JESK J Ergon Soc Korea 2025; 44(4): 491-509
https://dx.doi.org/10.5143/JESK.2025.44.4.491
https://jesk.or.kr eISSN:2093-8462

# An Information Retrieval System Framework with Item Attribute-based Filtering and Retrieval-Augmented Generation

**Yunseong Kim[1], Gyewon Jeon[1], Taehyun Ha[2], Sangwon Lee[1]**

[1]Korea University, Department of Industrial and Management Engineering, Seoul, 02841
[2]Sejong University, Department of Artificial Intelligence and Data Science, Seoul, 05006

# 아이템 속성 기반 필터링과 검색증강생성을 결합한 정보 검색 시스템 프레임워크

**김윤성[1], 전계원[1], 하태현[2], 이상원[1]**

[1]고려대학교 산업경영공학과
[2]세종대학교 인공지능데이터사이언스학과

**Corresponding Author**
Sangwon Lee
Korea University, Department of Industrial and Management Engineering, Seoul, 02841
Email : upcircle@korea.ac.kr

Received : June 26, 2025
Revised : July 16, 2025
Accepted : August 25, 2025

Copyright@2025 by The Human Factors & Ergonomics Society of Korea. All right reserved.

**Objective:** This study aims to propose an information retrieval (IR) system that enables dynamic control of item attributes through filtering, allowing more precise alignment with users' contextual needs. By explicitly manipulating attribute combinations, the framework is expected to improve personalization and contextual relevance in recommendation tasks.

**Background:** Existing IR systems, including those enhanced by Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG), primarily focus on capturing textual relevance between user queries and item descriptions, often relying on embedding-based similarity. While such models improve contextual understanding and semantic matching, they still struggle to fully reflect fine-grained user preferences or dynamically adapt to nuanced contextual needs. Incorporating attribute-based filtering offers a complementary mechanism to explicitly control item selection based on user-specific contextual and preference-related conditions.

**Method:** We developed a three-module IR framework consisting of contextual query expansion, search filter generation via LLM and RAG, and final recommendation using attribute-based filtering and ranking based on embedding relevance. To explore its feasibility, we applied the framework to a music domain dataset containing both numerical and categorical item attributes. The system was tested across four emotional scenarios to examine how well it adapts to different user contexts.

**Results:** Experiments conducted on a music recommendation dataset under emotion-based query scenarios show that the proposed system more effectively aligns attribute distributions with user context compared to traditional sparse and LLM-based IR baselines.

**Conclusion:** The proposed IR framework demonstrates the effectiveness of combining semantic understanding with attribute filtering for personalized and context-aware recommendations.

**Application:** The proposed IR framework can be extended to any domain where

structured item attributes are available, offering a flexible and scalable solution for building context-aware, personalized IR systems.

**Keywords:** Information retrieval, Recommendation, Item attribute-based filtering, Large language model, Retrieval-augmented generation

## 1. Introduction

정보 검색(Information Retrieval, IR) 시스템은 대규모 데이터 중 사용자 질의와 관련된 정보를 효과적으로 찾아주는 기술로 추천 시스템에서도 사용자가 요구하는 조건에 부합하는 최적의 아이템을 추천하는 데 활용되고 있다(Zhang et al., 2024; Deldjoo et al., 2024). 이러한 사용자의 정보 요구를 충족시키기 위해서는 단순히 키워드 일치를 넘어 질의에 담긴 의미와 맥락을 정확히 이해하고 이를 반영하는 것이 중요하다(Hanani et al., 2001). 특히 IR 시스템이 맥락을 반영하여 관련성이 높은 결과를 검색하고 추천하는 과정은 곧 사용자 경험에도 연결되므로 이를 정교하게 구현하는 것이 핵심이다(Hambarde and Proenca, 2023). 이러한 IR 기반 추천 접근은 맥락을 정교하게 반영하기 위해 사용자 쿼리와 아이템 설명 간의 임베딩 유사도를 계산하여 관련성이 높은 아이템을 선별하는 구조로 연구되어 왔으며(Li et al., 2024; Deldjoo et al., 2024), 해당 과정은 궁극적으로 사용자의 개별적 요구를 반영하는 개인화 추천과도 긴밀히 연결된다(Deldjoo et al., 2024).

기존에는 희소 표현(Sparse representation)을 중심으로 단어 수준의 유사도를 측정했으나, 이후 문장 수준의 의미 해석이 가능한 밀집 표현(Dense representation) 방식이 도입되면서 문맥 해석 능력이 향상되었고 최근에는 이러한 표현을 생성하고 활용하는 데 있어 대규모 언어 모델(Large Language Model, LLM)이 중심적인 역할을 수행하고 있다(Hambarde and Proenca, 2023). 특히, 검색증강생성(Retrieval-Augmented Generation, RAG) 구조는 LLM을 기반으로 쿼리 확장(Query expansion), 검색(Retrieval), 재정렬(Reranking) 전반에서 맥락을 정밀하게 반영하는 구조로 주목받고 있다(Zhu et al., 2023; Wang et al., 2025). 그러나 이러한 LLM 및 RAG 기반 IR 구조는 주로 아이템 설명 임베딩에 의존하는 방식이며, 이는 사용자 쿼리와 아이템 설명 간의 표현적 유사성만을 중심으로 관련성을 판단하는 한계를 지닌다. 이로 인해 각 아이템에 내재된 미묘한 맥락적 차이를 충분히 반영하지 못하며 감정 상태나 상황에 따라 달라지는 사용자 선호와 같은 세부적인 정보 요구를 정밀하게 반영하는 데에도 문제가 나타날 수 있다(Lei et al., 2024).

아이템 속성(Item attribute)은 아이템마다 내재된 맥락적 차이를 나타내는 지표로 사용자 요구를 보다 정밀하게 반영하기 위한 수단으로 활용될 수 있다. 이러한 아이템 속성은 정량적(예: 에너지, 춤 적합성) 또는 범주형(예: 장르, 가수, 발매 연도) 형태로 표현되며 사용자가 아이템을 인식하고 선호를 판단하는 데 핵심적인 기준이 된다(Zhai, 2024; Chen et al., 2020). 이러한 속성을 기반으로 사용자 취향이나 상황에 맞게 필터링하여 추천을 수행하려는 접근이 활발히 이뤄지고 있으며, 이는 사용자의 세부적인 선호나 맥락적 요구를 정밀하게 반영하는 데 효과적인 방식으로 간주된다(Chen et al., 2020). 특히, 아이템 속성은 아이템 설명보다 더 풍부한 맥락 정보를 내포하고 있어 이를 활용하면 사용자 쿼리와 더욱 밀접하게 연관된 아이템을 찾아낼 수 있다. 이를 통해 추천 결과가 사용자의 쿼리 의도에 더욱 부합하게 될 수 있으며, 사용자 경험을 개선하고 개인화 추천을 정교하게 실현하는 데에도 기여할 수 있다. 결국 속성 필터링은 맥락에 따라 변화하는 사용자 요구나 감정 상태를 정량적·범주형 속성 단위로 반영하고 이에 부합하는 아이템을 선별함으로써 다양한 상황에 유연하게 대응하는 추천을 가능하게 한다. 이러한 속성 필터링의 특징은 아이템 설명과 쿼리 간 임베딩 유사성에 의존하는 기존 LLM 및 RAG 기반 IR 시스템 구조가 세부 맥락을 반영하는 데 한계가 있다는 점을 속성 정보를 추가적으로 활용함으로써 효과적으로 보완할 수 있다.

본 연구는 이러한 속성을 LLM 기반 맥락 해석 과정에 결합함으로써 사용자의 감정·상황적 요구 맥락을 속성 수준에서 체계적으로 반영하는 새로운 IR 시스템 프레임워크를 제안한다. 이는 기존 구조들이 단순히 아이템 설명 임베딩의 유사성에 의존하는 것과 달리, 실시간으로 사용자 질의의 의미를 정밀하게 해석하고 맥락에 따라 아이템 속성을 유연하게 제어하는 점에서 차별성을 지닌다. 이러한 프레임워크는 LLM의 문맥 이해 능력을 활용하여 사용자 질의로부터 적절한 속성 조건을 도출하고 이를 검색 필터 조건으로 변환한 뒤 해당 조건에 부합하는 아이템을 벡터 스토어에서 검색하고 사용자 쿼리와의 유사도를 기준으로 최종 추천을 수행하는 방식으로 구성된다. 이때, 적절한 검색 필터를 생성하기 위해 RAG 구조를 활용하여 벡터 스토어에서 사용자 쿼리와 관련성이 높은 아이템을 검색하고 이들을 Few-shot 예시로 삼아 해당 아이템들의 속성 패턴과 사용자 메타데이터를 함께 분석함으로써 맥락에 부합하는 필터

조건을 도출한다. 이러한 구조는 기존 LLM 및 RAG 기반 IR 시스템이 아이템 설명 텍스트에 의존함으로써 발생하는 한계를 보완하고 감정 상태, 취향, 목적 등 복합적인 사용자 요구를 아이템 속성 단위로 세밀하게 반영하여 다양한 도메인에서 개인화된 추천을 제공하는 데 기여한다.

제안하는 IR 시스템 프레임워크가 맥락에 따라 속성을 얼마나 잘 반영하여 추천하는지를 분석하기 위해 본 연구에서는 음악 도메인 데이터셋을 활용한 실험을 수행하였다. 다양한 감정 기반 사용자 질의를 구성하고 제안하는 IR 시스템 프레임워크를 Sparse representation 기반 IR 시스템과 LLM 및 RAG 기반 IR 시스템을 베이스라인으로 선정 후 비교하였다. 각 시스템의 추천 결과에 대해 감정별로 선호되는 속성에 대한 기존 연구들을 바탕으로 추천된 아이템들이 해당 감정 맥락에 적절한 속성 분포를 나타내는지를 분석하였다. 이를 통해 제안하는 IR 시스템 프레임워크가 LLM 기반의 문맥 이해와 속성 기반 필터링을 효과적으로 결합하여 맥락에 따라 적절한 아이템 속성을 반영하고 사용자 선호에 맞춘 개인화된 추천을 제공할 수 있음을 실증적으로 확인한다.

## 2. Related Work

### 2.1 Retrieval-augmented generation in information retrieval

IR 시스템은 방대한 데이터에서 사용자의 질의에 부합하는 관련 정보를 효과적으로 찾아주는 기술로 사용자 쿼리에 포함된 자연어 표현의 의미와 맥락을 해석하고 이에 부합하는 정보를 탐색하는 데 특화되어 있다(Hanani et al., 2001; Zhang et al., 2024). 이러한 특성으로 인해 사용자의 질의에 담긴 의도를 정교하게 반영하여 상황적 맥락에 맞는 아이템을 검색할 수 있다는 점에서 추천 분야에 적극적으로 활용되고 있다(Li et al., 2024; Deldjoo et al., 2024).

IR 시스템 기반 추천에서는 단순히 관련 아이템을 찾아내는 것을 넘어 사용자 질의에 담긴 맥락을 얼마나 정밀하게 반영할 수 있는지가 중요한 주제로 다뤄진다. 이는 추천 결과가 사용자의 요구에 부합하는 정도와도 연결되며 나아가 추천을 받아들이는 과정에서 사용자의 추천에 대한 경험에도 영향을 줄 수 있다. 이러한 이유로 최근 IR 기반 추천 연구들은 사용자의 상황·정서·목적 등 맥락적 요구를 효과적으로 반영하여 사용자 경험을 향상시키는 방향으로 발전해 왔다(Hambarde and Proenca, 2023; Li et al., 2024). 또한 맥락 반영은 개별 사용자의 특성과 선호를 고려하는 개인화와도 밀접하게 연결되며 결과적으로 사용자 경험을 높이기 위해서는 맥락을 정교하게 반영하는 것이 중요하다는 점이 강조되고 있다(Deldjoo et al., 2024).

