

텍스트 기반 공간 지식 추론 프레임워크 연구: 소셜미디어 데이터 중심으로

전 원 표*

*바이브컴퍼니 스마트시티연구소 수석연구원

Framework for Text-based Geo-Semantic Knowledge Inference: Focusing on Social Media Data

Won-Pyo Jeon*

*Chief Researcher, Smart City Institute, VAIV company, Seoul 04419, Korea

[요 약]

최근에 개인 모바일과 소셜미디어에서 생산되는 콘텐츠는 메타버스 서비스와 연계하기 위한 공간과 시간의 표현 모델 연구와 공간 데이터를 활용한 서비스 개발이 활발해지고 있다. 하지만 소셜미디어 콘텐츠는 공간과 위치 정보의 보안성과 공간 표현의 구조가 미흡하여 동일명을 갖는 다른 정보와의 관계성을 정의하기 어렵다. 또한 공간에 대한 정의로 국가관심지점 공공 데이터가 있지만 공간명에 대한 정확도가 국가주요시설을 제외하고는 완전하지 않기 때문에 공간의 활동성을 연결하여 분석하기 곤란하다. 따라서 본 연구는 소셜미디어에서 수집한 데이터와 국가관심지점정보의 관계 속성을 고려하여 시공간의 흐름을 나타낼 수 있는 공간 지식 표현 모델과 해당 표현 모델을 분석할 수 있는 공간 지식 추론 프레임워크를 제시하고자 한다. 본 연구의 결과는 텍스트에서 구조화된 공간 지식 기반 추천 알고리즘과 지리정보 추적을 위한 인공지능 모델 등의 다양한 연구에 활용할 수 있는 기반을 마련하였다.

[Abstract]

Services utilizing space and expression models of space and time for content produced on personal mobile devices and social media have recently been developed to link them with metaverse services. However, this content lacks security and a spatial expression structure for acquiring spatial and location information, making it difficult to define the relationship of geographical information to other points with the same name. Additionally, because the accuracy of spatial names is not perfect, it is difficult to connect and analyze spatial activity. Therefore, this paper proposes a spatial knowledge expression model that can represent the flow of time and space and a spatial knowledge inference framework that can analyze the expression model, considering the relationship properties of data collected from social media and NGII-POI. The findings of this study establish a solid foundation for future research endeavors, including the development of recommendation algorithms based on structured spatial knowledge in text and artificial intelligence models for tracking geographic information.

색인어 : 공간지식, 텍스트마이닝, 데이터마이닝, 지식추출, 소셜미디어

Keyword : Geosemantic, Textmining, Datamining, Knowledge Extracting, Social Media

<http://dx.doi.org/10.9728/dcs.2023.24.11.2733>



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 17 October 2023; **Revised** 03 November 2023

Accepted 20 November 2023

*Corresponding Author; Won-Pyo Jeon

Tel: +82-2-565-0531

E-mail: jwp@vaiv.kr

1. 서론

소셜미디어는 개인의 성향과 의도를 다양한 표현 방법을 사용하여 게시하고 있으며, 최근에는 텍스트를 통한 메시지 전달 방법에서 확장되어 태그, 사진, 이미지, 동영상 등 비정형 데이터를 이용하고 있다. 이러한 소셜미디어의 데이터는 특정 지점에 대한 방문 기록이나 관심이 있을 때 해당 장소의 위치 정보 또는 명칭으로 기록한다. 하지만 명칭을 표시하는 방법은 다양한 태그와 별칭(Alias), 공간 어휘를 사용하는 경우가 많기 때문에 장소 기준으로 통합된 정보를 구성하기 어렵다. 이러한 정보의 통합이나 표현은 공간을 표현하는 메타버스나 디지털트윈에서 동일 공간에 대한 정보가 맞는지 확인하는 전처리 기술이나 공간 정보를 매핑하는 기준 데이터를 요구하고 있다. 이를 위해 국가에서는 공간 기준의 데이터를 지역의 건물, 상호, 위치 정보등을 모아서 국가관심지점정보(NGII POI, National Geographic Information Institute Point of Interest)로 제공하고 있으며, 공간 정보의 변화를 관리하는 데이터베이스로 자료를 제공하고 있다[1]. 이 데이터는 공간의 이름과 주소명에 대한 관계성을 포함하고 있기 때문에 소셜미디어의 장소명을 관심지점(POI, Point Of Interest)으로 통합하는 작업이 가능하다. 이렇게 통합된 데이터는 소셜미디어 공간에서 공간별로 시간의 흐름에 따른 상황과 현상을 분석하는데 용이하다. 다만 관심지점 내에서도 다른 지역, 동명이 존재하기 때문에 공간 구별을 위한 개체명 지식 관리가 필요하다.

참고 [2]의 연구에서는 관심지점 데이터 수집을 위하여 위치기반 사회관계망 서비스(LBSNS, Location Based Social Network Service)를 이용하였고, 참고 [3]의 연구에서는 GPS(Global Positioning System)가 내장되어 있는 모바일 폰과 같은 전자기기에서 수집된 데이터를 통해 공간 지식을 구축하고 이를 통해 2차 분석을 수행하였다. 참고 [4]는 기계학습을 통해 추출한 관심지점을 ‘근처, 주변, 앞, 맞은편’과 같은 공간어를 이용하여 기구축되어 있는 관심지점과 매핑 후 공간 어휘를 구분하여 공간 관계 데이터베이스를 구축하였다. 또한 동명이면서 다른 지역인 관심지점 데이터는 소셜미디어 데이터에서도 다른 상황이 적용되어 분석되기 때문에 공간 개체명으로서의 지역명 개선이 필요하다[5],[6]. 이 연구들은 소셜미디어 데이터를 형태소 분석하여 공간에 대한 키워드, 방향에 대한 키워드 등으로 구분하고, 관심지점 토픽으로 추출하여 공간 개체명에 대한 연관성 분석하고 있다. 또한 관심지점 토픽은 서로 다른 지역의 명칭, 별칭을 동일한 관심지점 토픽에 적용하는 인식의 문제를 해결하기 위해 주변 공간 어휘를 적용한 의미 기준의 어절 단위 공간 지식을 구성하고 있다. 그럼으로써 관심지점 토픽이 공간적 개념과 공간 어휘, 시간적 변화의 관계성을 정의하여 공간 지식과 검색에 활용되도록 하였다. 또한 최근에는 인공지능 딥러닝 분석 기술을 적용하여 소셜미디어의 사진, 이미지, 동영상 등의 비정형 데이

터에서 관심지점 공간 지식을 추출하는 연구를 통해 관심지점 개체의 관계성과 위치 연결성을 분석하는 연구가 진행되고 있다[7]. 그리고, 관심지점 데이터 간의 연결 유형, 사용자의 움직임이나 행동 유형을 분석하는 네트워크 공간에 대한 상호운용성을 제시하고 있다[7]-[9].

하지만 이러한 관심지점 공간 데이터는 위치 기반 서비스나 GPS가 내장되어 있는 전자기기를 통해 수집된 데이터에 한정되기 때문에 대부분의 소셜미디어 데이터는 제외되어 제한이 있다. 또한 텍스트마이닝을 통해 추출한 관심지점의 경우 위치는 다르지만 동일명을 갖는 공간 개체에 대하여 식별이 안되고, 시공간의 흐름에 대한 복잡성을 표현하기 어렵다.

따라서 본 연구는 공간에서 이루어지는 사회적 현상과 상황을 관계적으로 분석하기 위한 공간 지식 표현 모델과 이를 분석할 수 있는 공간 지식 추론 프레임워크를 제시한다. 공간 지식 표현 모델은 시간과 공간을 기준으로 동일한 데이터도 다양한 관계 속성을 고려하여 사용자의 움직임과 선호도를 비교적으로 분석할 수 있으며, 공간 지식 추론 프레임워크는 소셜미디어와 공간에 대한 관심지점의 관계성을 구조화하고 소셜미디어의 장소 개체명과 시각 정보를 추출하고 의미를 분석하는 모델 구성 및 추론 모델의 세부적 기능을 적시한다. 그리고 소셜미디어가 가지고 있는 장소 개체명에 대해 국가 관심지점과 관계성을 정의하는 비정형 데이터의 방향성을 개체명 검색과 상황에 대한 의사결정을 지원하는 시스템 구성 요소를 정의하였다. 또한 서로 다른 지점에 대한 동일명을 회피하기 위한 공간정보 및 국가관심지점의 관계성을 연관 모델에 대한 구조를 제시하고 공유를 위한 상호운용성(Interoperability)을 보장하는 관심지점 구조의 고려사항을 정리하였다. 이를 통해 국가관심지점과 소셜미디어에서 일어나는 이벤트에 대한 상황을 분석하고 사용자 단위의 재난 대응 및 회피에 빠른 의사결정과 효과적인 상황 전환을 제시하는 시스템을 구성할 수 있다.

