

기록정보서비스 지능화를 위한 논항구조 기반 챗봇 PoC 구현 및 이중 평가를 통한 검증

Developing an Argument-Structure-Based Chatbot PoC to Enhance the Intelligence of Archival Information Services and Validating through a Two-Phase Evaluation

최유리(You-Ree Choi)¹, 오효정(Hyo-Jung Oh)²

E-mail: choiulee@jbnu.ac.kr, ohj@jbnu.ac.kr

¹ 제 1저자 전북대학교 기록관리학과 석사과정

² 교신저자 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 공동연구원



논문접수 2026-01-12
최초심사 2026-01-15
게재확정 2026-02-20

ORCID

You-Ree Choi
https://orcid.org/0009-0006-9522-2504

Hyo-Jung Oh
https://orcid.org/0000-0001-8067-2832

© 한국기록관리학회

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.

*이 논문은 2025년 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2021R111A3047435).

https://jksarm.koar.kr

초 록

생성형 AI의 대중화로 이용자들이 자연어 기반 대화형 인터페이스에 익숙해지면서 기록정보서비스 분야에서도 단순 키워드 중심의 단발성 검색을 넘어, 원하는 정보를 쉽게 획득할 수 있는 지능형 인터페이스로서 챗봇 서비스가 효과적인 대안으로 주목받고 있다. 본 연구는 사전 연구에서 수립된 질문 유형 분류 체계와 응답 논리 구조를 기반으로, 지능형 기록정보서비스 챗봇 PoC(Proof-of-Concept)를 구현하고 실제 기록관리 기관에 적용 가능성을 검증하는 것을 목적으로 한다. 구현 단계에서는 다섯 가지 응답 안정화 규칙을 적용하였으며, 실제 기록정보서비스에 적용 가능성 검증을 위해 50건의 평가셋을 활용한 정량적 평가와 기록연구사 3인을 대상으로 한 정성적 평가(FGI)를 병행하였다. 정량 평가 결과, 질문 의도 분류 정확도는 96%로 나타났고, 문장 구조 적절성은 84%로 파악되었다. 정성 평가에서는 개발된 챗봇이 대민 행정기관에서 반복적으로 발생하는 질의 대응에 적합하다는 평가를 받았으며, 정보공개청구 전 단계 안내 도구로서의 활용 가능성에 대한 의견도 수렴되었다. 나아가 이중 평가 결과를 종합하여 질문 유형 정교화, 언어 규칙 보완, 복문 구조 고도화, 업무·데이터 연계 확장의 네 가지 개선 방향을 도출하였다. 본 연구는 질문 유형 기반 응답 구조를 실제 시스템 환경에서 구현하고, 정량·정성적 이중 평가를 통해 그 작동성과 성능을 검증함으로써 기록정보서비스 챗봇 개발을 위한 실증적 모델을 제시하였다는 데 의의가 있다.

ABSTRACT

As generative AI has popularized natural language-based conversational interfaces, chatbot services have gained attention as an intelligent alternative to single-turn, keyword-based searches in archival information services. This study implements an intelligent archival information service chatbot Proof of Concept (PoC) based on a previously established question-type classification framework and examines its applicability in archival institutions. The chatbot was developed with five response stabilization rules. Performance was evaluated through quantitative assessment using 50 test queries and qualitative focus group interviews (FGIs) with 3 records management professionals. Results showed 96% accuracy in question-type classification and 84% in sentence structure appropriateness. Qualitative findings indicated suitability for handling repetitive inquiries in public institutions and potential use as a preliminary guidance tool prior to filing an information disclosure request. Based on dual evaluations, four improvement directions were identified: question-type refinement, linguistic rule enhancement, complex sentence handling advancement, and task-data integration expansion. This study presents an empirical model for archival chatbot development by implementing a question-type-based response structure in an actual system environment and validating its functionality and performance through dual quantitative and qualitative evaluations.

Keywords: 기록정보서비스 지능화, 챗봇, 질문 유형, 논항구조분석, PoC
Intelligent Archival Information Service, Chatbot, Question Type, Argument Structure
Analysis, PoC

1. 서론

1.1 연구배경 및 목적

디지털 기술의 발전으로 정보 접근 환경은 크게 개선되었으나, 방대한 정보가 축적된 환경에서 이용자가 목적에 맞는 정보를 탐색하고 이해하는 과정은 점차 복잡해지고 있다. 이러한 환경에서는 단순히 정보를 나열하는 방식만으로 이용자가 원하는 기록에 효과적으로 도달하기 어렵다. 이에 따라 기록정보서비스는 이용자의 정보 요구를 보다 정확히 파악하고 탐색 과정에서 발생하는 어려움을 완화할 수 있도록 기능을 고도화할 필요가 있다. 이러한 정보환경의 변화 속에서 국가기록원은 국민의 기록 접근권을 보장하기 위한 제도적 기반을 마련하며 기록정보서비스를 운영하고 있다. 국가기록원은 이용자의 정보 열람을 지원하기 위한 제도적 근거를 갖추고 있으며, 국민의 다양한 열람 요구에 응답하는 역할을 수행한다. 1969년 국가기록원의 전신인 정부기록보존소 설립 당시에는 시민의 기록 이용이나 접근 지원이 설립 목적에 포함되지 않았으나, 이후 제도 개편을 거치며 열람 기능이 점진적으로 강화되었다. 「정부기록보존소조직제」는 정부의 영구보존 및 준영구기록물을 ‘수집·관리 및 보존하기 위하여’ 설치한다고 명시하였으며, 1983년 직제 개정을 통해 설립 목적에 ‘열람’을 포함하도록 변경되었다(국가기록원, 1969; 국가기록원, 1983). 이후 1999년 공공기록물법 제정을 통해 기록물 열람에 대한 법적 근거가 강화되었고, 비공개 기록물에 대해서도 제한적 열람을 지원하도록 규정함으로써 국가기록원의 기록정보서비스는 현재의 제도적 기반을 형성하게 되었다(설문원, 2022).