사용자 요구의 맥락을 반영하고자 하는 IR 기반 추천은 사용자의 쿼리와 아이템 설명을 벡터 형태의 임베딩으로 표현한 후 이들의 유사도를 정량적으로 계산하여 관련성이 높은 아이템을 탐색하고 추천하는 구조로 이루어진다(Deldjoo et al., 2024). 관련성 높은 아이템을 탐색하기 위해서는 사용자 쿼리와 아이템 설명의 비정형 텍스트에 내포된 의미를 정교한 임베딩 형태로 반영하는 것이 중요하며 이를 위해 다양한 임베딩 생성 방법론들이 연구되어 왔다. 전통적으로 TF-IDF, BM25 등과 같은 Sparse representation 기반 임베딩 방식은 전체 문서 집합에 등장하는 단어들을 기준으로 각 문서를 표현하며 각 문서는 해당 단어들의 출현 빈도에 따라 수치가 할당된 고차원 벡터로 변환된다(Hambarde and Proenca, 2023). 그러나 이러한 방식은 단어의 빈도 및 가중치만을 반영하기 때문에 단어 간 의미적 관계나 문장 내 순서를 고려하지 못하고, 다의어나 동의어와 같은 표현적 변이를 적절히 처리하지 못한다는 한계가 있다. 이로 인해 사용자의 복합적인 요구나 상황에 따라 달라지는 세부적인 맥락적 의미를 효과적으로 반영하는 데 제약이 존재한다(Zhu et al., 2023; Zhang et al., 2024). 이러한 한계를 극복하기 위해 단어 간 의미적 유사성과 문맥 정보를 반영할 수 있는 Dense representation 기반의 임베딩 방식이 도입되었다. Dense representation은 단어 또는 문장을 고정된 차원의 연속적인 숫자 값들로 이루어진 벡터로 표현하는 방식으로 의미적으로 유사한 표현들이 유사한 위치에 매핑되도록 학습된다. 즉, 단어 하나하나를 단순히 존재 여부나 빈도로 표현하는 것이 아니라 그 의미와 주변 문맥을 반영한 수치적 패턴으로 정교하게 변환하는 것이다. 이 방식은 단어 순서, 문맥 간 상호작용, 의미 흐름 등을 내재화할 수 있어 Sparse representation 임베딩 방식이 갖는 표현적 한계를 보완하며 보다 정밀한 의미 임베딩을 가능하게 한다(Hambarde and Proenca, 2023; Deldjoo et al., 2024). 이러한 Dense representation 임베딩 방식은 IR 시스템이 사용자 쿼리의 의미를 보다 정확하게 이해하고 복합적인 맥락에 대응할 수 있도록 성능을 향상시키는 데 효과적이다(Incitti et al., 2023; Hambarde and Proenca, 2023). 이처럼 의미 중심의 표현이 강조되면서 최근에는 LLM 및 RAG을 기반으로 한 IR 시스템에서도 Dense representation이 핵심적으로 활용되고 있다(Deldjoo et al., 2024).

LLM은 자연어 질의에 내포된 의미를 깊이 있게 이해하고 문장 전체의 논리적 흐름과 사용자 의도를 정교하게 해석하여 텍스트 생성을 비롯해 Dense representation 기반의 임베딩 생성에 활용된다(Zhu et al., 2023; Xu et al., 2025). 또한, RAG는 인코더 구조의 LLM를 통해 생성된 Dense representation 기반의 쿼리와 아이템 설명 임베딩 간의 유사도를 활용하여 벡터 스토어에서 관련성 높은 정보를 검색하고 이러한 정보와 사용자 메타데이터와 같은 추가 정보를 결합한 후 디코더 구조의 LLM에 입력함으로써 아이템 추천 리스트를 생성한다(Gao et al., 2023; Deldjoo et al., 2024). 이러한 LLM 및 RAG 기반의 IR 시스템 구조는 기존 IR 시스템의 핵심 단계인 Retrieval, Query expansion, Reranking 전반에 걸쳐 적용되어 문장의 의미 흐름과 논리를 반영한 정렬 결과를 제공하는 데 효과적이다(Wang et al., 2025; Zhu et al., 2023). 예를 들어, Tao et al. (2024)는 LLM을 활용하여 쿼리 및 문서의 임베딩을 생성하고 이를 기반으로 한 정보 검색 성능을 다양한 벤치마크에서 실험적으로 비교하였다. 이러한 결과는 LLM으로 생성된 임베딩이 RAG의 Retrieval 단계에 효과적으로 활용될 수 있음을 보여준다. 이와 더불어 Yao et al. (2025)는 LLM을 활용한 Query expansion 기법인 LLM-QE를 제안하여 기존 쿼리를 랭킹 선호도에 부합하도록 확장함으로써 검색 품질을 향상시켰다. 이처럼 LLM을 Query expansion 단계에서 사용하여 사용자의 초기 질의에 담긴 의도나 숨겨진 요구를 문맥적으로 파악하고 이를 보완하는 표현으로 재구성함으로써 IR 시스템이 관련성 높은 검색 결과를 도출하는 데 활용할 수 있음을 보여준다. 또한 Gao et al. (2025)은 LLM을 활용한 Reranking 프레임워크인 LLM4Rerank를 제안하여 프롬프트를 통해 후보 아이템들의 설명을 직접 해석하고 다양한 기준에 따라 Reranking하여 추천 결과를 조정하였다. 이는 LLM이 추천 항목의 다양한 평가 기준에 따라 정보를 종합적으로 판단하고 순위를 재조정할 수 있는 능력을 지니고 있음을 시사한다.

그러나 이러한 방식들은 아이템 설명 임베딩에 의존하는 구조로 구성되어 있어 주로 일반적인 텍스트 유사성만이 반영됨에 따라 사용자 요구에 따른 세부적이고 정밀한 아이템을 검색하는 데에는 한계가 있다(Kong et al., 2022; Shan et al., 2023; Lei et al., 2024). 이러한 한계는 각 아이템마다 내재된 미묘한 맥락적 차이를 충분히 반영하지 못하는 문제를 유발할 수 있다. 만약, 사용자가 "슬플 때 듣기 좋은 노래"를 추천받고자 할 경우 실제로는 정서적 맥락이 상이한 여러 곡이 존재할 수 있다. 예를 들어, 일부 곡은 잔잔하고 우울한 분위기를 강조하는 반면, 다른 곡은 슬픔을 위로하고 감정을 해소하는 데 초점을 맞춘 다소 희망적인 분위기를 지닐 수 있다. 그러나 이들 곡이 모두 곡 설명에 단순히 "슬플 때 듣기 좋은 노래"라고만 기재되어 있다면 텍스트 임베딩 기반 시스템은 이러한 세부 감정의 차이를 구분하지 못하고 동일한 맥락으로 처리하게 된다. 따라서 아이템 설명 텍스트의 품질에 따라 추천 결과의 품질 역시 달라질 수 있다. 이와 같이 추천 결과가 맥락적 차이를 반영하지 못할 경우 사용자는 결과를 부적절하다고 느낄 수 있으며 이는 곧 시스템 전반의 사용 경험 저하로 이어진다. 더 나아가 사용자별 요구에 대한 정서적 맥락을 정밀하게 반영하지 못해 개인화 추천 경험의 저하까지 연결될 수 있다. Aviss (2024)는 RAG 구조가 텍스트 임베딩 정보에 크게 의존함에 따라 검색된 아이템의 특성이 최종 응답에 영향을 줄 수 있다는 점을 지적하였다. 결과적으로, LLM이 사용자 쿼리의 맥락을 정교하게 해석하더라도 RAG에서 아이템 검색 단계의 구조적 한계로 인해 사용자의 감정 상태나 의도에 정확히 부합하는 추천을 제공하는 데에는 제약이 따를 수 있다.

## 2.2 Item attribute-based filtering for recommendation

아이템 속성은 아이템 간의 내재된 맥락적 차이를 명확하게 드러내어 각 아이템이 지닌 고유한 특성과 그에 따른 속성값의 차이를 효과적으로 표현할 수 있다(Chen et al., 2020). 이러한 이유로, 사용자는 아이템을 인식하고 선호를 판단하는 과정에서 속성 정보를 중요한 기준으로 활용한다(Zhai, 2024; Mehrdad et al., 2024). 속성 정보는 역할에 따라 정량적 또는 범주형으로 구성되며, 추천이 적용되는 도메인의 특성과 목적에 따라 그 유형과 활용 방식이 달라질 수 있다. 예를 들어, 음악 도메인에서는 추천의 목적에 따라 서로 다른 유형의 속성이 활용된다. 사용자의 감정 상태나 분위기에 부합하는 곡을 선별하고자 할 때는 '춤 적합성(Danceability)', '슬픔 정도(Sadness)', '에너지(Energy)' 등과 같은 정량적 속성이 곡 간의 미묘한 감정적 차이를 구분하는 데 중요하게 활용된다. 반면, 사용자의 장르 취향이나 특정 아티스트 선호와 같이 명시적 선호 기반의 추천을 목표로 할 경우 '장르', '가수', '발매 연도' 등의 범주형 속성이 사용자의 선호 패턴을 반영하는 주요 기준으로 활용된다.

이처럼 추천의 목적에 따라 서로 다른 유형의 속성이 다양한 방식으로 활용되면서 속성 정보를 기반으로 아이템을 탐색하고 사용자 선호에 부합하는 후보를 선별하려는 접근이 주목받고 있다. 이에 따라 아이템을 속성 수준에서 분석하고 사용자 요구에 맞는 속성 분포를 기준으로 필터링하여 보다 정밀한 추천을 제공하려는 방식이 제안되고 있다. Steck (2018)은 영화 추천 도메인에서 사용자 기대에 부합하는 결과를 도출하기 위해 사용자의 과거 시청 이력을 바탕으로 장르 속성 분포를 고려하고 추천 리스트가 이와 유사한 장르 속성 분포를 따르도록 구성하였다. Chen et al. (2013)은 음악 추천에서 오디오 특성이나 무드 태그 등 다양한 속성을 필터링 기준으로 활용하여 감정 문맥에 적합한 음악을 제공하고자 하였으며, Chen et al. (2021) 또한 속성 단위 필터링이 세부적인 맥락을 정량

적으로 반영하는 데 효과적임을 강조하였다. 이처럼 속성 기반 필터링은 사용자의 취향이나 상황을 세밀하게 반영할 수 있다는 점에서 개인화 추천을 위한 효과적인 수단으로 활용될 수 있다(Salehi et al., 2014).

