을 보여준다. Table 3과 같이 전처리가 완료된 Concat 컬럼을 이용하여 데이터의 속성 값을 비교하였다.

Table 2. Raw attributes of real estate data

A3	A6	A12
Jongno 3-ga, Jongno-gu, Seoul	148	Null
Sindang-dong, Jung-gu, Seoul	251-160	Dongdaemun Fashion Biz Center
Cho-dong, Jung-gu, Seoul	42	Asia Media Tower
Seongsu-dong 1-ga, Seongdong-gu, Seoul	656-1683	Sungsu-dong apartment-type factory Do & Can house

Table 3. Preprocessed attributes of real estate data

A3	A6	Concat
Jongno 3-ga, Jongno-gu, Seoul	148	Jongno 3-ga 148, Jongno-gu, Seoul
Sindang-dong, Jung-gu, Seoul	251-160	Sindang-dong 251-160, Jung-gu, Seoul
Cho-dong, Jung-gu, Seoul	42	Cho-dong 42, Jung-gu, Seoul
Seongsu-dong 1-ga, Seongdong-gu, Seoul	656-1683	Seongsu-dong 1-ga 656-1683, Seongdong-gu, Seoul

3.2 도로명주소건물데이터 전처리

도로명주소건물데이터는 행정안전부 한국지역정보개발원으로부터 획득하였다. 도로명주소건물데이터에서 공장정보만 추출하기 위하여 속성 BDTYP_CD를 공장 코드인 13000, 13100, 13200을 적용하여 추출하였다. 산업입지정보시스템 공장정보와 부동산건물데이터는 주소를 한글로 제공하는 반면, 도로명주소건물데이터의 경우 Table 4와 같이 주소를 시군구 코드(Sig_cd), 읍면동 코드(Emd_cd), 리 코드(Li_cd), 건물 본번(Buld_mnm),

건물 부번(Buld_slno)으로 각각 제공하고 있다. 위 코드를 한글 주소로 전환하기 위하여 10 자리 주소 코드로 통합하였다. 도로명주소 DB 사이트에서 제공하는 주소 코드별 시군구읍면동 이름 데이터를 이용하여 통합된 10 자리 주소 코드를 한글 주소로 전환하고 '시군구, 읍면동, 건물 본번, 건물 부번' 형식으로 맞춰 비교할 수 있도록 전 처리를 수행하였다. Table 4는 전처리를 수행하기 전 속성 테이블의 모습이고, Table 5는 전처리를 수행하여 한글 지번주소로 수정한 모습이다.

도로명주소건물데이터 또한 부동산건물데이터와 마찬가지로 한 공장이 여러 개의 공장 건물을 지닐 경우 한 공장으로 대표할 수 있도록 멀티폴리곤을 적용하였다. 그 결과 405,422 개의 공장 객체가 146,969 개로 멀티폴리곤화 되었다. 최종적으로 Table 5의 'Addr' 컬럼 값을 이용하여 다른 공장데이터와 속성정보 매칭을 수행하였다.

Table 4. Raw attributes of address name road data

Sig_cd	Emd_cd	Li_cd	Buld_mn_nm	Buld_sl_no	Buld_nm
42730	370	25	912	0	oriental blackboard co., ltd.
42730	360	32	974	0	Korea Farm Fertilizer Industry Co., Ltd.
42730	360	32	1319	12	Now Industry
42730	250	43	1606	0	Shinheung Rehabilitation Office

Table 5. Preprocessed attributes of address name road data

Code	Addr	Buld_nm
4117310400	Hogye-dong 901-39, Dongan-gu, Anyang-si, Gyeonggi-do	Dongwoo building
4117310400	Hogye-dong 897-8, Dongan-gu, Anyang-si, Gyeonggi-do	Teletech Co., Ltd.
4117310400	Hogye-dong 898-42, Dongan-gu, Anyang-si, Gyeonggi-do	Unicorn Technology Co., Ltd.

