

자연어 공간 질의 기반 공간분석 실행 자동화: RAG-AI Agent-MCP 연계 프레임워크를 중심으로

정진희*, 이경주**

*국립한국교통대학교 도시·교통공학과

**국립한국교통대학교 도시·교통공학전공

e-mail: jjh020729@naver.com

Automation of Spatial Analysis Execution Based on Natural Language Spatial Queries: Focusing on a RAG-AI Agent-MCP Integrated Framework

Jin-hee Jung*, Gyoung-ju Lee**

*Dept. of Urban & Transportation Engineering, Korea National University of Transportation

**Dept. of Urban & Transportation Engineering, Korea National University of Transportation

요약

본 연구는 자연어 공간 질의를 공간분석 실행으로 자동 전환하기 위한 RAG-AI Agent-MCP 연계 프레임워크를 제안한다. 기존 GIS 도구는 정형화된 입력 구조를 전제하기 때문에, 불명확한 조건, 복합적 의도, 구체적 시설 언급 등 비정형적 표현을 포함하는 자연어 질의를 직접 처리하는 데 한계가 있다. 이에 본 연구는 RAG, AI Agent, MCP를 계층적으로 연계함으로써 이러한 한계를 극복하고자 하였다. 제안된 프레임워크에서 RAG는 각 공간분석 방법론별로 사전에 구축한 예시 질의 데이터 (Seed Data)를 벡터 DB에 저장하고, 사용자 질의와의 유사도 검색을 통해 분석 전략 수립의 근거를 제공한다. AI Agent는 검색 결과를 바탕으로 분석 의도를 식별하여 실행 파라미터를 구성하며, MCP는 생성된 작업 명령에 따라 버퍼 분석, 공간 조인, 점수화 등 실제 공간 연산을 수행한다. 세 모듈은 각각의 역할을 분담하면서도 하나의 파이프라인으로 연결되어, 사용자의 일상적 표현이 구체적인 공간분석 결과로 이어지는 자동화 과정을 구현한다. 프레임워크의 적용 가능성을 확인하기 위해 보육시설 접근성 분석, 상권 활성화 분석, 범위 외 질의 처리의 세 가지 질의 유형을 대상으로 사례 분석을 수행하였다. 분석 결과, 프레임워크는 복합적 의도를 포함한 자연어 질의를 적절한 공간분석 방법론과 연계하고 후보 위치정보를 도출하는 데 유효하게 작동하였으며, 공간분석과 무관한 질의에 대해서는 유사도 임계값 기반 필터링을 통해 분석 범위를 스스로 제한하는 것을 확인하였다. 이를 통해 본 연구는 자연어 기반 공간분석의 활용 가능성을 확장하고, 공간분석의 기술적 접근성을 제고하는 데 기여하고자 한다.

1. 서론

1.1 연구의 배경 및 목적

빅데이터 기반의 인공지능(AI) 기술은 복잡한 도시 현상을 이해하고 도시 문제에 대응하는 핵심 수단으로 주목받고 있다(손정렬, 2020; Wang S. and Cao J., 2021; 제민희·정승현, 2024). 특히 대규모 언어모델과 자연어 처리 기술의 발전은 사용자가 일상적 언어로 표현한 질의를 분석 가능한 형태로 전환할 수 있는 새로운 가능성을 열어가고 있다(최대웅 외, 2023; 김규빈·유기윤, 2025). 이러한 흐름 속에서 RAG(Retrieval-Augmented Generation)는 외부 지식베이스의 검색 결과를 언어모델의 생성 과정에 결합함으로써 응답의 근거성을 제고하고 할루시네이

션(Hallucination)을 완화하는 방법론으로 주목받고 있다(Lewis P. and Perez E., 2020; 한국지능정보원, 2024). RAG는 지식 기반 질의응답 및 의사결정 지원 분야로 활용 범위를 확장해 나가고 있으며, 공간분석 분야에서도 자연어 질의에 내포된 사용자의 분석 의도를 식별하고 이를 적절한 분석 방법론 및 실행 절차와 연결하려는 시도가 확대되고 있다.