그러나 기록물 열람을 중심으로 구성된 기존 기록정보서비스는 정보 요구가 복잡화되고 비정형적 질문이 증가하는 최근의 이용 패턴을 충분히 포괄하기에는 한계를 지닌다. 디지털 전환이 가속화되면서 기록정보서비스를 둘러싼 이용 환경과 정보 접근 방식 역시 변화하고 있으며, 단순한 기록 검색만으로는 해결하기 어려운 유형의 문의가 점차 증가하고 있다. 예를 들어 특정 기록물이 어떤 절차를 거쳐 열람할 수 있는지, 관련 법령에 따라 비공개 대상에 해당하는지, 기록의 생산 주체와 행정적 맥락을 어떻게 해석해야 하는지를 묻는 문의처럼 단순 키워드 검색만으로는 의도를 파악하기 어려운 질문들이 지속적으로 늘고 있다. 이처럼 정보 이용의 즉시성과 복잡성이 강화됨에 따라, 질문에 담긴 의도와 맥락을 실시간으로 파악하여 적절한 정보를 제공하는 서비스의 필요성이 대두된다. 이러한 변화는 기존 기록정보서비스와는 다른 새로운 패러다임을 요구하며, 그 대안으로 지능형 기술 기반의 기록정보서비스가 부상하고 있다. 특히 AI 챗봇은 다양한 질문 유형을 자동으로 분석하고 관련 정보를 즉각 안내함으로써 서비스 접근성을 확장하고 이용 편의성을 향상시킬 수 있는 잠재력을 지닌다. 이는 정부가 AI를 핵심 전략 기술로 규정하고 공공서비스 전반의 디지털 전환을 추진하는 정책 방향(정책브리핑, 2025) 과도 궤를 같이한다.

본 연구는 사전 연구(최유리 외, 2025)에서 수립된 질문 유형 분류 체계와 응답 논리 구조를 기반으로 지능형 기록정보서비스 챗봇을 구현하고, 이를 실제 기록관리기관에 적용하여 그 적용 가능성을 검증하는 데 초점을 둔다. 사전 연구에서는 국가기록원의 실제 질의 데이터를 대상으로 논항구조 기반 의미역 분석을 적용하여 이용자 질의문의 구성 요소를 정밀하게 분석하였으며, 이를 통해 기록관리 도메인에 특화된 여섯 가지 질문 유형과 각 유형별 응답 템플릿을 설계하였다. 본 연구는 이러한 선행연구의 설계안을 실제 기록관리 현장에 적합한 PoC(Proof-of-Concept) 형태로 구현하고, 질문 유형 자동 분류 및 응답 생성 로직이 실제 기록정보서비스 맥락에서 적절하게 작동하는지를 검증하기 위해 정량적·정성적 평가를 병행한다. 이를 통해 질문 유형 기반 응답 구조가 기록관리기관의 업무 특성과 이용자의 정보 요구를 효과적으로 반영할 수 있음을 확인하고자 한다. 궁극적으로 본 연구는 국가기록원과 같은 기록관리기관이 디지털 전환 시대에 요구되는 지능형 서비스로 나아가기 위한 실질적 구현 모델을 제시하고, 향후 기록정보서비스의 접근성과 이용 가능성을 제고하는 데 기여하고자 한다.

1.2 관련 연구

최근 공공 및 기록관리 분야에서는 인공지능 기술을 활용한 기록정보서비스 고도화 방안에 대한 연구가 점차 확대되고 있다. 특히 챗봇과 생성형 AI를 중심으로 기록관리 업무 지원과 이용자 서비스 개선을 위한 다양한 적용 가능성이 논의된다. 강운아와 오효정(2023)은 생성형 AI 기술을 기록관리 실무에 적용하여 전자기록관리의 효율성을 제고하고 기록정보서비스의 만족도를 향상시킬 수 있는 방안을 제안하였다. 해당 연구는 기록물 생산현황 통보, 민원 및 정보공개청구 처리, 전시 콘텐츠 생성 등 기록관리 업무 전반에 생성형 AI를 활용할 수 있는 구체적인 적용 영역을 제시함으로써 기록관리 도메인에서 인공지능 기술의 활용 가능성을 이론적·실무적 차원에서 확장하였다.

이와 함께 기록관에 챗봇을 실제 서비스로 도입하려는 연구도 이루어졌다. 김선욱(2025)은 대학기록관을 대상으로 인공지능 기반 챗봇 도입 방안을 검토하고, 대학기록관 이용대장과 웹사이트 데이터를 분석하여 관리자용 및 이용자용 챗봇을 구현·평가하였다. 해당 연구는 기록관에서 활용 가능한 데이터를 선별하고 이를 기반으로 챗봇 시스템을 단계적으로 설계함으로써, 기록관리 현장에서 챗봇 서비스 구축의 가능성을 구체적으로 제시하였다. 또한 이창희 외(2018)는 기록정보서비스 강화를 목적으로 대학사료실과 대학기록관을 사례로 챗봇 프로토타입을 개발하고, 챗봇 도입을 위한 핵심 절차로 서비스 범위 설정, 요구 분석, 도구 선정, 대화 구성을 제시하였다. 이들 연구는 기록관 환경에서 챗봇이 정보 안내와 질의 대응을 지원할 수 있음을 보여주며, 기록정보서비스에서 챗봇 활용 가능성을 구체화하였다.

한편 공공기관 전반을 대상으로 한 연구에서는 챗봇 기술의 구조적 특성과 확장 방향에 대한 논의도 이루어졌다. 임상빈과 고영우(2017)는 경기도 지방세 상담봇 사례를 중심으로 공공분야 AI 챗봇의 가능성과 과제를 분석하고, 체계적인 AI 기반 지방세 브레인 시스템 구축 방안을 제시하였다. 해당 연구는 공공정보시스템이 주로 저장 중심 구조로 운영되어 자연어 처리 및 법률 상담 측면에서 제약이 있음을 지적하며, 공공분야 AI 챗봇의 확장을 위해서는 지식체계의 정비와 실무자 참여 기반 데이터 재구성이 필요함을 제안하였다.

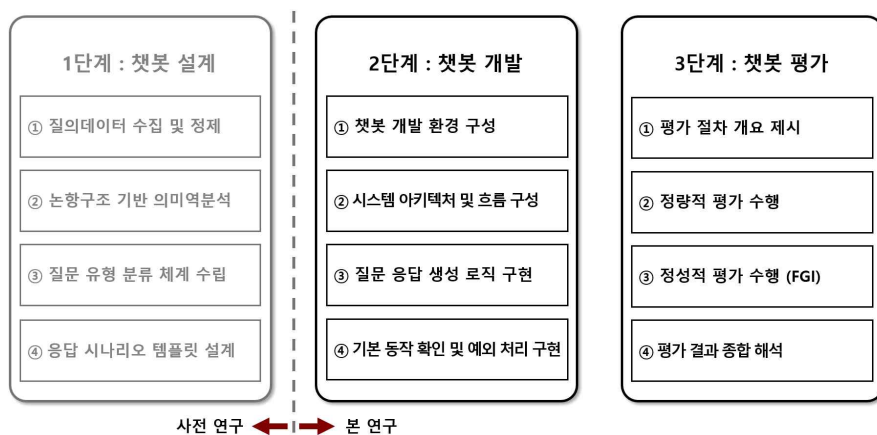
대화형 AI 챗봇 시스템의 평가 방법론과 관련하여, Gupta et al.(2025)는 챗봇 평가 프레임워크를 정확도 기반 평가, 사용자 만족도 평가, 태스크 완료율 평가 등으로 구분하여 제시하였다. 해당 연구는 챗봇의 기술적 성능과 사용자 경험을 균형 있게 평가하기 위해 다층적 접근이 필요함을 강조하였다. 본 연구는 이러한 평가 프레임워크를 참고하여, 프로토타입 단계 챗봇의 특성을 고려한 정량적 평가와 정성적 평가를 병행하는 이중 평가 체계를 설계하였다.