3.3 산업입지정보시스템 데이터 전처리

산업입지정보시스템 데이터 속성정보 중 공장주소 컬럼을 부동산건물데이터와 마찬가지로 공백 없이 '시군구, 읍면동, 건물 본번, 건물 부번' 형식으로 맞춰 공장 주소로 비교할 수 있도록 하였다. Table 6은 주소 컬럼을 수정한 산업입지정보시스템의 공장정보 모습의 일부이다. 'Addr1' 컬럼은 기존에 제공받은 공장의 지번 주소를 나

타내며, 'Addr2'는 비교를 위하여 형식을 통일한 모습이다. 최종적으로 공장의 속성정보를 비교할 때 'Addr2' 컬럼을 이용하였다.

Table 6. Preprocessed attributes of ILIS data

Addr1	Addr2	Conm
B2 at 118-1, Bongik-dong, Jongno-gu, Seoul	Bongik-dong 118-1, Jongno-gu, Seoul	S.M. Gold Exchange Co., Ltd.
Fifth floor, 118-1, Bongik-dong, Jongno-gu, Seoul	Bongik-dong 118-1, Jongno-gu, Seoul	Jewering Co., Ltd.
35-1, Bongik-dong, Jongno-gu, Seoul	Bongik-dong 35-1, Jongno-gu, Seoul	Korea Luff Diamond Co., Ltd.

두 번째로 공간 분석을 수행하기 위하여 공간 객체 전처리를 수행하였다. 산업입지정보시스템 공장정보는 점 데이터이므로 다각형 데이터의 부동산건물데이터와 도로명주소건물데이터와 공간 매칭을 수행할 때 겹치지 않는 문제가 있다. Fig. 4와 같은 문제를 보완하기 위하여 각 점 데이터의 건물 면적을 계산하여 버퍼를 수행하여 중첩할 수 있도록 하였다. 각 건물의 면적은 속성 테이블의 제조 면적과 부대시설 면적을 합하여 계산하였다. Fig. 9는 각 점 객체마다 건물의 면적을 계산하여 버퍼를 수행한 모습이다. Fig. 9와 같이 전처리가 완료된 산업입지정보시스템 공장정보 데이터를 부동산건물데이터, 도로명주소건물데이터와 비교하였다.

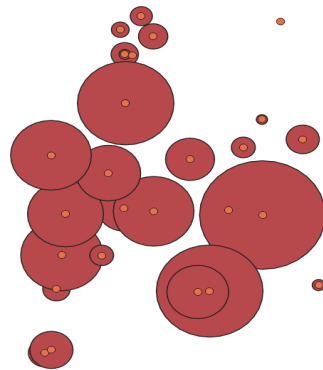


Fig. 9. Preprocessed ILIS data

3.4 데이터 매칭

공장 주소 형식이 통일된 세 종류의 데이터를 이용하여 속성정보로 매칭하였다. 하나의 행정구역 내에 공장이 많을 경우를 대비하여, 공장주소만 매칭하지 않고 공장 이름과 주소를 함께 그룹화 하여 매칭하였다. 공장 이름과 주소가 동일하게 매칭되는 공장데이터는 따로 추출하였다.

속성매칭을 진행한 후, 공장이름이나 주소가 등록되지 않아 매칭이 안 된 공장에 대하여 공간객체매칭을 수행하였다. 버퍼를 수행하여 다각형으로 전환된 산업입지정보시스템 공장정보를 부동산건물데이터와 도로명주소건물데이터와 비교하였다. 공간분석 도구 중 교차분석 도구인 intersection을 이용하여 겹치는 공장을 추출하고, 겹치는 공장을 제외한 공장을 신규 공장으로 확인하여 속성 값을 추출하였다.

4. 데이터 매칭 결과 및 정확도 평가

4.1 부동산건물데이터 비교

전처리가 완료된 부동산건물데이터와 산업입지정보시스템 공장이름과 주소를 비교하여 매칭하였다. 매칭한 결과는 Table 7과 같다. 부동산건물데이터에 등록된 공장 144,111개 중에서 공장주소와 이름으로 일치하는 부동산의 공장 수는 77,549 개로 확인되었다. 주소정보가 없거나 일치하지 않아 매칭되지 않은 66,562개 공장에 대하여 공간매칭을 수행한 결과 36,534개가 매칭되었다. 남은 공장정보 중에 30,028개 공장이 산업입지정보시스템 공장정보에 없는 부동산건물데이터 신규 공장으로 추출하였다. Fig. 10은 신규 공장정보를 지도에 나타낸 모습이고 Table 8은 신규 공장정보의 속성 값이다.