이러한 속성 기반 필터링은 LLM 및 RAG 기반 IR 시스템이 지닌 구조적 한계를 보완하는데 활용될 수 있다. RAG 구조는 주로 아이템 설명 임베딩에 의존하여 정보를 검색하기 때문에 사용자 질의에 내재된 미세한 감정 상태나 구체적인 맥락적 요구를 정확히 반영하는 데에는 제약이 존재한다. 반면, 속성 기반 필터링은 각 아이템의 정량적·범주형 특성을 기준으로 검색 조건을 명시적으로 조정함으로써 사용자 질의에 담긴 구체적인 요구를 보다 직접적으로 반영할 수 있다. 이를 통해 사용자 의도와 맥락을 보다 정밀하게 파악하고 보다 세분화된 추천 결과를 제공할 수 있는 보완적 역할이 가능하다. 또한, 사용자의 선호는 맥락에 따라 유동적으로 변화하며 이에 따라 선호되는 아이템 속성도 달라질 수 있다는 점에서 속성 기반 접근의 필요성이 더욱 강조된다(Eskandanian and Mobasher, 2019; Rashed et al., 2022). 예를 들어, Duman et al. (2022)은 음악의 경우 감정 회복, 집중, 분위기 전환 등 감상의 상황적 맥락에 따라 선호되는 정량적 속성(예: Energy, Danceability, Sadness 등)이 달라짐을 보여주었고, Canini et al. (2012)은 영화의 경우 장면이 전달하려는 감정 분위기(예: 따뜻함, 차가움, 정적 또는 역동성 등)와 같은 감정적 맥락에 따라 색감, 조명, 동작과 같은 시청각적 속성의 구성 방식이 달라짐을 보여주었다. 이는 정보 검색 과정에서 사용자 요구마다 내포된 맥락적 의미가 상이하며 이러한 맥락에 대한 정교한 이해를 바탕으로 사용자 선호에 부합하는 속성을 반영한 추천이 필요하다는 점을 보여준다. 이러한 측면에서, 변하는 맥락에 따라 사용자 쿼리의 의미를 유연하게 해석할 수 있는 LLM 및 RAG 기반 IR 시스템의 문맥 이해 능력과 속성 단위에서 세밀한 조정을 통해 맥락을 반영할 수 있는 속성 조건 필터링 방식을 결합하는 접근은 사용자 요구 맥락을 더욱 정밀하게 반영한 추천을 제공할 수 있는 효과적인 대안이 될 수 있다.

## 3. Development of an Information Retrieval System Framework

본 연구에서는 사용자의 쿼리에 내포된 맥락적 요구를 단순 텍스트 유사성 비교 차원을 넘어 속성 단위까지 정밀하게 반영할 수 있는 새로운 프레임워크를 제안한다. 이 프레임워크는 LLM 및 RAG 기반 IR 구조를 활용해 쿼리의 의미를 정교하게 해석하고 이를 바탕으로 상황별 맥락에 적합한 속성 조건을 동적으로 도출한다. 동시에 아이템의 정량적·범주형 속성값을 기준으로 검색 범위를 구조적으로 조정함으로써 맥락적 요구와 속성적 특성을 함께 반영하는 추천이 가능하다. 이러한 특성을 종합하면, 본 프레임워크는 맥락 해석을 통해 사용자의 요구 상황을 체계적으로 반영하는 장점과 속성 기반 제어를 통해 아이템 수준에서 세밀한 타겟팅을 통해 개인화를 실현하는 장점을 동시에 갖추고 있어 기존 시스템과 차별화된 추천 방식을 제공할 수 있다.

이 프레임워크는 두 가지 핵심 구성 요소를 중심으로 설계되었다. 첫째, LLM 및 RAG 기반의 IR 구조를 활용하여 단순 텍스트 유사성을 넘어 문장의 의미 흐름과 사용자 의도를 심층적으로 분석하고 쿼리의 문맥적 의미를 정확히 파악한다. 둘째, 속성 기반 필터링 모듈을 통해 쿼리에서 도출된 의미에 따라 적절한 속성 조건을 자동으로 생성하고 해당 조건에 부합하는 아이템을 필터링한다. 이때 사용되는 필터링 방식은 쿼리 맥락에 따라 실시간으로 속성 조합을 조정할 수 있도록 구성되며 이를 통해 사용자의 세부적인 요구나 감정 상태, 목적 등에 따라 변화하는 속성 중요도를 효과적으로 반영할 수 있다.

이러한 목적을 실현하기 위해 제안하는 IR 시스템 프레임워크는 전체 프로세스를 세 가지 모듈로 구성하여 설계하였다(Figure 1 참조). (1) Contextual query expansion 모듈은 사용자 쿼리, 사용자 메타데이터, 지시 프롬프트(Instruction prompt), 그리고 실시간으로 검색된 Few shot 예시 아이템 정보를 통합하여 상황 맥락을 반영한 LLM 프롬프트를 구성한다. (2) Search filter generation 모듈은 이 프롬프트를 바탕으로 사용자 맥락에 맞는 속성 조건을 도출하고 이를 기반으로 검색 공간을 동적으로 재구성한다. (3) Recommendation 모듈은 이처럼 구성된 후보군 내에서 사용자 쿼리와의 의미적 유사성을 기준으로 아이템을 정렬하고 최종적으로 추천 결과를 생성한다.

구체적으로, 사용자 쿼리가 입력되면 시스템은 먼저 해당 쿼리의 맥락을 파악하고 그에 적합한 속성 기반 필터를 생성한다. 이후 이 필터를 통해 검색 공간을 재정의하고 그 내부에서 사용자 쿼리와의 임베딩 유사도를 기준으로 아이템을 정렬하여 최종 추천 결과를 도출한다. 결과적으로, 추천 시스템은 사용자 쿼리의 의미를 해석하고 그 의미와 일치하는 속성 분포를 가진 아이템들 중에서 가장 관련성 높은 항목을 제공하는 구조로 작동한다.

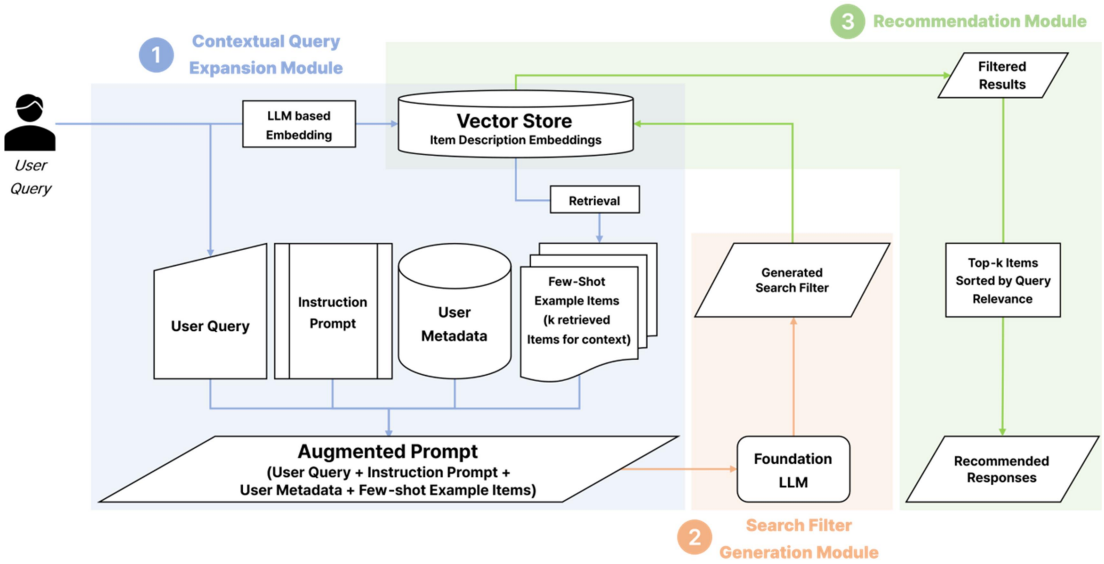


**Figure 1.** Overview of the proposed IR system framework

제안하는 IR 시스템 프레임워크는 OpenAI의 'text-embedding-3-large' 모델 API을 활용해 사용자 쿼리와 아이템 설명을 고차원 임베딩 공간으로 변환하고 이를 'Pinecone' 벡터 스토어에 저장 및 관리하여 속성 정보를 포함한 아이템 검색이 가능하도록 설계하였다. 또한 'gpt-4o' 모델을 사용하여 검색 필터를 자동으로 생성하도록 하였다. 전체 RAG 기반 시스템의 모듈 연계와 워크플로우 구현에는 LangChain 프레임워크를 적용하여 쿼리 처리, 벡터 검색, 필터링, 추천 결과 생성을 통합적으로 구성하였다.

## 3.1 Contextual query expansion module

Contextual Query Expansion Module은 사용자 쿼리에 담긴 의미적 맥락을 정확히 파악하고 이후 검색 필터 생성을 위한 입력 정보를 구조화하는 역할을 수행한다. 본 모듈의 핵심 목적은 사용자가 자연어로 입력한 쿼리를 기반으로 해당 요구 상황에 적합한 속성 조건을 추출할 수 있도록 프롬프트 구성 요소를 체계화하는 것이다.

사용자는 일반적으로 "슬플 때 듣기 좋은 노래를 추천해줘"와 같이 직관적인 자연어 쿼리를 입력한다. 이 쿼리는 LLM 기반 임베딩 모델을 통해 벡터로 변환되며 Vector Store 내 아이템 설명 임베딩들과의 코사인 유사도 기반 비교를 통해 상위 k개의 관련 아이템이 검색된다. 이렇게 검색된 아이템은 단순한 예시가 아닌 해당 쿼리의 의미를 반영하는 실제 사례로서 활용되며 이들을 Few-shot 예시 아이템이라 정의한다. 이 Few-shot 예시 아이템들은 자연어 기반 아이템 설명 텍스트와 정량적·범주형 속성값 정보를 모두 포함하고 있다. 전자는 문맥 해석을 위한 서술 기반 정보를 제공하고 후자는 필터 생성 시 기준이 되는 속성값 패턴 분석에 활용된다.

특히, 이처럼 쿼리와 의미적으로 유사한 예시 아이템들의 속성 분포를 분석함으로써 현재 사용자의 요구 상황과 유사한 맥락에서 주로 등장하는 속성의 경향성을 추출할 수 있다. 이는 단순히 사용자 쿼리에 대응하는 단어 수준의 매칭을 넘어 해당 맥락에서 의미 있는 속성 조합을 도출하는 데 중요한 단서를 제공한다. 예컨대, 슬픔이라는 정서적 맥락과 관련된 예시 곡들에서 공통적으로 높은 Sadness, 높은 Relaxed, 낮은 수준의 Danceability가 나타났다면 이러한 속성 경향이 필터링 조건으로 반영될 수 있다. 이러한 속성 기반 분석은 앞서 검색된 예시 아이템들이 실제 사용자 쿼리 맥락을 반영하는 사례라는 점에서 필터 조건 생성 시 해당 맥락에서 의미 있는 속성 조합을 유도할 수 있도록 돕는다. 결과적으로, LLM이 단순한 키워드 중심 응답이 아니라 맥락에 기반한 구조적인 속성 조건을 도출할 수 있는 기반을 마련해준다.

앞서 추출된 쿼리 맥락 기반의 속성 패턴 외에도 사용자 개인의 일반적인 선호가 함께 반영될 수 있도록 사용자 메타데이터(User

metadata)가 입력된다. 이러한 접근은 사용자 고유의 특성을 반영한 개인화된 필터 조건을 생성하는 데 유용하다(Ghorab et al., 2013). 예를 들어, 음악 도메인에서는 사용자가 자주 선호하는 장르, 최근 감상한 곡 등의 정보가 사용자 메타데이터에 포함될 수 있으며 이는 필터 생성 시 개인화를 위한 참고 요소로 작용한다.

앞선 정보들을 활용하여 검색 필터를 생성하도록 LLM에게 명시적으로 지시하기 위해 Instruction prompt가 함께 제공된다. 이 Instruction prompt는 단순한 입력값의 나열이 아니라 LLM이 어떤 방식으로 정보를 해석하고 처리할지에 대한 규칙과 방향성을 부여하는 역할을 한다. 구체적으로는, Few-shot 예시 아이템에서 공통적으로 나타나는 정량적·범주형 속성들의 경향을 분석하고, 이들을 기반으로 의미 있는 필터 조건을 생성할 것을 요구한다. 또한, 사용자 메타데이터를 고려해 사용자 선호에 기반한 조건이 함께 반영되도록 지시함으로써 생성되는 필터가 개인화된 특성을 갖추도록 돕는다.