이상의 관련 연구를 종합하면, 기록관리 및 공공 분야에서 챗봇과 인공지능 기술의 도입 가능성에 대한 논의는 점차 확대되어 왔으며, 실제 서비스 적용을 위한 설계 및 시범 구현 사례도 축적되고 있음을 확인할 수 있다. 다만 반복적으로 제기되는 이용자 질의의 구조를 분석한 결과를 실제 챗봇 시스템의 분류·응답 로직으로 구현하고, 그 동작과 성능을 체계적으로 검증한 연구는 아직 충분히 축적되지 않았다. 이러한 연구 공백을 보완하기 위해 본 연구는 국가기록원의 실제 질의 데이터를 기반으로 질문 유형 자동 분류와 템플릿 기반 응답 생성이 연계되어 수행되는 전체 처리 흐름을 구현하고, 기능 검증과 동작 확인을 통해 설계와 구현 간 일치성을 점검하였다. 또한 구축된 챗봇을 대상으로 정량적 평가와 정성적 평가를 병행하여, 응답 구조의 이해 용이성, 질문 유형 분류의 적합성, 응답 결과의 안정성을 종합적으로 검토하였다. 이는 선행연구에서 제안된 다층적 챗봇 평가 관점을 실제 구현 단계의 시스템에 적용하여 검증을 수행하였다는 점에서, 본 연구의 특징을 보여준다.

2. 연구방법 및 사전 설계

2.1 연구방법

연구방법은 <그림 1>과 같이 챗봇 설계, 챗봇 개발, 챗봇 평가의 세 단계로 구성되며, 본 논문에서는 챗봇의 개발과 평가를 중심으로 연구 내용을 전개한다. 먼저 사전 연구의 챗봇 설계 단계에서 국가기록원의 정보공개청구 질의와 기록관리 FAQ 데이터를 대상으로 논항구조 기반 의미역 분석을 적용하여 질문 유형 분류 체계를 도출하고, 질문 유형별 응답 시나리오 및 템플릿 구조를 설계하였다.



<그림 1> 지능형 기록정보서비스 챗봇 PoC 구현 및 평가를 위한 연구 절차

본 연구인 2단계 챗봇 개발 단계에서는 설계 단계에서 정리된 질문 유형 분류 체계와 응답 구조를 실제로 구현하는 데 초점을 두었다. 실제 기록관리 기관에 적용 가능한 챗봇을 개발하기 위한 환경을 구성하고, 논항구조 분석 및 실시간 정보 탐색을 위한 외부 API가 연동 가능한 형태로 전체 시스템 아키텍처를 설정함으로써, 이용자 질의 입력 시 논항구조 분석 결과가 실시간으로 반영되고, 분석 결과에 따라 질문 유형 분류와 응답 생성이 이루어 지도록 시스템을 구성하였다. 다음으로 논항구조 분석 결과를 활용한 질문 유형 자동 분류 로직과 템플릿 기반 응답 생성 로직을 구현하였다. 중심 동사와 주요 논항 정보를 기준으로 질문 유형을 판별하고, 유형별로 사전에 정의된 응답 템플릿이 적용되도록 구성하였다. 또한 의미역 누락이나 분석 오류와 같은 예외 상황에 대응하기 위한 기본적인 예외 처리 로직을 함께 구현하였으며, 테스트 질의를 통해 챗봇의 설계-구현 간 일치성을 점검하였다. 본 단계의 구체적인 구현 내용은 본 논문의 3장에서 상세히 제시한다.

3단계 챗봇 평가 단계에서는 구축된 챗봇이 설계 의도에 부합하게 작동하는지와 기록관리 실무 환경에서의 활용 가능성을 검증하기 위해 정량적 평가와 정성적 평가를 병행하였다. 먼저 평가 절차의 개요를 제시한 후, 정량적 평가에서는 테스트를 위한 평가셋을 활용하여 질문 유형 분류 결과와 응답 적합성을 확인한 뒤, 질문 유형별 응답 성공 여부와 오류 발생 양상을 분석함으로써 챗봇의 구조적 안정성과 성능 특성을 검토하였다. 이후 기록관리 실무에 종사중인 전문가 3명을 대상으로 그룹 인터뷰(FGI, Focus Group Interview)를 실시하여 질문 유형 분류의 타당성, 응답 구조의 이해 용이성, 기록관리 업무 특성과의 적합성에 대한 의견을 수집하였다. 이를 통해 챗봇 응답 과정에서 나타나는 문제 양상과 개선 요구를 도출하였다. 마지막으로 정량적 평가와 정성적 평가 결과를 종합적으로 해석하여 두 평가에서 나타나는 챗봇 구조의 강점과 한계, 향후 개선 방향을 도출하였다. 평가 결과는 4장에서 구체적으로 논의한다.

2.2 사전 설계: 이용자 질문 유형 분류 체계

사전 설계 단계(최유리 외, 2025)에서는 국가기록원 이용자의 정보 요구 특성을 반영한 질문 유형 분류 체계를 마련하기 위해 정보공개청구 제목 데이터와 기록관리 FAQ 데이터를 분석 대상으로 선정하였다. 정보공개청구 데이터는 일반 이용자가 기록물 접근을 목적으로 작성한 질의로, 구체적인 기록물 명칭이나 자료 형태, 열람 가능성 여부를 중심으로 구성되는 경향을 보인다. 반면 기록관리 FAQ 데이터는 국가기록원 실무자 관점에서 제기되는 질의를 바탕으로 구성된 자료로, 보존·이관·등록 등 기록관리 업무 절차와 관련된 법령 해석 및 행정적 판단을 포함하는 구조적 질의가 다수를 차지한다.

