

품셈 문서의 구조 기반 유사도 추천 시스템 설계 및 구현

장영훈*, 안방율*

*한국건설기술연구원 건설정책연구본부 공사비원가관리센터
e-mail:brahn@kict.re.kr

Design and Implementation of a Structure-Aware Similarity-Based Recommendation System for the Korean Standard Construction Cost Estimation Criteria

Young-Hoon Jang*, Bang-Yul Ahn*

*Construction Cost Engineering. & Management. Center, Dept. of Construction
Policy Research, Korea Institute of Civil Engineering and Building Technology

요약

본 논문은 「건설공사 표준품셈」을 구조적으로 분석하고, 이를 바탕으로 품셈 항목 간 유사도를 기반으로 추천하는 경량 추천 시스템을 구현하는 방법을 제시하였다. 품셈 문서를 코드 단위로 나누고, 각 항목을 제목, 설명, 주기사항으로 분리하여 임베딩 벡터를 생성하였다. 생성된 벡터는 가중 평균 방식으로 통합되며, 사용자 질의 문장과 유사도를 계산하여 관련성이 높은 품셈을 추천한다. 실험 결과, LLM 없이도 품셈 간 의미 기반 유사도 추천이 가능함을 확인하였으며, 시스템은 속도, 해석 가능성, 확장성 측면에서 높은 활용 가능성을 보였다.

1. 서론

건설공사 대부분의 단계에서 공사비 산정은 필수적으로 수행된다. 국내에서는 「건설공사 표준품셈」이 건설기술연구원에서 년 단위로 관리·발간되어 활용되고 있다[1]. 하지만 이러한 품셈 문서는 주로 PDF 또는 인쇄물 형태로 제공되고 있어, 사용자가 특정 공법이나 작업 조건에 부합하는 품을 직접 탐색하는 데 상당한 시간이 소요되고, 주관적인 판단에 의존하는 경우가 많다.

최근 자연어 처리 기술의 발달로 대규모 언어 모델(Large Language Model, LLM)을 활용한 문서 질의 응답 시스템이 제안되고 있으나, 이러한 접근은 그래픽카드 등의 연산 자원 소모가 크고, 할루시네이션 등의 환각 응답을 생성하는 등의 한계가 존재한다. 또한, LLM 기반 시스템은 학습된 지식에 의존하기 때문에 최신 문서나 특정 도메인에 대한 정확한 맥락 해석이 어려울 수 있다.

본 연구는 대형 LLM에 의존하지 않고, 품셈 문서의 구조적 특성을 활용하여 유사 항목을 효율적으로 추천할 수 있는 시스템을 구현하는 것을 목적으로 한다. 이를 위해 품셈 문서를 코드별로 청킹한 뒤, 각 항목을 제목, 설명, 주기사항 등으로

분리하고, 구조적으로 가중치를 부여한 임베딩을 생성하였다. 생성된 임베딩은 사용자의 질의 문장과 유사도 비교를 통해 가장 유사한 항목을 추천하는 방식으로 작동한다. 본 시스템은 가볍고 빠르며, 특정 도메인에 특화된 추천이 가능하다는 장점을 가진다.

2. 본론

2.1 품셈 문서 구조 분석 및 문제 정의

표준품셈은 공종 단위로 세분화되어 있으며, 각 항목은 고유 코드 / 제목 / 인력·장비 등의 산출표 / 주기사항 등으로 구성된다. 표준품셈은 다년간 축적된 데이터를 근거로 구축된 구조화된 기술 문서이며, 공통·토목·건축·기계설비 등 건설 분야 전반을 아우르는 많은 공종을 포함하고 있다. 건설 관련 숙련도에 따라 수작업으로 탐색하거나 정확하지 않은 검색어를 이용해 결과를 찾을 경우 어려움이 발생하고 있다.

기존의 키워드 기반 검색 시스템은 항목 간 의미적 유사성을 포착하기 어려우며, LLM을 사용하는 접근은 연산 비용과 복잡도가 크다는 단점이 있다. 따라서 본 연구에서는 품셈 문서의 구조적 특성을 반영하면서도 경량화된 임베딩 기반 유사도 검색 방식을 제안한다.

2.2 유사도 기반 추천 시스템 설계

표준품셈의 구조적 특성을 반영하여 유사 항목을 추천하는 경량 시스템을 설계하였다. 먼저 표준품셈 PDF 문서를 품셈 코드 단위로 청킹(chunking)하고, 각 항목을 제목(title), 인력·장비에 대한 산출표(body), 주기사항(notes)으로 분리하여 JSON 형태로 저장하였다. 표준품셈의 구조를 분리하여 텍스트의 각 구성 요소별 중요도를 다르게 반영할 수 있도록 하였으며, SentenceTransformer 기반 임베딩 모델(KR-SBERT)을 활용해 각 항목의 임베딩 벡터를 생성하였다.