Table 7. Results of real estate data

		Reference data 1	
		Matching	Non matching
Method	Attribute comparison	77,549	66,562
	Spatial Intersection	36,534	30,028
Total		144,111	

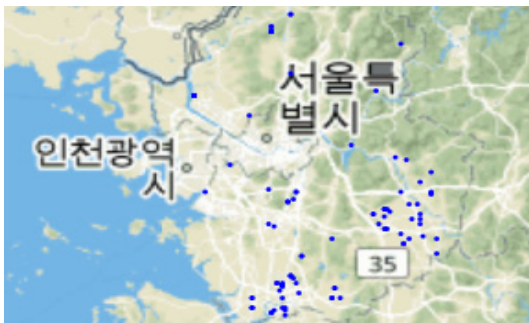


Fig. 10. Example of extracted new factory in real estate data

Table 8. A partial of extracted attributes of new factory in real estate data

Address	Name
Cheongpyeong-ri, Cheongpyeong-myeon, Gyeonggi-do	Cheongpyeong brewing Co., Ltd.
732-1 Goyang-si Deokyang-gu, Gyeonggi-do	Jichook vehicle base Co., Ltd.
Duri, Pyeongtaek, Gyeonggi-do	Jin Tae Electronic Co., Ltd.
Gung-ri 906-5, Pyeongtaek-si, Gyeonggi-do	Huneco Co., Ltd.
Seodun-dong, Suwon-si, Gyeonggi-do 110-1	taeyang moolsan Co., Ltd.

4.2 도로명주소건물데이터 비교

전처리가 완료된 도로명주소건물데이터와 산업입지정보시스템 공장정보를 매칭한 결과는 Table 9와 같다. 도로명주소건물데이터 총 공장 객체 수는 141,969개로 등록되어 있다. 도로명주소건물데이터에서 도로명 코드를 주소로 변환하여 산업입지정보시스템 주소와 비교한 결과, 일치하는 도로명주소건물데이터는 66,358개로 확인되었다. 남은 공장데이터 중에서 49,257개 공장데이터는 공간분석으로 매칭되었고, 26,354개 공장이 신규 공장으로 나타났다. Fig. 11은 추출된 신규 공장을 지도에 나타낸 모습이고 Table 10은 신규 공장 정보의 속성 값이다.

Table 9. Results of road name address data

		Reference data 2	
		Matching	Non matching
Method	Attribute comparison	66,358	75,611
	Spatial Intersection	49,257	26,354
Total		141,969	

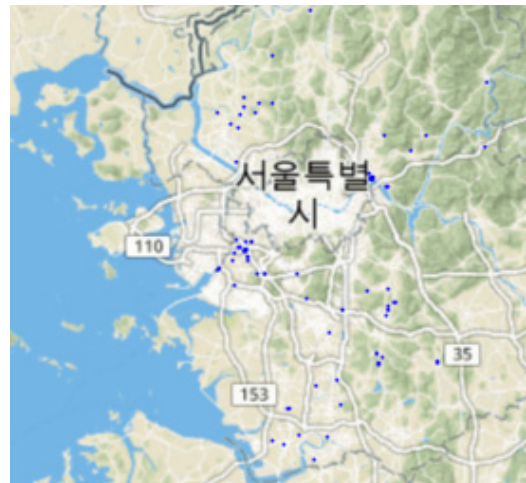


Fig. 11. Example of extracted new factory in road name address data

Table 10. A partial of extracted attributes of new factory in road name address data

Address of factory	Name of factory
Seo-ri 1-0, Yongin-si, Gyeonggi-do	Dae-hyun Industry
Sangiiseok-dong 1000-15, Paju, Gyeonggi-do	Silla Furniture
Jonam-dong 100-1, Siheung-si, Gyeonggi-do	Joo-young Industry
Janggok-ri 100-17, Jori-eup, Paju, Gyeonggi-do	Kyung-hoon co., Ltd.
Deok-myeon Dong 100-2, Go-ri, Pyeongtaek-si, Gyeonggi-do	D.S. food Co., Ltd.