그러나 자연어 기반 공간분석은 이러한 연구 흐름에도 불구하고 여전히 몇 가지 근본적인 한계를 내포한다. 첫째, 사용자가 제시하는 자연어 공간 질의는 모호하고 비정형적인 표현을 포함하는 경우가 많다. 가령 "아이 키우기 좋은 동네"나 "마포구 먹자골목"과 같은 질의는 사용자의 공간적 요구를 일상적 언어로 표현한 것으로, 이를 특정 공간분석 방법론에 직접 적용하기 어렵다. 이러한 질의에는 크게 세 가지 유형의 비정형성이 내재한다. 첫 번째는 '아이 키우기 좋은'과 같이 평가 기준이 명시되지 않은 불명확한 조건이며, 두 번째는 '이사하면서 육아 환경도 고려하고

싶다는 표현처럼 하나의 질의 안에 둘 이상의 분석 목적이 혼재하는 복합적 의도이다. 세 번째는 '어린이집 근처'와 같이 특정 시설을 직접 언급하되 공간적 조건으로 전환하기 위한 기준이 불분명한 구체적 시설 언급이다. 이처럼 다양한 비정형성은 사전에 정의된 규칙 기반 처리 방식으로는 대응하기 어렵다. 나아가 동일한 분석 목적을 지닌 질의라 하더라도 "아이 키우기 좋은 동네", "다자녀와 함께 살기 좋은 동네", "육아하기 좋은 지역"과 같이 다양한 표현 방식으로 나타날 수 있어, 표현의 다양성과 비표준성은 자연어 질의 처리의 구조적 난점을 한층 심화시킨다.

둘째, 자연어 질의에서 식별된 의도가 실제 공간 데이터 처리 및 분석 실행으로 안정적으로 이어지기 위해서는 체계적인 실행 구조가 필요하다. 사용자의 의도는 단순히 분류되는 데 그치지 않고, 필요한 공간 데이터의 선택, 분석 방법론의 적용, 공간 연산의 수행, 결과의 점수화 및 시각화로 이어지는 일련의 과정과 유기적으로 연결되어야 한다. 따라서 자연어 공간 질의를 실행 가능한 공간분석 과정으로 전환하기 위해서는 의도 식별, 방법론 매칭, 공간 연산, 결과 도출을 통합적으로 연계하는 프레임워크가 요구된다.

이에 본 연구는 자연어 공간 질의를 기반으로 공간분석 실행을 자동화하기 위한 'RAG-AI Agent-MCP 연계 프레임워크'를 제안한다. 본 프레임워크에서 RAG는 사용자의 질의와 유사한 공간 분석 사례 및 방법론을 검색하여 분석 전략 수립의 근거를 제공하고, AI Agent는 검색 결과를 바탕으로 사용자의 의도를 식별하며 필요한 공간 데이터와 분석 절차를 구성한다. MCP는 생성된 작업 명령을 바탕으로 공간 데이터를 불러오고, 버퍼 분석, 공간 조인, 거리 계산, 점수화 등 실제 공간 연산을 수행한다. 이를 통해 본 연구는 자연어 질의가 공간분석 방법론과 연결되고 최종적으로 상위 후보 위치정보가 도출되는 자동화 과정을 제시한다. 나아가 본 연구는 자연어 기반 GIS의 활용 가능성을 확장하고, 공간분석의 기술적 접근성을 제고하는 데 기여하고자 한다.

2. 분석 방법

2.1 분석 프레임워크 구성

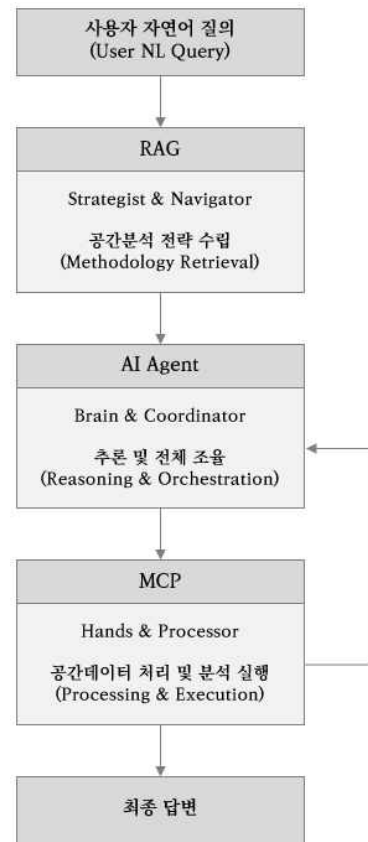
본 연구는 자연어 공간 질의를 공간분석 실행으로 자동 전환하기 위해 RAG-AI Agent-MCP 세 모듈을 계층적으로 연계한 프레임워크를 구성하였다(그림 1 참조). 각 모듈은 고유한 기능적 역할을 수행하면서도 순차적으로 연결되어 하나의 통합된 분석 파이프라인을 형성한다.