본 논문의 구성은 소셜미디어 상에서 텍스트에서 추출된 공간정보를 대상으로 관심지점을 생성하고 연계하는 관련 연구와 기반 기술을 이용한 공간 지식 추론 프레임워크를 구성하였다. 그리고 공간 지식 추론 프레임워크가 가지는 각 요소의 특징과 시스템 구축에 필요한 단위 모델 및 분석 모델을 기술하였다.

II. 관련연구

2-1 소셜미디어 관심지점 데이터 특징

소셜미디어 데이터는 다양한 키워드와 태그를 통해 사용자의 의사와 장소를 표시하고 있다. 소셜미디어 데이터에서 게시된 장소는 GIS(Geographic Information System) 정보를 가지고 표시할 수 있지만 이름만으로 표시되는 경우가 존재

한다. 이는 기존의 장소에 새로운 장소명이 등록되거나 장소의 성격이 변경되는 경우에 자주 발생한다. 특히 국내에서는 사용자의 개인정보에 위치 데이터도 포함될 수 있기 때문에 개인정보를 동의하거나 위치 정보를 동의한 상태에서 GIS 관련 정보를 수집할 수 있다. 그렇기 때문에 소셜미디어에서 위치 정보는 개인정보로 마스킹 되거나 위치 정보가 삭제되어 수집되고 있는 경우가 많다. 그래서 위치 기반 서비스를 구성하기 위해서는 텍스트마이닝에서 위치 정보와 관련된 키워드를 분류하거나 개체명으로 구분하여 태그를 정의하는 관계성을 정의하는 것이 중요하다. 특히 소셜미디어의 관심지점은 시설물에 대한 정보 갱신 주기, 공간에 머물러 있었던 시간을 시간의 흐름에 따라 공간의 변화나 장소의 성격이 변화되는 정보를 재구축하는 것이 필요하다. 이런 상태에서 소셜미디어의 관심지점 정확도가 증가하여 시설물의 위치 변화와 국가 관심지점정보의 오류 정보를 최소화한 상태의 위치 한계점을 보완하는 기능이 필요하다[10].

소셜미디어 공간 데이터의 기준이 되는 국가관심지점정보는 국가, 공공기관, 지자체, 기업 등에서 취득한 공간 또는 지명 데이터를 가공하고 정제하여 별도의 작업 없이 국가 기본 지도에 적용하는 정보이다[1]. 국가관심지점정보는 997만 건의 공간과 지명 데이터를 구축하여 보관하고 있으며, 전국의 공공기관, 관광지 등의 지역 랜드마크를 중심으로 구성하고 있다. 국가관심지점정보는 구축대상이 주요 건축물, 구조물, 국가주요시설, 행정 경계, 도로, 수계, 지형, 관광지 등의 정보로 구성되어 있다. 국가관심지점의 위치 정보는 GIS를 포함하면서 국내 주소 표기법을 모두 반영하고 있기 때문에 공간의 위치 정보를 대상으로 관심지점의 위치 정보를 적용할 수 있다. 다만 많은 사람들이 이동하는 시간과 장소의 속성 정보가 부족한 공간도 존재하고 있기 때문에 이에 대한 보완은 계속적으로 필요하다.

그림 1은 국가관심지점정보에 대한 구성 내용이며, 텍스트와 코드 중심의 데이터를 제공하고 있어 관심지점에 대한 공간정보에 대한 기본 정보, 종류별 분류, 주소, 공간의 모자관계, 속성, 명칭, 별칭, 메타 및 원본, 외국어로 구성되어 Open API(Application Programming Interface)로 제공하고 있다. 종류별 관심지점을 분류하는 구조는 기관, 지명, 시설, 상호명 등으로 구분하고 있으며 지속적인 업데이트를 하고 있지만 실시간으로 반영하고 있지 않다. 또한 관심지점 간의 관계를 정의하는 모자 관계는 대등 건물, 부속건물, 출입구 등과 같은 시설이나 건물에 속한 명칭이나 건물 간의 관계를 정의하고 있다. 그리고 대등건물, 부속시설, 출입구 등과 같은 시설이나 내부 지점에 대한 관계성은 소셜미디어의 데이터와 연계된 고리를 조회하여 통합하는 기본 속성으로 적용할 수 있다. 또한 관심지점의 관계에서 모자 관계는 기관, 지명, 시설, 상호명 등으로 구분하고 지속적인 업데이트를 반영하고 있다. 그리고 별칭으로 최대 20개까지 등록하고 관리하면서 시간의 변화에 따라 다른 별칭으로 사용할 수 있도록 국가관심지점의 데이터 속성을 확대하고 있다. 하지만 국가관심지점

정보는 공간 어휘를 표현하는 주변 관심지점과의 관계성, 방향성 등을 정의하는 구조가 없기 때문에 소셜 미디어 관심지점과 연계된 공간 어휘나 공간 개체명을 추론하는 공간 지식 추론 모델이나 공간 학습 모델이 필요하다.

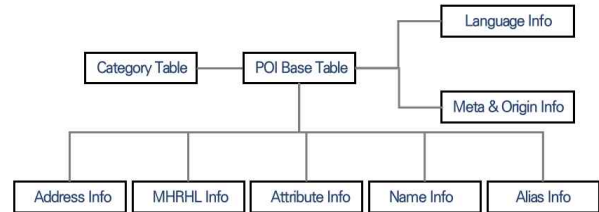


그림 1. 국가관심지점정보 테이블간 관계
 Fig. 1. Table relation of National Point of Interest Information

소셜미디어 데이터를 기준으로 관심지점 구축을 위한 공간 어휘 추론 모델은 포스퀘어[10], 텍스트마이닝[11],[12]을 적용한 연구가 있다. 포스퀘어는 사용자가 직접 자신의 위치를 입력하는 서비스로, 특정 공간이나 장소에 대한 정보를 입력하고 평판 정보를 공유하는 방식이다. 포스퀘어의 표현은 베뉴(Venues), 체크인(Check-in), 팁(Tips), 메이어십(Membership), 슈퍼유저(Super user), 뱃지(Badge) 등의 용어를 사용하면서 공간에 대한 활동 정보를 기록하고 있다. 특히 트위터의 경우에는 포스퀘어와 연계하여 상세한 위치 정보를 공유하는 서비스를 제공하고 있으며, 페이스북은 주변 친구들을 찾는 서비스(Nearby Friends)를 적용하여 GIS 상에서 가까운 친구들을 찾아 대략적인 위치 정보나 공간 정보를 제공하고 있다. 또한 인스타그램은 포스퀘어와 플레이스 기능을 연계하여 소셜 네트워크에서 공간에 대한 체크인을 공유하도록 하였다. 이러한 공간 체크인 서비스는 사용자의 위치기반 추천 서비스나 개인화된 서비스로 발전시키기 위해 포스퀘어 모델을 적용한 서비스 개발이 가능하다.

소셜미디어 데이터에서 키워드 연관이나 의미적인 구성을 분석하는 텍스트마이닝은 지역명이나 건물, 상호 등에 대한 개체명을 지정하고 이를 관심지점으로 처리하는 과정을 빠르게 적용할 수 있다. 그림 2는 트위터에서 지역명을 처리하는 텍스트마이닝의 관심지점 처리 워크플로우이다[11]. 소셜미디어 데이터는 위치태그를 대상으로 지역명과 관련된 주요 명사 키워드를 추출하고 카테고리화하여 공간적 분포를 분석하여 주요한 벤치마크 데이터 셋을 이용하여 공간 관심지점과 연관성을 분석하였다. 또한 텍스트마이닝을 이용하여 주요 지역명 키워드를 추출하고 각각의 키워드를 카테고리로 분류하고 각 카테고리별로 재분류된 키워드의 지역별 분포를 검증하는 방법과 격자 기반 클러스터링을 통해 공간적 패턴 분석 모델을 연구하였다. 그림으로써 특정 지역에서 발생한 키워드와 실제 공간이 가지는 관심지점의 관련성을 확인하였고 시간과 사용자 활동 영역을 고려한 선호도 학습 모델과 카테고리

리별 장소의 중요도를 계산한 관심지점 추천 연구를 진행하였다[12]. 이 연구에서는 사용자의 지역별 체크인 이력들을 시간대로 분할하여 유사도와 선호도를 분석하고 사용자 카테고리 방문 통계를 활용하는 모델을 제시하였다. 그리고 장소별 방문에 따라 선호도 차이가 존재하기 때문에 희귀성을 고려한 카테고리 선호도를 정규화한 후 각 카테고리별 방향성 가중치 그래프를 기반으로 상대적 중요도와 카테고리 선호도를 결합하여 최종 장소를 추천하도록 모델링하였다.

