

이용자 프로파일링 기반 학술정보 서비스 프레임워크 설계 및 타당성 검증*

Design and Validity Verification of User Profiling-based Academic Information Service Framework

이석형 (Seok-Hyoung Lee)**

정서영 (Seo-Young Jeong)***

초 록

본 연구는 연구자의 연구 맥락을 반영한 학술정보 서비스 고도화를 위해 연구 단계와 행동 수준을 핵심 기준으로 하는 이용자 프로파일링 프레임워크를 제안한다. 기존 학술정보 서비스는 정적 프로필과 단순 행동 데이터에 의존하여 연구자의 실제 정보 요구와 연구 과정 단계를 반영하기 어렵다는 한계가 있다. 이를 보완하기 위해 본 프레임워크는 정적 정보, 동적 탐색 정보, 의미 기반 속성, 연구 단계, 행동 수준을 통합한 다차원 프로파일 구조를 설계하였다. 또한 임베딩·지식그래프·토픽 모델링 기반의 의미 분석 계층을 통해 논문·특허·보고서 등 이기종 학술정보를 의미적으로 연결하고, 연구 단계와 행동 수준에 따라 차별화된 추천을 제공하도록 구성하였다. 타당성 검증을 위해 문헌정보학, 학술정보 플랫폼, AI 개발 전문가로 구성된 3인의 패널 평가를 실시한 결과, 프레임워크는 이론적 타당성, 구조적 정합성, 실무적 유용성에서 높은 평가를 받았다. 다만 연구 단계 및 행동 수준 판별을 위한 레이블 데이터 확보의 어려움이 주요 고려사항으로 제기되었으며, 이를 위해 규칙 기반 초기 레이블링과 능동 학습 전략을 제안하였다. 본 연구는 연구자 중심의 개인화된 학술정보 서비스의 구현 가능성을 제시하며, 향후 지능형 학술정보 플랫폼 구축의 핵심 기반으로 활용될 수 있다.

ABSTRACT

This study proposes a user-profiling framework for scholarly information services that integrates research stages, behavior levels, and semantic analysis to overcome the limits of keyword-based systems. The framework unifies static attributes, dynamic behavioral signals, semantic topic representations, and contextual research states to support adaptive, researcher-centered information services. Five core components structure the model: research stage modeling (SE-1), behavior-level classification (SE-2), embedding- and knowledge-graph-based semantic analysis (SE-3), log-driven dynamic profiling (SE-4), and a modular architecture for interoperability (SE-5). Structural validation confirmed that the framework satisfies ten requirements derived from prior studies, including semantic linkage, quality filtering, adaptive interfaces, and predictive recommendations. Expert evaluation by three specialists yielded an average score of 4.6 out of 5, highlighting strong theoretical grounding and practical usefulness. The main challenge identified was the scarcity of labeled data for inferring contextual user states. To address this, heuristic bootstrapping and active-learning strategies are proposed.

키워드: 이용자 프로파일링, 의미기반 추천, 연구 단계, 행동 수준, 학술정보 서비스

user profiling, semantic recommendation, research stage, behavior level, academic service

* 본 연구는 충남대학교 학술연구비에 의해 지원되었음.

** 충남대학교 사회과학대학 문헌정보학과 부교수(skyi@cnu.ac.kr) (제1저자, 교신저자)

*** 충남대학교 사회과학대학 문헌정보학과 박사과정(eswhy77@gmail.com) (공동저자)

■ 논문접수일자: 2025년 11월 22일 ■ 최초심사일자: 2025년 12월 7일 ■ 게재확정일자: 2025년 12월 11일

■ 정보관리학회지, 42(4), 349-374, 2025. <http://dx.doi.org/10.3743/KOSIM.2025.42.4.349>

※ Copyright © 2025 Korean Society for Information Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

디지털 학술정보 환경은 빠른 속도로 진화하고 있으며, 연구자와 학습자들은 이전보다 훨씬 복잡하고 세분화된 정보 요구를 보이고 있다. 과거의 정보탐색은 주로 키워드를 기반으로 한 단순 검색 중심이었고, 이용자는 검색 결과 중 일부를 선별해 스스로 필요한 정보를 판단하는 방식에 익숙했다. 그러나 최근의 연구 환경에서는 문헌의 양적 팽창, 연구프로세스의 복잡성 증가, 인공지능 기반 지식 처리기술의 확산 등으로 인해 기존의 단순 검색 방식만으로는 충분한 탐색 효율을 확보하기 어려워졌다. 이와 같은 변화는 연구자가 어떤 목적을 가지고 어떤 단계에서 어떤 유형의 정보를 필요로 하는지를 이해하고 이를 반영한 적응형 정보 서비스가 필요하다는 요구로 이어지고 있다(감미아, 이지연, 2023; Jeong et al., 2020).

특히 권나현 외(2012)과 이정연 외(2019)은 과학기술 분야의 R&D 전주기 및 연구 콘텐츠 라이프사이클을 분석하며, 연구 활동이 아이디어 발굴, 실험 및 분석, 성과 확산 등 일련의 단계로 구성되며 각 단계마다 생성되고 소비되는 정보의 특성이 상이함을 실증하였다. 연구를 막 시작하는 단계에서는 문제 영역 정의와 기초 개념 이해에 도움이 되는 배경 자료가 필요하며, 나은엽(2023)이 자연과학 연구자들을 대상으로 분석한 바와 같이 연구계획서 작성을 위한 초기 단계에서는 기존 연구와의 차별성을 확보하고 연구의 타당성을 증명하기 위해 폭넓고 탐색적인 정보추구행동이 두드러진다. 이후 중간 단계에서는 관련 문헌의 심층 분석과 핵심 개념 간 관계 파악이 중요해지며, 보다 고도화된 단

계에서는 연구 아이디어를 구체화하기 위한 트렌드 분석, 연구 간 비교를 위한 데이터 기반 정보 등이 요구된다. 이러한 연구 단계별 정보 요구는 단순히 검색어 수준의 입력만으로는 파악하기 어려우며, 이용자의 행동 패턴, 연구 맥락, 반복적인 정보 접근 양상 등을 종합적으로 분석해야만 정확히 파악할 수 있다.

그러나, 국내외의 주요 학술정보 플랫폼은 이러한 복합적 요구를 충족하기엔 구조적 한계를 갖는다. 예를 들어 Google Scholar나 Semantic Scholar는 방대한 문헌 데이터베이스에 기반한 검색을 제공하고 있으나, 이용자의 연구 수준이나 연구 과정 단계와 같은 맥락적 요인을 반영한 추천이나 개인화 기능은 매우 제한적이다(Gusenbauer & Haddaway, 2020). 최근 등장한 SciSpace, Elicit, Perplexity 등 생성형 AI 기반 서비스들도 주로 입력된 문헌이나 질의를 중심으로 텍스트 요약, 연관 문헌 추천, 질의응답을 제공하고 있는데, 연구 단계 기반 맥락(Context), 행동 수준(Level), 의미정보(Semantic Profile)를 통합적으로 고려한 정교한 프로파일링 기능은 존재하지 않는다(Scherbakov et al., 2025).

대다수 도서관 도서관 시스템은 열람·대출 정보와 기본적인 검색 로그 정도만을 수집하고 있어 의미 기반 이용자 모델링을 수행하기 어렵다. 또한 개인정보 문제로 인해 상세 로그 데이터를 활용하기 어려워 행동 기반 추천 서비스나 고급 연구지원 기능의 구현이 사실상 불가능한 환경에 놓여 있다. 이러한 서비스들은 특정 문헌이나 키워드를 중심으로 의미 기반 요약이나 문서 간 유사도 분석을 수행하는 데 강점이 있지만, 이용자 개인의 연구 목적, 연구

수준, 탐색 행동 패턴을 반영하여 정보 흐름을 조정하는 수준의 서비스로 발전하지는 못하고 있다.

이러한 문제를 극복하기 위해서는 이용자의 정적 정보뿐 아니라 동적 행동 데이터와 의미 기반 정보까지 통합한 새로운 이용자 프로파일링 모델이 필요하다. 연구자는 연구 과정에서 수많은 문헌, 개념, 문제 상황을 접하고 이를 반복적으로 탐색·선택·배제·비교하면서 의미 구조를 형성해 나간다. 따라서 이용자가 반복적으로 조회한 주제, 관심 키워드의 변화, 문헌 간 연속적 열람 관계, 주제 영역 간 이동 패턴 등을 수집하고 이를 의미적으로 재해석할 수 있어야 한다. 또한 이용자의 행동 수준과 연구 단계를 명시적으로 정의함으로써, 프로파일링의 기준을 정형화하고 서비스 엔진이 이를 기반으로 한 지능형 추천을 수행할 수 있도록 만들어야 한다.

본 연구는 이러한 배경을 바탕으로, 연구 단계 및 행동 수준 기반 이용자 프로파일링 프레임워크를 설계하고 그 타당성을 분석하는 것을 목적으로 한다. 이 프레임워크는 정적 프로필, 행동 기반 동적 프로필, 의미 기반 분석결과, 연구 단계별 컨텍스트 정보를 통합적으로 결합하여 이용자 모델을 구성한다. 이를 통해 이용자가 어떤 목적으로 어떤 정보를 찾고 있는지를 이해하고, 연구 과정 단계에 따라 필요한 정보를 선행적으로 제공할 수 있도록 설계되었다. 본 연구는 단순 기술 제안에 머무르지 않고, 요구사항 기반 분석, 아키텍처 설계의 논리적 정합성 검증, 글로벌 AI 기반 플랫폼과의 비교분석, UML 기반 데이터 흐름 검증, 연구 시나리오 기반 기능 검증 등을 통해 프레임워크의 학

술적 타당성을 충분히 확보하고자 한다.

2. 관련 연구

본 장에서는 먼저 이용자 프로파일링의 개념과 연구 동향을 검토하고, 이어서 정보행태와 정보요구 이론을 설명한다. 또한 최근 빠르게 발전하고 있는 AI 기반 학술정보 서비스들, 특히 SciSpace, Elicit, Perplexity 등 생성형 모델 기반 플랫폼의 기술적 특성과 한계를 분석한다. 마지막으로 이러한 선행연구 및 서비스 분석을 바탕으로 기존 접근법의 한계를 도출하고, 본 연구가 제안하는 프레임워크의 차별성과 의의에 대해 논의한다.

2.1 이용자 프로파일링 연구

이용자 프로파일링은 정보시스템이 이용자의 특성을 파악하여 개인에게 적합한 정보를 제공하기 위한 핵심 기술로서, 1990년대 후반부터 검색엔진과 추천시스템 분야에서 본격적으로 연구되기 시작하였다. Roy와 Dutta(2022)의 추천시스템의 문헌고찰 결과, 초기 추천 시스템은 주로 연령, 성별, 전공과 같은 정적 속성에 기반하거나 단순한 협업 필터링을 적용하는 데 그쳤다. 이러한 접근은 이용자의 복잡한 정보요구나 변화하는 이용 맥락을 충분히 반영하기 어렵다는 한계가 있었다. 이후 2000년대에는 웹 로그 데이터, 클릭 패턴과 같은 동적 행위 기반 데이터가 포함되면서 적응형 모델링의 가능성이 열렸다.

동적 프로파일링이 확산되었음에도 불구하고

고, 단순 통계 기반의 행동 분석은 이용자 행동의 의미적 맥락을 해석하는 데 여전히 한계를 보였다. Purificato et al.(2024)는 이용자 모델링 및 프로파일링 기술의 진화 과정을 포괄적으로 분석한 연구에서, 최근의 프로파일링이 단순한 데이터 수집을 넘어 암묵적 데이터와 다중 행동을 통합적으로 모델링하는 방향으로 패러다임이 전환되고 있음을 강조하였다. 특히 이들은 최신 딥러닝 기술과 그래프 데이터 구조의 결합이 이용자 표현의 정확도를 높이는 핵심 요인을 지적하였으며, 더 나아가 기술적 성능뿐만 아니라 설명 가능성과 공정성과 같은 윤리적 가치가 차세대 이용자 모델링의 필수적인 요구 사항으로 대두되고 있다고 하였다.