RAG(Strategist & Navigator)는 사용자 질의에 내포된 공간적 의도를 식별하고, 이에 부합하는 공간분석 사례를 검색하는 역할을 수행한다. 각 공간분석 방법론별로 사전에 구축한 예시 질의 데이터(Seed Data)를 벡터 DB에 저장하고, 사용자 질의와

의 유사도 검색을 통해 분석 전략 수립의 근거를 제공한다. 이 과정에서 문맥 기반의 벡터 검색과 키워드 기반의 BM25 검색을 결합한 하이브리드 검색 방식을 채택함으로써 검색 결과의 정확성을 제고하였으며, 검색된 유사 사례는 유사도 점수, 분석 의도, 실행 함수 정보 등 메타데이터 형태로 구조화되어 AI Agent로 전달된다.

AI Agent(Brain & Coordinator)는 RAG의 검색 결과와 원본 질의를 프롬프트 템플릿에 결합하여 LLM에 전달하고, 그 출력을 JSON 형식으로 파싱하여 분석 대상 지역, 필요 레이어, 버퍼 거리 등의 MCP 실행 파라미터로 구조화한다. 이 과정에서 AI Agent는 사용자의 모호한 표현을 분석 가능한 조건으로 변환하고, 어떤 공간 데이터가 필요한지, 어떤 순서로 연산을 수행할지를 결정하는 추론 역할을 담당한다. 이렇게 생성된 작업명령서는 MCP로 전달되어 실제 공간 연산의 입력값으로 기능한다.

MCP(Hands & Processor)는 전달받은 작업 명령에 따라 공간 데이터를 로드하고, 버퍼 분석, 공간 조인, 점포 수 집계 등의 연산을 순차적으로 수행한다. 산출된 분석 결과는 0-100점 척도로 정규화되며, 상위 후보 지점이 최종적으로 도출되어 지도 형태로 시각화된다. MCP는 LLM이 직접 처리하기 어려운 대규모 공간 연산을 전담함으로써 분석의 정확성과 실행 안정성을 제고한다.



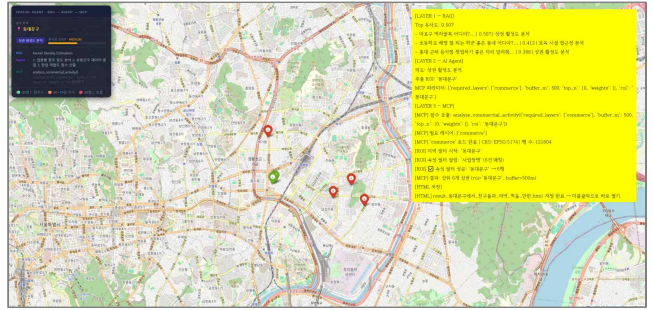
[그림 1] RAG-AI Agent-MCP 연계 프레임워크

3. 분석 결과

3.1 분석 결과 및 고찰

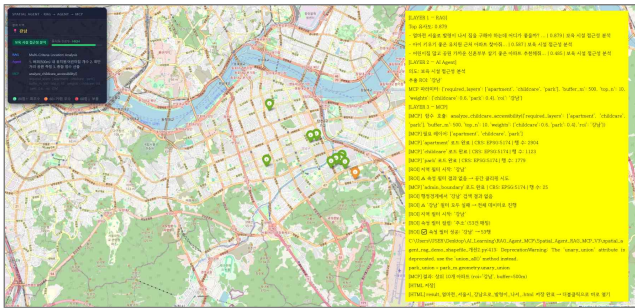
본 연구는 제안된 프레임워크의 실효성을 검증하기 위해 세 가지 유형의 자연어 질의를 입력하고 그 결과를 분석하였다.