4.3 신규 공장정보 갱신

부동산건물데이터와 도로명주소건물데이터 신규 공장 정보를 추출한 결과를 산업입지정보시스템에 갱신하기 위해서는 유사한 속성 값과 공간객체 좌표를 포함하는 지오매트리 값이 필요하다. 산업입지정보시스템 공장 정보는 직원 수, 공장크기와 같이 공장에 관한 세세한 정보를 포함한다. 산업입지정보시스템 공장정보의 속성 33개 중에서 부동산건물데이터로부터 갱신할 수 있는 정보는 건물명, 건물주소, 건물제조면적과 연면적이 있다. 도로명주소건물데이터 속성정보는 건물주소와 구성에 집중한 데이터이므로, 건물주소, 지번코드와 도로명코드 속성 정보를 제공한다. 추출된 도로명주소건물데이터 공장정보 중에서 산업입지정보시스템에 갱신할 수 있는 정보는 건물주소, 건물지상층수와 지하층수가 있다. 따라서 신규 공장 위치정보인 지오매트리와 속성정보인 공장이름, 공장주소, 공장면적과 건물층수를 갱신할 수 있고 나머지 29개 정보에 대해서는 수동으로 입력해야한다. 그러나 산업입지정보시스템의 목적은 전국의 공장 위치와 이름을 나타내는 것이므로 본 연구에서 제안한 방법을 이용하여 먼저 공장의 위치와 이름을 나타낸 뒤, 부가적인 29개 정보에 대해서는 후처리를 이용하여 추가할 수 있다.

4.4 정확도 평가

점 데이터와 다각형 데이터로 서로 다른 형식의 공간 데이터를 비교하기 위하여 점 데이터를 면적만큼 버퍼하여 이용하였다. 위 방법의 정확도를 평가하기 위하여 세종시를 예시로 샘플링 하여 정확도 평가를 수행하여 보았다. 샘플링 된 데이터에 대하여 공간분석을 수행하고, 공간분석으로 일치하다고 추출된 건물데이터의 속성 값에서 실제로 건물주소와 이름이 일치하는지 평가하였다.

그 결과, 부동산건물데이터에 대해서는 총 665개 공장건물 중에서 478개 공장이 실제 공장주소와 이름이 일

치하여 71.88% 일치율을 보였다. Table 11은 공간 분석을 수행하여 매칭한 건물주소와 실제 주소가 일치하는지 판단한 결과이다. Concat 컬럼은 공간 분석으로 매칭한 주소이고, Addr2 컬럼은 실제 공장 주소를 나타낸다.

도로명주소건물데이터의 경우 총 234개의 공장건물과 공간매칭을 하였고, 실제 공장주소와 이름이 일치하는 공장은 203개로 확인되었다. 도로명주소건물데이터에 대해서는 86.75%의 일치율을 보였다. 기존의 산업입지정보시스템의 공장정보는 2018년 정보만 확인이 가능하지만, 본 연구에서 제안한 방법을 이용하면 2020년 공장 정보에 대하여 약 79%의 정확도로 공장정보를 갱신할 수 있어 의미가 있다고 확인하였다. 매칭 오류의 경우로는, 공간분석으로 매칭 되었지만 주소가 다른 경우로 확인되었다. 예를 들어, Fig.12의 A factory와 같이 공장면적이 넓어 근처에 있는 다른 공장까지 같은 공장으로 매칭시키는 경우에 오차가 생기는 것으로 확인하였다.

Table 11. Example of matched factory in Sejong data

concat	addr2	conm
Nam-ri 60-1, Jochiwon-eup, Sejong-si	Nam-ri 60-1, Jochiwon-eup, Sejong-si	Han-lim Paper
Shinheung-ri 62-8, Jochiwon-eup, Sejong-si	Shinheung-ri 62-8, Jochiwon-eup, Sejong-si	Jochiwon Food
Bunam-ri 116-1, Jochiwon-eup, Sejong-si	Bunam-ri 116-1, Jochiwon-eup, Sejong-si	International High-tech
Bunam-ri 115-4, Jochiwon-eup, Sejong-si	Bunam-ri 117, Jochiwon-eup, Sejong-si	Dong-yang Food
Bunam-ri 116-1, Jochiwon-eup, Sejong-si	Bunam-ri 117, Jochiwon-eup, Sejong-si	Dong-yang Food