이러한 흐름 속에 구체적인 모델링 기법에 대한 연구도 진행되고 있다. Liu et al.(2023)는 이용자의 구매 기록이나 블로그 글처럼 이용자가 실제로 남긴 행동 자료와, 서로 비슷한 이용자들 간의 연결 관계를 나타내는 이용자 간 관계 정보를 함께 활용하여 나이·성별 같은 기본적 특성을 예측하는 새로운 모델을 제안하였다. 이 연구에서는 행동 데이터 속에서 중요한 패턴은 강조하고 불필요한 정보는 줄이는 기법을 사용하여 기존보다 훨씬 높은 정확도를 얻었으며, 행동 정보와 이용자 간 관계를 동시에 고려하는 방식이 프로파일링 성능을 높일 수 있다고 주장하였다.

또한, 지식그래프를 활용한 의미 기반 접근도 주목받고 있다. Jung et al.(2023)은 지식그래프에 담긴 연구 주제 간의 의미 관계를 유지하면서 이용자의 선호를 표현할 수 있는 방법을 제안하였다. 연구에서는 이용자가 관심을 가진 항목들 사이에서 반복적으로 나타나는 의

미 있는 패턴을 추출하여 이용자 프로필로 구성하고, 이를 추천 모델과 결합해 추천 품질을 높였다. 이는 Zhang et al.(2024)과 Di Scala et al.(2024)이 지적한 바와 같이, 지식그래프 임베딩을 통해 데이터 희소성을 해결하고 개념 간의 심층적 관계를 활용할 때 추천 시스템의 성능과 이용자 만족도가 향상된다는 연구 결과와 맥을 같이 한다.

한편, 학술정보 분야에서는 의미적 적합성뿐만 아니라 정보의 질적 가치와 이용자 맥락이 중요한 요소로 작용한다. Zhang et al.(2023)은 학술 추천 시스템 서베이에서 연구자는 단순한 연관성뿐만 아니라 논문의 권위와 신뢰성을 중요시한다고 지적하였으며, Sabiri et al.(2025)은 인용도나 저널 영향력과 같은 질적 지표를 추천 알고리즘에 통합한 하이브리드 품질 기반 추천 시스템을 제안하였다. Hussain et al.(2018)은 이용자의 경험과 숙련도에 따라 UI가 적응적으로 변화하는 모델을 제시하여, 프로파일링이 단순 추천을 넘어 인터페이스 차원의 개인화로 확장되어야 함을 시사하였다.

2.2 정보요구 및 정보행위 이론

정보학 분야에서 이루어진 정보요구와 정보행위 연구는 이용자 프로파일링을 이해하는 데 중요한 이론적 기반을 제공한다. Taylor(1968)는 정보요구가 처음에는 막연하고 불확실한 상태에서 출발하여 점차 구체적인 요구로 발전한다고 설명하였다. 그는 이용자가 느끼는 지식의 공백과 불확실성이 새로운 정보 탐색을 촉발한다는 점을 강조하며 정보시스템이 이용자의 상황과 맥락을 이해해야 한다는 점을 시사하였다.

Belkin(1980)의 ASK(Anomalous State of Knowledge) 이론 역시 정보요구의 핵심이 이용자의 지식의 부족이나 이상 상태에 있다고 설명하였다. Belkin은 이용자가 실제로 어떤 정보가 필요한지 명확히 알지 못한 상태에서 탐색을 시작하는 경우가 많다고 보았다. 이러한 관점은 단순히 사용자가 입력한 검색어만으로는 이용자의 요구를 파악하기 어렵다는 점을 의미하며, 시스템이 이용자의 행동과 탐색 흐름 속에서 숨겨진 요구를 추론할 필요가 있음을 알 수 있다.

Kuhlthau(1991)의 정보탐색과정 모델은 탐색 과정을 시작, 선택, 탐색, 형성, 수집, 제시의 6단계로 구분하며, 각 단계마다 이용자의 인지적 상태와 정보 요구가 달라짐을 입증하였다. 또한 Vakkari(2001: 2003)는 연구자가 과제를 수행함에 따라 초기에는 일반적 정보를 찾지만, 후기로 갈수록 구체적이고 방법론적인 정보를 찾는다는 단계별 정보원 선호 변화를 실증적으로 분석하였다. 이는 학술정보 서비스가 단편적인 검색 로직이 아닌 연구 단계별로 차별화된 추천 로직을 가져야 한다는 본 연구의 방향과 부합한다고 볼 수 있다.

Wilson(1996)의 정보행위 모델은 정보탐색이 단순히 검색을 수행하는 활동에 그치지 않고, 동기, 장벽, 인지적 특성, 사회적·환경적 요인 등 여러 요소가 함께 작용하는 복합적 과정이라고 설명하였다. Wilson은 정보 추구·탐색·이용의 단계뿐 아니라, 이용자의 역할, 주변 환경, 다른 사람과의 상호작용까지도 정보행위에 영향을 미친다고 보았다. 이러한 관점은 학술정보 시스템이 단순히 검색 결과를 제공하는 수준을 넘어, 연구 목적, 연구 단계, 이용자의 수준

과 배경까지 고려해야 한다는 본 연구의 접근과 관련성이 높다고 할 수 있다.

2.3 AI 기반 학술정보 서비스의 발전과 한계

최근 몇 년간 AI 기반 학술정보 서비스는 빠른 속도로 발전하고 있으며, 특히 생성형 언어 모델(Large Language Model, LLM)을 활용한 플랫폼이 문헌 요약, 질의응답, 문헌 추천 등 다양한 기능을 제공하면서 연구자의 주목을 받고 있다. 특히 Zhao et al.(2024)이 지적한 바와 같이, 기존의 딥러닝 기반 추천 시스템이 이용자의 관심사를 심층적으로 이해하거나 추천 결과에 대해 추론하는 데 한계를 보임에 따라, 뛰어난 언어 이해 및 추론 능력을 갖춘 거대언어모델(LLM)을 학술정보 서비스에 접목하려는 시도가 급증하고 있다. 박정훈(2025)의 연구에 따르면 SciSpace, Elicit, Perplexity 등의 플랫폼은 문헌 자동 요약, 개념 정리, 자연어 질의응답 등에서 높은 성능을 보이며 연구 효율성을 크게 향상시켰다.

이러한 생성형 AI 기반 플랫폼들은 문헌 요약의 정확도, 텍스트 재구성 능력, 질의응답 속도 면에서 매우 높은 성능을 보이며 연구자에게 새로운 형태의 정보 접근 방식을 제공하고 있으나 이용자 모델링 측면에서는 근본적인 한계를 갖는다. Dessi et al.(2021)은 AI 분야의 방대한 문헌을 분석하여 자동으로 지식그래프를 생성하는 연구를 통해 비정형 텍스트를 구조화된 지식으로 변환하는 것이 가능함을 보였으나, 현재의 상용 서비스들은 이러한 구조화된 지식과 이용자의 연구 흐름을 유기적으로

결합하지 못하고 있다.

실제로 생성형 모델은 입력된 문장 또는 문헌을 기반으로 텍스트를 생성하는 데 최적화되어 있기 때문에 이용자가 어떤 연구 단계에 있는지, 연구 경험이 어느 정도인지, 현재 무엇을 하고 싶은지, 어떤 문맥에서 질문을 던지는지, 연구자가 속한 지역이나 연구 커뮤니티의 특성이 어떠한지 등을 장기적으로 추적하거나 이해하는 기능은 제한적이다. LLM은 대화 중 맥락을 유지할 수 있지만, 이는 세션 내부에서의 일시적 맥락 유지에 불과하며, 이용자의 장기적 탐색 행태나 연구 패턴을 기반으로 한 이용자 프로파일링과는 다른 개념이다. 즉, LLM이 유지하는 맥락은 텍스트 연속성이고, 학술정보서비스가 필요로 하는 맥락은 이용자 상태와 연구 단계의 이해라는 점에서 근본적으로 차이가 있다.

또한 데이터의 파편화와 상호운용성 부족 역시 중요한 과제로 남아있다. Wilkinson et al. (2016)은 과학 데이터의 관리와 공유를 위한 FAIR 원칙(Findable, Accessible, Interoperable, Reusable)을 제안하며, 데이터가 기계적으로 처리 가능하고 상호 운용되기 위해서는 표준화된 메타데이터와 식별 체계가 필수적임을 강조하였다. 하지만 현재의 대다수 학술정보 플랫폼은 여전히 데이터 사일로 문제에 직면해 있어, 논문, 특허, 데이터셋 등 이기종 콘텐츠 간의 의미적 연계가 원활하지 않다.

결과적으로 현재의 AI 기반 학술정보 서비스는 강력한 텍스트 생성 기능을 제공하지만, 문헌 중심 문맥(document-centric context)은 유지하지만, 이용자 중심 문맥(user-centric context)은 반영하지 못하고 있다. 또한, 데이터의 상호운용성과 Hussain et al.(2018)이 제안한 이용자

경험 기반의 적응형 인터페이스, 그리고 Zhang et al.(2024)이 제시한 지식그래프 기반의 심층적 의미 연결을 통합적으로 구현하는 데에는 이르지 못하고 있다. 이는 연구자가 단순한 정보 소비자가 아니라 지식을 생성하고 해석하는 역할을 수행한다는 문헌정보학적 관점에서 볼 때 한계점이라고 볼 수 있다.

2.4 기존 연구와 서비스의 한계

기존 이용자 프로파일링 연구는 정적 정보 또는 단순 행동 데이터를 기반으로 한 모델링에 집중되어, 연구자가 어떤 맥락에서 정보를 찾고 있는지, 연구 과정에서 무엇을 필요로 하는지를 충분히 반영하지 못했다. 또한, 최근의 AI 기반 학술정보 서비스는 문헌 요약이나 분석 능력은 뛰어나지만, 이용자가 어떤 단계의 연구를 진행 중인지, 어떤 배경이나 목적을 가지고 탐색하고 있는지를 이해하고 이에 맞는 정보를 제공하는 기능은 미흡하다. 기존 연구들은 연구 단계를 세부적으로 나누거나 이용자의 탐색 행동 수준을 체계적으로 모델링하는 시도가 부족하였다. 연구자가 정보 탐색 과정에서 어떤 종류의 도움을 필요로 하는지, 어떤 수준의 안내가 적합한지를 구체적으로 파악하기 어려웠다.

본 연구는 이러한 문제를 해결하기 위해 기존 연구에서 논의했던 정적·동적·의미 기반 정보뿐만 아니라, 연구 단계와 행동 수준이라는 새로운 기준을 추가하여 이용자 프로파일링 구조를 재설계하고자 한다. 이를 통해 기존 연구와 서비스가 갖고 있던 한계를 보완하고, 연구자 중심의 학술정보 서비스 모델을 제시하는데 그 목적이 있다.

3. 요구사항 분석 및 설계요소 도출

본 장에서는 이용자 프로파일링 기반 학술정보 서비스 프레임워크를 설계하기 위한 요구사항을 구조적으로 분석하고, 이를 토대로 핵심 설계 요소를 도출한다. 학술정보 서비스는 연구자의 지식 탐색·해석·생성 활동을 직접적으로 지원하는 플랫폼이므로, 요구사항 분석은 기능적 관점뿐 아니라 연구자의 행동 패턴, 탐색 맥락, 콘텐츠 구조, 기술적 환경 등 다층적 요소를 포괄해야 한다. 특히 기존 학술정보 서비스가 정적 정보 중심으로 이용자를 이해하거나 단순 키워드 기반 검색에 의존해온 한계를 고려할 때, 연구 단계와 행동 수준을 명시적으로 반영한 요구사항 도출이 필요하다.