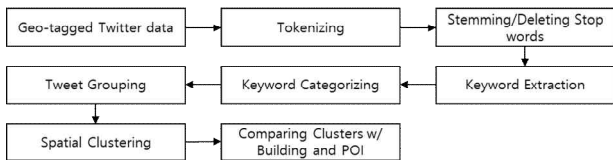


그림 2. 텍스트마이닝 기반 관심지점 처리 워크플로우
 Fig. 2. Workflow of textmining based POI processing

2-2 공간 지식 구성모델

공간 지식은 국내 국토연구원에서 공간과 시간속에 존재하는 구조(Structure), 과정(Process) 및 관계(Relationship)에 대한 지식을 의미하는 것으로 정의하고 있으며, 공간 자기상관성(Spatial autocorrelation)에 의해 만들어지는 패턴이나 위상관계를 포함하는 것으로 정의하고 있다. 또한 공간 지식 플랫폼은 누구나 쉽게 지리 공간자료를 공유하고 분석하여 도시생태계와 관련된 국가 현안을 진단하고, 집단 지성을 기반으로 창의적이고 합리적인 정책 대안을 창출할 수 있는 정보환경으로 정의하였다[13]. 공간 지식 플랫폼 연구는 공간정보를 기반으로 플랫폼의 활용성과 구축에 대한 방향성을 수요와 공급의 관점에서 상호작용하는 것으로 정의하고 다양한 전문가가 협업하는 공간 지식 창출을 위한 도구로 공간 지식 플랫폼을 연구하고 있다. 이는 공간 지식을 구성하는 형태가 공간의 구성체인 물체, 사람, 시설물 등과 같은 개체정보가 포함되고 물체, 사람 등과 같은 이동체에 대한 방향성이 정의되어 시공간적으로 변화되는 정보를 표현하는 프로세스를 그림 3과 같이 정의하였다.

공간 지식(GeoSemantic) 프레임워크는 데이터와 분석 및 시각화 모델간의 통합하는 관점에서 구조를 설계한 모델이다 [14]-[16]. 공간 지식 프레임워크는 분석 모델에서 룬테일 데이터를 사용하면서 기하급수적으로 늘어나는 데이터를 지식으로 구성하기 위한 문제 해결을 시도하였다. 룬테일 데이터와 모델은 반도를 기준으로 낮은 결과를 가져올 수 있지만 사용자의 의도와 가까워지고자 구체적인 정보를 포함하는 모델이다. 그렇기 때문에 좁은 범위에서도 공간 지식을 구성하는데 자동화된 통합을 통해 공간 데이터를 통합하고 지식의 격차를 줄일 수 있는 구조를 가지고 있다. 이 연구의 목표는 지구과학 자원의 통합을 위한 주석을 기록하고 연결하여 통합함으로써 추론하는 분산형 프레임워크를 개발하는 것이다. 데이터 리소스는 모델이 자동으로 시맨틱 웹 활용, 자동화된 시맨틱 중재 도구 개발,

룬테일 모델에서의 의미론적 지식 발견, 수집, 큐레이션 등의 전체 수명주기를 지원하는 SEAD(Sustainable Environment Actionable Data)와 상호 운용이 가능한 플러그 앤 플레이 시스템으로 변환하는 것을 지원하고 있다.

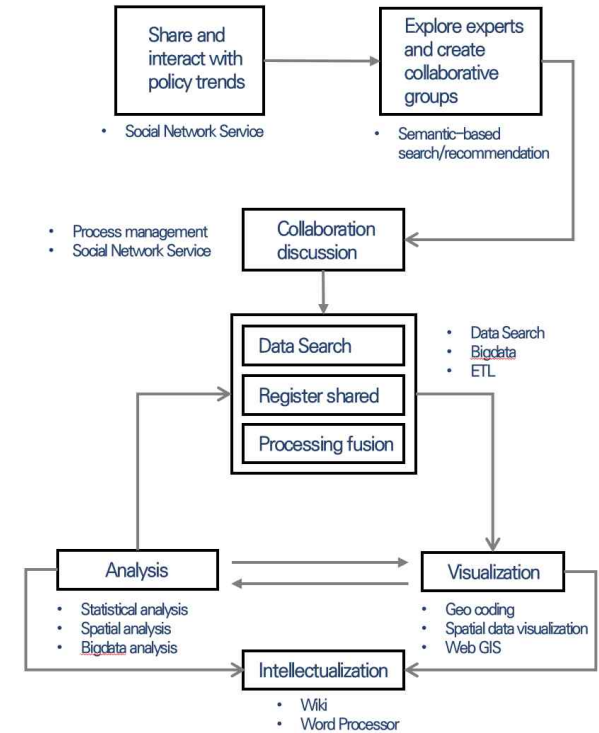


그림 3. 공간 지식 창출 과정을 지원할 수 있는 정보 기술
 Fig. 3. The information technology for processing of spatial knowledge creation

공간 지식 프레임워크의 기술적인 접근은 연결된 데이터와 함께 서비스를 기반으로 재사용이 가능한 프레임워크 구조를 설계함으로써 복잡한 구조의 데이터를 수집하고 이질적인 데이터를 특성화하여 동일한 개념을 가진 구조로 통합을 시도하였다. 이기종의 이질적인 데이터는 차원 정보, 좌표계, 배열, 변수, 제공자, 사용자, 상황정보 등으로 공간 지식을 구성하기 위한 정보를 목록화 하였다. 또한 다양한 서버를 대상으로 서비스를 제공하고 데이터 연동을 위한 RESTful(Representational State Transfer) 구조의 표준 인터페이스를 사용하여 리소스 지향의 아키텍처를 구성하였다. 그림 4는 공간 지식 프레임워크의 아키텍처를 나타낸다.

공간 지식 프레임워크는 데이터를 송수신하는 구조에서 RESTful 인터페이스를 적용한 JSON-LD(JavaScript Object Notation for Linked Data) 데이터 형식을 사용한다. 또한 Play Web APP(Application) Framework와 Apache Jena 미들웨어를 사용하여 데이터를 송수신하기 위한 웹 인터페이스 보안 구조를 만족하도록 설계하였다. 그리고, 아키텍처에서 지식 기반, 서비스 모델, 지식기반

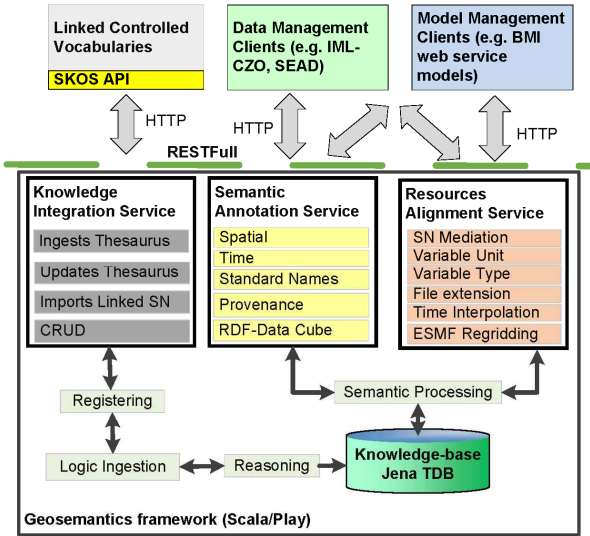


그림 4. 공간 지식 프레임워크 아키텍처
 Fig. 4. Architecture of HydroComplexity-Geosemantics framework