선행연구에서는 이와 같은 데이터의 성격 차이를 반영하여 두 질의 집합을 대상으로 논항구조 기반 의미역 분석(SRL, Semantic Role Labeling)방법을 제안하였다. 의미역 분석이란 문장에서 동사나 술어를 중심으로 그에 요구되는 논항 간의 의미적 관계를 체계적으로 분석하는 방법으로, 각 논항은 행위자, 대상, 수혜자 등 특정한 의미역을 갖는다. 의미역 분석 결과로 제시되는 격(格, Case) 정보는 크게 필수격과 부가격으로 구분된다. 필수격에 포함되는 논항은 ARG0(행위주체), ARG1(대상), ARG2(수혜자 또는 시작점) 등 술어가 반드시 요구하는 핵심 논항 정보이며, 부가격은 ARGM-CND(조건), ARGM-LOC(장소) 등 술어의 의미를 보완하는 선택적 정보이다. 사전 설계에서는 기록관리 분야 특성을 고려하여 ARG0, ARG1, ARG2, ARGM-CND에 중점을 두었다. 필수적인 ARG0, ARG1, ARG2는 기록관리 업무의 핵심인 ‘누가-무엇을-누구에게’ 구조를 드러내며, 부가적인 ARGM-CND는 ‘법적 요건 충족 시’, ‘보존기간 만료 시’ 등 조건부 처리 상황을 효과적으로 표현할 수 있다.

상기한 논항구조 분석을 통해 질의문에 포함된 중심 동사와 주요 논항을 추출하고, 논항 구성 양상과 동사 사용 특성을 비교·분석함으로써 일반 이용자 질의와 실무자 질의 간의 구조적 차이를 파악하였다. 또한 중심 동사 분석 결과를 바탕으로 정보 접근을 목적으로 하는 질의와 업무 처리 또는 판단을 요구하는 질의를 구분할 수 있는 기준을 도출하였다. <표 1>은 이러한 논항구조 기반 의미역 분석 결과와 중심 동사 분석을 토대로 기록정보서비스 도메인에 적용 가능한 여섯 가지 질문 유형 분류 체계를 정의한 결과이다.

<표 1> 기록정보서비스 질문 유형 분류 체계(사전 연구 결과)

유형(Intent)	설명	중심 동사	주요 논항	예시	
정보 요구형	1) 정보 요청형	정보나 기록물을 요구	요청, 청구, 신청, 받, 관련, 찾	ARG1	기록물 목록, 문서 사본, 특정 기관의 자료
	2) 사실 확인형	사실이나 현황 확인	확인, 알	ARG1	보존연한, 생산연도, 담당부서명
	3) 정보 열람형	열람 여부 확인	열람, 알리	ARG1	토지조사부, 교원 말령 기록물, 원문자료
업무 지원형	4) 업무 처리형	기록 생산, 보존, 이관 등 절차 문의	작성, 보존, 등록, 이관, 도입, 설치, 생산	ARG1, ARG2	이관 절차, 폐기 기준, 등록 시스템
	5) 조건 확인형	특정 조건에서의 가능성 질문	가능	ARGM-CND	보존기간 만료 여부, 법적 요건 충족 여부
	6) 권한 설정형	열람, 접근 권한 관련 질문	설정, 포함	ARG1	개인정보 포함 여부

사전 설계 단계에서는 전체 질의 집합에서 출현 빈도가 높은 중심 동사를 기준으로 질문 유형 분류 체계를 구성하였다. 정보공개청구 제목 데이터(8,531건)와 기록관리 FAQ 데이터(299건)를 대상으로 중심 동사 출현 빈도를 분석한 결과, 정보공개청구 질의에서 확인된 전체 433개 동사 중 상위 10위 미만 동사는 개별 동사 기준 출현 빈도가 0.5% 미만으로 나타났으며, 기록관리 FAQ 질의의 경우 전체 144개 동사 중 상위 10위 미만 동사는 1.3%

미만에 해당하였다. 이는 소수의 고빈도 동사가 전체 질의에서 반복적으로 사용되고 있음을 의미한다. 이에 본 연구에서는 두 데이터셋에서 높은 출현 빈도를 보인 상위 20개의 중심 동사를 분석 대상으로 선별하였으며, 이들이 기록정보서비스 환경에서 반복적으로 나타나는 핵심 기능과 이용자 요구를 대표한다고 판단하였다. 해당 분류 체계는 이용자 질의에 반복적으로 나타나는 중심 동사와 주요 논항의 결합 양상을 기준으로 구성되었으며, 질의에 내재된 의도와 기능적 목적을 중심으로 유형화했다는 점에서 특징을 지닌다.

도출된 질문 유형은 일반 이용자의 정보 접근 요구를 중심으로 한 ‘정보 요구형’과 기록관리 실무자의 업무 절차 수행 및 판단을 지원하기 위한 ‘업무 지원형’으로 구분된다. 정보 요구형에는 정보 요청형, 사실 확인형, 정보 열람형이 포함되며, 업무 지원형에는 업무 처리형, 조건 확인형, 권한 설정형이 해당된다. 각 유형은 요구되는 정보의 성격과 응답 목적이 상이하므로, 응답 설계 단계에서 서로 다른 응답 구조와 정보 제공 방식을 필요로 한다. 아울러 사전 연구에서는 질문 유형 분류 체계를 기반으로 각 유형별 핵심 의미 요소를 반영한 응답 시나리오 및 템플릿의 기본 구조를 함께 설계하였다.

3. 챗봇 PoC 구현

본 장에서는 앞 절의 선행연구를 통해 설계된 논항 기반 응답 논리 구조를 실제 지능형 기록정보서비스로 적용 가능한 형태의 챗봇 프로토타입으로 구현한 과정을 기술한다. 이를 위해 챗봇의 전체 시스템 구성과 데이터 처리 흐름을 보이고, 이용자 질의에 대한 논항구조 분석 결과를 활용하여 질문 유형을 분류한 뒤 템플릿 기반으로 응답을 생성하는 과정을 설명한다. 또한 구축된 챗봇의 주요 동작 특성을 사전 설계 단계의 처리 흐름과 대조하여 제시함으로써, 질문 유형 분류와 응답 생성 과정이 설계 의도에 부합하게 구현되었는지를 확인한다.