3.1 요구사항 분석

3.1.1 이용자 관점에서의 요구사항

학술정보 이용자는 일반적인 정보 서비스 이용자와는 다른 정보행태를 보인다. 연구자는 특정 주제에 대한 이해를 확장하기 위해 반복적이고 구조화된 탐색을 수행하며, 각 탐색 단계에서 필요로 하는 정보의 성격이 다르다. 연구자의 정보요구는 단일 시점에서 고정된 것이 아니라 연구 진행 상황에 따라 지속적으로 변화하는 동적 특성을 갖는다. 따라서 이용자 요구를 도출하기 위해서는 연구자의 활동을 탐색 초기 단계, 문제 정의 단계, 관련 연구 수집 단계, 비교·분석 단계, 연구 설계 단계와 같은 연구 과정 단계로 구분하여 각 단계에서의 요구를 세밀하게 파악해야 한다.

연구자가 탐색을 수행하는 초기 단계에서는 문제 영역 설정을 위한 폭넓은 문헌 탐색과 개념 이해가 중심이 된다. 이 단계에서는 검색 과정에 대한 지식이 부족하거나 주제에 대한 이해도가 낮아 관련 자료를 효율적으로 찾기 어렵다. 따라서 이용자가 입력한 검색어가 충분하지 않을 때라도 연관된 개념을 확장하여 기본 자료를 제공할 수 있어야 한다. 반면 중간 단계에서는 반복적인 검색과 비교 행위가 나타나며, 이용자는 보다 구조화된 정보 탐색을 수행한다. 이 단계에서는 주요 키워드의 상관관계, 문헌 간 의미적 유사성, 주제별 연구 흐름을 이해할 수 있는 정보가 필요하다. 고급 연구 단계에서는 연구 문제를 정교하게 정의하고 근거 기반을 마련하기 위한 자료가 요구되며, 해당 분야의 핵심 연구자, 주요 학술지, 최신 트렌드, 데이터 기반 분석 결과 등이 필요하다.

이처럼 연구자 요구는 단순히 더 많은 정보를 제공하는 것이 아니라, 연구 단계와 수준에 적합한 정보를 제공해야 한다는 점에서 기존 서비스와 차이가 있다. 이에 본 프레임워크는 이용자의 행동 로그로부터 연구 단계(예: 탐색 단계, 문제 정의 단계, 자료 축적 단계, 비교·분석 단계, 정리 단계)를 추론하고, 이용자의 탐색 방식에 따라 행동 수준-기초 탐색(L1), 반복 탐색(L2), 분석 중심 탐색(L3) 등을 판별하는 것을 설계 목표로 하였다. 이를 통해 연구자의 상황적 요구를 동적으로 파악하고, 단계에 적합한 탐색 지원을 제공하는 기반을 마련하였다.

3.1.2 콘텐츠 관점에서의 요구사항

학술정보 서비스가 제공하는 콘텐츠는 논문,

특히, 연구보고서, 데이터셋 등 다양한 형식과 구조를 가지며, 이들은 주제적 성격·메타데이터 구조·생산 주체·인용 관계 등 여러 차원에서 서로 다른 특성을 지닌다. 그러나 기존 학술정보 서비스는 이러한 이질적인 자원을 단순 키워드 기반 검색이나 최신순·정확도순과 같은 일차원적 정렬에 의존해 제공하고 있다. 이러한 방식은 자료 유형의 차이, 의미적 맥락, 연구 흐름과 같은 심층적 관계를 반영하지 못하여 연구자가 실제로 필요로 하는 지식을 충분히 연결해주기 어렵다.

우선, 이종(Heterogeneous) 콘텐츠 간 의미적 연계의 부족이 중요한 문제로 지적된다. 예를 들어 동일한 연구 주제를 다루더라도 논문은 이론적 배경과 실험 결과를 제공하며, 특허는 기술적 구현 방식과 응용 가능성을 중심으로 구조화되어 있고, 연구보고서는 정책적·산업적 함의를 포함하는 등 콘텐츠의 목적과 구조가 상이하다. 그럼에도 불구하고 기존 서비스는 이러한 자원들을 개별적 객체로 처리하여, 연구자가 주제와 관련된 다양한 정보원을 한번에 이해하는 데 필요한 의미 기반 연결을 제공하지 못한다. 따라서 콘텐츠 간의 개념적·주제적 유사성을 반영하여 의미적으로 연관된 자원을 통합적으로 제시할 수 있는 구조가 요구된다.

둘째, 학술정보는 동일한 주제를 다루더라도 질적 가치의 차이가 뚜렷하게 나타난다. 인용도, 저자 영향력, 학술지 수준, 기관·프로젝트 규모, 이용량 등은 정보의 신뢰성과 전문성을 판단하는 데 중요한 지표라고 할 수 있다. 기존 시스템은 이러한 질적 요소를 충분히 반영하지 못해, 연구자가 핵심 문헌과 주변 문헌을 구별

하기 어렵다는 한계가 있다. 따라서 콘텐츠의 품질과 평판을 반영하는 다차원적 평가 기준을 도입하여 의미 기반 탐색을 보완하고 신뢰도 높은 정보 선택을 지원해야 한다.

셋째, 학술정보는 단일 문헌보다 문헌 집합과 연구흐름을 파악하는 것이 더 중요한 경우가 많다. 연구자는 특정 문헌만 확인할 뿐만 아니라 주제 영역과 주제의 변천 과정, 핵심 연구자 네트워크 등을 통해 연구 지식을 구조적으로 파악하기도 한다. 따라서 문헌 간 인용 구조, 개념의 확장 관계 등을 반영한 집합적·맥락적 콘텐츠 제공 기능이 요구된다.

이러한 관점에서 본 프레임워크는 다음과 같은 콘텐츠 중심 요구사항을 도출하였다. 첫째, 논문·특허·보고서 등 이종 콘텐츠 간 주제적·의미적 연관성을 분석하여 교차 추천이 가능한 구조를 마련해야 한다. 둘째, 콘텐츠의 질적 가치(예: 인용도, 저자 영향력, 학술지 수준, 이용량 등)를 반영한 신뢰 기반 정보 제공 모델을 구축해야 한다. 셋째, 의미 기반 임베딩, 지식그래프, 토픽 모델링 등 고급 분석 기법을 활용하여 문헌 집합 단위의 구조적 탐색을 지원하는 기능을 포함해야 한다.

이와 같은 콘텐츠 중심 요구사항은 사용자 요구 분석에서 도출된 연구 단계·행동 수준 기반 요구와 결합하여, 보다 정교한 의미 기반·맥락 기반 학술정보 서비스 설계를 가능하게 한다.

3.1.3 서비스 관점에서의 요구사항

학술정보 서비스는 단순한 검색 기능을 제공하는 시스템을 넘어, 연구자의 탐색 목적·연구 단계·행동 수준을 고려하여 정보 제공 방식을 능동적으로 조정하는 지능형 서비스 구조를 요

구한다. 기존 서비스는 동일한 검색 화면과 동일한 추천 방식이 모든 이용자에게 일괄적으로 제공되었기 때문에, 연구자의 상황에 따른 탐색 요구를 충분히 차별화하지 못하였다. 이러한 문제는 학술정보 탐색이 구조화된 연구 과정 속에서 수행된다는 점을 고려하면, 서비스 제공 방식의 고도화가 필수적임을 알 수 있다.

따라서, 서비스는 이용자의 상태에 따라 동적으로 반응할 수 있어야 한다. 예를 들어 초기 연구 단계에서는 기본 개념, 핵심 용어, 대표 문헌 등 기초 자료 중심의 안내가 요구되며, 어느정도 연구가 진행되는 단계에서는 인용 네트워크, 연구 트렌드, 핵심 저자·기관 분석 등 고도화된 분석 기능을 필요로 한다. 따라서 이용의 연구 단계에 따른 행동 수준(L0~L3) 별로 UI 구성, 탐색 흐름, 추천 방식이 달라지는 수준 기반 인터페이스가 필수적이다.

또한, 학술정보 탐색 과정은 이용자가 인지하지 못하는 정보 요구를 선제적으로 파악할 필요가 있다. 연구자는 특정 문헌을 열람하거나 반복적으로 유사한 키워드를 검색하는 과정에서 암묵적 연구 목적을 드러내는데, 기존 시스템은 이러한 행동 패턴 기반 정보 요구 예측 기능이 부족하였다. 따라서 서비스는 이용자의 과거 행동, 탐색 흐름, 반복 조회 패턴 등 행동 기반 데이터를 분석하여 다음 단계에서 필요할 가능성이 높은 정보를 선제적으로 제공하는 예측형 추천 기능을 갖추어야 한다.

학술정보 이용자들은 단일 문헌보다는 문헌 집합, 연구분야 구조, 지식의 흐름과 같은 맥락적·집합적 정보를 필요로 하는 경우가 많다. 따라서 연구자들은 주제별 정보, 개념 간 관계를 탐색하는 기능이 필요하며, 서비스는 이러

한 요구를 반영한 지식 탐색 도구를 제공해야 한다. 문헌 간 의미적 분류, 인용구조 기반 연구 흐름 등 다양한 지표를 활용한 구조적 탐색 기능은 연구자 요구 충족에 필수적이다.

서비스의 상호연동성은 필수적인 요구사항으로 볼 수 있다. 학술정보 서비스는 다양한 메타데이터 구조, 외부 데이터, API 등 여러 정보 및 시스템들과 연동되기 때문에, 서비스 구조는 모듈 간 호환성과 확장성을 확보하는 형태로 설계되어야 한다. 특히 의미 기반 추천과 연구 단계 기반의 서비스 흐름을 결합하기 위해서는 각 서비스 모듈(탐색·추천·시각화·분석)이 일관된 정보 구조를 공유해야 한다.

3.1.4 기술적 관점에서의 요구사항

학술정보 서비스 프레임워크는 다양한 데이터 유형과 복잡한 분석 기법을 통합하여 운영되기 때문에, 이를 안정적으로 지원하기 위한 기술적 요구조건이 필요하다. 기술적 요구사항은 데이터 수집·처리·저장·분석·배포 등의 전 과정에서 시스템이 효율적으로 작동하기 위한 기반을 구성한다. 우선, 로그 데이터 수집 범위의 제한 및 개인정보 보호 정책 등으로 인해 시스템은 제한된 로그 데이터만으로도 의미 있는 프로파일을 생성할 수 있어야 한다. 검색 질의, 클릭 이력, 페이지 이동, 문헌 열람 여부 등 제한적인 행동 정보만으로도 이용자의 연구 단계와 행동 수준을 추정할 수 있는 분석 모듈이 요구된다.

또한, 의미 기반 탐색 및 분석을 위해서는 대규모 언어모델, 임베딩 모델, 지식 그래프 연동 등 고급 분석 기술이 필요하기 때문에, 시스템은 이러한 분석 모듈을 외부 API 또는 내부 모

들로 유연하게 연동할 수 있는 구조를 갖추어야 한다. 메타데이터는 연구 문헌, 특허, 보고서, 데이터셋 등 다양한 자료 유형을 처리하기 위해, 국제 표준(Dublin Core, DataCite, CERIF 등)을 기반으로 다양한 자료 유형을 확장 가능하게 설계하여 기관 간 상호운용성을 확보해야 한다. 개인정보 보호 역시 중요한 기술적 요구 사항이다. 사용자 프로파일은 개인의 탐색 패턴, 연구 관심사 등 민감한 정보를 포함하므로 이를 안전하게 처리하기 위한 비식별화, 암호화, 접근제어 등의 개인정보보호 기술이 필수적이다.

3.2 요구사항 종합 분석을 통한 설계요소 도출

이용자, 콘텐츠, 서비스, 기술의 네 가지 관점에서 도출된 개별 요구사항들을 분석한 결과, 이들은 상호 독립적인 것이 아니라 기능적 유사성에 따라 5가지 핵심 범주로 자연스럽게 연결된다. 구체적으로 살펴보면, ① 연구 맥락 파악 요구는 연구 단계 모델링과 행동 수준 판별로, ② 이중 데이터의 연결 및 가치 평가 요구는 의미 기반 분석으로, ③ 동적 서비스 제공 요구는 프로파일 갱신 메커니즘으로, ④ 시스템 연동성 요구는 모듈화 아키텍처 필요하다.