RDF(Resource Description Framework) 트리플, 추론 및 조작을 위한 파이프라인으로 빌딩 블록을 구성하였다. 지식베이스는 등록된 요소의 URI(Uniform Resource Identifier)를 그래프 노드로 저장하고 서비스되지 않는 요소에 대한 URI를 생성하여 제공할 수 있다. 그리고 분산 온톨로지의 엔티티는 공간 지식에 의해 만들어지고 활용되도록 하였다. 서비스계층은 세 가지 모델을 통해 프레임워크의 하위 구성 요소에 대한 인터페이스를 제공하고, 각 서비스의 RESTful 웹 서비스를 통해 독립적인 실행 기능을 구성하였다. 지식베이스(Knowledge-base)는 공간 지식 프레임워크에서 온톨로지를 모아서 사용하고 Apache Jena Tuple Database(Jena TDB)를 이용하여 영구 RDF 그래프 저장소를 사용하도록 하였다. 공간 지식 프레임워크에서 지식관리파이프라인(Knowledge Management Pipeline)은 온톨로지를 처리하는 4가지 구성요소로, 쿼리문, 정보 수집관리기능, 지식추론기능, 시맨틱 프로세스 기능으로 설계하였다. 쿼리문은 온톨로지의 URI를 저장하고 검색하는 질의어를 저장하고 관리하는 기능, 로직수집구성 기능은 의미 추론을 위한 링크 수집과 SKOS(Simple Knowledge Organization System) 온톨로지 일관성 점검 및 SKOS 온톨로지 기반의 일관성을 유지하는 기능, Jena 추론기능은 규칙과 추론엔진에 대한 시맨틱 추론과 지식 그래프에서 의미론적 지식 발견 및 지식 기반 추론 관리 기능, 마지막으로 지식 기반의 시맨틱 중재, Sparql 쿼리, 시맨틱 유사성 및 RDF 태그 구성을 위한 시맨틱 프로세스 기능을 제공하도록 한다. 서비스 레이어(Service Layer)는 SAS(Semantic Annotation Service), KIS(Knowledge Integration Service), RAS(Resource Alignment Service)의 3가지 서비스로 구성하였다. SAS(Semantic Annotation Service)는 시공간 컨텍스트, 변수 이름, 관계성을 위해 리소

스에 주석을 달고 데이터 파일에 GeoTiff나 CSV (Comma-Separated Values) 타입의 MIME(Multipurpose Internet Mail Extensions) 유형을 기반으로 자동 추출기를 설계하였다. 지식통합서비스(KIS, Knowledge Intergration Service)는 Controlled Vocabularies 및 W3C 표준을 프레임워크의 Knowledge-base에 수집, 등록 및 체크인하고 관련 데이터 관리 시스템에서 시맨틱 연합 검색을 제공하도록 하였으며, 자원정렬 서비스(RAS, Resource Alignment Service)는 통합 전에 의미론적 일관성을 보장하기 위해 두 개의 지리 자원과 연결된 정보 프로필을 관리한다. 이를 통해 공간 지식을 추론하고 관리하는 컨텍스트의 모호성을 의미론적으로 구성하여 다양한 분야의 상호 작용이 가능한 구조로 개선하고자 하였다.

2-3 공간 지식 추론모델

공간 지식 추론 모델은 자연어를 대상으로 공간의 명칭(location)과 방향(direction), 포함속성(containment), 관계성(relationship) 등을 이용하여 응용에 필요한 개념(Concept)들을 추론하고 관계성을 지식으로 표현하는 것이다. 이를 위해 온톨로지 지식(Ontology knowledge)을 이용하여 객체 간의 관계성(relationship), 공유 어휘(common vocabulary), 추론 규칙(inference rule) 등을 체계적으로 구성하는 작업을 수행한다. 온톨로지를 표현하는 대표적인 언어는 RDF(Resource Description Framework), OWL(Web Ontology Language) 등이 사용되고 있다[17]. 공간 지식을 표현하는 연구는 RDF와 OWL을 사용하고 있으며, 이를 확장한 연구로 시간과 지식을 표현하는 연구로 tRDF, tOWL, stRDF, STOWL[18]가 대표적이며, 공간을 지식으로 표현하는 연구는 SOWL, PelletSpatial, CHOROS, QUSAR 등이 대표적인 공간 추론(Spatial Resoner) 연구들이다[19]. 각각의 연구는 시맨틱 웹 발전에 따른 웹에서 자연어로 구성된 정보를 추출하고 그 정보를 시간과 공간으로 지식정보를 표현하고 구성하는 공간 추론 모델들이다.

공간 추론 모델은 RDF/OWL을 기초로 하는 시공간 지식을 4차원 서술자(4-D fluent)와 다자 관계(N-ary relation)으로 표현하고 CSD-9와 RCC-8 이론을 각각 적용하였다. 공간 지식베이스는 각 문장(statement) 또는 사실(fact)들로 구성되며, 두 공간 사이의 방향, 경계, 위상관계 등을 CSD-9와 RCC-8에서 정의한 공간 서술자(spatial property)들을 이용하여 표현하는 형태로 적용하였다. 공간 추론은 모든 공간에 대한 9개의 방향 중 하나로 표현하고, 위상 관계를 8개의 관계로 구성하여 영역(region) 관점의 실세계 공간들 사이의 다양한 관계를 표현하고 추론하는데 상호 보완적으로 이용할 수 있도록 설계하였다. 공간 추론 규칙은 9개의 방향을 대상으로 규칙을 만들고, 두 공간 사이의 방향 관계를 정의하였으며, 방향 관계를 명확히 할 수 없는 상태를 이접 관계들(disjunctive relations)로 정의하였다. 이러한 방법으로

RCC 공간 추론규칙들을 조합하여 두 개의 공간에 대한 조합 표를 이용하여 새로운 사실들을 유도하는 추론과정인 이행적 조합(transitive composition)으로 정의하였다. 이 방법은 서로 다른 관점에서 공간 지식 표현과 독립적인 이론을 방향과 위상에 대하여 통합적으로 추론하는 모델의 구성 가능성을 확인하였다. 각각의 모델은 방향 관계의 정의에서 CSD-9를 적용하고, 위상 관계를 표현하기 위해서는 RCC-8을 이용하여 새로운 사실을 유추하도록 하였다. 그렇게 함으로써 다양한 관계에 대한 정의 관계(defined relation) 또는 인접 관계들(adjacent relations)을 지식베이스에서 발견하고 이들이 암시하는 새로운 관계들로 변환할 수 있도록 하였다. 공간 지식들은 서로 간의 연관성을 강하게 하고 추론을 위한 조인 작업을 위해 부하분산 모델로 이행적 조합 추론 작업의 부하를 줄일 수 있도록 그림 5와 같이 설계하였다.

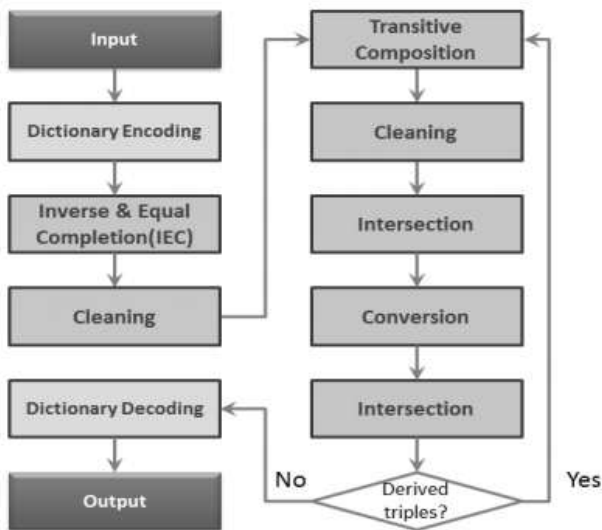


그림 5. 맵리듀스 기반 공간 추론의 작업 흐름도
 Fig. 5. Job flow the MapReduce-based spatial reasoner

시간 지식표현이나 시간 질의표현의 연구에서는 시간을 하나의 점(Time Point)이나 하나의 시간 간격(Time Interval)을 갖는 한 쌍의 시작점(Starting Point)와 끝점(Ending Point)로 구성하여 표현하였다. 그리고 이러한 형태의 시간 지식에 대한 질의를 처리하기 위해 시간 데이터에 관한 연산(computation of temporal data)에 의존하는 정량적 시간 추론을 수행하였다. 시간의 관계를 표현하기 위해 문장에서 사실(fact)들의 집합을 구성하고 이를 트리플(triple) 형태로 표현하여 유효한 시간 정보를 부여해 주는 시간 이벤트(Time Event)를 가정하였다. 이를 통해 하나의 시간 이벤트가 단순한 시간 이벤트인 하나의 시간 점(time point)이 될 수 있으며, 하나의 시간 간격(Time Interval)이 되어 시작 점과 끝점을 연결할 수 있도록 하였다. 이 구조에서 강조하는 것은 두 개의 시간 이벤트들 사이의 순서 관계가 존재하고 이 관계를 정의함으로써 대용량 데이터에서 시간 관계를 추론할 수 있

음을 확인하였다[18]. 그림 6 같이 13가지의 시간 프로퍼티(Temporal Property)들을 정의하고 서로 다른 시간 프로퍼티의 관계성을 정성적으로 표현하였다.

