건설공사 표준품셈 기반 AI 시스템 구축을 위한 데이터 구조와 저장방식 적용성 분석

장영훈*, 송태석*
*한국건설기술연구원 건설정책연구소 공사비원가관리센터
e-mail: songtaeseok@kict.re.kr

Data Structure and Storage Strategy for AI Systems Based on the Construction Standard Estimating System

Young-Hoon Jang*, Taeseok Song*
*Cost Engineering & Management Center, Department of Construction Policy Research, Korea Institute of Civil Engineering and Building Technology

요 약

건설 분야에서 AI 기반 검색 및 질의응답 시스템의 활용 가능성이 확대되고 있으나, 건설공사 표준품셈은 주로 문서 형태로 관리되어 AI 활용에 적합한 데이터 구조를 충분히 갖추고 있지 않다. 표준품셈은 계층구조, 적용조건, 주기사항, 공통 기준 등 정형·비정형 정보가 복합적으로 결합된 데이터로서 단순 텍스트 저장만으로는 한계가 있다. 본 연구는 표준품셈의 구조적 특성을 분석하고, AI 활용을 위한 데이터 요구사항을 도출한 후, 관계형, 문서형, 그래프형, 벡터형 데이터 저장방식의 적용성을 비교하였다. 분석 결과, 표준품셈의 AI 활용을 위해서는 구조화된 속성 관리와 의미 기반 검색을 함께 고려한 하이브리드 저장전략이 적절한 것으로 나타났다. 본 연구는 향후 표준품셈 기반 AI 시스템 구축을 위한 데이터 구조화 방향을 제시하고자 한다.

1. 서론

최근 건설 분야에서도 인공지능(AI)을 활용한 검색, 질의응답, 의사결정 지원 기술에 대한 관심이 높아지고 있다. 건설공사 표준품셈은 공사비 산정과 품셈 적용 판단의 기초자료로 활용되지만, 주로 문서 형태로 관리되어 AI 기반 시스템에 직접 적용하기에는 한계가 있다. 표준품셈은 장·절·항의 위계 구조, 작업 단위와 수량 기준, 적용 조건과 예외사항, 주기사항 및 공통 적용기준 등이 결합된 복합적 정보체계이기 때문이다.

건설 원가산정 분야에서도 BIM 및 온톨로지를 활용하여 적정 내역항목을 자동 추론하려는 연구가 수행되어 왔으며[1], 이는 공사비 산정 기준의 AI 활용을 위해 항목 속성, 조건, 관계 정보를 구조화하는 접근이 필요함을 시사한다. 따라서 표준품셈의 AI 활용을 위해서는 단순 문서 저장이나 키워드 검색을 넘어, 데이터의 구조적 특성을 반영한 저장방식과 데이터 구조화 방향에 대한 검토가 필요하다.

본 연구는 건설공사 표준품셈의 구조적 특성을 분석하고, 주요 데이터 저장방식의 적용성을 비교함으로써 표준품셈 기반 AI 시스템을 위한 데이터 구조화 방향을 검토하는 것을 목적으로 한다. 본 연구는 특정 데이터베이스 제품의 성능 비교가 아니라, 표준품셈 데이터 특성과 AI 활용 목적에 비추어 적합한 저장전략을 개념적으로 검토하는 데 초점을 둔다.

2. 표준품셈 데이터의 구조적 특성과 AI 활용 요구사항

건설공사 표준품셈은 코드 체계와 항목 구조를 기반으로 정형 정보와 비정형 설명이 결합된 복합 데이터이다. 예를 들어 “2-7-1 강관비계 설치 및 해체” 항목은 코드, 항목명, 단위, 작업조 기준 시공량, 조건별 시공량, 주기사항을 함께 포함한다. 해당 항목에서는 작업 높이와 규모 조건에 따라 시공량이 달라지며, 주기사항을 통해 적용 기준, 포함 작업, 별도계상 대상, 손료 적용 여부 등이 제시된다.

이는 표준품셈 데이터가 단순 수치표가 아니라 복수의 정보 유형이 결합된 구조임을 보여준다. 코드와 항목명은 품셈 체계 내 위치를 식별하는 정보이고, 단위·수량·시공량은 구조화 가능한 정형 데이터이다. 반면 주기사항은 적용 기준, 포함·제외 범위, 별도계상 대상 등 실제 적용 판단에 필요한 비정형 해석 정보를 담고 있다. 또한 작업 조건에 따라 시공량이 달라질 수 있으므로 조건 기반 필터링과 해석이 필요하다.