앞서 구성된 각각의 정보들(사용자 쿼리, 사용자 메타데이터, Few-shot 예시 아이템, Instruction prompt)은 하나의 증강 프롬프트(Augmented prompt)로 통합된다(Figure 2 참조). Augmented prompt는 LLM이 사용자 쿼리의 의미적 맥락을 해석하고 해당 맥락과 속성 정보를 반영한 검색 필터를 생성할 수 있도록 필수적인 정보 구조를 제공한다. 결과적으로 본 모듈은 다음 단계인 검색 필터 생성 모듈에서 LLM이 보다 정밀하고 의미 있는 속성 조건을 생성할 수 있도록 기초를 제공한다.

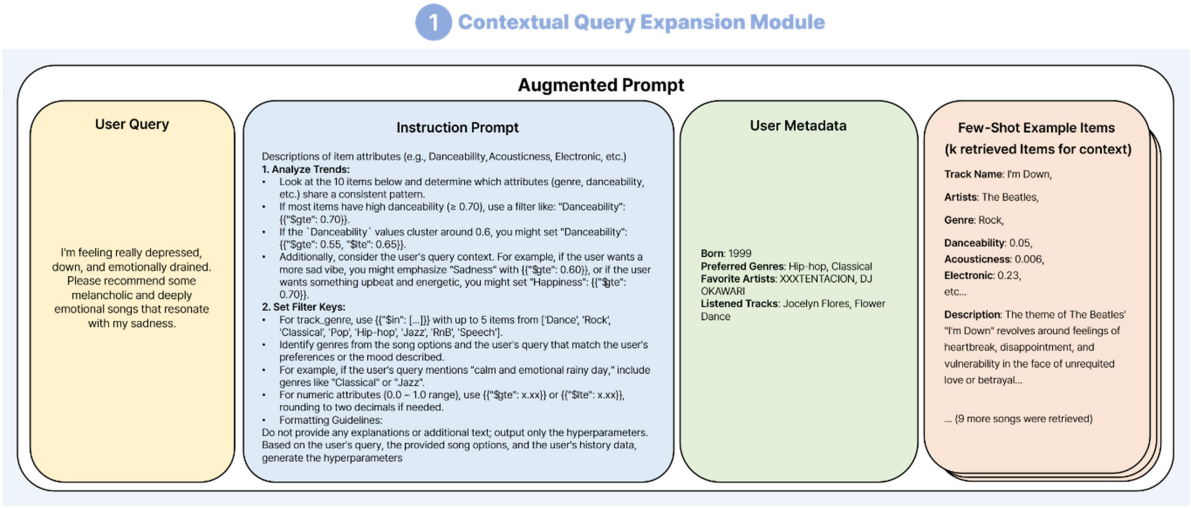


**Figure 2.** Example of an augmented prompt constructed for the music domain in the contextual query expansion module

## 3.2 Search filter generation module

Search Filter Generation Module은 사용자의 요구를 반영하여 검색 대상이 될 아이템의 속성 조건을 구체화하고 전체 아이템 중 의미 있는 후보군을 선별할 수 있도록 속성 기반의 검색 필터를 생성하는 단계이다. 이 모듈의 핵심 목적은 단순히 관련성 높은 아이템을 찾는 데 그치지 않고 사용자의 쿼리 맥락과 연관된 정량적·범주형 속성 조건을 명시적으로 도출함으로써 더 정밀한 추천이 가능하도록 검색 공간을 재정의하는 것이다.

이 과정에서는 앞 단계에서 구성된 Augmented prompt가 입력으로 사용된다. 이 프롬프트는 사용자 쿼리, 메타데이터, 그리고 의미 기반으로 선택된 예시 아이템의 설명과 속성 정보를 포함하여 LLM이 필터 조건을 유추하는 데 필요한 맥락 정보를 제공한다. LLM은 단순한 자연어 처리에 그치지 않고 정형 구조 데이터 내 변수 간의 경향성과 패턴을 효과적으로 요약 및 추론하는 능력을 갖추고 있으며(Fang et al., 2024) 이와 같은 특징은 정량적 조건 기반의 필터 생성에 적합하게 활용한다. LLM은 이러한 입력을 바탕으로 예시 아이템들의 공통 속성 경향을 분석하고 이를 속성 조건의 형태로 필터를 구성한다. 예를 들어, Figure 3에 나타난 바와 같이

Sadness, Relaxed, Danceability 등의 속성이 특정 방향(예: Sadness ≥ 0.8)으로 집중되는 경우 LLM은 해당 조건들을 조합하여 '사용자 요구에 부합하는 아이템이 갖춰야 할 속성 조건'을 정의한다. 이 조건은 수치 기반(≥, ≤)의 범위 설정과 Genre와 같은 범주형 속성의 집합 조건으로 구성된다.

이와 같은 구조는 사용자의 자연어 쿼리에 따라 속성 조건이 유동적으로 결정된다는 점에서 개인화 및 맥락 반영 측면에서 중요한 차별성을 지닌다. 검색 필터는 사용자 요구에 따라 검색 공간을 유연하게 조정하며 이로써 사용자의 요구에 더욱 정밀하게 부합하는 속성 단위의 추천이 가능해진다.

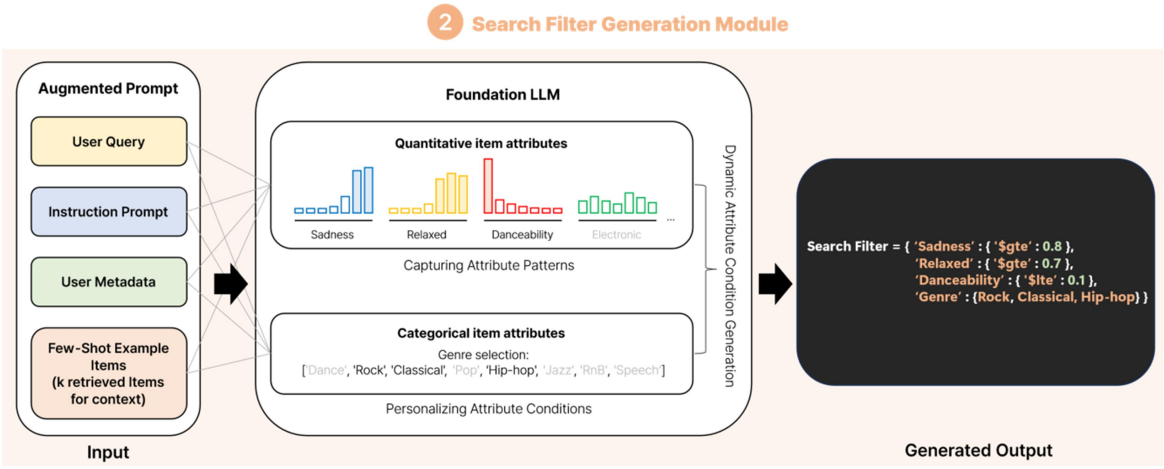


**Figure 3.** Example of search filter generation from an augmented prompt in the search filter generation module

## 3.3 Recommendation module

Recommendation Module은 앞선 단계에서 생성된 검색 필터를 기반으로 사용자 요구에 부합하는 아이템을 최종적으로 선별하는 역할을 수행한다. 이 모듈의 핵심 목적은 필터를 통해 속성 조건에 맞는 아이템을 먼저 추린 뒤 이 후보군 내에서 사용자 쿼리와의 의미적 유사도를 기준으로 재정렬하여 보다 정확한 추천을 제공하는 데 있다.

우선, Search Filter Generation Module에서 생성된 속성 필터는 Vector store에 적용되어 전체 아이템 중 필터 조건에 부합하는 하위 집합 아이템만을 선별한다. 예를 들어 Figure 4에서는 Sadness ≥ 0.8, Relaxed ≥ 0.7, Danceability ≤ 0.1 등의 조건을 만족하는 아이템들만을 검색 공간으로 재정의한다. 이 과정은 사용자의 요구 맥락과 일치하는 후보군을 사전에 선별하기 위한 필터링 단계이다.

그 다음 단계에서는 축소된 후보군 내에서 사용자 쿼리와 각 아이템 설명 텍스트의 임베딩 간 코사인 유사도를 기반으로 관련성 높은 순으로 정렬이 이루어진다. 이는 Contextual Query Expansion Module에서 생성된 사용자 쿼리 임베딩을 기반으로 하며 의미적 유사성이 높은 아이템일수록 우선적으로 추천 결과에 포함된다. 결과적으로, 최종 Top-k 개의 아이템이 사용자에게 추천되며 이들은 속성 조건과 의미 기반 유사도를 모두 만족하는 형태로 구성된다.

이러한 속성 기반 필터링과 의미 유사도 기반 재정렬을 결합한 구조는 기존 텍스트 임베딩 유사도 기반 IR 시스템들이 단순히 텍스트 정보 관련성만을 기준으로 정렬하여 추천했던 방식과 달리 사용자 쿼리의 상황적 맥락과 세부 속성 조건을 동시에 만족하는 결과를 추천할 수 있도록 한다. 이를 통해 사용자는 명확한 의도를 반영한 상황 맞춤형 추천을 경험 가능하며 단순히 설명 정보만 유사한 아이템이 아니라 실제 요구와 취향에 부합하는 결과를 아이템 속성 수준까지 고려하여 보다 만족스럽게 추천받을 수 있도록 한다.

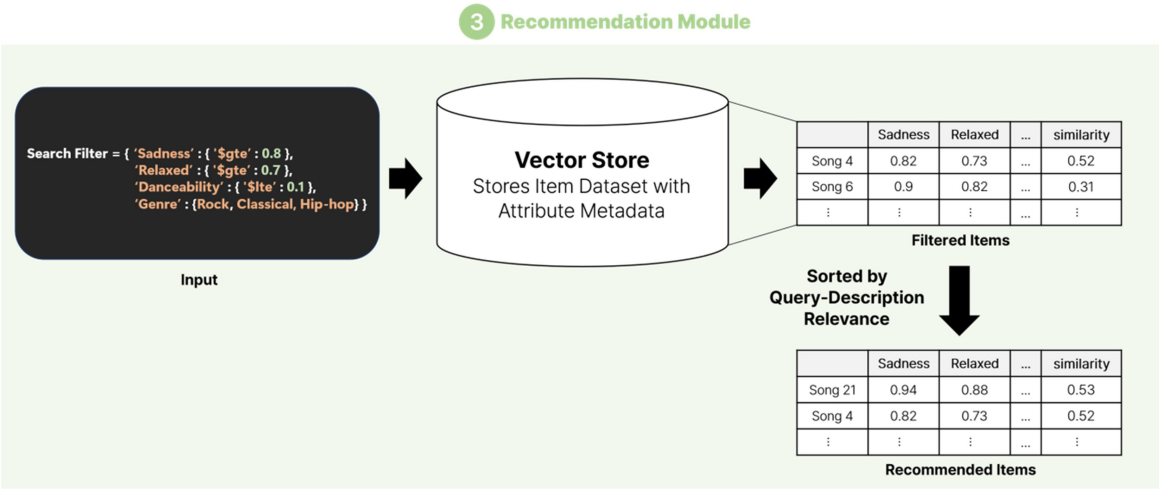


**Figure 4.** Example of recommendation based on attribute-based item filtering in the recommendation module

## 4. Evaluation of Validity of the Proposed Framework

### 4.1 Dataset and preprocessing

본 연구에서는 제안하는 IR 시스템 프레임워크의 실증 분석을 위해 음악 아이템에 대한 정량적 속성값을 포함하는 AcousticBrainz high-level 데이터셋(Porter et al., 2015)을 활용하였다. 해당 데이터셋은 실제로 존재하는 음악에 대해 'Track name', 'Artists', 'Genre'와 같은 기본적인 메타 정보와 'Danceability', 'Acousticness', 'Electronic', 'Aggressive', 'Happiness', 'Party', 'Relaxed', 'Sadness'와 같이 음악적 특성을 정량적으로 표현한 고차원 속성값들을 포함하고 있다(Table 1 참조). 이러한 정량적 변수들은 모두 0과 1 사이의 실수값으로 구성되어 있어 음악의 감정적 또는 리듬적 특성을 수치적으로 비교 분석하는 데 유용하다.