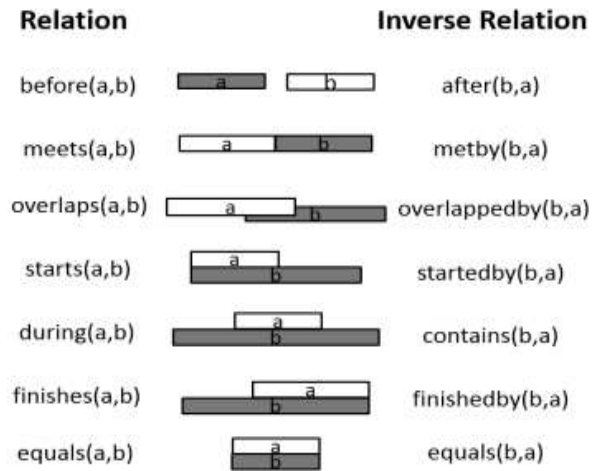


그림 6. 두 개의 이벤트 사이의 순서 관계성의 표현
 Fig. 6. Illustration of possible temporal ordering relationships between two events

정성적 시간 추론은 이벤트들간의 순서 관계들이 정성적으로 정의된 시간 지식베이스를 이용하여 시간 순서 관계들 사이의 모순(contradiction) 혹은 불일치성(inconsistency)이 존재하는지 검사하고, 정성 시간 추론(qualitative temporal reasoning)을 정의하여 이행 관계 추론 규칙, 동일 관계 추론 규칙, 역 관계 추론 규칙, 정제 추론 규칙을 정의하였다. 그래서 주어와 목적어로 구성된 이벤트들 사이의 시간적 순서 관계에서 새로운 사실을 유도하는 결정적 규칙(deterministic rule)과 비-결정규칙(non-deterministic rule)을 이행 관계 추론 규칙을 생성하였다. 이 과정은 또 다른 이행 관계를 추론하면서 다수의 이접 관계 사실들이 유도되고 확대 생산되기 때문에 추론 과정의 복잡성을 해소하기 위한 정의 관계 조합표가 있어야 효율적인 추론을 유도할 수 있다.

III. 공간 지식 표현 모델

공간 지식 표현 모델은 3차원 표현 기술을 이용한 디지털 트윈과 메타버스 서비스가 확장되면서 우리가 살고 있는 공간에 대해 시간의 흐름에 따라 표현하고 의미를 전달할 수 있는 구조로 표현하기 위한 세부적인 기술의 중요도가 증가하고 있다. 또한 특정한 개인을 대상으로 서비스를 하는 개인 맞춤형 서비스도 시간의 흐름에 따라 변화되는 공간의 흐름을 분석하여 공간의 상황별 서비스를 구성하고 표현하는 기술의 중요도가 함께 증가하고 있다. 다만 개인 맞춤형 서비스를 위해서는 각 개인의 정보를 하나로 통합하고 관리하는 개

인정보에 대한 문제와 개인별 데이터를 처리하기 위한 고성능 분산 병렬처리 프로세스가 요구되고 있다. 이 관점에서 시간의 흐름에 따라 공간의 변화를 분석하는 공간 지식 표현 모델과 공간을 중심으로 하는 시간의 흐름에 따른 상황을 분석하는 공간 지식 표현 모델을 구성하는 프레임워크를 구조화하였다. 표 1은 공간과 시간을 중심으로 각각 공간과 시간의 흐름에 따라 구성할 수 있는 지식 정보를 표현한 것이다.

표 1. 공간과 시간의 흐름에 따른 공간 지식 분석 모델
Table 1. A model of spatial knowledge by time-line and spatial

	model of spatial-based expression	model of time-based presentation
Spatial series	<ul style="list-style-type: none"> · Spatial directional expression · Spatial relational expression 	<ul style="list-style-type: none"> · Analysis/expression of spatial movement relationships · Spatial time series analysis/expression · Analysis of spatial preferences · Space movement optimization expression · User or collaborative filtering recommendations, etc
Time series	<ul style="list-style-type: none"> · Analysis/expression of situation in space · Expression of relationships between spaces · Expression of spatial preferences · Item-based recommendations 	<ul style="list-style-type: none"> · Event flow analysis

공간 지식 표현 모델은 공간과 시간의 흐름에 따라 수집되는 데이터가 동일해도 분석 기준이 다르면 표현 방법이 달라진다. 공간 지식은 공간 개체명과 공간 관계 속성 정보로 구성되어 공간정보의 핵심 주체를 정의한다. 공간 개체명은 수집된 소셜미디어 데이터에서 공간 데이터 추출하여 공간 개체명으로 정의하고, 공간 관계 속성 정보는 수집 데이터에서 분석 내용을 속성 정보(이동도구, 방향성, 공간의 연관성, 관계성, 선호도, 추천, 트래킹, 트렌드 등)를 공간 관계 속성으로 정의할 수 있다.

3-1 공간의 흐름 기준 공간 중심 표현 모델

공간 중심 표현 모델에서 공간의 흐름을 기준으로 분석하는 분석 모델은 수집된 데이터에서 각 문장이 가지는 시작 위치 정보와 다음 위치 정보를 분석하여 그 결과를 공간 개체명으로 적용하고, 해당 문장에서 공간 관계 속성 정보를 추출하거나 유추한다. 예를 들면, 수집된 문장에서 “제주도 서귀포로 워케이션을 떠나요!! 지금 어디? 김포공항”의 문장을 분석하면, 공간 개체명은 제주도, 서귀포, 김포공항으로 추출된다. 공간정보의 추출 기준은 국가관심지점정보를 대상으로 공간정보와 지역명, 건물명, 상호명을 이용하여 공간 개체명을

추출하고, 공간 개체명이 갖는 속성을 추출하여 공간 속성을 정의한다. 공간 개체명으로 추출된 제주도, 서귀포, 김포공항에 대한 지역적 범위로 포함관계를 구성한다. 그리고 포함관계로 묶인 집합에 대해 공간 관계 속성으로 시작 지점과 종료 지점으로 분석한다. 제주도와 서귀포에 해당하는 공간 개체명은 같은 범주의 지역명 포함(종속)관계에 속한다. 그러므로 최소단위의 지역명인 서귀포를 공간 개체명으로 표출하고, 상위에 해당하는 공간 개체명인 제주도에 대해서는 공간 관계 속성 정보로 재정의한다. 그리고 김포공항에 대한 공간 개체명은 제주도와 서귀포를 대상으로 포함관계가 성립되지 않기 때문에 다른 지역명으로 정의되어 공간의 시작과 끝에 대한 연결속성을 정의할 수 있다. 이렇게 정의된 공간 개체명을 기준으로 문장 분석으로 공간 관계 속성 정보를 구성한다. 표 2와 같이 공간 관계 속성 정보는 공간 개체명에 대한 포함관계 정보와 출발지, 도착지, 방향, 거리 등의 정보를 구성해서 공간 방향성과 관계성을 표현한다.

예시 문장에서 공간 개체명을 제외하고, 행위를 나타내거나 목적성을 나타내는 키워드는 “떠나요”와 “지금”이라는 단어를 활용할 수 있다. 인공지능 딥러닝 학습모델을 통해 자연어 상의 표현을 기준으로 출발지와 목적지를 충분히 추출할 수 있으며, 이와 관계된 정보를 공간 관계 속성 정보로 개체 구성이 가능하다. 휴리스틱 학습모델을 통해서도 출발지와 도착지에 대한 흐름을 정의할 수 있다. 예시문을 기준으로 출발지는 김포공항이며, 도착지는 서귀포라는 공간 개체명의 속성 정보를 획득할 수 있다. 그리고, 이 두 개의 공간 개체명을 이용하여 이동에 대한 속성 추론이 가능하다. 이렇게 구성된 공간 개체명과 공간 관계 속성 정보를 정리하면, “Start : 서울김포공항”, “Goal : 서귀포”, “Direction : South”, “Tools : Airplane”, “Distance : 479km”, “TS : 2021-10-11T12:30:28”로 정리할 수 있다.