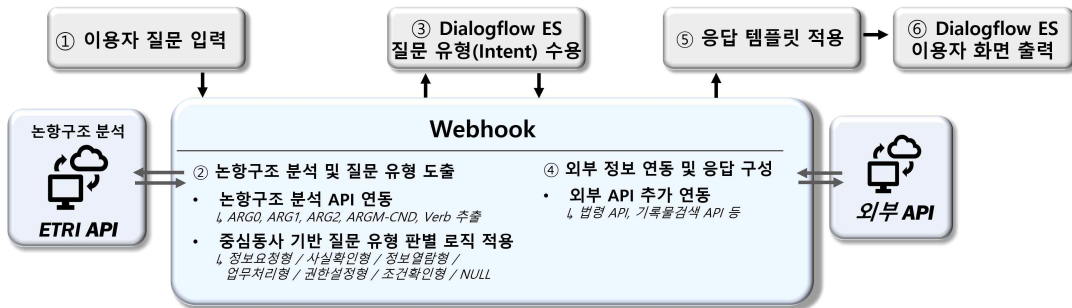
3.1 시스템 구성 및 개발 환경

본 연구에서는 기록정보서비스 챗봇 구축을 위해 이용자 질의문의 논항구조 분석 결과를 실시간으로 반영하고, 이를 템플릿 기반 응답 구조에 안정적으로 매핑할 수 있는 개발 환경을 구축하고자 하였다. 이를 위해 한국어 자연어 질의 처리의 안정성과 의미역 분석 및 정보 제공을 위한 검색 등 외부 API와의 연동 가능성을 핵심적인 요구사항으로 설정하였다. 이에 따라 상용 챗봇 플랫폼 3개와 오픈소스 프레임워크 2개를 대상으로 한국어 지원 수준, 외부 API 연동 가능성, 구축 난이도, 비용 및 접근성 등을 주요 기준으로 비교·검토하였다.

먼저, Google Dialogflow ESTM(Google, n.d.)는 한국어 자연어 처리 지원과 Webhook 기반 외부 API 연동 기능 측면에서 모두 우수한 평가를 보였다. IBM Watson AssistantTM(IBM, n.d.)는 다양한 외부 API 통합이 가능하다는 장점이 있으나, 한국어 질의 처리의 완성도가 상대적으로 낮아 의미역 분석 결과를 활용한 질문 유형 분류에 한계가 있었다. Microsoft Bot FrameworkTM(Microsoft, n.d.)는 API 확장성이 높으나 초기 구축 및 관리 과정이 복잡하여 프로토타입 구현 단계에서의 적용 부담이 컸다. Rasa Technologies에서 개발한 오픈소스 기반의 RasaTM(Rasa, n.d.)는 접근성과 커스터마이징 측면에서 장점이 있으나, 한국어 자연어 처리 지원이 제한적이라는 한계가 확인되었다. 반면 Kakao i 오픈빌더TM(Kakao, n.d.)는 한국어 처리에 최적화되어 있으나 외부 API 호출 기능이 제한적이어서 논항구조 분석 기반 응답 구조 구현에는 제약이 있었다. 이와 같은 비교 결과를 종합적으로 고려하여, 본 연구에서는 한국어 자연어 처리의 안정성과 외부 API 연동의 유연성을 동시에 충족하는 Google Dialogflow ES를 기록정보서비스 챗봇 개발 도구로 최종 선정하였다.

본 연구에서 제안하는 챗봇 시스템은 이용자의 질문 입력부터 응답 출력까지 단계적인 처리 절차에 따라 작동하며, 논항구조 분석 결과를 중심으로 질문 유형 분류와 응답 생성이 이루어지는 구조를 갖는다. <그림 2>는

최종 선정된 Google Dialogflow ES 내 Webhook 기능을 활용한 처리 절차를 도식화한 것으로, 사용자 질의 입력 이후 의미역 분석 API 연동, Dialogflow ES의 질문 유형(Intent) 수용, 외부 정보 API 연계, 템플릿 기반 응답 생성 및 출력에 이르는 전체 데이터 처리 흐름을 제시한다.



〈그림 2〉 기록정보서비스 챗봇 프로토타입의 Webhook 기반 처리 절차

제안한 챗봇 시스템의 핵심은 Webhook을 매개로 한 외부 API 연동 구조에 있다. 사용자 질의는 Dialogflow ES를 거쳐 Webhook으로 전달되며, Webhook은 외부 ETRI 의미역 분석 API와 연동하여 논항구조 분석을 수행한다. 본 연구에서는 다양한 유형의 구어체 질문에 강건한 Exobrain 기반 ETRI 의미역 분석 API(배경만b, 2024)를 활용하였다(박주영, 2020). 분석된 논항 정보는 Webhook 내부에서 중심 동사 기반 규칙을 통해 질문 유형 판별에 활용되며, 판별 결과는 Dialogflow ES로 전달되어 질문 유형으로 수용된다. 이후 Webhook은 질문 유형과 논항구조 정보를 바탕으로 ETRI 법률 QA API(배경만a, 2024), 나라기록물 검색서비스 OpenAPI(국가기록원, 2025) 등 외부 정보원과 연계하여 응답 구성에 필요한 정보를 수집한다. 최종적으로 구성된 응답 정보는 Dialogflow ES로 전달되어 사전 정의된 템플릿이 적용되며, 생성된 응답은 사용자 화면에 출력된다. 이러한 Webhook 기반 구조를 통해 논항구조 분석과 외부 API 연동이 유기적으로 결합된 응답 생성이 가능하다.

3.2 논항구조 분석 기반 응답 처리 흐름

3.2.1 질문 유형 기반 인텐트 구성

차용한 Google Dialogflow ES는 사용자 질의 유형을 ‘인텐트(intent)’라는 자료구조를 통해 처리한다. 본 연구에서는 기록정보서비스의 특성을 반영하여 사용자 의도를 질문 유형 단위로 정의하고 이를 인텐트 구조로 구현하였다. 본 챗봇에서 인텐트는 사용자 질의가 수행하고자 하는 기록관리 행위 목적을 질문 유형 단위로 구조화한 개념을 의미한다. 이에 따라 사전 설계 단계에서 도출한 여섯 가지 질문 유형(<표 1> 참조)을 각각 독립된 인텐트 단위로 구성하였다.

각 인텐트는 ‘정보요청형’, ‘사실확인형’, ‘권한설정형’ 등 기록정보서비스에서 반복적으로 나타나는 행위 목적을 기준으로 정의되었으며, 이를 통해 사용자 질의의 의미적 목적에 따라 응답 흐름이 구분되도록 설계하였다. 이 여섯 가지 질문 유형 외에도 서비스 이용 과정에서 발생하는 인사말이나 대화 시작 발화 등을 처리하기 위해 ‘인사’ 유형 인텐트를 보조적으로 포함하였다. 또한 특정 질문 유형으로 분류되지 않거나 의미 분석이 어려운 입력을 처리하기 위해 ‘NULL’ 인텐트를 별도로 정의하여, 예외 상황에서도 대화 흐름이 중단되지 않도록 하였다.