첫째, 연구 단계 모델링(SE-1) 구조에서 연구 단계는 연구자가 어떤 상황에서 어떤 목적을 갖고 탐색을 수행하는지 설명하는 핵심 변수이며, 탐색 범위, 추천 전략, 정보 제공 방식 등 서비스 전반의 동작을 규정하는 기준이라고 할 수 있다. 기존의 정적 프로파일이나 단순 클릭 기반 이해 방식으로는 연구자의 실제 정보 요

구를 포착하기 어렵기 때문에, 연구 단계 모델을 프로파일 구조, 추천 알고리즘, 서비스 전반에 적용하는 것이 필수적이다. 이를 통해 시스템은 기초 정보 탐색부터 고급 분석 단계까지 맞춤형으로 지원할 수 있다.

둘째, 행동 수준 기반 이용자 모델링(SE-2)의 행동 수준은 탐색 방식과 이용자의 숙련도를 정량적으로 구분하는 개념으로, 예를 들어 L1은 기초 탐색 중심, L2는 반복적 비교 탐색 중심, L3는 분석 중심 탐색으로 설정할 수 있다. 행동 수준은 단순한 이용빈도나 클릭 횟수를 넘어, 탐색 과정에서 나타나는 패턴과 정보 활용 방식을 반영하므로 개인화 서비스의 정밀도를 향상시키는 핵심 요소이다. 특히 연구 단계와 행동 수준이 결합되면, 이용자의 현재 상황에 가장 적합한 정보와 기능을 선별해 제시할 수 있어 고도화된 맥락 기반 탐색이 가능하다.

셋째, 학술정보는 단순 키워드 일치만으로는 내용적 관련성을 충분히 설명하기 어렵기 때문에, 의미 임베딩, 지식그래프, 토픽 모델링 등 의미 기반 분석 기술(SE-3)이 필수적으로 포함되어야 한다. 이러한 기술은 문헌 간 의미적 유사성 계산, 주제 확장, 핵심 연구자 및 연구 흐름 분석 등 다양한 고급 기능을 지원하며, 연구자가 단순 검색 이상의 구조적 탐색을 수행할 수 있도록 돕는다. 이는 특히 이중 콘텐츠 간 의미적 차이를 해소하여 논문·특허·보고서·데이터셋 간 유기적 연결을 가능하게 한다는 점에서 중요하다.

넷째, 연구자의 탐색은 정적이 아니라 지속적으로 변화하는 동적 과정이기 때문에, 프로파일 역시 고정된 형태가 아니라 행동 로그에 따라 지속적으로 업데이트되는 프로파일 생성

및 갱신 메커니즘(SE-4)이 필요하다. 검색 질의, 클릭 흐름, 조회 문헌, 페이지 이동 패턴 등 최소한의 로그만으로도 연구 단계와 행동 수준을 재추정하고, 장기적 관심사와 연구 흐름을 파악할 수 있는 메커니즘이 요구된다. 이러한 동적 갱신 구조는 상황 변화에 따라 적절한 정보를 제시하는 적응형 서비스를 구현하는 데 핵심 역할을 한다.

다섯째, 학술정보 서비스는 기관별 환경, 데이터 종류, 외부 연동 조건 등이 다양하기 때문에, 프레임워크는 프로파일링 모듈, 의미 분석 모듈, 추천 모듈 등 기능 단위를 독립된 계층으로 구성하고 API 기반으로 연동할 수 있도록 모듈화·확장성 기반 아키텍처(SE-5)가 필요하다. 이러한 계층적 구조는 유지보수와 확장성을 높일 뿐만 아니라, 새로운 분석 엔진이나 데이터 소스가 추가될 때도 시스템의 일관성을 유지할 수 있게 한다.

마지막으로, 본 연구에서는 네 가지 관점에서 도출된 요구사항이 앞서 제시한 설계 요소

와 어떤 방식으로 연결되는지를 요구사항 추적 매트릭스를 통해 정리하였다. 이를 통해 프레임워크 설계의 타당성, 요구사항 충족 여부, 구성요소 간 일관성을 검증할 수 있으며, 다음 장에서 제시할 시스템 아키텍처와 프로파일 구조 설계의 근거를 명확히 제시한다(〈표 1〉 참조).

4. 이용자 프로파일링 기반 프레임워크 설계

4.1 프레임워크 개요

제안하는 프레임워크는 연구자의 탐색 과정 전반을 하나의 연속된 흐름으로 바라보며, 이를 지원하기 위해 입력 계층, 프로파일링 계층, 의미 분석 계층, 추천 계층, 서비스 제공 계층으로 구성되는 다층적 구조를 채택하였다. 이러한 계층 구조는 각 모듈의 역할과 책임을 명확히 분리하여 시스템의 복잡성을 낮추는 동시에,

〈표 1〉 요구사항 추적 매트릭스

요구사항 ID	요구사항 내용	설계요소	설계요소 설명
REQ-U-01	연구 단계 식별	SE-1 연구 단계 모델링	탐색 흐름·추천 전략을 연구 단계 기반으로 구조화
REQ-U-02	행동 수준 판별	SE-2 행동 수준 모델링	탐색 행동 차이에 따른 개인화 정보 제공
REQ-C-01	콘텐츠 의미 기반 연계	SE-3 의미 기반 분석 모듈	임베딩·지식그래프 기반 의미 연관 분석
REQ-C-02	콘텐츠 질적 가치 반영	SE-3 의미 기반 분석 모듈	평판·인용도 기반 신뢰성 있는 콘텐츠 추천
REQ-S-01	동적 UI 구성	SE-1 연구 단계 모델링 SE-2 행동 수준 모델링	단계/수준별 인터페이스 및 탐색 구성
REQ-S-02	선제적 추천	SE-4 프로파일 갱신/추론 메커니즘	행동 기반 예측 추천 및 실시간 요구 반영
REQ-T-01	로그 기반 분석	SE-4 프로파일 갱신 메커니즘	최소 로그 기반 연구 단계·행동 수준 추정
REQ-T-02	외부 시스템 연동성	SE-5 모듈화·확장성 아키텍처	내부/외부 API 연동 기반 구조
REQ-T-03	메타데이터 표준 호환	SE-5 모듈화·확장성 아키텍처	데이터 구조 표준화 및 상호운용성 확보
REQ-T-04	개인정보 및 보안	SE-4·SE-5 결합	안전한 프로파일 저장/처리 및 접근 제어

다양한 분석 기능이 긴밀하게 연동되도록 설계된 것이 특징이다.

입력 계층은 이용자의 검색 질의, 클릭 히스토리, 문헌 열람 기록 등 탐색 과정에서 발생하는 모든 행동 로그를 수집하는 역할을 담당한다. 이러한 행동 데이터는 이후 프로파일 생성과 의미 분석의 기초 자료가 되며, 로그 접근이 제한적인 기관 환경에서도 최소한의 행동 정보만으로 의미 있는 분석이 가능하도록 구성되었다.

프로파일링 계층은 입력된 데이터를 기반으로 이용자의 정적 특성, 동적 탐색 패턴, 의미 기반 관심 영역, 연구 단계, 행동 수준을 종합적으로 모델링하는 핵심 계층이다. 이 계층은 제3장에서 도출된 이용자 관점 요구사항인 연구 단계 식별과 행동 수준 판별을 직접적으로 구현하며, 이용자의 맥락적 정보를 구조화하여 이후 분석 모듈에 전달하는 중심적 역할을 수행한다.

의미 분석 계층은 문헌 임베딩, 지식그래프, 토픽 모델링 등 다양한 의미 분석 기법을 활용하여 이중 콘텐츠 간의 주제적·개념적 관계를 파악한다. 이를 통해 논문, 특허, 보고서 등 다양한 콘텐츠의 의미적 연결고리를 도출하고, 특정 주제가 어떤 개념망 속에서 확장되는지를 분석한다. 이 계층은 제3장에서 확인된 콘텐츠 중심 요구사항을 반영하여 학술정보의 구조적 이해를 가능케 한다.

추천 계층은 프로파일링 계층과 의미 분석 계층에서 생성된 정보를 통합하여 연구자에게 적합한 정보를 선별하여 제시하는 역할을 한다. 추천 방식은 연구 단계와 행동 수준에 따라 차별화되며, 의미 기반 분석과 예측형 추천 기법을 결합하여 연구자의 다음 탐색 단계에서 필요

할 정보를 선제적으로 제공하도록 설계되었다.

마지막으로 서비스 제공 계층은 사용자 인터페이스, 시각화 도구, API 기반 모듈 간 연동 구조 등을 포함한다. 이 계층은 연구자의 수준과 탐색 단계에 따라 화면 구성이 변하는 동적 인터페이스를 제공하며, 학술정보의 구조적·관계적 특성을 직관적으로 전달하기 위한 시각화 기능을 갖춘다. 또한 이용자의 행동 결과는 다시 입력 계층으로 전달되어 프로파일이 갱신되는 순환 구조를 이룬다.

4.1.1 프로파일링 계층

프로파일링 계층은 본 프레임워크의 중심을 이루는 구성요소로서, 이용자의 정적 특성과 동적 특성, 의미 기반 정보, 연구 단계, 행동 수준을 일관된 구조로 통합한다. 프로파일은 단순한 정적 프로필 정보의 저장 뿐만 아니라 연구자 탐색 행동의 의미적 변화를 지속적으로 반영하는 동적 구조로 설계되었다.

정적 프로필은 연구자의 소속, 연구 분야, 직급, 관심 영역 등 상대적으로 변화가 적은 기본 정보를 포함한다. 이는 학술정보의 의미적 필터링이나 추천의 초기 기준으로 활용된다. 반면 동적 프로필은 검색 질의 변화, 문헌 조회 패턴, 문헌 난이도 수준의 변화 등 시간이 지남에 따라 달라지는 탐색 행동을 반영한다. 연구자는 특정 주제를 반복적으로 탐색하거나 새로운 주제 영역으로 확장하는 등 동적 행동을 보이기 때문에, 이러한 변화는 연구자의 정보 요구를 판단하는 중요한 단서가 된다.

의미 기반 프로필은 임베딩 벡터, 지식그래프 분석 결과 등 연구 주제의 의미적 구조를 반영하여 구성된다. 이를 통해 연구자가 어떤 개

념 클러스터를 중심으로 탐색하는지, 특정 주제가 어떤 인접 개념과 연계되는지 등을 파악할 수 있다. 이러한 의미 기반 프로파일은 단순 텍스트 매칭이 아닌 학술적 의미 단위에 기반한 탐색 지원을 가능하게 한다.

연구 단계 상황은 이용자의 행동 패턴을 기반으로 분석하여 연구자가 탐색, 문제 정의, 자료 수집, 비교·분석, 정리 단계 중 어느 단계에 위치해 있는지를 추론한다. 예를 들어 기초 개념을 집중적으로 탐색하는 경우 탐색 단계로 판단하며, 반복적 비교·심층 분석이 나타나는 경우 분석 단계로 분류된다.

행동 수준 프로파일은 이용자의 탐색 방식의 복잡성을 반영하는 지표로서, 단순 조회 중심의 기본 수준에서부터 심층 분석 중심 탐색까지 L1~L3의 수준으로 분류된다. 행동 수준은 연구 단계와 함께 결합하여 이용자에게 적합한 콘텐츠 유형과 기능을 제공하는 데 활용된다.

4.1.2 의미 기반 분석 계층

의미 기반 분석 계층은 학술정보 간의 의미적 구조를 파악함으로써 연구자에게 단순 목록 이상의 탐색 경험을 제공하는 역할을 담당한다. 문헌 임베딩 모델은 문헌의 내용을 벡터 공간에 매핑하여 문헌 간 유사성을 정밀하게 계산할 수 있도록 하며, 이는 특정 문헌과 의미적으로 유사한 문헌을 탐색하는 데 활용된다.