한편, IR 기반 추천 시스템에서는 아이템을 검색 및 추천하기 위해 각 아이템에 대한 설명 변수가 필수적이다. 그러나 AcousticBrainz 데이터셋은 이러한 설명 정보를 포함하고 있지 않으며 모든 곡에 대해 수작업으로 설명을 수집하는 데에는 현실적인 제약이 존재한다. 이를 보완하고자 본 연구에서는 Acharya et al. (2023)이 제안한 LLM 기반의 아이템 설명 생성 방식을 차용하였다. 구체적으로 각 곡의 'Track name', 'Artists', 'Genre' 정보와 외부에서 수집한 'Lyrics' 데이터를 결합하여 LLM에 입력하고 이를 기반으로 곡의 의미적 특징을 요약한 설명 텍스트를 생성하였다. 이와 같이 곡별 설명 텍스트를 생성한 뒤, 전체 데이터셋에 대해 전처리 과정을 수행하였다. 그 과정에서 중복 항목과 결측치를 제거하여 최종적으로 총 20,021개의 음악 데이터를 확보하고 이를 실험에 활용하였다.

**Table 1.** Description of the dataset

| Features | Description |
|---|---|
| Track_name | Song title |
| Artists | Artist name |
| Genre | Song genre (Dance, Rock, Classical, Pop, Hip-hop, Jazz, RnB, Speech) |
| Danceability | Degree of suitability for dancing (0.0 to 1.0) |
| Acousticness | Degree of acoustic characteristics of the sound (0.0 to 1.0) |
| Electronic | Degree of electronic sound (0.0 to 1.0) |

**Table 1.** Description of the dataset (Continued)

| Features | Description |
|---|---|
| Aggressive | Degree of aggressiveness in expression (0.0 to 1.0) |
| Happiness | Degree related to happiness (0.0 to 1.0) |
| Party | Degree of suitability for a party atmosphere (0.0 to 1.0) |
| Relaxed | Degree of relaxation or calmness (0.0 to 1.0) |
| Sadness | Degree of sadness (0.0 to 1.0) |
| Description | Textual description of the song including interpretation of track name, artist name, genre, and lyrics |

제안하는 IR 시스템 프레임워크를 실증적으로 적용하기 위해 본 연구에서는 이와 같은 정량적 음악 데이터셋의 구조에 맞춰 Instruction prompt를 구성하였다. 구체적으로, LLM이 검색 필터를 생성할 수 있도록 각 정량적 변수('Danceability', 'Acousticness', 'Electronic', 'Aggressive', 'Happiness', 'Party', 'Relaxed', 'Sadness')가 의미하는 바를 설명하고 이 값들이 0부터 1 사이의 연속형 수치임을 명시적으로 제공하였다. LLM은 Few-shot 예시 아이템들의 속성값을 기반으로 특정 속성에서 공통된 패턴이 나타날 경우 해당 조건을 포함한 검색 필터를 생성(예: 'Sadness' ≥ 0.7, 'Danceability' ≤ 0.3) 하도록 설계되었다. 아울러, 장르별 추천 요구를 반영하기 위해 'Genre' 변수 또한 필터 조건에 포함되었으며 Dance, Rock, Classical, Pop, Hip-hop, Jazz, RnB, Speech의 8개 장르를 선택지로 하여 사용자의 쿼리, 선호 메타데이터, 10가지 Few-shot 예시 아이템들의 장르 패턴에 따라 해당 장르에 대한 필터 조건이 함께 생성(예: 'Genre' = 'Jazz', 'RnB') 하도록 Instruction prompt 내에서 지시하였다.

## 4.2 Baselines

본 연구에서는 제안한 속성 기반 IR 시스템의 유효성을 평가하기 위해 IR 분야에서 주요하게 활용되는 두 가지 유형의 베이스라인 시스템을 비교 대상으로 설정하였다. 이는 각각 Sparse representation 기반의 전통적인 키워드 중심 IR 시스템과 Dense representation 기반의 LLM 및 RAG 기반 IR 시스템이다. 두 방식은 텍스트의 표현 구조와 정보 탐색 전략 측면에서 근본적으로 상이하며 이를 비교함으로써 본 연구가 제안하는 IR 시스템 프레임워크의 상대적 강점과 기여 지점을 분석하였다.

첫 번째 베이스라인은 대표적인 Sparse representation 기반 모델인 BM25 (Robertson and Walker, 1994)이다. BM25는 단어의 빈도(Term frequency)와 역문서 빈도(Inverse document frequency)를 조합하여 문서와 쿼리 간의 관련도를 산출하는 랭킹 함수로 IR 분야에서 오랜 기간 표준 모델로 자리 잡아왔다(Rosa et al., 2021). 이 방법론은 키워드 빈도 기반 검색으로 단순하면서도 준수한 성능을 보여 최근 연구에서도 여전히 강력한 벤치마크로 널리 사용되고 있다(Svore and Burges, 2009; Rosa et al., 2021; Abolghasemi et al., 2022). 이러한 특징은 다른 IR 접근 방식과의 성능 차이를 명확하게 드러낼 수 있는 비교 기준을 제공하므로 본 연구에서는 전통적인 Sparse representation 모델의 표준 사례로 BM25를 베이스라인에 선정하였다.

두 번째 베이스라인은 최근 주목받고 있는 LLM 기반의 Dense representation으로 구성된 LLM 및 RAG 기반 IR 시스템이다. 이 접근은 LLM의 문맥 이해 능력을 활용하여 쿼리와 문서 간의 의미적 유사도를 기반으로 검색을 수행하는 구조를 따른다. 본 연구에서는 가장 일반적으로 활용되는 LLM 및 RAG 기반 IR 시스템의 구조를 반영하여 베이스라인을 설계하기 위해 Zhu et al. (2023)이 제안한 LLM 및 RAG 기반 IR 시스템의 핵심 구성 요소를 참고하여 Query expansion, Retriever, Re-ranker 모듈을 포함하였다. Query expansion 단계에서는 LLM을 활용하여 사용자의 자연어 쿼리를 맥락적으로 확장하였으며, Retriever 단계에서는 LLM 기반의 Dense representation 임베딩 모델을 사용해 쿼리와 아이템 설명 간의 의미적 유사도를 계산하였다. Re-ranker 단계에서는 디코더 기반 LLM을 활용하여 검색된 상위 아이템들을 의미 기반으로 재정렬하였다. 각 단계에서는 RAG 구조를 통해 벡터 스토어로부터 유사한 아이템 정보를 검색하고 LLM이 이를 입력 받아 최종 추천을 생성할 수 있도록 설계하였다. 이러한 설계는 Sparse representation 기반의 임베딩 모델에 비해 문장 수준의 의미 해석과 맥락 정보를 더 잘 반영할 수 있어 복잡한 사용자 쿼리에서도 상대적으로 높은 검색 성능을 보인다.

두 가지 베이스라인 시스템 모두 제안하는 IR 시스템 프레임워크와 동일하게 OpenAI의 "text-embedding-3-large" 및 "gpt-4o" 모델 API, Pinecone 벡터 스토어, LangChain 프레임워크를 기반으로 구현하였다.

### 4.3 Test process

기존의 IR 시스템은 주로 사용자 쿼리와 문서 간의 표면적인 유사성에 기반하여 검색 및 추천을 수행하기 때문에, 정서적 맥락이나 세부 속성 수준에서의 사용자 요구를 충분히 반영하기에는 한계가 있다는 논의가 이어져 왔다(Mehrdad et al., 2024). 이에 본 연구의 실험은 제안하는 IR 시스템 프레임워크가 실제 환경에서 발생하는 음악 검색 맥락 속 사용자 요구를 얼마나 충실히 반영할 수 있는지, 그리고 각 상황에서 적합한 정량적 속성 조건을 얼마나 효과적으로 도출하고 적용할 수 있는지를 확인하는 데 목적을 두었다. 이러한 목적을 달성하기 위해 실험은 음악 검색 과정에서 실제로 나타날 수 있는 대표적인 상황을 일반화한 네 가지 시나리오를 구성하였으며 또한 실제 음악 속성과 메타데이터를 기반으로 한 데이터셋을 활용하여 현실적인 맥락을 반영하고자 하였다.

실험은 총 세 가지 시스템(제안하는 IR 시스템 프레임워크, BM25, LLM 및 RAG 기반 IR 시스템)을 비교하는 방식으로 진행되었다. 모든 시스템은 각 시나리오에서 동일한 자연어 기반 사용자 쿼리를 입력으로 사용하였으며 4.1절에서 설명한 음악 도메인의 데이터셋을 기반으로 추천 결과를 생성하였다. 또한 실험에 사용된 쿼리는 사용자가 음악을 탐색하는 과정에서 흔히 직면할 수 있는 맥락을 대표하도록 구성되어 정서적 상태에 따라 선호되는 음악 속성 분포의 변화를 관찰할 수 있도록 설계하였다.

감정 기반 쿼리 시나리오는 Russell's Circumplex Model (Russell, 1980)을 이론적 배경으로 삼아 구성되었다. 해당 모델은 Valence(긍정–부정)와 Arousal(흥분–차분)의 두 축을 기반으로 감정 상태를 연속적인 원형 공간에 배치할 수 있도록 하며 각 감정 상태는 특정 음악 속성과 연관될 수 있음을 설명한다. 이에 따라 본 연구는 모델의 각 사분면을 대표할 수 있는 네 가지 감정(Happy, Angry, Sad, Relaxed)을 실험 시나리오로 채택하였다. 각 감정 시나리오에 대해서는 감정을 직관적으로 표현한 자연어 쿼리를 아래와 같이 설계하였다. 이러한 네 가지 감정 시나리오에 대한 쿼리를 세 시스템에 동일하게 입력하였다.

- **Happy :** I'm feeling thrilled and full of energy! Please recommend some upbeat, happy mood songs that amplify my excitement.
- **Angry :** I'm feeling frustrated, tense, and full of stress. Please recommend some aggressive, intense, and high-energy songs that match my anger and amplify my raw emotions.
- **Sad :** I'm feeling really depressed, down, and emotionally drained. Please recommend some melancholic and deeply emotional songs that resonate with my sadness.
- **Relaxed:** I'm feeling light and content, and I want to stay in this peaceful mood. Please recommend some joyful yet calm songs that are both uplifting and relaxing.

기존 실제 사용자를 대상으로 수행된 연구들에서는 각 감정 상태에 따라 청취자가 선호하는 음악 속성이 달라지는 경향이 보고되어 왔다(Rentfrow and Gosling, 2003; Baltazar and Västfjäll, 2020; Duman et al., 2022). 본 연구는 이러한 선행 연구에서 확인된 속성과 감정의 관계를 바탕으로 각 시스템이 감정 시나리오에 대해 속성 수준에서 선행 연구와 얼마나 일치하는 결과를 보여주는지 분석하였다. 즉, 추천된 곡 목록이 감정 맥락에 따라 예상되는 속성 분포가 실제 결과에 반영되는지를 살펴봄으로써 시스템이 맥락을 반영하는 정도를 확인하였다. 이를 위해 네 가지 감정 시나리오 각각에 대해 각 시스템으로부터 총 30개의 추천 결과를 수집하였다.

분석 과정에서는 추천 결과가 감정 맥락을 반영하는 방식을 단계적으로 검토하였다. 먼저, 제안하는 IR 시스템 프레임워크에서 속성 기반 필터링 전후의 결과를 비교하여 필터링이 속성 분포를 어떻게 조정하는지 확인하였다. 이어서 감정 시나리오 간 속성 분포 차이를 관찰하여 서로 다른 맥락적 요구가 실제 추천 패턴에 반영되는지를 분석하였다. 마지막으로 제안 시스템과 두 베이스라인 시스템의 추천 결과를 비교함으로써 각 시스템이 쿼리의 맥락을 어떤 방식으로 처리하고 반영하는지에 대한 차이도 함께 분석하였다.