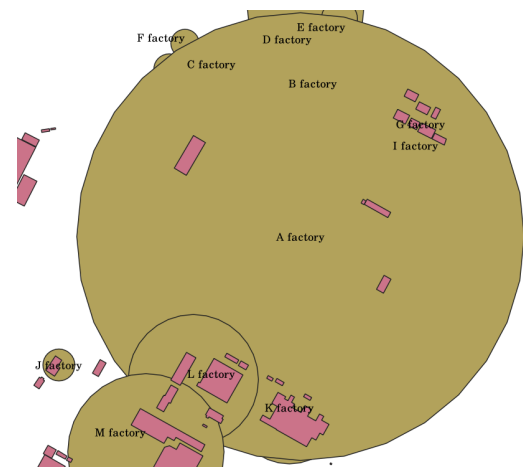


Fig. 12. Example of unmatched error in Sejong data

5. 결론 및 향후 계획

산업입지정보시스템 공장정보 최신화하기 위한 방법으로 공공데이터를 이용하여 공장정보를 매칭하고 갱신하는 방안을 제시하였다. 또한 서로 다른 형태의 데이터를 매칭하고 비교할 때, 공통 속성 값인 공장이름과 주소로 비교하고 매칭되지 않는 공장은 공간 분석을 이용하여 매칭 하였다. 공간정보를 포함하여 서로 다른 데이터를 비교하고 갱신할 때 일반적으로 모든 데이터가 공통적으로 갖고 있는 속성 값으로 비교한다. 그러나 공장이름이나 주소와 같이 속성 값이 통일되지 않아 정확히 매칭되지 않는 점은 공간 매칭으로 보완할 수 있음을 확인하였다. 추후에는 데이터의 속성 값이 통일된다면 좀 더 높은 일치율로 매칭 할 수 있을 것이라고 기대된다.

또한 공간분석 매칭을 수행할 때, 형식이 다른 점 데이터와 다각형 데이터를 비교하기 위하여 각 공장의 면적을 계산하여 버퍼를 적용하여 매칭하였다. 부동산건물데이터와 도로명주소건물데이터를 이용하여 비교한 결과 약 79% 정확도를 보였다. 그러나 건물 면적이 주변 건물에 비해 월등히 넓은 경우, 다른 건물까지 매칭하는 오차가 발생한다. 향후에는 각 공장 포인트 별로 면적을 계산하여 버퍼를 수행할 때 임계값을 정하여 비교적 넓은 면적을 갖는 공장에 대해서는 주변 건물의 면적비를 계산하는 방법을 사용하는 연구가 필요할 것으로 보인다.