지식그래프는 다양한 개념 간의 의미적 관계를 구조적으로 표현하여, 사용자가 탐색하는 주제가 어떤 개념망 속에서 위치하는지를 이해할 수 있도록 한다. 예를 들어 특정 개념과 관련된 상위·하위 개념, 인접 개념, 특정 분야의 핵심 연구자 등을 탐색할 수 있다.

토픽 모델링은 대규모 문헌 집합에서 주제 분포를 추출하여 연구 동향을 분석하고, 특정 주제가 어떤 방식으로 확장되거나 변화해왔는지를 이해하는 데 활용된다. 이러한 분석 결과는 연구자가 보다 넓은 연구 흐름 속에서 자신의 탐색 위치를 파악하도록 돕는다.

이 계층은 프로파일링 계층과 긴밀하게 연동되어 의미 기반 프로파일을 갱신하며, 추천 계층의 입력 데이터로 활용된다. 따라서 의미 기반 분석 계층은 콘텐츠 관점에서 제기된 주제적 유사성, 질적 가치, 구조적 탐색의 필요성을 충족하는 핵심 모듈로 기능한다.

4.1.3 추천 계층

추천 계층은 프로파일 정보와 의미 분석 결과를 결합하여 연구자의 탐색 맥락에 적합한 콘텐츠를 제공하는 역할을 한다. 추천 방식은 규칙 기반 추천, 의미 기반 추천, 행동 수준 기반 추천, 예측형 추천으로 구성된다.

초기 연구 단계나 기초 수준 행동을 보이는 이용자에게는 검색어 기반이나 대표 문헌 중심의 규칙 기반 추천이 주로 제공된다. 이에 반해 의미 기반 추천은 임베딩 기반 문헌 유사도와 지식그래프 기반 개념 관계를 활용하여 보다 정교한 의미적 추천을 수행한다.

행동 수준 기반 추천은 이용자의 숙련도에 따라 추천의 깊이와 복잡성을 조정한다. 예를 들어 L1 수준의 이용자에게는 기초적인 개념 설명과 핵심 문헌을 제공하는 반면, L3 수준의 이용자에게는 인용 네트워크, 연구자 네트워크, 트렌드 분석 등 고급 기능을 제공한다.

예측형 추천은 반복 조회 패턴, 문헌 이동 경로, 주제 집중도 등 행동 기반 로그를 활용하여

이용자가 다음 탐색 단계에서 필요할 가능성이 높은 정보를 선제적으로 제시한다. 이러한 기능은 연구자가 탐색 과정에서 인지하지 못한 정보 요구를 발견하는 데 도움을 준다.

4.1.4 서비스 계층

서비스 계층은 추천된 정보를 이용자에게 시각적으로 전달하는 역할을 수행한다. 학술정보는 내용 구조가 복잡하고 의미적 관계망이 깊기 때문에, 단순 텍스트 기반의 화면 구성만으로는 정보의 구조적 의미를 충분히 전달하기 어렵다. 이에 따라 본 프레임워크는 연구 단계와 행동 수준에 따라 화면 구성이 동적으로 변경되는 인터페이스를 제공하도록 설계하였다.

초급 수준의 이용자에게는 단순한 문헌 리스트와 기본적 개념 설명을 중심으로 한 인터페이스가 제공되며, 고급 수준의 이용자에게는 지식그래프 기반 개념 지도, 연구자 네트워크, 주제 흐름 시각화 등 구조적 탐색 도구가 제공된다. 이러한 UI의 적응적 변화는 연구 과정 단계와 행동 수준을 반영하여 이용자의 탐색 효과를 극대화하기 위한 조치이다.

또한 서비스 제공 계층은 이용자의 클릭, 문헌 선택, 비교 행동 등의 피드백을 다시 입력 계층으로 전달하여, 프로파일의 지속적으로 갱신되는 순환 구조를 적용하였다. 이는 연구자의 탐색 과정이 정적이 아니라 변화하는 과정이라는 점을 반영하여 시스템이 동적으로 적응할 수 있도록 하는 핵심 설계 원리이다.

4.2 프로파일 구조 설계

이 절에서는 제안 프레임워크에서 활용되는

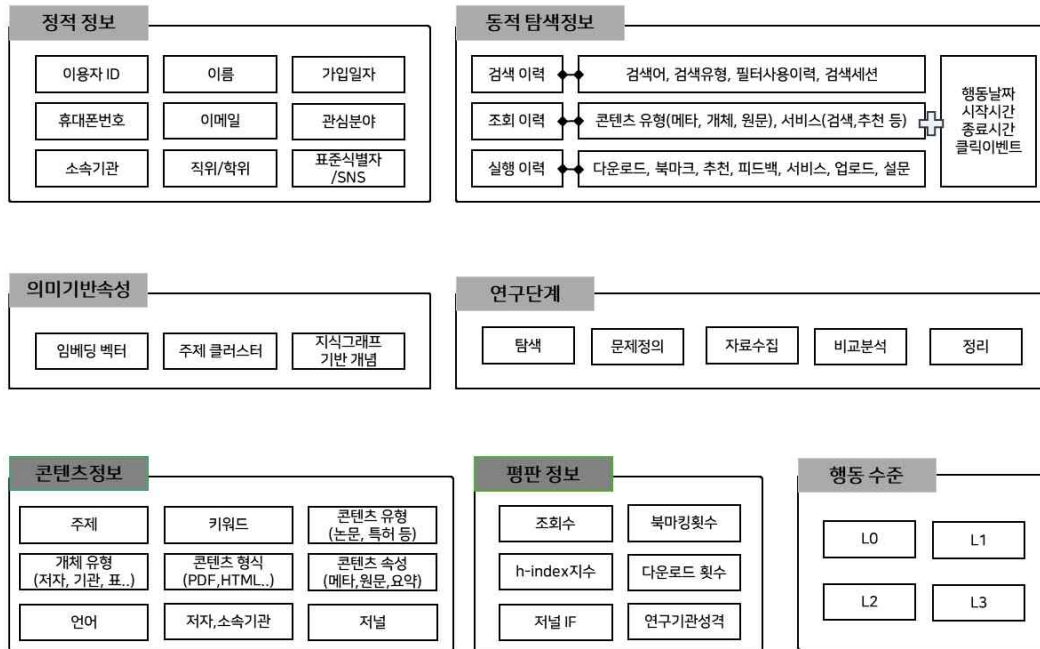
이용자 프로파일의 구조를 체계적으로 정의한다. 본 연구에서 제시하는 프로파일은 이용자가 입력하는 정적정보나 로그 기반 패턴을 저장하는 수준을 넘어 <그림 1>과 같이 이용자의 연구 단계, 탐색 행동 수준, 의미 기반 관심 영역, 콘텐츠 상호작용 결과 등을 통합적으로 반영하는 다차원 구조로 구성된다. 프로파일은 정적·동적·의미·행동·평판·콘텐츠 기반 요소가 상호 결합된 형태를 갖추고 있으며, 이는 프레임워크의 추천·탐색·상황인지 기능의 핵심 입력으로 활용된다.

4.2.1 정적 정보

정적 정보 영역은 이용자의 ID, 이름, 연락처, 소속기관, 직위/학위, 관심분야 등과 같은 비교적 변화가 적은 속성으로 구성된다. 본 정보는 연구자 특성을 기반으로 한 초기 필터링, 분야 기반 분류, 콘텐츠 적합성 판단의 기준으로 활용된다. 특히 관심분야·직위와 같은 속성은 의미 기반 프로파일과 결합하여 콘텐츠 클러스터링 및 탐색 범위 조절에 활용된다.

4.2.2 동적 탐색 정보

동적 탐색 정보는 이용자의 검색 행동과 콘텐츠 이용 패턴을 포괄하며, 검색어·검색유형·필터 사용 이력, 열람 콘텐츠 유형(논문, 특허, 원문 등), 다운로드·북마크·추천·피드백·설문 등 서비스 상호작용 정보를 포함한다. 또한 행동 시작-종료 시점, 클릭 이벤트 등 시간 기반 로그가 포함되어 탐색의 연속성, 집중도, 반복성 분석이 가능하도록 설계하였다. 이 정보는 행동 수준 판별과 연구 단계 추정의 핵심 근거가 된다.



〈그림 1〉 사용자 프로파일 구조 설계

4.2.3 의미기반 속성

의미 기반 속성은 임베딩 벡터, 주제 클러스터, 지식그래프 기반 개념 등 의미 분석 계층에서 생성되는 결과값을 저장한다. 임베딩 벡터는 이용자가 반복적으로 조회한 문헌들의 의미적 중심을 반영하며, 주제 클러스터는 연구자가 집중하고 있는 주제가 학문적 개념 공간에서 어디에 위치하는지를 나타낸다. 지식그래프 기반 개념은 특정 개념과 연관된 상·하위 개념, 인접 개념을 포함하여 추후 주제 확장·연관 추천 등에 활용된다.

4.2.4 연구단계

연구단계는 탐색 → 문제정의 → 자료수집 → 비교분석 → 정리의 다섯 단계로 구성되며, 이용자의 탐색 행동·콘텐츠 이용 흐름·반복

패턴 등을 기반으로 자동 추론된다. 예를 들어 넓은 범위의 기초 문헌을 다량 조회하면 탐색 단계, 특정 주제의 문헌을 반복 비교하거나 인용 네트워크를 집중적으로 조회하면 비교분석 단계로 판단한다. 연구단계 정보는 적합한 추천 방식과 UI 구조를 선택하는 데 직접적으로 활용된다.

4.2.5 행동 수준

행동 수준은 이용자의 탐색 숙련도를 나타내며, L0~L3까지 네 단계로 분류된다.

L0 : 기초 정보 탐색 중심

L1 : 단일 주제 기반 탐색

L2 : 반복·비교 탐색 중심

L3 : 분석·심층 탐색 중심

행동 수준은 동적 탐색정보와 의미기반 속성

분석 결과를 결합하여 산출되며, 연구 단계와 함께 개인화 탐색 흐름을 결정하는 핵심 프로파일 요소이다.

4.2.6 콘텐츠 정보

콘텐츠 정보 영역은 연구자가 상호작용한 콘텐츠의 주제, 키워드, 콘텐츠 유형(논문, 특허, 보고서 등), 형식(PDF, HTML 등), 메타데이터 속성(초록, 원문, 요약), 언어, 저자·기관 정보, 저널 정보 등을 포함한다. 이는 연구자가 어떤 형태의 자료에 선호를 보였는지, 어떤 주제·기관·저널 중심으로 탐색이 이루어졌는지를 구조적으로 분석할 수 있도록 한다.

4.2.7 평판 정보

평판 정보는 조회수, 북마킹 횟수, 다운로드 횟수, h-index, 저널 IF, 연구기관 성격 등 콘텐츠의 질적 가치를 판단할 수 있는 다차원 지표로 구성된다. 본 정보는 연구자가 선호하는 품질 수준을 파악하는 데 활용되며, 추천 계층에서는 콘텐츠의 의미적 적합성과 질적 가치가 결합된 하이브리드 추천을 수행할 수 있다.

4.3 계층별 컴포넌트 명세 및 프레임워크 아키텍처

본 절에서는 앞선 4.1절에서 도출된 프레임워크의 계층별 구조와 4.2절에서 설계한 프로파일 구조를 기반으로 각 계층을 구성하는 핵심 컴포넌트를 체계적으로 정의하고 그 기능적 역할을 명세한다. 이는 요구사항 분석에서 정립된 설계 요소(SE-1~SE-5)가 실제 프레임워크 구조로 어떻게 구현되는지를 명확히 보여

주기 위한 절차로서, 프레임워크의 구조적 타당성과 구성요소 간 일관성을 검증하는 단계이다. 또한 전체 프레임워크의 흐름을 아키텍처 다이어그램으로 제시함으로써 각 계층과 컴포넌트 간의 상호작용을 직관적으로 파악할 수 있도록 하였다.