표 2. 공간 지식 표현을 위한 속성 정보 구성의 예
Table 2. A example of configuring attribution for spatial-based presentation

Tag	Description	Type	Note
Start	Start location spatial elements	String	
Goal	End location spatial elements	String	
Direction	Eight cardinal points	String	
Tools	How to move	String	
Distance	Distance information between start and goal	Number	
TS	Destination arrival time	Number	
POI	Point of Interest	String	
TBS	Time range index	String	
Event	Outlier & event	String	
Rate	Status reading rate	Number	
Range	Location range	Number	
Preference	Preference attribute	Number	
Relation	Relation attribute	String /Number	

3-2 시간의 흐름 기준 공간 중심 표현 모델

공간 중심 표현 모델에 시간의 흐름을 적용한 표현 모델은 공간 중심으로 사용자의 이동을 모형화하고, 공간 이상 상황 탐지, 공간의 밀집도, 공간의 선호도, 공간 간의 관계성 등을 분석 할 수 있는 공간 지식 구성이 가능하다. 이 모델은 공간 이라는 아이টে임을 중심으로 공간에서 인식된 사용자 밀집도를 이용하여 선호도를 구하고, 그 선호도를 기준으로 공간 이상 상황을 탐지할 수 있다. 그렇기 때문에 관심지점 기준의 공간 정보 리스트와 소셜미디어의 공간정보를 인식하는 공간 사전과 상황관리가 중요하다. 표 3은 관심지점을 대상으로 공간 기준의 변화를 정리한 것이며, 소셜미디어 데이터로 제시된 예시글을 기준으로 관심지점 공간 정보에서 인식된 사용자를 표현한 것이다. 그리고, 관심지점 목록은 변경된 정보를 관리하고 신규 공간을 재구성하기 위한 공간 사전으로 통합하여 관리하는 기능이 추가되어야 한다.

표 3에서 보는 것과 같이 관심지점을 대상으로 인식된 사용자들에 대해 시간 단위로 그룹핑한다. 각각의 공간은 공간 개체명으로 공간 사전과 함께 적용되어 공간 상황에 대한 이벤트와 분석 정보를 공간 관계 속성 정보로 등록하게 된다. 예시로 제시한 문장의 사용자를 A라고 가정하고, A를 인식된 공간에 표시하면 문장에서 등록된 정보를 기준으로 정리할 수 있다.

표 3. 공간 중심 표현 모델의 시간 흐름 적용 모델
Table 3. A model of spatial-based presentation by time series

Hour(24) \ POI	GMP	CJU	Seogwipo Hotel	Jungmun Tourist Complex	Shilla Hotel
00
...
12	A, B, C, ...	R, T, X,
13	K, F, R, ...	A, B, C,
14	S, U, P, ...	K, F, R,	A, C, X...	...
...

- “POI : 김포공항”, “Start : 서울역”, “Goal : 김포공항”, “TS : 12:30:28”, “Direction : West”, “Event : ”, “Tools : Subway”
- “POI : 제주공항”, “Start : 김포공항”, “Goal : 제주공항”, “TS : 13:40:15”, “Direction : South”, “Event : ”, “Tools : Air Plane”
- “POI : 중문단지”, “Start : 김포공항”, “Goal : 제주공항”, “TS : 14:50:55”, “Direction : South”, “Event : ”, “Tools : Bus”

이렇게 구성되는 정보는 각각의 관심지점 공간 사전을 기준으로 공간의 상황 분석과 밀집도, 공간간의 관계성 등을 분석하여 사용자들의 다음 공간에 대한 추천이 가능하다.

3-3 공간의 흐름 기준 시간 중심 표현 모델

시간 중심 표현 모델에 공간의 흐름을 적용한 모델은 시간 중심으로 사용자의 이동공간에 대한 정보를 구조화하고, 사용자 중심의 공간 선호도, 공간간 관계성, 사용자 상태변화, 시계열 기준의 공간이동 및 최적화, 사용자의 상태 모니터링 등을 분석할 수 있는 공간 지식 구성이 가능하다. 이 모델은 시간의 흐름에 따라 변화되는 공간을 연결하기 때문에 사용자에게 선호도를 장소에 대한 머문 시간이나 반복적인 방문이 이루어진 장소에 대해 시간 이벤트를 구성하고 공간별 클러스터링을 통해 개인 맞춤형 공간 관계 속성 정보 구성이 가능하다. 표 4는 제시된 예문을 이용하여 시간 중심 표현 모델에서 공간 흐름을 샘플링한 표이다. 이 모델은 시간과 사용자를 기준으로 사용자가 움직인 공간 개체명을 공간 사전으로 구성하여 정보를 표현하고, 상황을 분석하는 모델이다. 그리고, 관심지점 구조는 개인 사용자의 용어를 기준으로 연관 관심지점을 포함하는 관계정보를 통합하여 관리하는 기능이 필요하다.

표 4와 같이 사용자와 시간을 기준으로 사용자의 공간 이동에 대한 정보를 반영한다. 이동 공간에 대한 정보는 공간 개체명과 공간과 연계되는 공간 관계 속성 정보를 통합하여 구성한다.

표 4. 시간 중심 표현 모델의 공간 흐름 적용 모델
Table 4. A model of time-based presentation by spatial series

Hour(24) \ USER	A	B	E	T	...
00
...
12	GMP	GMP
13	CJU	CJU	...	Hallasan	...
14	Jungmon Tourist Complex	Hallasan	...	Jungmon Tourist Complex	...
...

예로 제시된 문장을 이용하여 시간 중심의 표현 모델을 대상으로 공간 흐름을 정리하면,

- “TBS : 12”, “Start : 서울역”, “Goal : 김포공항”, “TS : 12:30:28”, “Direction : West”, “Event : ”, “Tools : Subway”
- “POI : 제주공항”, “Start : 김포공항”, “Goal : 제주공항”, “TS : 13:40:15”, “Direction : South”, “Event : ”, “Tools : Air Plane”

- “POI : 중문단지”, “Start : 김포공항”, “Goal : 제주공항”, “TS : 14:50:55”, “Direction : South”, “Event : ”, “Tools : Bus”

이렇게 구성되는 정보는 각각의 관심지점 공간 사전을 기준으로 공간의 상황 분석과 밀집도, 공간 사이의 관계성 등을 분석하여 사용자들의 다음 공간에 대한 추천이 가능하다.

3-4 시간의 흐름 기준 시간 중심 표현 모델

시간 중심 표현 모델에 시간의 흐름을 적용한 표현 모델은 시간 중심으로 사용자에게 일어나는 사건 이벤트를 분석할 수 있다. 그러므로 사건이 일어난 공간에 대한 정보와 사건의 형태나 상황에 대한 정보를 수집하여 이벤트의 정확성을 판단하고 의사결정을 위한 정보를 공간 지식으로 구성해야 한다. 또한 사건 이벤트는 사용자를 행위를 특징하는 조건의 데이터가 함께 수집되거나 공간에서 행위를 유추할 수 있는 관심지점 공간 주석을 구성하여 이벤트에 대한 추론 근거를 제시해야 한다. 그리고 이벤트에 대한 다양성을 보장하고 수집된 데이터에서 행위 이벤트를 분석하는 클러스터링 모델이 요구된다. 표 5는 시간 중심 표현 모델에서 시간의 흐름으로 사용자의 행위를 유추한 이벤트를 공간 관계 속성값으로 체계화해야 한다.

표 5에서 보는 것과 같이 사용자는 공간에 속해 있지만 그 공간에서 이루어지는 행위를 이벤트로 활동을 그룹핑한다. 그 이유는 시간대별로 사용자의 공간보다는 시간에 집중된 사용자의 행위를 판단해야 하기 때문이다. 사용자의 행위는 이벤트로 구성하여 공간 관계 속성 정보에 연결하여 공간 지식을 이용한 사용자의 맞춤형 서비스에 적용이 가능하다. 예시로 제시된 문장을 기준으로 사용자에 대한 행위 분석과 유추한 시간 흐름을 정리하면,

- “Event : 교통, 이동”, “Start : 서울역”, “Goal : 김포공항”, “TS : 12:30:28”, “Direction : West”, “Tools : Subway”
- “Event : 교통, 이동, 비행”, “Start : 김포공항”, “Goal : 제주공항”, “TS : 13:40:15”, “Direction : South”, “Event : ”, “Tools : Air Plane”
- “Event : 교통, 이동”, “Start : 김포공항”, “Goal : 제주공항”, “TS : 14:50:55”, “Direction : South”, “Tools : Bus”

이렇게 구성되는 정보는 이벤트 단위의 해석을 통해 사용자의 상황에 맞는 추천과 시간대별 공간에서 발생된 사용자들의 행위를 그룹핑한 공간사용 서비스를 추천하는 프로그램이 가능하다.

표 5. 시간 중심 표현 모델의 시간 흐름 적용 모델

Table 5. A model of time-based presentation by time series

USER	A	B	E	T	...
Hour					
00
...
12	Flight	Flight
13	Flight	Flight	...	Hiking	...
14	Falls, Walking	Hiking	...	Photo, Falls	...
...