### 4.4 Results

제안하는 IR 시스템 프레임워크의 내부 변화를 살펴보기 위해 사용자의 쿼리 입력을 바탕으로 Few-shot 예시 아이템을 Retrieval하고,

해당 예시 아이템들의 속성 분포와 이후 추천된 결과 아이템들의 분포를 비교하였다. Retrieval된 10개의 Few-shot 예시 아이템은 모듈 1의 결과에 해당하며 이로부터 생성된 필터를 바탕으로 추천된 30개의 아이템은 모듈 3의 결과에 해당한다. 이러한 과정에서 각 감정 시나리오(Happy, Angry, Sad, Relaxed)별로 필터링 전(Pre-filtering)과 필터링 후(Post-filtering)의 평균값 및 표준편차를 Table 2에 요약하였다.

**Table 2.** Mean and standard deviation of audio features pre- and post-filtering across scenarios

| Audio features | Happy | | | | Angry | | | | Sad | | | | Relaxed | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pre-Filtering (N=10) | | Post-Filtering (N=30) | | Pre-Filtering (N=10) | | Post-Filtering (N=30) | | Pre-Filtering (N=10) | | Post-Filtering (N=30) | | Pre-Filtering (N=10) | | Post-Filtering (N=30) | |
| | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| Danceability | 0.463 | 0.461 | 0.972 | 0.032 | 0.270 | 0.410 | 0.521 | 0.395 | 0.357 | 0.400 | 0.133 | 0.172 | 0.436 | 0.414 | 0.950 | 0.078 |
| Acousticness | 0.155 | 0.294 | 0.030 | 0.042 | 0.044 | 0.045 | 0.001 | 0.001 | 0.285 | 0.316 | 0.809 | 0.247 | 0.503 | 0.374 | 0.105 | 0.123 |
| Electronic | 0.744 | 0.286 | 0.711 | 0.237 | 0.739 | 0.305 | 0.893 | 0.038 | 0.597 | 0.307 | 0.417 | 0.309 | 0.632 | 0.326 | 0.741 | 0.272 |
| Aggressive | 0.342 | 0.345 | 0.288 | 0.276 | 0.487 | 0.493 | 0.987 | 0.011 | 0.181 | 0.236 | 0.077 | 0.190 | 0.104 | 0.219 | 0.361 | 0.465 |
| Happiness | 0.474 | 0.344 | 0.883 | 0.045 | 0.338 | 0.274 | 0.140 | 0.170 | 0.290 | 0.334 | 0.063 | 0.048 | 0.288 | 0.287 | 0.770 | 0.038 |
| Party | 0.556 | 0.370 | 0.878 | 0.067 | 0.369 | 0.397 | 0.848 | 0.110 | 0.211 | 0.222 | 0.049 | 0.079 | 0.189 | 0.245 | 0.161 | 0.269 |
| Relaxed | 0.465 | 0.342 | 0.376 | 0.269 | 0.422 | 0.380 | 0.042 | 0.023 | 0.605 | 0.367 | 0.960 | 0.051 | 0.785 | 0.251 | 0.835 | 0.050 |
| Sadness | 0.174 | 0.126 | 0.168 | 0.093 | 0.073 | 0.092 | 0.076 | 0.074 | 0.350 | 0.235 | 0.809 | 0.095 | 0.462 | 0.238 | 0.326 | 0.138 |

Happy 시나리오에서는 Danceability의 평균이 0.463에서 0.972로 상승하고 표준편차는 0.461에서 0.032로 감소하였다. 이는 높은 Danceability를 가진 곡들로 필터링 되어 결과가 정제되었음을 보여준다. Happiness 속성의 평균은 0.474에서 0.883으로 Party는 0.556에서 0.878로 각각 증가하였고 두 속성의 표준편차도 현저히 감소하여 정서적 일관성을 가진 음악들이 선별된 것을 확인할 수 있다. 추천된 곡들의 주요 장르는 Pop, Dance, Hip-hop으로 나타났다.

Angry 시나리오에서는 Aggressive 속성의 평균이 0.487에서 0.987로 급격히 상승하고 표준편차도 0.493에서 0.011로 감소하였다. Relaxed 속성은 평균 0.422에서 0.042로 낮아졌으며 전반적으로 높은 Arousal과 부정적인 감정을 반영하는 곡들이 선별되었다. 해당 시나리오에서는 주로 Rock 장르가 추천되었다.

Sad 시나리오의 경우 Happiness 속성이 평균 0.290에서 0.063으로 감소하고 Sadness는 평균 0.350에서 0.809로 증가하며 슬픈 감정이 증가되었다. 동시에 Relaxed 속성의 평균도 0.605에서 0.960으로 증가하며 차분하고 감성적인 분위기의 음악으로 추천 결과가 이동함을 확인할 수 있었다. 추천된 음악들은 주로 Jazz, Classical, RnB 장르로 구성되었다.

Relaxed 시나리오에서는 Happiness 속성이 0.288에서 0.770으로 상승하고 Relaxed 속성의 표준편차는 0.251에서 0.050으로 감소하였다. 이는 긍정적이며 안정된 분위기의 음악들이 일관되게 추천되었음을 보여준다. 주요 장르로는 Pop, RnB, Jazz가 포함되었다.

전반적으로 각 시나리오에서 필터링 이후 오디오 속성값의 평균과 표준편차가 사용자의 맥락에 부합하도록 변화한 양상을 확인할 수 있었다. 이러한 결과는 본 연구에서 제안한 검색 기반 추천 시스템이 사용자의 쿼리와 상황적 맥락을 반영하여 의미 있는 속성 기반 필터링을 수행할 수 있음을 보여준다. 이어지는 절에서는 제안 시스템과 기존 베이스라인 시스템들 간의 추천 결과 분포를 비교함으로써 본 시스템이 베이스라인 시스템 대비 어떠한 특성을 가지는지를 분석하고자 한다.

Figure 5의 제안한 시스템의 결과 분포를 살펴보면 각 감정 시나리오 별로 사용자 의도에 적합한 오디오 속성들이 비교적 뚜렷하게

분포되어 있는 양상을 보인다. 예를 들어, Happy 시나리오에서는 Danceability, Happiness, Party와 같은 긍정(High valence) 및 흥분(High arousal)과 관련 속성들이 높은 값에 밀집된 분포를 나타낸다. Sad 시나리오에서는 Sadness와 Relaxed 속성이 모두 높은 값에 분포하여 부정(Low valence) 및 차분한(Low arousal) 상태를 반영한다. Angry 시나리오에서는 Aggressive 속성이 높은 값 영역에서만 형성된다. Relaxed 시나리오에서는 Relaxed 속성이 0.8 이상 값으로 집중된 동시에 Sadness 속성은 전반적으로 낮은 수준을 유지한다.

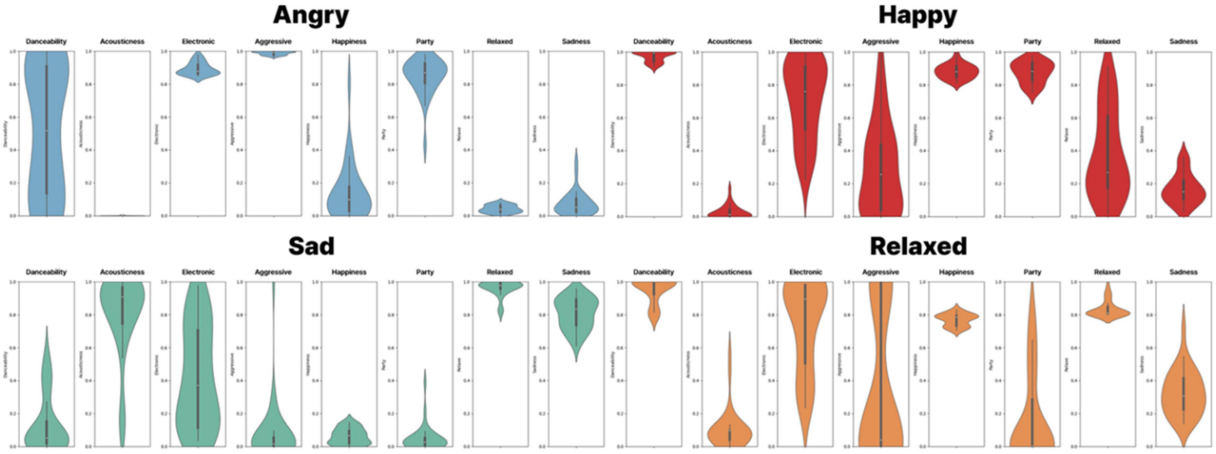


**Figure 5.** Distribution of audio features in the proposed IR system across four scenarios

반면, Figure 6의 BM25 시스템의 경우 전반적으로 분포의 경향성이 뚜렷하지 않거나 시나리오 별 차이가 명확하지 않은 것으로 나타난다. 특히 Happy 시나리오에서는 Danceability가 전반적으로 넓은 범위에 걸쳐 퍼져 있으며 기대되는 긍정(High valence) 속성인 Happiness나 Party 역시 낮은 값에 밀집되어 있어 시나리오의 정서적 특성과의 부합도가 낮은 양상을 보인다. 또한 Angry 시나리오의 경우 흥분(High arousal)과 관련된 핵심 속성인 Aggressive가 특정 영역에 집중되지 않고 넓게 분포되어 있으며 Sad 시나리오의 부정(Low valence) 속성인 Sadness 또한 뚜렷한 경향 없이 고르게 퍼져 있다. Relaxed 시나리오에서도 차분(Low arousal) 상태를 반영해야 할 Relaxed 속성이 고르게 분산되어 있고 동시에 Sadness와 같은 부정적 속성이 낮은 값으로 나타나지 않아 감정 시나리오 별 기대 속성과의 연관성이 낮은 분포 특성을 나타낸다.

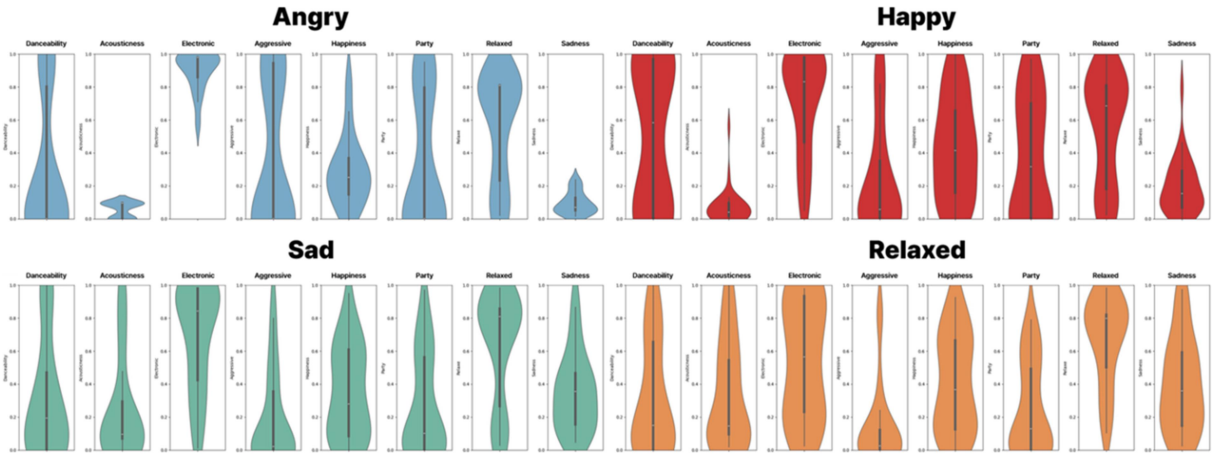


**Figure 6.** Distribution of audio features in the BM25 system across four scenarios

Figure 7의 LLM 및 RAG 기반 IR 시스템 분포에서는 BM25 대비 개선된 경향이 관찰된다. 예를 들어 Happy 시나리오에서는 Danceability가 상대적으로 높은 값에서 형성되며, Relaxed 시나리오에서도 Relaxed 속성의 평균 값이 높아지고 분산 값 또한 감소하는 분포로 나타난다. 다만, Sad 시나리오의 경우 BM25 분포에서 나타난 것과 같이 Sadness 속성이 분산된 양상으로 나타나며 Angry 시나리오에서도 Aggressive 속성이 다소 넓은 범위로 퍼져 있다.