REFERENCES

- [1] J. D. Lim, S. G. Kim & C. Y. Shin. (2019). A study on the Improvement for Industrial Land Information System. *The Korean Journal of The Korea Convergence Society*, 10(1), 35-43. DOI : 10.15207/JKCS.2019.10.1.035
- [2] M. S. Chung. (2017). *A Study on the GIS Application for Industrial Complex Permit System*. Sejong : KRIHS.
- [3] H. T. Kim, C. S. Kang, B.Y. Choi & Y. D. Eo. (2010). Experimental Research of Business Practice Utilizing Geospatial Data. *Proceedings of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography Conference*. (pp. 111-112). Seoul
- [4] K. R. Kim, Y. Huh & K. Y. Yu. (2010). Study on method of linking with navigation map and new address map for updating navigation Map buildings. *Proceedings of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography Conference*. (pp. 23-25). Seoul
- [5] Y. S. Kim. (2005). A Study on the Design for Contents System of Real Estate based on Web Engineering. *Journal of the Korea Society of Computer and Information*. 10(6), 55-64.
- [6] J. Y. Kim & K. Y. Yu. (2012). Automatic Detection of the Updating Object by Areal Feature Matching Based on Shape Similarity. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*. 30(1), 59-65. DOI : 10.7848/ksgpc.2012.30.1.059
- [7] S. C. Shin & S. J. Lee. (2010). The Design of UFID Management System Structure for Next Generation Digital Map. *Journal of the Korean Society for Geospatial Information Science Conference*. (pp. 73-74). Seoul
- [8] B. N. Choe & H. C. Jin. (2015). Correlation Analysis Between Casual and Outcome Factors for Developing Land Information System. *Journal of the Korea Spatial Information Society*. 23(6), 1-8. DOI : 10.12672/kisis.2015.23.6.001
- [9] D. B. Shin & S. C. Yu. (2008). Experiments on Transform to World Geodetic System of Korea Land Information System Database. *Proceedings of the Korean Society for Geospatial Information Science*. (pp. 405-410). Seoul
- [10] S. G. Jung. (2011). *A Study on the Ubiquitous based Integrated Urban Information Center*. Gwangju : GJI.
- [11] S. K. Nam, Y. S. Oh, T. H. Kim, J. A. Kang & J. W. Kim. (2009). Developing an Urban Ground Facilities Management System based on u-GIS using an Open Source GIS software. *Journal of the Korean Society for Geospatial Information Science*. 17(4), 67-74.
- [12] S. A. Park, K. Y. Yu & W. J. Park. (2014). Updating Building Data in Digital Topographic Map Based on Matching and Generation of Update History Record. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*. 32(4-1), 311-318. DOI : 10.7848/ksgpc.2014.32.4-1.311
- [13] S. S. Kim, S. H. Goo, W. S. Kim & H. H. Yoo. (2006). Building Layer Update of Digital Map using LIDAR Data. *Proceedings of the Korean Society of Surveying, Geodesy, Photogrammetry, and Cartography Conference*. (pp. 411-417). Chuncheon
- [14] H. J. Kim, S. C. Yang, C. O. Ga & K. Y. Yu. (2008). A method of Update and Inspection for Building Layer of Digital Maps 2.0 using e-AIS Data. *Korea Society of Geospatial Information Science*. (pp. 427-433). Seoul
- [15] J. H. Yeom, Y. Heo & J. B. Lee. (2014). Building Matching Analysis and New Building Update for the Integrated Use of the Digital Map and the Road Name Address Map. *Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography*. 32(5), 459-467. DOI : 10.7848/ksgpc.2014.32.5.459
- [16] M. P. Lin, Y. He, V. W. Hsiao, R. Chang & S. Lee. (2013). Common-Centroid Capacitor Layout Generation Considering Device Matching and Parasitic Minimization.

IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 32(7), 991-1002.
DOI : 10.1109/TCAD.2012.2226457

- [17] D. Huang, R. Zhang, Y. Yin, Y. Wang & Y. Wang. (2016). Local Feature Approach to Dorsal Hand Vein Recognition by Centroid-based Circular Key-point Grid and Fine-grained Matching. *Image and Vision Computing*. 58, 266-277.
DOI : 10.1016/j.imavis.2016.07.001
- [18] J. J. Ruiz-Lendinez, M. A. Urena-Camara & F. J. Ariza-Lopez. (2017). A Polygon and Point-Based Approach to Matching Geospatial Features. *International Journal of Geo-Information*. 6(12), 399-422.
DOI : 10.3390/ijgi6120399

최 유 정(Yu-Jeong Choe) [정회원]



- 2019년 2월 : 세종대학교 대학원 지구정보공학과(공학석사)
- 2019년 4월 ~ 2019년 9월 : 세종대학교 산학협력단 석사후연구원
- 2020년 3월 ~ 현재 : (주)지앤티솔루션 기술연구소 연구원
- 관심분야 : 정보융합, GIS, ICT, ITS

· E-Mail : yjchoe@gntsolution.com

임 재 덕(Jae-Deok Lim) [정회원]



- 2008년 2월 : 한서대학교 대학원 컴퓨터정보학과(이학석사)
- 2008년 2월 ~ 2009년 12월: 한국건설기술연구원 연구원
- 2009년 12월 ~ 2016년 1월 : 한국교통연구원 위촉연구원
- 2016년 2월 ~ 현재 : (주)지앤티 솔루션 기술연구소 책임연구원

· 관심분야 : 정보융합, 정보시스템, ICT, ITS

· E-Mail : jdlim@gntsolution.com

김 성 건(Seong-Geon Kim) [정회원]



- 2001년 2월 : 울산대학교 대학원 건축공학과(공학석사)
- 2002년 7월 ~ 2006년 2월: (주)선도소프트 교통사업본부 과장
- 2006년 4월 ~ 2007년 3월 : (주)엠앤소프트 기술연구소 차장
- 2007년 4월 ~ 현재 : (주)지앤티솔루션 기술연구소 연구소장

· 관심분야 : 정보융합, GIS, ICT, ITS

· E-Mail : sgkim@gntsolution.com