공간 지식 표현 모델을 이용하여 공간 지식 추론을 위한 공간 지식베이스 구성은 RDF/OWL의 시맨틱 웹 표준 온톨로지 언어로 구성한다. 문장은 트리플 문장(N-triples statement)으로 구성요소인 Subject, Predicate, Object를 표현하고, 지식베이스에 등장하는 각 장소는 국가관심지점정보를 확장한 공간 개체명 클래스에 속하는 한 원소로 정의한다. 공간 지식베이스를 구성하는 각 문장(Statement) 혹은 사실(fact)들은 그림 7과 같다. 공간 개체명 또는 장소 사이의 방향, 구역, 고도, 위상, 이벤트, 도구, 거리 관계 등을 정의한 공간과 시간 속성(spatial property)들을 이용하여 표현하는 형태이다. 또한 공간에 대한 추론을 위한 참조모델로 CSD-9과 RCC-8 공간 추론 모델을 통해 공간과 공간 사이의 관계를 표현하는 공간 관계 속성 정보를 표현한다.

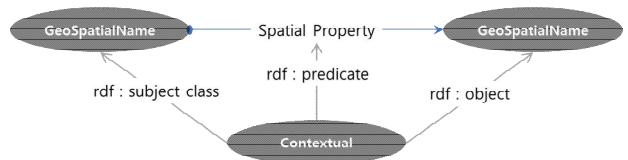


그림 7. N-트리플 상태 표현

Fig. 7. A Graph presentation of RDF(N-Triples statement)

IV. 공간 지식 추론 프레임워크

공간 지식 추론 프레임워크는 공간 지식을 구성하고 분석된 지식 정보를 표현하기 위한 구조적 재사용성을 갖는 프로세스 기준의 개발 플랫폼으로 정의할 수 있다. 공간 지식 추론 프레임워크는 공간 지식을 구성하고 관리하는 데이터 수집, 저장관리 기능과 수집된 데이터를 분석하여 공간 지식 구성 및 지식분석을 위한 처리 기능, 서비스를 구성하기 위한 서비스 구성 기능으로 구성되며, 각 기능에 대한 상태와 접근성을 관리하는 모니터링 기능으로 구성된다. 그림 8은 공간 지식 추론 프레임워크의 기본 구성을 위한 구조이다.

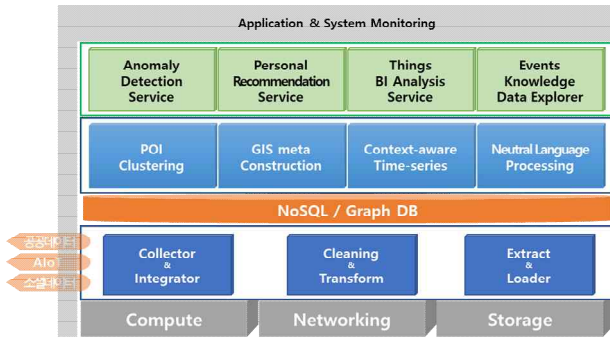


그림 8. 공간 지식 추론 프레임워크 구성 개요
 Fig. 8. A framework for spatial acknowledgment reasoning

데이터 수집 및 저장관리 기능은 공간의 상태와 상황에 대한 추론과 함께 사용자를 대상으로 의사결정을 지원하기 위한 데이터를 수집하는 환경의 구축과 관리 기능의 구성이다. 데이터 수집은 크롤링 에이전트(Agent)를 이용한 수집 기능이 대표적이며, 데이터베이스에서 필요데이터를 정제하고 기록하는 ETL(Extract, Transform, Load) 도구를 이용하는 방법 등 다양한 도구를 사용하고 있다. 이러한 수집 기능은 힙 메모리(Heap Memory)를 관리하면서 데이터를 오류 없이 전송하는 것이 중요하다. 그리고 가비지(Garbage) 없이 수집된 데이터를 전송하는 기능도 중요하다. 특히 프로파일링 인터페이스를 통해 수집 중에 영향을 받은 개체에 대한 자세한 정보를 제공하는 것이 중요하다. 이것은 수집대상이 증가하면서 생성되는 수집 로봇 관리와 수집 대상을 모니터링하면서 제어하는 관리도구의 필요성이 증가하면서 클라우드 가상화 솔루션과 동기화 프로그램의 사용이 강화되고 있다. 또한 공통의 프로파일링 인터페이스를 통해 수집 데이터가 버퍼링 소프트웨어로 전송되는 상태를 모니터링할 수 있도록 정보를 제공하는 것도 에이전트 통신언어의 구성도 중요하다. 수집된 데이터는 비정형과 반정형 구조를 갖는 데이터가 많은 비중을 차지하는 대용량 데이터이며, 데이터에 대한 시간과 공간에 대한 흐름을 통제할 수 있는 구조로 관리하는 그래프 구조의 데이터 저장 관리도로 구성하는 것이 중요하다. 그래프 데이터베이스는 개별 개체의 데이터를 나타내는 노드(Node), 성질이 비슷한 객체들을 묶어주는 라벨(Label), 객체들의 관계를 표시하는 엣지(Edge)로 구성되기 때문에 공간 지식을 표현하는 구조와 유사한 기능을 포함하고 있다. 그래서 직관적인 분석이 가능하다.

그래프 데이터베이스에 저장된 수집 데이터는 공간에 대한 상황 분석과 이벤트 분석 기능, 공공 데이터와 국가관심지점에 대한 업데이트와 자연어 처리를 통한 연관성 분석 기능이 포함되어 공간 지식을 구성한다. 공간 지식은 인텍스된 관심지점 상의 공간 지식 지도를 구성하고 각 관심지점 키워드를 대상으로 공간 관계 속성 정보를 연결한다. 공간 관계 속성 정보는 서비스를 처리하기 위한 2형 분석 데이터로 사용된다. 2형 분석 데이터는 이벤트의 유형별 패턴 분석과 상황에 대한

이해, 문맥 분석 등을 통해 공간을 제어하고 서비스에 활용하는 범주를 정의할 수 있다. 그리고 개인 맞춤형 서비스를 구성하여 공간과 사용자를 대상으로 맞춤형 추천 서비스 구성이 가능하다. 공공 데이터 분석은 공간과 관련된 국가관심지점을 대상으로 지역의 인구, 교통, 경제 활동 지표를 이용하여 지역별 관심지점을 기준으로 통합된 정보를 구성하여 2차 가공 데이터로 구성한다. 그리고, 각 데이터가 갖는 고유한 특징이자 문제인 업데이트 일자를 동기화하는 최소단위 타임동기화가 이루어져야 한다. 국가관심지점은 지역에서 상호의 변경이나 다른 이름으로 불리는 경우가 많기 때문에 이명 처리에 대한 기준이 명확해야 하며, 소셜미디어 데이터로부터 이명과 연관된 정보를 적용하는 이명처리 모델이 수립되어야 한다. 보통은 공간 클러스터링을 통해 지역명과 이명을 하나로 그룹핑하는 작업을 통해 구성하고 있다. 상황인식과 자연어 처리 분석은 인공지능 알고리즘을 사용한 분석이 가장 많이 사용되고 있으며, 그 중에서도 한국어 처리가 우수한 KoBERT 모델이 대표적이다. 상황인식(Context-aware) 분석은 시간의 흐름을 기준으로 공간에서 발생하는 이벤트를 클러스터링하고, 이벤트의 중요도에 따라 소셜미디어 상이나 공간을 관리하는 관리자를 대상으로 2차 서비스 구성을 위한 정보로 사용된다. 그래서 상황인식은 공간에서 인식되는 공공 데이터 또는 센서 데이터를 통합해서 사용하고, 자연어 처리를 함께 처리하는 구조의 딥러닝 모델을 구성하는 것이 중요하다. 자연어 처리는 공간에 대한 정보를 클러스터링하고 시계열 기준의 사건을 추론할 수 있는 공간 관계 속성에 대한 정보를 다중 분류하여 소셜 데이터에서 공간 기준의 감성분석과 상황에 대한 다중 클래스로 분류해야 한다.

이렇게 분석된 2차 가공 데이터는 빠른 접근과 해석을 지원하는 데이터 검색엔진을 적용하여 공간 단위 맞춤형 서비스 제공에 필요한 성능을 보장할 수 있다.