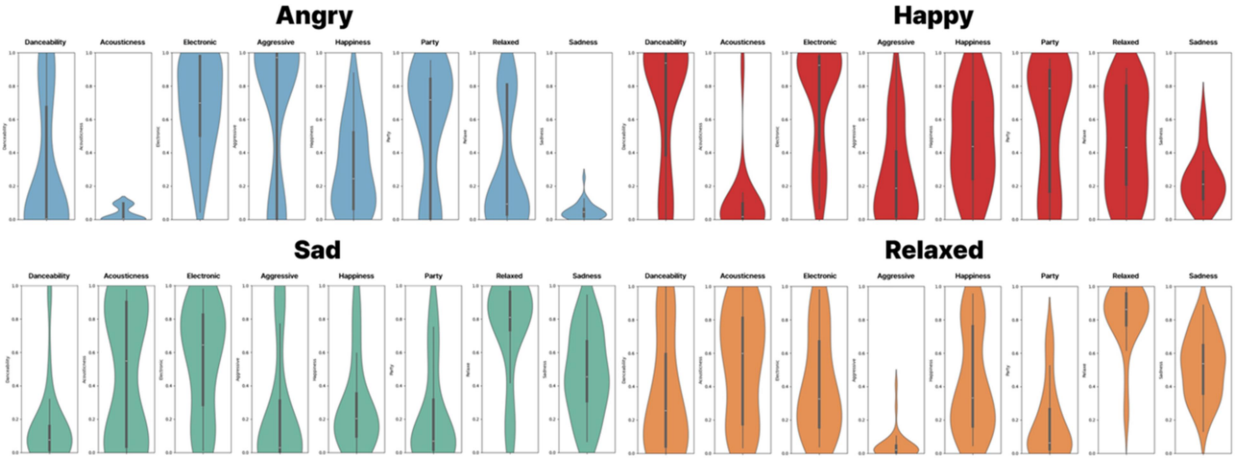


**Figure 7.** Distribution of audio features in the LLM and RAG based IR system across four scenarios

## 5. Discussion

본 연구의 핵심은 제안하는 IR 시스템 프레임워크가 실제 음악 검색 맥락 속에서 사용자의 요구를 얼마나 충실히 반영할 수 있는지, 그리고 각 상황에서 적합한 정량적 속성 조건을 얼마나 효과적으로 도출 및 적용할 수 있는지에 대한 타당성을 검토하는 것이다. 이를 위해 추천 결과의 속성 분포를 중심으로 맥락 반영 정도와 개인화 가능성이라는 두 가지 핵심 관점에서 결과를 논의한다. 특히 대표적인 네 가지 감정 시나리오를 설정하고 각 시나리오 별 추천 결과를 기존 문헌에서 보고된 속성–감정 간 연계성과 비교함으로써 추천 결과가 어떠한 속성적 특징을 보이는지 논의한다. 이러한 두 가지 관점의 논의를 통해 본 연구는 기존 프레임워크와 비교했을 때 제안 모델이 지니는 구조적 차별성과 방법론적 진보성을 탐색하며 동시에 후속 연구의 방향을 제시한다.

첫째, 제안하는 IR 시스템 프레임워크는 사용자 쿼리에 담긴 다양한 정서적 맥락을 반영하여 추천 결과의 속성 분포를 유연하게 조정해내는 특징을 보였다. 예를 들어, "Happy" 시나리오에서는 'Danceability', 'Happiness', 'Party'와 같은 속성이 상대적으로 높고 분산은 낮은 양상을 보였으며, 이는 Duman et al. (2022)이 제시한 'Upbeat' 음악 특성과 유사하다. "Angry" 시나리오에서는 'Aggressive', 'Energy' 속성이 높고 'Relaxed'는 낮게 나타나 Rentfrow and Gosling (2003)의 'Intense and Rebellious' 음악 유형과의 속성 일치를 보였다. "Sad" 시나리오에서는 'Sadness', 'Relaxed' 속성이 높고 'Happiness'는 낮게 나타나 Krols et al. (2023)이 제시한 감정-속성 간 회귀 분석 결과와 방향이 유사했다. 마지막으로 "Relaxed" 시나리오에서는 'Relaxed'와 'Happiness'가 높고 'Sadness'는 낮게 나타나 Baltazar and Västfjäll (2020)이 설명한 평온한 음악의 정량적 특성과 연결된다. 이러한 결과들은 사용자 쿼리의 감정 맥락에 따라 실제로 추천된 아이템들의 속성 분포가 어떻게 달라지는지를 보여주는 사례이며 제안하는 IR 시스템 프레임워크가 의미 기반 필터링과 속성 중심 필터링을 결합함으로써 각 시나리오에 부합하는 추천 결과를 생성할 수 있었음을 보여준다. 반면, BM25 기반 시스템은 시나리오 간 속성 분포의 차이가 명확하지 않았으며 LLM 및 RAG 기반 IR 시스템 또한 일부 속성에서는 의미 있는 변화를 보였으나 전반적으로 감정 시나리오 별 분포 차이는 제한적이었다. 이는 기존 시스템들이 단순히 아이템 설명 임베딩과 쿼리 간 유사도에 의존하여 추천 결과를 정렬하는 수준을 넘어, 감정적 맥락에 따라 추천 리스트의 속성 분포 자체를 능동적으로 조정할 수 있다는 점에서 기존 프레임워크 대비 구조적 차별성을 가진다.

둘째, 제안하는 IR 시스템 프레임워크는 사용자 특성과 쿼리 맥락을 함께 반영하여 개인화된 추천을 제공할 수 있는 가능성을 보여주었다. Hagen et al. (1999)은 개인화를 "개인의 선호와 행동에 대한 정보를 바탕으로 콘텐츠와 서비스를 맞춤 제공하는 능력"이라고 정의하였다. 이러한 정의에 따라, 본 연구에서는 사용자 쿼리에 내포된 맥락을 보다 정교하게 반영하기 위해 LLM의 문맥 이해 능력과 아이템의 정량적 속성 정보를 결합하였으며 여기에 더해 사용자의 선호 정보를 담은 메타데이터를 함께 활용하여 개인화 수준을 높이고자 하였다. 이러한 사용자 메타데이터에 선호 장르 정보를 포함한 후 동일한 쿼리 조건에서도 해당 장르가 선택적으로 적절히 반영되어 추천 결과에 영향을 미치는 것을 확인할 수 있었다. 예를 들어 Angry 시나리오에서 사용자의 선호 장르가 Rock과 Classical로 설정되었을 때 최종 추천 결과에서는 Rock 장르가 포함되는 경향이 나타났다. 이는 Rentfrow and Gosling (2003)의 연구에서 보고한 정서 상태별 음악 장르 선호 경향과 유사한 결과로 제안하는 IR 시스템 프레임워크가 사용자의 선호 정보를 맥락에 따라 효과적으로 반영할 수 있음을 보여준다. 나아가, 이 프레임워크는 자연어 쿼리를 통해 정형화된 입력 없이도 다양한 사용자 요구를 유연하게 반영할 수 있어 여러 시나리오에서 개인화된 추천을 효과적으로 제공 가능하다는 점에서 의미가 있다.

앞서 제안하는 IR 시스템 프레임워크가 사용자의 정서적 맥락을 해석하여 추천 리스트의 속성 분포를 능동적으로 제어하고 개인의 선호 정보까지 결합하여 개인화된 결과를 제공할 수 있음을 두 가지 핵심 관점에서 논의하였다. 다만, 본 연구에서는 이러한 추천 결과에 대해 실제 사용자가 만족하는지에 대한 사용자 평가는 수행하지 않았다. 따라서 향후 연구에서는 실제 사용자를 대상으로 한 실험을 진행하여 제안하는 IR 시스템 프레임워크가 통계적으로 유의미한 사용자 경험을 이끌어내는지 검증함으로써 연구의 객관성을 강화할 수 있을 것이다.

## 6. Conclusion

본 연구는 기존의 LLM 및 RAG 기반 IR 시스템이 아이템 설명 임베딩에 의존함으로써 사용자 쿼리에 내포된 세부적인 감정 상태나 속성 수준의 요구를 정밀하게 반영하지 못하는 구조적 한계를 주목하였다. 이러한 문제를 해결하기 위해 본 연구는 LLM 및 RAG 기반의 문맥 해석 능력에 아이템 속성 기반 필터링을 결합한 새로운 IR 프레임워크를 제안하였다. 제안하는 IR 시스템 프레임워크는 사용자 쿼리로부터 LLM이 적절한 속성 조건을 도출하고 RAG 구조를 통해 벡터 스토어에서 관련 아이템들의 속성 패턴을 수집한 뒤 이를 기반으로 필터 조건을 생성하고 해당 조건에 부합하면서 사용자 쿼리와 높은 관련성을 가지는 아이템을 최종적으로 추천하는 방식으로 구성된다.

해당 구조는 사용자 쿼리에 포함된 정서적 맥락이나 목적, 취향 등의 복합적인 요구를 아이템 속성 단위로 반영할 수 있다는 점에서 기존 아이템 설명 임베딩 정보에 의존하는 IR 시스템보다 직관적이고 상황에 맞는 추천 결과를 생성하는 데 기여한다. 실험적으로는 음악 추천 도메인에서 감정 기반 시나리오를 구성하여 추천 결과의 속성 분포가 사용자의 정서적 맥락과 얼마나 일치하는지를 비교·분석하였다. 그 결과, 제안한 시스템은 베이스라인 시스템에 비해 각 감정 시나리오에 따라 주요 속성이 뚜렷하게 구분되는 양상을 나타냈으며 이는 감정 맥락에 따라 적절한 속성이 효과적으로 조정되어 추천 결과에 반영되었음을 보여준다. 또한 사용자 선호 정보를 메타데이터로 함께 활용함으로써 사용자 특성에 따라 추천 결과를 조정할 수 있는 개인화 추천이 가능함을 확인하였다.

향후 연구에서는 실제 사용자를 대상으로 하는 실험을 통해 제안하는 IR 시스템 프레임워크의 효과성을 평가하고 추천 결과가 사용자 경험에 미치는 영향을 다각도로 분석하고자 한다. 이러한 후속 연구를 통해 본 프레임워크가 사용자 중심의 개인화 추천 시스템에 실질적으로 활용될 수 있음을 검증할 수 있을 것이다.

## Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00343599).

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2022-0-00078, Explainable Logical Reasoning for Medical Knowledge Generation).

## References

Abolghasemi, A., Askari, A. and Verberne, S., "On the Interpolation of Contextualized Term-Based Ranking with BM25 for Query-by-Example Retrieval", *Proceedings of the 2022 ACM SIGIR International Conference on Theory of Information Retrieval* (pp. 161-170), 2022, https://doi.org/10.1145/3539813.3545133.

Acharya, A., Singh, B. and Onoe, N., "LLM Based Generation of Item-Description for Recommendation System", *Proceedings of the 17th ACM Conference on Recommender Systems* (pp. 1204-1207), 2023, https://doi.org/10.1145/3604915.3610647.