4.3.1 계층별 컴포넌트 명세

프레임워크는 입력 계층-프로파일링 계층-의미 분석 계층-추천 계층-서비스 제공 계층으로 구성되며, 각 계층은 고유 기능을 수행하는 복수의 컴포넌트로 구성된다. 계층별 컴포넌트와 그 기능은 <표 2>와 같다. 각 컴포넌트는 제3장에서 도출된 요구사항과 설계 요소(SE-1 연구 단계 모델링, SE-2 행동 수준 모델링, SE-3 의미 기반 분석 모듈, SE-4 프로파일 갱신·추론 메커니즘, SE-5 모듈화·확장성 아키텍처)에 대응되도록 설계하였다.

본 연구는 프레임워크의 실무적 구현 가능성을 확보하기 위해 핵심 모듈 간의 데이터 흐름과 입출력 형식을 명확히 정의하였다. 구체적으로 행동 로그 수집기와 문서 임베딩 엔진은 사용자 아이디, 타임스탬프, 클릭 이벤트 등과 같은 원시 로그 및 문헌 메타데이터를 입력받아 세션 단위 행동 특징 벡터, 시계열 로그 데이터셋, 연구 단계 상태값 등을 출력데이터로 얻을 수 있으며, 이를 바탕으로 추론 엔진과 분류기가 연구 단계 레이블(Label) 및 숙련도 점수(L0~L3)를 산출할 수 있다. 이러한 표준화된 데이터 구조는 맥락 인지 추천 엔진이 사용자 상황에 최적화된 문헌 목록을 제공하는 기반이 되며, 향후 시스템 구현 시 명확한 개발 가이드라인으로 활용될 수 있다.

〈표 2〉 이용자 프로파일링 기반 프레임워크 컴포넌트 정의서

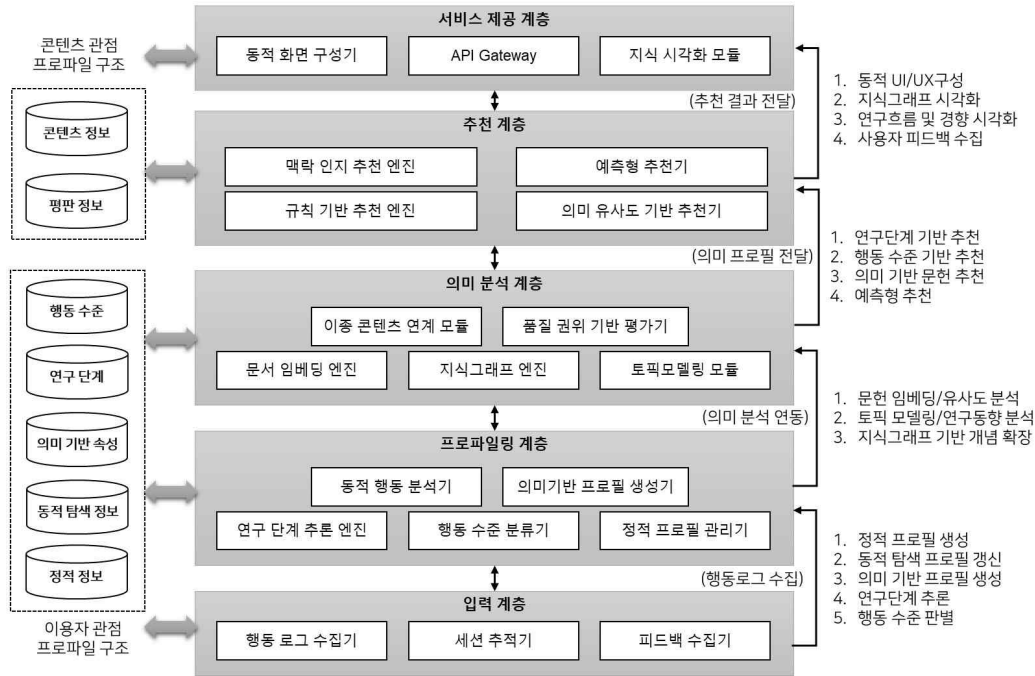
계층	설계요소	컴포넌트 이름	주요 기능
입력 계층	SE-4	행동 로그 수집기	검색 질의, 클릭, 조회, 스크롤, 체류시간 등 미세 행동 데이터 수집
	SE-4	세션 추적기	이용자의 탐색 세션과 행동 흐름을 구조적으로 기록
	SE-4	피드백 수집기	원문저장, 즐겨찾기, 공유, 비교하기 등 명시적 피드백 수집
프로파일링 계층	SE-1	연구 단계 추론 엔진	연구주제 탐색, 아이디어 발굴, 문제정의, 수집, 분석, 정리 단계 자동 판별
	SE-2	행동 수준 분류기	L1~L3 행동수준 판별(탐색량, 반복 탐색, 비교 행위 등 분석)
	SE-4	정적 프로필 관리기	소속, 분야, 직급, 연구자 역할 등 기본 프로필 저장
	SE-4	동적 행동 분석기	검색·열람 패턴 기반 관심도 변화 분석 및 탐지
	SE-3 & SE-4	의미 기반 프로필 생성기	임베딩·지식그래프 기반 주제 클러스터/개념 지도의 개인화 프로필 생성
의미 분석 계층	SE-3	문서 임베딩 엔진	문헌·특허·보고서 임베딩 벡터 생성하여 의미적 유사도 측정
	SE-3	지식그래프 엔진	개념 간 관계, 연구자 네트워크, 기술 흐름 구조 분석
	SE-3	토픽 모델링 모듈	대규모 문헌 집합의 주제 분포·연구 동향 추출
	SE-3 & SE-5	의미 기반 이중 콘텐츠 연계 모듈	논문-특허-보고서 간 의미 기반 교차연계 정보 생성
	SE-3	품질·권위 기반 평가기	인용도, 저자 영향력, 이용량 등 콘텐츠 평판 점수 산정
추천 계층	SE-1	규칙 기반 추천 엔진	기초 검색 및 초기 탐색 단계에 대한 규칙 기반 추천
	SE-3	의미 유사도 기반 추천기	임베딩 및 KG 기반 연관 문헌 추천
	SE-1 & SE-2	맥락 인지 추천 엔진	연구 단계·행동 수준 기반 차별화 추천(기초→고급)
	SE-4	예측형 추천기	다음 탐색 단계에서 필요할 정보 자동 예측
서비스 제공 계층	SE-1 & SE-2	동적 화면 구성기	연구 단계/행동수준에 따른 화면 구성 자동 변경
	SE-5	API Gateway	내부 모듈 및 외부 시스템 간 연동 관리
	SE-3	지식 시각화 모듈	KG 기반 개념지도·연구자 네트워크·주제 흐름 시각화

4.3.2 프레임워크 아키텍처 및 프로파일 갱신 절차

〈그림 2〉는 본 연구에서 제안하는 이용자 프로파일링 기반 학술정보 서비스 프레임워크의 전체 아키텍처와 프로파일 생성·갱신 절차를 통합적으로 제시한 것이다. 본 프레임워크는 입력 계층-프로파일링 계층-의미 분석 계층-추천 계층-서비스 제공 계층으로 이어지는 계층적 구조를 기반으로 하며, 각 계층은 상호 독립적인 기능을 수행하면서도 이용자의 탐색 행위를 중심으로 유기적으로 연결된다. 특히 〈그

림 2〉는 이용자의 행위 데이터가 어떻게 수집되고, 프로파일로 변환되며, 의미 분석과 추천을 거쳐 서비스로 전달되는지를 단계적으로 시각화하여 보여준다.

입력 계층은 행동 로그 수집기, 세션 추적기, 피드백 수집기 등으로 구성되며, 이용자의 검색 질의, 클릭 흐름, 문헌 열람, 다운로드, 북마크 등 다양한 상호작용 데이터를 실시간으로 수집한다. 이 계층은 프로파일 생성의 출발점으로서 이후 분석을 위한 최소한의 피드백 정보를 체계적으로 제공하는 역할을 수행한다.



<그림 2> 프레임워크 아키텍처 및 프로파일 흐름도

프로파일링 계층은 입력 계층에서 전달된 행동 데이터를 기반으로 정적 정보, 동적 탐색 정보, 의미 기반 속성, 연구 단계, 행동 수준을 통합하여 사용자 프로파일을 구성한다. 정적 프로필 관리기는 이용자의 소속, 연구 분야, 관심 영역과 같은 변하지 않는 속성을 관리하며, 동적 행동 분석기는 검색 패턴·열람 흐름·반복 탐색 등의 행동 신호를 분석하여 이용자의 탐색 방향 변화를 반영한다. 연구 단계 추론 엔진과 행동 수준 분류기는 이용자의 연구 상황을 탐색-문제정의-자료수집-비교분석-정리 등의 단계로 분류하고, 탐색의 복잡성과 목적성을 기반으로 L0~L3의 행동 수준을 결정한다. 의미 기반 프로필 생성기는 임베딩, 주제 클러스터, 지식그래프 기반 개념 등을 활용하여 이용자의 의미적 관심 공간을 구축한다.

의미 분석 계층은 프로파일링 계층에서 생성된 정보와 콘텐츠 정보를 통합하여 문헌 간의 의미 구조를 해석한다. 문서 임베딩 엔진은 문헌 간 의미적 유사도를 계산하고, 지식그래프 엔진은 개념 간 구조적 관계를 도출하며, 토픽모델링 모듈은 문헌 집합에서 주제 흐름을 추출한다. 이중 콘텐츠 연계 모듈은 논문·특허·보고서 등 서로 다른 형태의 자료 간 의미적 연계를 파악하고, 품질·권위 기반 평가기는 인용지표, 저자 영향력, 활용도 등을 결합하여 콘텐츠의 신뢰도를 평가한다.

추천 계층에서는 앞선 의미 분석 결과와 이용자 프로파일이 결합되어 실제 탐색 상황에 최적화된 추천이 이루어진다. 규칙 기반 추천 엔진은 초기 단계 이용자를 위한 기본적인 추천을 제공하며, 의미 유사도 기반 추천기는 의미 공

간에서 가까운 문헌을 제시한다. 맥락 인지 추천 엔진은 연구 단계와 행동 수준에 따라 추천의 깊이와 범위를 조정하고, 예측형 추천기는 이용자의 탐색 흐름을 분석하여 다음에 필요할 가능성이 높은 정보를 선제적으로 제시한다.

서비스 제공 계층은 API Gateway, 지식 시각화 모듈, 동적 화면 구성기를 통해 이용자에게 탐색 결과를 전달한다. 이 계층은 이용자의 연구 단계 및 행동 수준에 따라 UI 구성을 변경하고, 지식그래프·연구자 네트워크·주제 흐름 등의 시각화 기능을 제공하여 학술정보의 구조적 의미를 직관적으로 이해할 수 있도록 지원한다. 또한 이용자의 피드백은 다시 입력 계층으로 반환되어 프로파일이 지속적으로 갱신되는 순환 구조를 형성한다.

이와 같이 제안된 프레임워크는 입력-프로파일-의미 분석-추천-서비스로 이어지는 일련의 데이터 처리 흐름을 통해 이용자의 탐색 맥락을 반영하고, 연구 과정 단계와 행동 수준에 적응적으로 반응하는 구조를 갖춘다. 이는 기존의 키워드 기반 정적 검색 중심 서비스를 넘어, 연구자 중심의 지능형 학술정보 서비스를 실현하기 위한 핵심 기반을 제공한다.

5. 프레임워크 타당성 검증

본 장에서는 제안된 이용자 프로파일링 기반 학술정보 서비스 프레임워크 설계가 이론적 배경과 요구사항을 논리적으로 충족하는지를 분석하는 구조적 적합성 검증과 해당 분야 전문가 그룹을 대상으로 한 전문가 평가를 수행하였다.