V. 결론 및 향후연구

본 연구에서는 소셜미디어 데이터로부터 추출한 공간 개체를 국가관심지점정보를 이용하여 지리정보와 관계성을 강화하는 공간 지식 표현 모델과 공간 지식 추론을 위한 프레임워크를 제안하였다. 공간 지식 표현 모델은 시간과 공간의 기준에 따라 4가지 모델로 제안하였으며, 각 모델은 시간과 공간의 기준에 따라 수집되는 데이터가 동일하더라도 분석 기준에 따라 이동도, 방향성, 연관성, 선호도 등의 관계 속성을 정의하여 표현을 달리 할 수 있는 구조를 정의하였다. 이러한 표현 모델은 공간적, 시간적 데이터를 통합하여 사용자의 움직임과 선호도에 대하여 통합적으로 분석하여 표현할 수 있다. 공간 지식 추론 프레임워크는 공간 지식 표현 모델을 구조화하여 분석하기 위해 설계된 프레임워크로 공간 상태 및 상황에 대한 의사결정에 중요한 데이터 수집 및 저장 관리,

모니터링 기능 등으로 구성되어 있으며, 공간 지식 표현과 유사한 직관적인 분석 구조를 제공한다. 또한 분석된 데이터는 검색 엔진을 통한 빠른 접근 및 해석이 가능하도록 설계되어 맞춤형 서비스의 성능 요구를 충족시킬 수 있도록 지원한다.

국가관심지점은 공간이 변경되는 정보를 실시간으로 반영하지 않기 때문에 이에 대한 관심지점 정보를 관리하기 위한 연관 공간 개체명을 생성하고 관리하는 기능도 함께 수집된 정보의 정제과정에서 확보하여 공간 지식을 표현하는 공간 지식 프레임워크에 반영한다. 공간의 연관성을 군집화하는 인공지능 알고리즘을 적용해 다양성을 확보해야 한다. 이렇게 정제된 정보와 연관성을 이용해 소셜미디어에서 언급되는 공간과 상황 정보를 공간 지식 표현 모델에 적용하여 공간 지식에 대한 지식베이스를 구축함으로써 공간에 대한 상황정보와 필요한 정보를 제공한 의사결정을 지원하는 서비스 구축으로 확장할 수 있다. 특히, 공간에 대한 개념을 공간과 시간으로 분리하여 개인이나 공간에 맞는 서비스를 추천하거나 공간에서 일어나는 상황을 실시간에 가까운 상황을 모니터링 하는 서비스로 활용할 수 있다. 향후 연구는 앞에서 제시한 공간 지식 추론 프레임워크를 기반으로 공간 지식 표현 모델과 챗봇 기반 공간 지식 추론 플랫폼을 개발하여 메타버스 서비스와 연동하는 것이다. 그럼으로써 물리적 공간에서 이루어지는 사건이나 공간의 특별한 상황을 논리적 공간으로 이동하여 개인에 맞는 서비스를 논리적으로 구성하여 추천할 수 있다.

감사의 글

본 논문은 2022년도 정부(국토교통부)의 재원으로 국토교통과학기술진흥원의 지원을 받아 수행된 연구임(RS-2022-00143336, 공간 지식추론 엔진 기술개발 사업)

참고문헌

- [1] NGII (National Geographic Information Institute). Point of Interest [Internet]. Available: <https://map.ngii.go.kr/mi/dwldSvc/poiDwldSvcInfo.do>.
- [2] E. Y. Kim and C. Y. Ku, "Analysis of Geographic Information Using Social Network Service Data and Topic Modelling: A Case of Instagram Data of Seongsu-Dong, Seoul," *The Geographical Journal of Korea*, Vol. 56, No. 3, pp. 189-202, September 2022. <https://doi.org/10.22905/kaopqj.2022.56.3.4>
- [3] J. Park, S. Kim, and S. Lee, "Analysis of Determining Factors of Urban Vitality with Mobile Phone Location-Based Origin-Destination Bigdata by Travel Purpose: Using the PageRank Algorithm and SHAP Machine Learning," *Journal of Korea Planning Association*, Vol. 57, No. 5, pp. 72-89, October 2022. <https://doi.org/10.17208/jkpa.2022.10.57.5.72>
- [4] M. G. Kim and S. H. Park, "Construction and Application of POI Database with Spatial Relations Using SNS," *Journal of Korea Spatial Information Society*, Vol. 22, No. 4, pp. 21-38, August 2014. <https://doi.org/10.12672/ksis.2014.22.4.021>
- [5] K. Yi, R. Yamagishi, T. Li, Z. Bai, and Q. Ma, "Recommending POIs for Tourists by User Behavior Modeling and Pseudo-Rating," arXiv:2110.06523, November 2022. <https://doi.org/10.48550/arXiv.2110.06523>
- [6] C. Zhuang, Q. Ma, and M. Yoshikawa, "SNS User Classification and Its Application to Obscure POI Discovery," *Multimedia Tools and Applications*, Vol. 76, No. 4, pp. 5461-5487, February 2017. <https://doi.org/10.1007/s11042-016-4034-6>
- [7] A. B. Shhaim and J. Berri, "Context-Aware Recommender Systems for Social Networks: Review, Challenges and Opportunities," *IEEE Access*, Vol. 9, pp. 57440-57463, April 2021. <https://doi.org/10.1109/ACCESS.2021.3072165>
- [8] S. W. Ha, K. W. Nam, and K. H. Ryu, "A Technique for Extracting GeoSemantic Knowledge from Micro-blog," *Spatial Information Research*, Vol. 20, No. 2, pp. 129-136, April 2012. <https://doi.org/10.12672/ksis.2012.20.2.129>
- [9] S. Kong, C. Chung, S. Ju, and J. J. Whang, "Knowledge Graph Embedding with Entity Type Constraints," *Journal of KIISE*, Vol. 49, No. 9, pp. 773-779, September 2022. <https://doi.org/10.5626/JOK.2022.49.9.773>
- [10] Y. Jee and C. Y. Ku, "Application Location-Based Social Network Data for the Construction of Point of Interest (POI): Using Stirling Edit Distance (Levenshtein Distance) Method," *The Geographical Journal of Korea*, Vol. 57, No. 2, pp. 101-110, June 2023. <https://doi.org/10.22905/kaopqj.2023.57.2.1>
- [11] W. J. Park and K. Y. Yu, "Spatial Clustering Analysis Based on Text Mining of Location-Based Social Media Data," *Journal of the Korean Society for Geospatial Information Science*, Vol. 23, No. 2, pp. 89-96, June 2015. <https://doi.org/10.7319/kogsis.2015.23.2.089>
- [12] K. Lee, J. Lim, K. Bok, and J. Yoo, "Recommending Personalized POI Considering Time and User Activity in Location Based Social Networks," *The Journal of the Korea Contents Association*, Vol. 18, No. 1, pp. 64-75, January 2018. <https://doi.org/10.5392/JKCA.2018.18.01.064>
- [13] D. Kim, K. Seo, R. Im, G. Shin, Y. Kim, and I. Han, A Study on Building and Utilization of Spatial Knowledge Platform for Collaborative Policy Making Research, Korea

Research Institute for Human Settlements, Anyang, BR
2016-30, December 2016.

- [14] Y. Hu, "Geospatial Semantics," arXiv:1707.03550, August 2017. <https://doi.org/10.48550/arXiv.1707.03550>
- [15] European Commission and Directorate-General for Digital Services, New European Interoperability Framework: Promoting Seamless Services and Data Flows for European Public Administrations, Publications Office of the European Union, Luxembourg, NO-07-16-042-EN-N, March 2017.
- [16] University of Illinois Urbana-Champaign. HydroComplexity-Geosemantics Framework [Internet]. Available: <http://ecgs.ncsa.illinois.edu/index.html>.
- [17] W3C Recommendation. OWL Web Ontology Language Semantics and Abstract Syntax [Internet]. Available: <https://www.w3.org/TR/owl-semantics>.
- [18] J. Kim and I. Kim, "MRQUTER: A Parallel Qualitative Temporal Reasoner Using MapReduce Framework," *KIPS Transactions on Software and Data Engineering*, Vol. 5, No. 5, pp. 231-242, May 2016. <https://doi.org/10.3745/KTSDE.2016.5.5.231>
- [19] S. H. Nam and I. C. Kim, "Design and Implementation of a Large-Scale Spatial Reasoner Using MapReduce Framework," *KIPS Transactions on Software and Data Engineering*, Vol. 3, No. 10, pp. 397-406, October 2014. <https://doi.org/10.3745/KTSDE.2014.3.10.397>



전원표 (Won-Pyo Jeon)

2014년 : 강원대학교 일반 대학원
(공학석사)

2014년 ~ 현 재: 바이브컴퍼니 스마트시티연구소 수석연구원
※ 관심분야 : 문서 분류, 개체명 인식, 텍스트마이닝, 실시간 분석, 스마트시티(Smart City) 등