종합하면, 표준품셈 항목은 코드 기반 식별 구조, 구조화 가능한 정형 정보, 주기사항 중심의 비정형 설명, 조건별 적용 기준, 포함·제외 및 별도계상 관계가 결합된 복합 데이터이다. 따라서 표

준품셈의 AI 활용을 위해서는 원문 보존과 핵심 속성의 구조화가 동시에 필요하며, 이러한 특성은 데이터 저장방식 비교의 주요 기준이 된다.

3. 데이터 저장방식별 비교 및 적용성 분석

표준품셈의 구조적 특성을 기준으로 할 때, 관계형 데이터베이스, 문서형 데이터베이스, 그래프 데이터베이스, 벡터 데이터베이스는 각각 다른 장단점을 가진다. 본 연구의 비교는 특정 제품의 처리 성능평가가 아니라, 각 데이터베이스 유형의 일반적 데이터 모델 특성과 표준품셈 데이터 요구사항 간의 정성적 적합성 검토이다. 평가는 구조화 속성 관리, 원문 및 주기사항 보존, 조건 기반 검색, 항목 간 관계 표현, 의미 기반 검색, 버전 관리의 대응 정도를 기준으로 하였다.

최근 RAG 기반 질의응답에서는 외부 데이터베이스의 지식을 검색하여 생성 과정에 활용하는 방식이 논의되고 있으며[2], 이에 따라 표준품셈 기반 AI 시스템에서도 구조화 저장방식과 의미 기반 검색방식의 역할을 구분하여 검토할 필요가 있다.

표 1. 표준품셈 데이터 요구사항에 따른 저장방식별 정성적 적합성 비교

비교 항목	관계형 DB	문서형 DB	그래프 DB	벡터 DB
계층구조 표현	보통	높음	높음	낮음
조건 기반 검색	높음	보통	보통	낮음
원문 보존	낮음	높음	보통	보통
항목 간 관계 표현	낮음	보통	높음	낮음
의미 기반 검색	낮음	낮음	낮음	높음
버전 관리	높음	보통	낮음	낮음
AI 연계성	보통	보통	보통	높음

표 1에서 보는 바와 같이, 단일 저장방식만으로 표준품셈의 복합적 특성을 충분히 반영하기는 어렵다. 관계형 또는 문서형 데이터베이스는 기본 속성과 원문 관리에 적합하고, 벡터 데이터베이스는 자연어 기반 검색과 질의응답 기능을 보완하는 데 유리하다. 그래프 데이터베이스는 참조관계나 공통 기준 연계 분석이 필요한 경우 선택적으로 활용할 수 있다. 따라서 표준품셈 기반 AI 시스템에서는 관계형 또는 문서형 구조를 기본 저장소로 활용하고, 의미 기반 탐색을 위해 벡터 검색을 결합하는 하이브리드 저장전략이 현실적인 대안으로 판단된다.

4. 결론

본 연구는 건설공사 표준품셈의 AI 활용을 위한 기초 검토로서, 표준품셈 데이터의 구조적 특성을 분석하고 주요 데이터 저장방식의 적용성을 비교하였다. 분석 결과, 표준품셈은 코드 체계, 정형 속성, 주기사항, 조건별 적용 기준, 포함·제외 관계가 결합된 복합 데이터로 확인되었다. 이에 따라 단순 텍스트 저장이나 단일 의미검색 방식만으로는 충분한 활용이 어렵고, 구조화된 속성 관리와 의미 기반 검색을 함께 고려하는 저장전략이 필요하다.

특히 특정 데이터베이스의 우열을 단순 비교하기보다, 데이터 특성과 활용 목적에 맞는 저장구조를 설계하는 관점이 중요하다. 현 단계에서는 관계형 또는 문서형 구조를 기본 저장소로 활용하고, 질의응답 및 유사 항목 탐색을 위해 벡터 검색을 결합하는 방식이 적절한 대안으로 판단된다. 향후 실제 표준품셈 항목을 대상으로 구조화 실험과 검색 시나리오 적용 검토를 수행할 필요가 있다.

감사의글

본 연구는 국토교통부 기술혁신과 공사비산정기준관리운영사업(과제번호: 20260014)의 일환으로 수행된 연구임을 밝히며 이에 감사를 드립니다.

참고문헌

[1] S. K. Lee, K. R. Kim, J. H. Yu, “BIM and ontology-based approach for building cost estimation”, Automation in Construction, Vol.41, pp.96-105.

[2] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, “Retrieval-Augmented Generation for Large Language Models: A Survey”, arXiv preprint arXiv:2312.10997, 2023. Available From: https://arxiv.org/abs/2312.10997