각 인텐트에는 해당 질문 유형을 대표하는 발화 예시를 등록하여, 실제 이용자들이 동일한 행위를 다양한 표현 방식으로 질문하는 특성을 반영하였다. 예를 들어 ‘요청합니다’, ‘요청드립니다’, ‘요청해요’와 같이 동사의 활용형이나 문장 종결 방식이 상이하더라도, 의미역 분석 결과를 통해 동일한 행위로 인식될 수 있도록 구성하였으며,

이를 통해 질의에 내재된 행위 목적을 중심으로 인텐트가 분류되도록 하였다. 본 연구에서 질문의도는 질문 유형과 의미역 분석 결과를 결합한 구조화된 정보 단위로 표현된다. 하나의 질문의도는 다음의 세 가지 구성 요소로 이루어진다(<그림 3> 참조). 첫째, 질의가 수행하고자 하는 행위 목적을 나타내는 질문 유형인 인텐트, 둘째, 질의 문 내에서 행위를 대표하는 중심 동사, 셋째, 중심 동사를 기준으로 도출된 주요 논항이다. 이러한 구조는 사용자 질의가 수행하고자 하는 행위의 유형과 그 대상, 조건을 함께 포함하도록 설계하였다.

인텐트 분류 과정에서는 챗봇 플랫폼의 기본 인텐트 매칭 기능에 의존하지 않고, 외부 논항구조 분석 결과를 활용한 분류 구조를 적용하였다. ‘인사’ 유형을 제외한 모든 인텐트에 대해 Webhook 기능을 활성화하여 사용자 질의가 외부 서버로 전달되도록 구성하였으며, 외부 서버에서는 중심 동사와 주요 논항 정보를 기준으로 질문 유형을 판별한 뒤 적합한 응답 템플릿을 선택하도록 하였다. 이를 통해 동일한 의미 구조를 가진 질의에 대해 일관된 인텐트 분류와 응답 생성이 가능하도록 하였다. 이와 같은 인텐트 구성 방식은 질문 유형 분류체계와 응답 생성 로직을 긴밀하게 연계함으로써, 기록정보서비스에 요구되는 응답의 정확성과 일관성을 확보하는 데 목적이 있다.

3.2.2 의미역 분석 결과 기반 응답 흐름

본 연구에서 개발한 기록정보서비스 챗봇은 사용자 질의에 대한 논항 분석 결과를 토대로 응답을 생성하는 구조를 갖는다. 사용자가 질문을 입력하면 해당 문장은 외부 논항구조 분석 API를 통해 중심 동사와 주요 논항으로 분석되며, 이러한 결과는 질문 유형 판별과 응답 구성의 근거로 활용된다. 논항 분석 결과 중 중심 동사는 사용자 질의에서 ‘무엇을 하려는가’라는 행위 목적을 가장 직접적으로 나타내는 요소이다. 기록정보서비스 질의의 경우 동일한 기록물이나 주제를 대상으로 하더라도 열람을 원하는지, 제공을 요청하는지, 절차를 확인하려는지에 따라 질문의 성격이 달라진다. 본 연구에서는 이러한 행위 목적의 차이가 주로 동사를 통해 표현된다는 점에 주목하여, 질문에 사용된 중심 동사를 기준으로 질문 유형을 분류하도록 설계하였다. 이를 통해 질의에 내재된 행위 중심의 의미 구조를 바탕으로 질문 유형을 판별할 수 있다.

질문 유형이 결정되면 해당 유형에 대응하는 응답 템플릿이 선택된다. 응답 템플릿은 질문 유형별로 사전에 정의되어 있으며, 논항 분석을 통해 도출된 논항 정보가 템플릿 구성 요소로 활용된다. 논항 분석 결과 중 ARG1은 사용자가 요청한 기록물이나 정보의 대상을 나타내며, ARGM-CND가 추출된 경우에는 해당 조건 정보가 응답 문장에 포함되도록 설계하였다. 또한 질문 유형에 따라 외부 정보원과의 연계가 활용된다. 본 연구에서는 ETRI 법률 QA API와 나라기록물 검색서비스 OpenAPI를 외부 정보원으로 활용하였다. ETRI 법률 QA API는 관련 법령 조항을 조회하는 데 활용되었으며, 나라기록물 검색서비스 OpenAPI는 기록물 검색 결과의 전체 건수 정보를 제공하는 용도로 사용되었다. 예컨대 ‘권한설정형’ 질문의 경우 ARG1로 추출된 요청 대상에 대한 열람 가능 여부를 판단하기 위해 관련 법령을 조회하며, ‘정보요청형’과 ‘사실확인형’ 질문에서는 기록물의 존재 여부를 확인하기 위해 기록물 검색 API와 연계하도록 설계하였다. 이러한 유형별 외부 정보 연계 방식은 질문의 성격에 부합하는 근거 정보를 응답에 반영하기 위함이다. 템플릿 기반 응답 생성 방식은 동일한 의미 구조를 가진 질의에 대해 일관된 형식의 응답을 제공할 수 있다는 특징을 갖는다. 특히 기록정보서비스와 같이 절차적 정확성과 정보 전달의 명확성이 요구되는 환경에서는 논항 분석 결과를 구조적으로 반영한 응답 생성 방식이 효과적으로 적용될 수 있다.

3.2.3 응답 안정화를 위한 처리 규칙 생성

본 연구에서 구축한 기록정보서비스 챗봇은 논항구조 분석 결과를 기반으로 응답을 생성하도록 설계되었으나, 실제 사용자 질의 환경에서는 문장 구조의 불완전성, 의미역 누락, 복합 문장 입력 등으로 인해 응답 생성 과정의 불안정성이 발생할 수 있다. 이에 본 연구에서는 실제 질의 데이터를 활용한 반복적인 테스트 과정을 통해 응답 생성 과정에서 발생하는 주요 문제 양상을 확인하고, 이를 보완하기 위한 응답 안정화 처리 규칙을 적용하였다.

(<표 2> 참조).