5.1 선행연구 및 요구사항 기반 구조적 적합성 검증

프레임워크의 설계가 기존 연구의 한계를 극복하고 정보학적 이론 토대 위에 구축되었음을 입증하기 위해 선행연구-요구사항-설계요소 간의 추적성 분석을 수행하였다. 첫째, 이용자 연구 맥락의 시스템적 구현이다. Kuhlthau(1991)와 Vakkari(2003)는 정보 탐색이 시간의 흐름에 따라 단계적으로 변화함을 규명하였으며, 이는 ‘연구 단계 식별’ 요구사항을 거쳐 이용자 로그 기반으로 5단계를 판별하는 ‘연구 단계 추론 엔진’으로 구체화되었다. 또한 Purificato et al.(2024)과 Wilson(1996)은 이용자의 숙련도에 따라 정보 행태에 차이가 있음을 지적하였는데, 본 연구는 이를 ‘행동 수준 판별’ 요구사항으로 도출하고 복잡도 기반으로 수준을 분류하는 ‘행동 수준 분류기’로 설계하여 이론적 모델을 충실히 반영하였다.

둘째, 콘텐츠의 의미적 연계와 질적 가치의 반영이다. Di Scala et al.(2024), Zhang et al.(2024), Zhao et al.(2024)은 지식그래프의 의미적 풍부성이 추천 시스템의 성능을 좌우하며 이기종 자원의 연결이 필수적임을 입증하였다. 이는 ‘콘텐츠 의미 연계’ 요구사항의 근거가 되었으며, ‘의미 기반 분석 모듈’을 통해 구현되었다. 아울러 Sabiri et al.(2025)와 Zhang et al.(2023)이 강조한 학술정보의 권위와 신뢰성 중요성은 ‘질적 가치 반영’ 요구사항으로 연결되어, 인용도 등 다차원 지표를 활용하는 ‘품질·권위 기반 평가기’ 설계로 이어졌다.

셋째, 서비스의 동적 적응성과 선제적 지원이다. Belkin(1980)의 ASK 이론과 Hussain et

al.(2018)의 적응형 인터페이스 연구는 이용자의 불확실한 요구를 지원하기 위한 '동적 UI 구성'의 이론적 토대가 되었으며, 이는 상황에 따라 가이드형 UI를 제공하는 '동적 화면 구성기'로 구체화되었다. 또한 Roy와 Dutta(2022), Liu et al.(2023), Jung et al.(2023) 등이 지적한 관심사 변화 대응의 필요성은 '로그 기반 분석'과 '선제적 추천' 요구사항을 도출하였고, 이는 세션 간 변화를 추적하고 예측하는 '동적 행동 분석기'로 구현되었다.

넷째, 시스템의 기술적 확장성과 윤리적 고려이다. Dessi(2021)의 자동화된 지식그래프 생성 연구와 Wilkinson et al.(2016)의 FAIR 원

칙은 데이터의 연동성과 호환성 문제를 해결하기 위한 '외부 시스템 연동', '메타데이터 호환' 요구사항의 직접적 근거가 되며, 이는 '모듈화·확장성 아키텍처'로 설계되었다. 마지막으로 본 연구는 이용자의 행동 정보와 맥락적 상황에 기반한 학술정보 서비스를 고려하므로 프라이버시 보호 및 윤리가 고려되어야 한다. 이는 '개인정보 및 보안' 요구사항으로 반영되어, 안전한 데이터 처리를 위한 '보안 모듈'로 구현됨으로써 프레임워크의 윤리적 타당성을 확보하였다.

상세한 추적 분석 결과는 <표 3>과 같으며, 분석 결과 본 프레임워크의 핵심 설계 요소들은 선행연구에서 도출된 총 10개의 요구사항을

<표 3> 선행연구 및 요구사항 기반 프레임워크 설계 추적 매트릭스

이론적 배경 및 선행 연구	핵심 요구 사항	프레임워크 설계 요소	검증 결과
Kuhlthau(1991), Vakkari(2003) 정보 탐색 과정의 단계적 변화	[REQ-U-01] 연구 단계 식별 진행 상황(탐색/분석 등) 파악	[SE-1] 연구 단계 모델링 5단계 자동 추론 알고리즘 탑재	충족
Purificato et al.(2024), Wilson(1996) 이용자 행동 수준차이	[REQ-U-02] 행동 수준 판별 초급/고급 이용자 구분 지원	[SE-2] 행동 수준 분류기 복잡도 기반 수준(L0~L3) 분류	충족
Di Scala et al.(2024), Zhao et al.(2024), Zhang et al.(2024) 지식그래프 풍부성과 의미적 연결	[REQ-C-01] 콘텐츠 의미 연계 이종 자원 간 의미적 연결	[SE-3] 의미 기반 분석 모듈 지식그래프/임베딩 기반 연계	충족
Sabiri et al.(2025), Zhang et al.(2023) 학술정보의 권위와 신뢰성	[REQ-C-02] 질적 가치 반영 평판/인용 기반 필터링	[SE-3] 품질·권위 기반 평가기 h-index 등 다차원 지표 활용	충족
Belkin(1980), Hussain et al.(2018) 불확실한 요구(ASK)와 적응형 UI	[REQ-S-01] 동적 UI 구성 상황에 맞는 가이드형 UI 제공	[SE-1 & SE-2] 동적 화면 구성기 단계/수준별 UI 자동 변경	충족
Roy와 Dutta(2022), Liu et al.(2023), Jung et al.(2023) 관심사 변화대응	[REQ-S-02] 선제적 추천 [REQ-T-01] 로그 기반 분석	[SE-4] 동적 행동 분석기 세션 간 변화 추적 및 예측	충족
Dessi et al.(2021), Wilkinson et al.(2016) 자동화된 KG 생성 및 FAIR 원칙	[REQ-T-02] 외부 시스템 연동 [REQ-T-03] 메타데이터 호환	[SE-5] 모듈화·확장성 아키텍처 API Gateway 및 표준 준수	충족
이용자 행동 정보 기반 분석 프라이버시 보호 및 윤리	[REQ-T-04] 개인정보 및 보안 안전한 데이터 수집 처리	[SE-4 & SE-5] 보안 모듈 비식별화 및 접근 제어 적용	충족

빠짐없이 충족하고 있음을 확인하였다.

5.2 전문가 평가를 통한 타당성 검증

본 연구에서 제안한 프레임워크가 학술적 이론에 부합하는지, 실제 대규모 학술정보 플랫폼 환경에서 운영 가능한지, 그리고 최신 AI 기술로 구현 가능한지를 검증하기 위해 전문가 평가를 실시하였다.

5.2.1 평가 개요 및 전문가 패널 구성

전문가 패널은 프레임워크의 핵심 구성 요소(SE-1~SE-5)의 특성을 고려하여 이론(Theory), 서비스(Service), 기술(Technology) 분야를 각각 대표할 수 있는 총 3인으로 구성하였다. 문헌정보학 전문가 1인은 연구 단계 모델링(SE-1)과 행동 수준 모델(SE-2)의 이론적 적합성을 평가하기 위해 선정하였다. 학술정보 플랫폼 개발 및 운영을 담당하는 책임연구원 1인은 이기종 학술정보(논문·특허·보고서)의 통합 운영 경험 및 개인화 서비스 구현을 바탕으로 실무적 적용 가능성을 검증하였다. 셋째, LLM 모델 개발 연구를 수행 중인 선임연구원 1인은 의미 기반 분석(SE-3), 동적 로그 분석(SE-4)의 기술적 실현 가능성을 평가하였다. 전문가 패널의

구성은 <표 4>와 같다.

전문가 평가는 사전 검토, 구조화된 설문, 심층 인터뷰로 구성된 3단계 혼합 방법을 통해 수행되었다. 우선, 전문가들에게 프레임워크 명세서 및 아키텍처, 프로파일 명세서 등 관련 문서를 제공하여 충분한 사전 검토 시간을 부여하였다. 그리고, 프레임워크의 (1) 이론적 타당성, (2) 구성 요소 간 구조적 정합성, (3) 기술적 실현 가능성, (4) 실무적 유용성의 네 가지 영역에 대하여 5점 리커트 척도로 정량 평가를 실시하였다. 이후, 각 전문가와 60분의 심층 인터뷰를 수행하여 정량 평가에서 드러나지 않은 세부적 의견과 개선사항을 수집하였다.

5.2.2 평가 결과 논의

전문가 평가 결과, 본 연구에서 제안한 프레임워크는 이론적 정합성, 실무적 적용 가능성, 기술적 실현 가능성 측면에서 <표 5>와 같이 평균 4.6점(5점 만점)의 평가를 받았다.

우선 이론적 타당성(평균 4.7점)에서는 연구 단계 모델과 행동 수준 모델이 Kuhlthau, Vakkari, Wilson 등 주요 정보행태 이론에 기반하여 안정적으로 구성되어 있다는 점에서, 구성 요소 간 구조적 정합성(평균 4.5점)에서는 프레임워크가 SE-1(연구 단계), SE-2(행동 수준), SE-3

<표 4> 타당성 검증을 위한 전문가 패널 구성

ID	전문 분야	경력 및 직위	주요 검증 역할
A	문헌정보학	문헌정보학과 교수 (J대학, 경력 21년)	연구 단계·행동 수준 모델의 이론적 타당성 검증
B	학술정보 플랫폼	학술정보 플랫폼 개발·운영 담당 책임연구원 (K연구원, 경력 22년)	이기종 데이터 연계 및 플랫폼 적용 가능성 검증
C	AI 및 데이터 분석	AI 모델 개발 담당 선임연구원 (K연구원, 경력 7년)	의미 기반 분석 및 동적 로그 분석 모듈의 구현 가능성 검증

〈표 5〉 전문가 평가 영역별 결과 및 주요 의견 요약

평가 영역	평가 항목 요약	점수	주요의견
이론적 타당성	- 연구 단계(5단계) 구분 적절성 - 행동 수준(L0~L3) 분류 타당성	4.8	- 정보행태 이론(ISP)을 프레임워크 설계에 반영한 것에 의미 - 맥락 반영: 정적 프로파일의 한계를 고려함
구조의 적합성	- 계층 간 데이터 흐름 논리성 - 피드백 루프(순환) 구조 적절성	4.5	- 입력-프로파일 분석-의미 연결- 추천의 체계적 흐름을 가짐 - 정적/동적 프로파일 결합을 통한 서비스 품질 향상 가능
기술적 실현성	- 의미 기반 분석(LLM/KG) 구현성 - 동적 갱신 메커니즘 가용성	4.3	- sLLM 기반, 지식그래프 중심의 구현 전략 - 학습데이터 확보 문제 등 현실적 문제 존재
실무적 유용성	- 기존 검색 대비 차별적 효용 - 플랫폼(ScienceON 등) 적용성	4.7	- 이기종 학술정보의 의미적 통합 가능성 - 연구 단계 기반 추천을 통한 연구 효율 향상
종합	전체 평균	4.6	

(의미 기반 분석), SE-4(로그 기반 프로파일링), SE-5(모듈화 아키텍처)로 구성된 다층적 구조임에도 불구하고 전체적인 논리적 일관성과 모듈 간 연계성이 잘 유지되고 있다는 점에서 긍정적으로 평가되었다. 다만 플랫폼 전문가는 실제 대규모 학술정보 서비스에서 모듈이 동시에 운영될 경우 시스템 복잡도가 증가할 수 있다는 현실적 우려를 제시하였다.

기술적 실현 가능성은 평균 4.3점으로 다른 평가 영역 대비 상대적으로 낮은 점수를 기록하였다. 전문가들은 LLM, 임베딩 모델, 벡터DB, 지식그래프를 활용하는 의미 기반 분석 구조가 기술적으로 적합하다는 점은 인정했으나, 실제 도입 과정에서는 컴퓨팅 자원 소모, 대규모 학습 데이터 확보의 어려움 등 기술적 제약이 존재한다는 점을 지적하였다. 특히 AI 전문가는 SE-4 동적 프로파일링 구조가 최신 세션 기반 추천 시스템과 높은 정합성을 갖고 있음에도 불구하고, 일반적인 포털서비스와 달리 학술정보 서비스에서의 이용자의 맥락(연구 단계·행동 수준)을 정확히 추론하기 위해 필요한 레이블링 데이터 확보가 현실적으로 쉽지 않다는 점을 기술적 한계로 지적하였다.