임베딩 벡터는 각 구성 요소에 대해 가중 평균을 적용하여 생성되며, 제목에 대한 가중치 0.5, 인력·장비 산출표 영역에 대한 가중치 0.2, 주기사항에 대한 가중치를 0.3으로 설정하였다. 표준품셈에서 산출표는 수치적으로 중요한 의미를 포함하고 있으나, 의미적인 중요도는 낮으며 질문과의 유사도 계산을 포함하지 않는 것으로 분석되었다. 사용자가 질의하는 문장의 핵심 키워드가 대개 제목과 주기사항에 포함되는 것으로 가정하고 연구를 진행하였다.

생성된 임베딩 벡터는 별도로 저장하여 재활용되며, 사용자의 질의 문장은 실시간으로 임베딩되어 코사인 유사도 기반으로 비교된다. 유사도가 가장 높은 N개의 품셈 항목을 추천하는 방식으로 동작하며, 추천 결과는 품셈 코드, 제목, 주기사항 요약과 함께 사용자에게 제공된다.

이 시스템은 높은 GPU 자원 없이도 구동 가능한 경량 구조를 갖추고 있으며, 품셈 문서 외에도 메뉴얼, 규정집 등 다른 구조적 기술 문서에도 쉽게 확장 적용이 가능하다.

2.3 시스템 적용 및 결과 분석

시스템의 유효성을 검증하기 위해, 실제 「2025년 건설공사 표준품셈」 PDF 파일을 입력 데이터로 활용하고, 표준품셈의 전체 항목을 대상으로 임베딩 벡터를 생성하였다. 이후 사용자가 자주 탐색하는 시나리오에 기반한 총 10개의 질의 문장을 구성하고, 각각의 질문에 대해 Top-3 유사 품셈 항목을 추천하였다.

실험 결과, 예를 들어 "크레인을 사용해서 방음문을 해체하는 작업은?"이라는 질의에 대해 시스템은 "2-10-3 터널방음문 설치 및 해체"를 가장 높은 유사도로 추천하였다. 유사도 점수는 0.85 이상을 기록하였고, 주기사항에도 "조립 설치 및 해체", "용접 작업 포함"과 같은 핵심 키워드가 포함되어 있어 의미 기반 추천이 성공적으로 이루어졌음을 확인할 수 있었다.

다만, 질의가 지나치게 모호하거나 명확한 키워드를 포함

하지 않을 경우에는 임베딩만으로는 맥락 추론에 한계가 있음을 확인하였다.

3. 결론

본 연구에서는 「건설공사 표준품셈」의 구조적 특성을 반영하여, LLM의 도움 없이도 유사 품셈 항목을 효과적으로 추천할 수 있는 경량 유사도 기반 추천 시스템을 설계하고 구현하였다. 표준품셈 문서를 항목별로 분리하고, 각 항목을 제목, 설명, 주기사항으로 구조화하여, 임베딩 벡터를 가중 평균 방식으로 생성함으로써 문서의 의미적 요소를 반영하고자 하였다.

또한, 사용자의 질의 문장과 품셈 항목 간의 유사도를 코사인 유사도 기반으로 계산하여 실시간으로 유사도가 높은 품셈을 추천하는 기능을 구현하였다. 시스템은 품셈 항목 간의 유사성을 수치화하여 추천할 수 있으며, LLM 기반 시스템과 비교했을 때 경량성, 속도, 비용 면에서 실용성을 가진다.

본 연구의 결과는 건설 분야 외에도 메뉴얼, 지침서, 규정집 등 구조화된 문서를 활용하는 다양한 산업군에 적용 가능성이 높으며, 향후에는 표 정보까지 포함한 임베딩 개선, 사용자 피드백 기반 성능 고도화 등의 후속 연구를 진행하고자 한다.

감사의 글

본 연구는 국토교통부 기술혁신과 공사비산정기준관리운영사업(과제번호: 20250039) 및 한국건설기술연구원의 주요사업(건설정책 및 건설관리 발전전략)의 일환으로 수행된 연구임을 밝히며 이에 감사를 드립니다.

참고문헌

- [1] 한국건설기술연구원(2025), 2025 건설공사 표준품셈, 국토교통부