Aviss, T., "Improving Embedding Accuracy for Document Retrieval Using Entity Relationship Maps and Model-Aware Contrastive Sampling", *arXiv preprint*, arXiv:2410.18105, 2024, https://doi.org/10.48550/arXiv.2410.18105.

Baltazar, M. and Västfjäll, D., "Songs perceived as relaxing: Musical features, lyrics, and contributing mechanisms", in International Conference: Psychology and Music-Interdisciplinary Encounters, Faculty of Music, University of Arts in Belgrade, 2020.

Canini, L., Benini, S. and Leonardi, R., "Affective Recommendation of Movies Based on Selected Connotative Features", *IEEE Transactions on Circuits and Systems for Video Technology*, 23(4), 636-647, 2012, https://doi.org/10.1109/TCSVT.2012.2211935.

Chen, C.M., Tsai, M.F., Liu, J.Y. and Yang, Y.H., "Using Emotional Context from Article for Contextual Music Recommendation", *Proceedings of the 21st ACM International Conference on Multimedia* (pp. 649-652), 2013, https://doi.org/10.1145/2502081.2502170.

Chen, H., Qian, F., Chen, J., Zhao, S. and Zhang, Y., "Attribute-based Neural Collaborative Filtering", *Expert Systems with Applications*, 185, 115539, 2021, https://doi.org/10.1016/j.eswa.2021.115539.

Chen, W.H., Hsu, C.C., Lai, Y.A., Liu, V., Yeh, M.Y. and Lin, S.D., "Attribute-Aware Recommender System Based on Collaborative Filtering: Survey and Classification", *Frontiers in Big Data*, 2, 49, 2020, https://doi.org/10.3389/fdata.2019.00049.

Deldjoo, Y., He, Z., McAuley, J., Korikov, A., Sanner, S., Ramisa, A., Vidal, R., Sathiamoorthy, M., Kasrizadeh, A., Milano, S. and Ricci, F., "Recommendation with Generative Models", *arXiv preprint*, arXiv:2409.15173, 2024, https://doi.org/10.48550/arXiv.2409.15173.

Duman, D., Neto, P., Mavrolampados, A., Toiviainen, P. and Luck, G., "Music We Move To: Spotify Audio Features and Reasons for Listening", *PLOS ONE*, 17(9), e0275228, 2022, https://doi.org/10.1371/journal.pone.0275228.

Eskandanian, F. and Mobasher, B., "Modeling the Dynamics of User Preferences for Sequence-Aware Recommendation Using Hidden Markov Models", *The Thirty-Second International Florida Artificial Intelligence Research Society Conference* (pp. 425-430), 2019.

Fang, X., Xu, W., Tan, F.A., Zhang, J., Hu, Z., Qi, Y., Nickleach, S., Socolinsky, D., Sengamedu, S. and Faloutsos, C., "Large Language Models (LLMs) on Tabular Data: Prediction, Generation, and Understanding—A Survey", *arXiv preprint*, arXiv:2402.17944, 2024, https://doi.org/10.48550/arXiv.2402.17944.

Gao, J., Chen, B., Zhao, X., Liu, W., Li, X., Wang, Y., Wang, W., Guo, H. and Tang, R., "LLM4Rerank: LLM-based Auto-Reranking Framework for Recommendations", *Proceedings of the ACM on Web Conference 2025* (pp. 228-239), 2025, https://doi.org/10.1145/3696410.3714922.

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J. and Wang, H., "Retrieval-Augmented Generation for Large Language Models: A Survey", *arXiv preprint*, arXiv:2312.10997, 2023, https://doi.org/10.48550/arXiv.2312.10997.

Ghorab, M.R., Zhou, D., O'connor, A. and Wade, V., "Personalised Information Retrieval: Survey and Classification", *User Modeling and User-Adapted Interaction*, 23(4), 381-443, 2013, https://doi.org/10.1007/s11257-012-9124-1.

Hagen, P., Manning, H. and Souza, R., "Smart personalization", Forrester Research, Cambridge, MA, 1999.

Hambarde, K.A. and Proenca, H., "Information Retrieval: Recent Advances and Beyond", *IEEE Access*, 11, 76581-76604, 2023, https://doi.org/10.1109/ACCESS.2023.3295776.

Hanani, U., Shapira, B. and Shoval, P., "Information Filtering: Overview of Issues, Research and Systems", *User Modeling and User-Adapted Interaction*, 11(3), 203-259, 2001.

Incitti, F., Urli, F. and Snidaro, L., "Beyond Word Embeddings: A Survey", *Information Fusion*, 89, 418-436, 2023, https://doi.org/10.1016/j.inffus.2022.08.024.

Kong, W., Khadanga, S., Li, C., Gupta, S.K., Zhang, M., Xu, W. and Bendersky, M., "Multi-Aspect Dense Retrieval", *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3178-3186), 2022, https://doi.org/10.1145/3534678.3539137.

Krols, T., Nikolova, Y. and Oldenburg, N., "Multi-Modality in Music: Predicting Emotion in Music from High-Level Audio Features and Lyrics", *arXiv preprint*, arXiv:2302.13321, 2023, https://doi.org/10.48550/arXiv.2302.13321.

Lei, Y., Lian, J., Yao, J., Wu, M., Lian, D. and Xie, X., "Aligning Language Models for Versatile Text-based Item Retrieval", *Companion Proceedings of the ACM Web Conference 2024* (pp. 935-938), 2024, https://doi.org/10.1145/3589335.3651468.

Li, Y., Lin, X., Wang, W., Feng, F., Pang, L., Li, W., Nie, L., He, X. and Chua, T.S., "A Survey of Generative Search and Recommendation in the Era of Large Language Models", *arXiv preprint*, arXiv:2404.16924, 2024, https://doi.org/10.48550/arXiv.2404.16924.

Mehrdad, N., Rathi, V. and Rajanala, S., "Session Context Embedding for Intent Understanding in Product Search", *arXiv preprint*, arXiv:2406.01702, 2024, https://doi.org/10.48550/arXiv.2406.01702.

Porter, A., Bogdanov, D., Kaye, R., Tsukanov, R. and Serra, X., "ACOUSTICBRAINZ: A Community Platform for Gathering Music Information Obtained from Audio", *16th International Society for Music Information Retrieval Conference*, 2015.

Rashed, A., Elsayed, S. and Schmidt-Thieme, L., "Context and Attribute-Aware Sequential Recommendation via Cross-Attention", *Proceedings of the 16th ACM Conference on Recommender Systems* (pp. 71-80), 2022, https://doi.org/10.1145/3523227.3546777.

Rentfrow, P.J. and Gosling, S.D., "The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences", *Journal of Personality and Social Psychology*, 84(6), 1236, 2003, https://doi.org/10.1037/0022-3514.84.6.1236.

Robertson, S.E. and Walker, S., "Some Simple Effective Approximations to the 2-Poisson Model for Probabilistic Weighted Retrieval", *SIGIR '94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, Organised by Dublin City University* (pp. 232-241), 1994, https://doi.org/10.1007/978-1-4471-2099-5_24.

Rosa, G.M., Rodrigues, R.C., Lotufo, R. and Nogueira, R., "Yes, BM25 is a Strong Baseline for Legal Case Retrieval", *arXiv preprint*, arXiv:2105.05686, 2021, https://doi.org/10.48550/arXiv.2105.05686.

Russell, J.A., "A circumplex model of affect", *Journal of Personality and Social Psychology*, 39(6), 1161, 1980.

Salehi, M., Nakhai Kamalabadi, I. and Ghaznavi Ghoushchi, M.B., "Personalized Recommendation of Learning Material Using Sequential Pattern Mining and Attribute Based Collaborative Filtering", *Education and Information Technologies*, 19(4), 713-735, 2014, https://doi.org/10.1007/s10639-012-9245-5.

Shan, H., Zhang, Q., Liu, Z., Zhang, G. and Li, C., "Beyond Two-Tower: Attribute Guided Representation Learning for Candidate Retrieval", *Proceedings of the ACM Web Conference 2023* (pp. 3173-3181), 2023, https://doi.org/10.1145/3543507.3583254.

Steck, H., "Calibrated Recommendations", *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 154-162), 2018, https://doi.org/10.1145/3240323.3240372.

Svore, K.M. and Burges, C.J.C., "A Machine Learning Approach for Improved BM25 Retrieval", *Proceedings of the 18th ACM Conference on Information and Knowledge Management* (pp. 1811-1814), 2009, https://doi.org/10.1145/1645953.1646237.

Tao, C., Shen, T., Gao, S., Zhang, J., Li, Z., Tao, Z. and Ma, S., "LLMs are Also Effective Embedding Models: An In-depth Overview", *arXiv preprint*, arXiv:2412.12591, 2024, https://doi.org/10.48550/arXiv.2412.12591.

Wang, S., Fan, W., Feng, Y., Lin, S., Ma, X., Wang, S. and Yin, D., "Knowledge Graph Retrieval-Augmented Generation for LLM-based Recommendation", *arXiv preprint*, arXiv:2501.02226, 2025, https://doi.org/10.48550/arXiv.2501.02226.

Xu, Z., Mo, F., Huang, Z., Zhang, C., Yu, P., Wang, B., Lin, J. and Srikumar, V., "A Survey of Model Architectures in Information Retrieval", *arXiv preprint*, arXiv:2502.14822, 2025, https://doi.org/10.48550/arXiv.2502.14822.

Yao, S., Huang, P., Liu, Z., Gu, Y., Yan, Y., Yu, S. and Yu, G., "LLM-QE: Improving Query Expansion by Aligning Large Language Models with Ranking Preferences", *arXiv preprint*, arXiv:2502.17057, 2025, https://doi.org/10.48550/arXiv.2502.17057.

Zhai, C., "Large Language Models and Future of Information Retrieval: Opportunities and Challenges", *Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 481-490), 2024, https://doi.org/10.1145/3626772.3657848.

Zhang, W., Liao, J., Li, N., Du, K. and Lin, J., "Agentic Information Retrieval", *arXiv preprint*, arXiv:2410.09713, 2024, https://doi.org/10.48550/arXiv.2410.09713.

Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., Chen, H., Liu, Z., Dou, Z. and Wen, J.R., "Large Language Models for Information Retrieval: A Survey", *arXiv preprint*, arXiv:2308.07107, 2023, https://doi.org/10.48550/arXiv.2308.07107.

## Author listings

**Yunseong Kim:** yunseongkim@korea.ac.kr

**Highest degree:** B.S. in Industrial and Management Engineering, Hanyang University
**Position title:** Graduate Student, Department of Industrial and Management Engineering, Korea University
**Areas of interest:** Information Retrieval System, LLM Agent, Recommender system, Human-AI Interaction

**Gyewon Jeon:** gwjeon@korea.ac.kr
**Highest degree:** B.S. in Industrial Engineering, Seoul National University of Science and Technology
**Position title:** Graduate Student, Department of Industrial and Management Engineering, Korea University
**Areas of interest:** Interactive Machine Learning, Serendipitous Recommendation System, eXplainable AI, Graph Analytics, Anomaly Detection

**Taehyun Ha:** taehyunha@sejong.ac.kr
**Highest degree:** Ph.D. in Interaction Science, Sungkyunkwan University
**Position title:** Assistant Professor, Department of Artificial Intelligence and Data Science, Sejong University
**Areas of interest:** Human-Computer Interaction, User Experience, Agentic AI

**Sangwon Lee:** upcircle@korea.ac.kr
**Highest degree:** Ph.D in Industrial Engineering, Pennsylvania State University
**Position title:** Professor, Department of Industrial and Management Engineering, Korea University
**Areas of interest:** User Experience, Human-Computer Interaction, Explainable Artificial Intelligence, Affective Computing, User Modeling