<표 2> 챗봇 구현 과정에서 적용한 응답 안정화 규칙

구분	적용 규칙	규칙 적용 목적
①	Webhook 로그 점검 및 외부 API 호출 결과 확인	Dialogflow와 외부 API 간 연동 상태 점검
②	ARG0, ARG1, ARG2, ARGM-CND를 명령창에 출력	인식 오류나 누락된 논항을 직접 확인
③	논항 종성 여부에 따라 조사(은/는, 이/가, 을/를 등) 자동 선택	문법적으로 자연스러운 응답 생성
④	고빈도 동사를 핵심 행위로 우선 인식*	기본 템플릿(완전형) + 축소 템플릿(간소형) 이중 구조
⑤	기본 템플릿(완전형) + 축소 템플릿(간소형) 이중 구조	의미역 누락 상황에서도 안정적 응답

* 고빈도 동사 우선순위는 사전 연구의 동사 빈도 분석 결과를 기준으로 함.

첫째, 시스템 연동의 안정성을 확보하기 위해 사용자 질의 전달 및 외부 API 호출 과정을 지속적으로 점검하였다. 이용자가 입력한 질의가 챗봇 플랫폼을 통해 외부 서버로 정상적으로 전달되고, 의미역분석 API의 분석 결과가 응답 생성 단계까지 정확히 반영되는지를 확인하기 위해 Webhook 로그를 점검하였다. 이를 통해 시스템 간 연동 오류나 데이터 전달 과정에서의 문제를 조기에 파악하고 수정할 수 있었다. 둘째, 논항구조 분석 결과의 확인 가능성을 높이기 위한 처리 규칙을 적용하였다. 논항 분석 과정에서 추출된 중심 동사와 주요 논항을 명령창에 출력하도록 설정함으로써, 각 질의에 대해 어떤 논항이 도출되었는지를 직접 확인할 수 있도록 하였다. 이 과정은 의미역 누락이나 오인식 사례를 식별하는 데 활용되었으며, 이후 응답 구조 보완과 평가 단계의 분석 기준으로도 활용되었다.

셋째, 응답 문장의 문법적 자연스러움을 확보하기 위한 규칙을 적용하였다. 초기 구현 단계에서는 조사 선택이 문맥과 일치하지 않는 문제가 확인되었으며, 이를 해결하기 위해 의미역 값에 따라 ‘은/는’, ‘이/가’, ‘을/를’ 등의 조사가 자동으로 선택되도록 템플릿 규칙을 수정하였다. 이를 통해 응답 문장이 보다 자연스러운 한국어 문장 구조를 유지할 수 있도록 하였다. 넷째, 복합 문장 질의에 대응하기 위한 처리 규칙을 도입하였다. 실제 사용자 질의에서는 하나의 문장에 두 개 이상의 동사가 포함되는 사례가 빈번하게 나타났으며, 이로 인해 질문 유형 분류가 불안정해지는 문제가 발생하였다. 이에 본 연구에서는 질의 데이터 분석 결과를 바탕으로, 복수의 동사가 출현하는 경우 핵심 행위를 우선적으로 인식하도록 규칙을 설정하여 질문 유형 분류의 일관성을 유지하도록 하였다.

마지막으로, 논항의 누락 상황에 대비한 템플릿 구조를 적용하였다. 반복적인 테스트 과정에서 일부 질의는 ARG0나 ARG2와 같은 논항이 안정적으로 추출되지 않는 경우가 확인되었으며, 이로 인해 응답 문장이 어색하게 생성되는 문제가 발생하였다. 이를 보완하기 위해 일부 질문 유형에 대한 논항이 정상적으로 추출된 경우 적용되는 기본 템플릿과, 일부 논항만을 활용하는 축소 템플릿을 함께 설계하였다. 논항 누락이 발생한 경우에는 축소 템플릿을 적용함으로써 응답 오류가 확산되는 것을 방지하고, 질의의 핵심 의도를 유지한 응답을 제공하도록 하였다. 이와 같은 응답 안정화 처리 규칙을 통해 챗봇은 문장 구조가 불완전하거나 일부 논항이 생략된 질의에서도 비교적 일관된 응답을 생성할 수 있는 구조를 갖추게 되었다. 이러한 처리 규칙은 다음 장에서 수행한 정량적·정성적 평가에서 응답의 안정성과 일관성을 검증하기 위한 전제 조건으로 작용한다.

3.3 챗봇 PoC 구현 결과

본 절에서는 3.1장에서 제시한 Webhook 기반 처리 절차(<그림 2>)를 토대로 챗봇의 실제 동작 과정을 설명한다. 이를 위해 수집된 국가기록원 질의로그로부터 기록정보서비스에서 가장 빈번하게 발생하는 자료 요청 형태의 자연어 질의 중 하나를 선정, “국가기록원 소장 창원기제공단 관련 사진자료를 요청합니다”라는 예제를 활용해

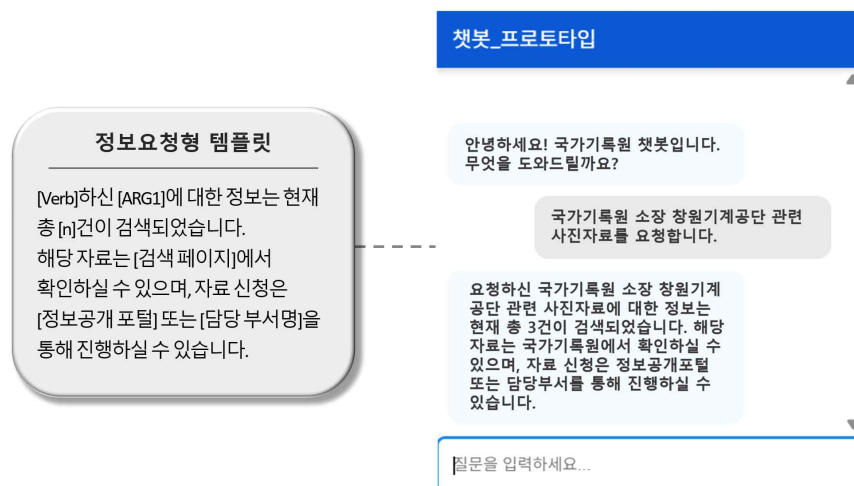
이용자 질의가 시스템 내에서 처리되는 흐름을 단계별로 보인다.