실무적 유용성(평균 4.7점)은 네 영역 중 가

장 높은 점수를 기록하였다. 전문가들은 이기종 학술정보(논문·특허·보고서)의 파편화를 해결하는 지식그래프 기반 의미 연계(SE-3)와 연구 단계 기반 추천이 이용자에게 원하는 가치를 제공할 수 있다고 하였다. 학술정보 플랫폼 운영 전문가는 이러한 구조가 현재의 학술정보 제공 방식에서 발생하는 이용자의 탐색 효율을 크게 향상시킬 수 있다고 분석하였다.

종합적으로, 제안된 프레임워크는 이론적 기반, 기술적 구성요소, 사용자 중심 서비스 설계, 실무 적용 가능성 등 다양한 차원에서 균형 있게 평가되었다. 다만 전문가들은 행동 수준과 연구 단계를 로그 분석 및 콘텐츠 활용 정보를 통해 추론해야 한다는 점에서 레이블링된 학습 데이터 구축이 실제 구현 과정의 핵심 고려사항이라고 공통적으로 지적하였다.

6. 결론

본 연구는 연구자의 탐색 맥락을 정교하게 반영하기 위해 연구 단계와 행동 수준이라는 개념을 중심으로 이용자 프로파일링 기반의 학술정보 서비스 프레임워크를 제안하였다. 기존

학술정보 서비스가 정적 프로필이나 단순 로그 분석에 의존함으로써 연구자의 변화하는 요구를 반영하지 못했던 문제를 보완하기 위해, 본 연구는 정적 정보, 동적 탐색 정보, 의미 기반 속성, 연구 단계, 행동 수준을 통합한 다차원 프로파일 구조를 설계하였다. 더불어 임베딩· 지식그래프 기반 의미 분석을 통해 논문·특허·보고서 등 이기종 학술정보를 맥락적으로 연결함으로써 기존의 키워드 중심 검색이 제공하지 못했던 맥락 기반 탐색을 구현할 수 있는 기반을 제시하였다.

프레임워크 타당성 검증을 위해 구조적 추적 평가와 전문가 평가를 병행한 결과, 프레임워크는 요구사항 정합성, 의미 기반 분석의 적절성, 연구 단계·행동 수준 모델링의 학술적 근거 등에서 높은 평가를 받았다. 전문가들은 특히 연구자의 상태를 반영한 동적 UI·추천 조정 기능과 지식그래프 기반 연결 구조가 실제 학술정보 서비스 품질 향상에 크게 기여할 수 있다는 점을 긍정적으로 평가하였다. 기술적 실현 가능성은 상대적으로 낮은 점수(4.3점)를 받았으나, 이는 연구 단계·행동 수준 판별을 위한 레이블 데이터 구축의 선형적 어려움에 기인한 것으로 분석된다. 본 연구는 이를 보완하기 위해 규칙

기반 초기 라벨링과 능동 학습 전략을 제안하였으며, 이러한 단계적 접근은 데이터 부족 상황에서도 모델 개선을 가능하게 하는 방법으로 판단된다.

다만 본 연구는 전문가 검증 중심으로 이루어져 실제 서비스 환경에서의 대규모 이용자 기반 실증 검증이 이루어지지 못했다는 점, 그리고 제안된 프레임워크의 구성 요소가 복잡적이기 때문에 실제 플랫폼에 적용될 때 발생할 수 있는 시스템 부하나 운영 비용을 직접적으로 평가하지 못했다는 점에서 방법론적 한계를 가진다. 규칙 기반 초기 레이블링과 능동 학습 전략을 통해 기술적 실현 가능성을 제시하였으나, 복잡한 실제 데이터 환경에서 규칙 기반 추론 정밀도의 한계를 실증적으로 검증하지 못했다는 점은 연구의 제한점이다. 또한, 타당성 검증이 각 분야 전문가 3인에 의해 수행되어 초기 타당성 검토 수준에 머물렀으며, 패널 수의 제한으로 인해 평가 결과를 일반화하기에는 한계가 있다.

이러한 한계를 보완하기 위해 향후 연구에서 프로토타입 구현을 통해 실제 사용자 로그를 수집하고, 이를 기반으로 맥락 추론의 정확도와 시스템 성능을 정량적으로 평가하는 실증 연구가 뒤따라야 할 것이다.

참 고 문 헌

- 감미아, 이지연 (2023). 다차원 메타데이터 공간을 활용한 학술 문헌 추천기법 연구. 정보관리학회지, 40(1), 121-148. <https://doi.org/10.3743/KOSIM.2023.40.1.121>
- 권나현, 이정연, 정은경 (2012). 과학기술분야 R&D 전주기 연구: 국내 생명 및 나노과학기술연구자를 중심으로. 한국문헌정보학회지, 46(3), 103-131. <https://doi.org/10.4275/KSLIS.2012.46.3.103>

- 나은엽 (2023). 자연과학분야 연구자들의 연구계획서 작성을 위한 정보추구행동의 탐색적 연구. *한국비블리아학회지*, 34(1), 53-74. <https://doi.org/10.14699/kbiblia.2023.34.1.053>
- 박정훈 (2025). LLM 기반 대화형 AI 학술정보서비스 현황 분석 및 서비스 전략 연구. *한국도서관·정보학회지*, 56(1), 85-110. <https://doi.org/10.3743/KOSIM.2025.42.1.029>
- 이정연, 윤소영, 박지영, 황혜경, 김재훈, 이해림 (2019). 연구콘텐츠라이프 사이클 모델 개발: 한국과학기술정보연구원 콘텐츠 큐레이션 센터를 중심으로. *정보관리학회지*, 36(3), 203-228. <https://doi.org/10.3743/KOSIM.2019.36.3.203>
- Belkin, N. J. (1980). Anomalous states of knowledge as a basis for information retrieval. *Canadian Journal of Information Science*, 5, 133-143.
- Dessi, D., Osborne, F., Recupero, D. R., Buscaldi, D., Motta, E., & Sack, H. (2021). AI-KG: An automatically generated knowledge graph of artificial intelligence. *Scientific Data*, 8, 185.
- Di Scala, D. L., Wang, X., & Brewster, C. (2024, October 18). The effect of semantic knowledge graph richness on recommender systems. *KaRS 2024: Knowledge-aware and Conversational Recommender Systems Workshop*, Bari, Italy.
- Gusenbauer, M. & Haddaway, N. R. (2020). Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods*, 11(2), 181-217. <https://doi.org/10.1002/jrsm.1378>
- Hussain, J., Ul Hassan, A., Muhammad Bilal, H. S., Ali, R., Afzal, M., Hussain, S., Bang, J., Banos, O., & Lee, S. (2018). Model-based adaptive user interface based on context and user experience evaluation. *Journal on Multimodal User Interfaces*, 12(1), 1-16. <https://doi.org/10.1007/s12193-018-0258-2>
- Jeong, C., Jang, S., Park, E., & Choi, S. (2020). A context-aware citation recommendation model with BERT and graph convolutional networks. *Scientometrics*, 124, 1907-1922. <https://doi.org/10.1007/s11192-020-03561-y>
- Jung, H., Park, H., & Lee, K. (2023). Enhancing recommender systems with semantic user profiling through frequent subgraph mining on knowledge graphs. *Applied Sciences*, 13(18), Article 10041. <https://doi.org/10.3390/app131810041>
- Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 42(5), 361-371. [https://doi.org/10.1002/\(SICI\)1097-4571\(199106\)42:5%3C361::AID-ASI6%3E3.0.CO;2-%23](https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C361::AID-ASI6%3E3.0.CO;2-%23)
- Liu, X., Zhu, Y., & Wu, X. (2023). Joint user profiling with hierarchical attention networks.

- Frontiers of Computer Science, 17, Article 173608.
<https://doi.org/10.1007/s11704-022-1437-6>
- Purificato, E., Boratto, L., & Luca, E. W. (2024). User modeling and user profiling: A Comprehensive Survey. <https://doi.org/10.48550/arXiv.2402.09660>
- Roy, D. & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9, 59. <https://doi.org/10.1186/s40537-022-00592-5>
- Sabiri, B., Khtira, A., El Asri, B., & Rhanoui, M. (2025). Hybrid quality-based recommender systems: A systematic literature review. *Journal of Imaging*, 11(1), 12. <https://doi.org/10.3390/jimaging11010012>
- Scherbakov, D., Hubig, N., Jansari, V., Bakumenko, A., & Lenert, L. A. (2025). The emergence of large language models as tools in literature reviews: A large language model-assisted systematic review. *Journal of the American Medical Informatics Association*, 32(6), 1071-1086. <https://doi.org/10.1093/jamia/ocaf063>
- Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College & Research Libraries*, 29(3), 178-194.
- Vakkari, P. (2001). A theory of the task-based information retrieval process. *Journal of Documentation*, 57(1), 44-60. <https://doi.org/10.1108/EUM0000000007075>
- Vakkari, P. (2003). Task-based information searching. *Annual Review of Information Science and Technology*, 37(1), 413-464. <https://doi.org/10.1002/aris.1440370110>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wilson, T. D. (1996). Information behaviour: An interdisciplinary perspective. *Information Processing & Management*, 33(4), 551-572. [https://doi.org/10.1016/S0306-4573\(97\)00028-9](https://doi.org/10.1016/S0306-4573(97)00028-9)
- Zhang, J. C., Zain, A. M., Zhou, K. Q., Chen, X., & Zhang, R. M. (2024). A review of recommender systems based on knowledge graph embedding. *Expert Systems with Applications*, 250, 123876. <https://doi.org/10.1016/j.eswa.2024.123876>
- Zhang, Z., Patra, B. G., Yaseen, A., Zhu, J., Sabharwal, R., Roberts, K., Cao, T., & Wu, H. (2023). Scholarly recommendation systems: A literature survey. *Knowledge and Information Systems*, 65(11), 4433-4478. <https://doi.org/10.1007/s10115-023-01901-x>
- Zhao, Z., Fan, W., Li, J., Liu, Y., Mei, X., Wang, Y., Wen, Z., Wang, F., Zhao, X., Tang, J.,

& Li, Q. (2024). Recommender systems in the era of large language models (LLMs). *IEEE Transactions on Knowledge and Data Engineering*, 36(11), 6889-6907. <https://doi.org/10.1109/TKDE.2024.3392335>

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Kam, Miah & Lee, Jeeyeon (2023). A study on the method of scholarly paper recommendation using multidimensional metadata space. *Journal of the Korean Society for Information Management*, 40(1), 121-148. <https://doi.org/10.3743/KOSIM.2023.40.1.121>
- Kwon, Na-Hyun, Lee, Jung-Yeon, & Jung, Eun-Kyung (2012). Full cycle research in science and technology R&D: Focusing on domestic life and nanoscience technology researchers. *Journal of the Korean Society for Library and Information Science*, 46(3), 103-131. <https://doi.org/10.4275/KSLIS.2012.46.3.103>
- Lee, Jung-Yeon, Yoon, So-Young, Park, Ji-Young, Hwang, Hye-Kyung, Kim, Jae-Hoon, & Lee, Hye-Rim (2019). Development of a research content lifecycle model: Focused on the content curation center of the Korea institute of science and technology information. *Journal of the Korean Society for Information Management*, 36(3), 203-228. <https://doi.org/10.3743/KOSIM.2019.36.3.203>
- Na, Eun-Yeop (2023). An exploratory study on the information seeking behavior for research proposal writing among natural science researchers. *Journal of the Korean Biblia Society for Library and Information Science*, 34(1), 53-74. <https://doi.org/10.14699/kbiblia.2023.34.1.053>
- Park, Jung-Hun (2025). Analysis of current trends and service strategies for LLM-based conversational AI scholarly information services. *Journal of the Korean Library and Information Science Society*, 56(1), 85-110. <https://doi.org/10.3743/KOSIM.2025.42.1.029>