첫 번째 질의는 "얼마전 서울시 강남으로 발령이 나서 집을 구해야 하는데 어디가 좋을까? 참고로 8세 아이가 있어."로, 주거지 탐색과 육아 환경을 복합적으로 요구하는 질의이다(그림 2 참조). RAG는 해당 질의를 보육시설 접근성 분석 의도로 식별하였으며(유사도: 0.879), AI Agent는 강남구를 분석 대상 지역으로 설정하고 어린이집·공원 등 공간 데이터 레이어에 항목별 가중치를 부여하여 MCP 실행 파라미터를 생성하였다. MCP는 이를 바탕으로 보육시설의 반경 내 개수와 가장 가까운 시설까지의 거리를 종합하여 점수를 산출하였으며, 강남구 내 상위 10개 아파트 후보지를 지도로 시각화하였다. 이는 단순한 행정구역 단위의 입지 탐색을 넘어, 실제 보육 인프라의 공간적 분포를 반영한 미시적 수준의 주거 입지 평가가 자연어 질의만으로 구현될 수 있음을 보여준다.



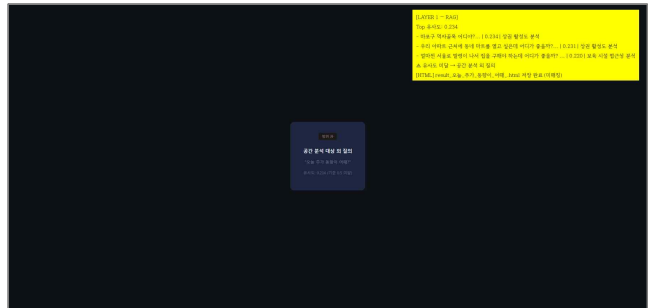
[그림 3] 자연어 질의 기반 상권 활성화 분석 및 시각화

세 번째 질의인 "오늘 주가 동향이 어때?"는 공간분석과 무관한 금융 정보 질의로, RAG의 최고 유사도가 0.234에 그쳐 설정된 임계값(0.5)에 미달하였다(그림 4 참조). 이에 프레임워크는 해당 질의를 공간분석 대상 외로 판별하고 분석을 수행하지 않았으며, 그 결과를 사용자에게 명시적으로 안내하였다. 이는 RAG에 사전 정의된 Seed Data와 방법론의 범위가 분석 가능한 질의와 그렇지 않은 질의를 구별하는 기준으로 기능할 수 있음을 보여주는 결과로, 시스템의 신뢰성과 운용 안정성 측면에서 중요한 기능적 요건을 충족한다고 평가할 수 있다.



[그림 2] 자연어 질의 기반 보육시설 접근성 분석 및 시각화

두 번째 질의는 "동대문구에서 친구들과 저녁 먹을 만한 곳 어디야?"로, 특정 지역 내 상권 밀집도를 탐색하는 질의이다(그림 3 참조). RAG는 해당 질의를 상권 활성화도 분석 의도로 식별하였으며(유사도: 0.507), MCP는 동대문구 상업 데이터를 대상으로 점포 밀도 분석과 종합 적합도 점수를 산출하여 상위 6개 상권을 도출하였다. 첫 번째 질의(0.879)에 비해 유사도가 다소 낮게 산출되었는데, 이는 해당 질의가 상권 탐색이라는 공간적 맥락보다 일상적 언어의 성격이 강하게 나타났기 때문으로 해석된다. 그러나 사전에 설정된 임계값(0.5)을 상회하는 유사도가 산출됨에 따라, 프레임워크는 분석 의도를 적절히 식별하고 결과를 도출하였다.



[그림 4] 공간분석 범위 외 질의에 대한 처리 결과

4. 결론

본 연구는 자연어 공간 질의를 기반으로 공간분석 실행을 자동화하기 위한 'RAG-AI Agent-MCP 연계 프레임워크'를 제안하고, 이를 실제 질의 사례에 적용하여 그 실효성을 검증하였다. 이를 위해 RAG는 사용자 질의와 유사한 공간분석 사례를 검색하여 분석 전략 수립의 근거를 제공하고, AI Agent는 검색 결과를 바탕으로 분석 의도를 식별하여 실행 파라미터를 구성하며, MCP는 생성된 작업 명령에 따라 실제 공간 연산을 수행하는 계층적 구조를 구축하였다. 사례 분석 결과, 프레임워크는 복합적 의도를 포함한 자연어 질의를 적절한 공간분석 방법론과 연계하고 후보 위치정보를 도출하는 데 유효하게 작동하였으며, 공간분석과 무관한 질의에 대해서는 유사도 임계값 기반의 필터링을 통해 분석 범위를 스스로 제한하는 것을 확인하였다. 본 연구의 핵심적인 기여는 RAG의 학습 자료를 구성하는 방식에

있다. 복잡한 이론적 분류 체계를 사전에 설계하는 대신, 일상적인 공간 질의부터 전문적인 분석 요구에 이르기까지 다양한 사례를 점진적으로 축적하는 방식을 채택하였다. 초기 예시 데이터(Seed Data)를 기반으로 사례를 단계적으로 확장하는 이 접근은 당장 필요한 질의부터 해결하면서 시스템의 커버리지를 넓혀가는 실용적 전략으로, 정교한 규칙 설계보다 한 조각씩 쌓아온 데이터 구축이 실질적인 성능 향상으로 이어질 수 있음을 시사한다. 한편 본 연구는 몇 가지 한계를 내포한다. 현재 프레임워크에서 활용 가능한 공간분석 방법론의 범위는 사전에 구축된 Seed Data에 의존하기 때문에, 데이터가 충분히 축적되지 않은 분석 유형에 대해서는 의도 식별의 정확도가 저하될 수 있다. 또한 본 연구에서 적용한 사례는 서울시 내 특정 지역을 대상으로 한정되어 있어, 다양한 도시 맥락과 공간 데이터 환경에서의 범용성은 추가적인 검증이 필요하다.

이러한 한계에도 불구하고 본 연구는 자연어와 공간분석 사이의 기술적 간극을 좁히고, 일상적 언어로 표현된 도시 공간 문제를 정량적 분석 결과로 전환하는 새로운 방법론적 가능성을 제시하였다. 이 점에서 의의를 지닌다. 향후 연구에서는 Seed Data의 지속적 확충과 분석 방법론의 다양화를 통해 프레임워크의 적용 범위를 확장하고, 실제 도시계획 및 공간 의사결정 지원 시스템로의 발전 가능성을 모색할 필요가 있다.

참고문헌

- [1] 김규빈·유기운, “임베딩 기반 토픽 모델링을 활용한 GeoQA 데이터셋의 의미 구조 및 질의 유형 비교 분석 연구”, 한국측량학회지, 제 43권 6호, pp. 789-798, 12월, 2025년.
- [2] 손정렬, “빅데이터 연구에서의 빅마켓, 빅네임 그리고 빅네트 워크: 국제학술지를 통해 본 도시 빅데이터 연구동향”, 대한지리학회지, 제 55권 2호, pp. 161-179, 1월, 2020년.
- [3] 이승석·이상준, “AI 할루시네이션에 대한 빅데이터 분석: LDA 토픽 모델링 및 감성 분석을 중심으로”, 한국산업보안연구학회, 제 14권 2호, pp. 153-168, 8월, 2024년.
- [4] 제민화·정승현, “빅데이터와 인공지능 기술 도입에 관한 도시계획 분야 직종 간 인식 분석”, 한국산학기술학회, 제 25권 1호, pp. 469-477, 1월, 2024년.
- [5] 최대웅·이석용·양종현·유기운, “공간 질의응답 시스템에서의 개체링킹과 딥러닝 모델을 활용한 멘션 탐지”, 대한공간정보학회지, 제 31권 4호, pp. 83-93, 4월, 2023년.
- [6] 한국지능정보원, “검색증강생성(RAG) 기술의 등장과 발전 동향”, 한국지능정보사회진흥원(NIA), 12월, 2024년.
- [7] Lewi P., Perez E., Piktus A., Petroni F., Karpukhin V., Goyal N., Kiittler H., Lewis M., Yih W., Rocktaschel T., Riedel S. and Kiela D., “Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks”, <https://doi.org/10.48550/arXiv.2005.11401>, 5월, 2020년.
- [8] Wang S. and Cao J., “AI and Deep Learning for Urban Computing”, Urban Informatics, pp. 815-844, 4월, 2021년.