먼저 ① 사용자 질의 입력 단계에서는 예제 질의문이 챗봇 사용자 화면을 통해 입력되며, 시스템은 이를 자연어 질의로 인식한다. ② 논항구조 분석 및 질문 유형 도출 단계에서는 입력된 질의가 Dialogflow ES의 Webhook 기능을 통해 외부 처리 서버로 전달되고, Webhook은 외부의 논항구조 분석 API와 연동하여 의미역 분석을 수행한다. 이 과정에서 ETRI 의미역 분석 API가 호출되어 질의문에 포함된 중심 동사 ‘요청’과 요청 대상에 해당하는 ‘창원기계공단 관련 사진자료’가 ARG1으로 추출되며, 질의에 포함된 기타 논항 정보 역시 의미역 분석 결과로 구조화된다(<그림 3> 참조). 이렇게 도출된 논항 분석 결과를 바탕으로 중심 동사 기반 규칙을 적용하여 사용자 질의의 질문 유형을 <표 1>의 ‘3) 정보요청형’으로 판별한다. ③ Dialogflow ES 질문 유형(Intent) 수용 단계에서는 Webhook에서 판별된 질문 유형과 논항구조 정보가 Dialogflow ES로 전달되며, Dialogflow ES는 이를 인텐트 처리 결과로 수용하여 이후 응답 생성 절차와 대화 흐름을 제어한다.



<그림 3> 인텐트 구조 개념도(좌)와 Webhook 기반 실제 처리 결과 예시(우)

④ 외부 정보 연동 및 응답 구성 단계에서는 ‘정보요청형’ 질문을 해결하기 위한 기록물 검색 API를 연계하고 응답 생성에 필요한 정보를 종합한다. ⑤ 응답 템플릿 적용 단계에서는 Webhook을 통해 구성된 응답 정보가 다시 Dialogflow ES로 전달되며, Dialogflow ES는 사전에 정의된 질문 유형별 응답 템플릿을 적용하여 최종 응답 문장을 생성한다. 이 과정에서 문장 구성, 조사 선택, 문장 배열과 같은 규칙이 함께 적용되어 응답 문장이 정제된다(<그림 4> ‘정보요청형’ 응답 템플릿 참조). ⑥ 응답 출력 단계에서는 최종적으로 생성된 응답이 Dialogflow ES 사용자 화면을 통해 출력되며, 이용자는 요청한 자료에 대한 안내 문장을 확인할 수 있다(<그림 4> 챗봇 PoC UI 화면 참조).



<그림 4> ‘정보요청형’ 응답 템플릿(좌)과 챗봇 PoC UI 화면(우)

이와 같은 과정을 통해 챗봇은 사용자 질의를 입력받아 논항구조 분석, 질문 유형 판별, 외부 정보 연계, 템플릿 기반 응답 생성 및 출력에 이르는 일련의 흐름에 따라 동작한다. 이를 통해 Webhook 기반 처리 절차가 실제 챗봇 구현 환경에서도 사용자 질의 입력부터 응답 출력까지 일관된 구조로 작동함을 확인할 수 있다.

4. 챗봇 PoC 평가 및 검증

4.1 평가 설계 및 절차

기록정보서비스 챗봇의 질문 유형 분류체계 타당성과 응답 템플릿의 이해 용이성 그리고 응답의 정확성을 다각도로 검증하기 위해 정량 및 정성 평가를 병행하였다. 정량적 평가는 구축된 챗봇에 실제 질의를 입력하여 인텐트 분류 정확도, 템플릿 적용 정확도, 문장 구조의 적절성을 확인하는 방식으로 진행하였다. 이를 위해 학습데이터 통계를 바탕으로 유형별 비율을 조정해 평가셋을 구성하였으며, 이를 챗봇에 입력해 응답이 설계된 템플릿 구조와 일치하는지, 의미역 매핑이 정확하게 수행되는지, 문장 흐름이 자연스러운지를 중심으로 검증하였다.

평가셋에 포함된 질의 유형은 전체 질의 집합 중 사전 설계 단계에서 선별한 상위 20개 중심 동사가 포함된 질의를 대상으로 구축되었으며 여섯 가지 질문 유형의 통합비율 합계는 약 46.88%에 해당한다(<표 3> 참조). 나머지 질의는 출현 빈도가 낮은 중심 동사를 포함한 질의로서 본 연구의 질문 유형 분류 체계 구성 범위에는 포함되지 않았으나, 이러한 저빈도 질의는 응답문 형식이 아닌 단순 검색 결과 제시 방식으로 대응 가능하다.

정성적 평가는 기록관리 분야 실무 경험을 보유한 전문가를 대상으로 FGI를 실시하여 질문 유형 분류 체계가 적절한지, 제시한 응답 템플릿의 이해가 용이한지, 현장 활용 가능성이 있는지를 중심으로 검토하였다. 이러한 이중 평가 방식은 상호보완적으로 작용하여, 챗봇의 실효성과 응답 품질을 다각적인 관점에서 검증할 수 있는 평가 틀을 제공하였다는 점에서 의의가 있다. 정량적 평가는 실제 출력 결과를 기반으로 응답 정확성과 구조적 적절성을 수치적으로 확인하는 역할을 하며, 정성적 평가는 전문가의 경험적 판단을 통해 실무적 적절성과 활용 가능성을 검토하는데 기여한다.

4.2 정량적 평가

정량적 평가는 챗봇이 실제 질의에 대해 설계된 구조를 얼마나 정확하게 수행하는지를 확인하기 위해 수행되었다. 이를 위해 각 유형별 대표 문항으로 구성된 평가셋을 마련하고, 해당 문항을 챗봇에 입력하여 응답 결과를 비교검증하였다. 정량적 평가는 다음 세 가지 항목을 중심으로 진행되었다. 첫째, 인텐트 분류 정확도는 문장의 중심 동사를 기반으로 챗봇이 질문 유형을 올바르게 판별했는지를 평가하였다. 둘째, 템플릿 적용 정확도는 도출된 논항이 응답 템플릿의 지정 위치에 정확하게 매핑되었는지를 점검하였다. 셋째, 문장 구조의 적절성은 조사 선택, 문장 배열, 주술 호응 등 문법적 요소가 자연스럽게 구성되었는지를 확인하였다. 정량적 평가는 규칙 기반 응답 구조가 실제 질의에 대해 얼마나 안정적으로 작동하는지를 객관적으로 검증하는 과정이며, 정성적 평가에서 제시된 전문가 의견과 함께 챗봇 응답 구조의 타당성과 실효성을 종합적으로 판단 할 수 있는 근거를 제공한다.

4.2.1 챗봇 평가셋 구축

본 연구의 평가셋은 학습 데이터의 실제 분포를 반영하되, 유형 간 비율 편중을 완화할 수 있도록 정규화 과정을 거쳐 구축하였다. 학습 데이터는 정보공개청구 8,531건과 기록관리 FAQ 299건으로 구성되어 있으며, 두 출처

간 데이터 규모의 차이가 크기 때문에 단순 비율 평균이 아닌 가중평균(weighted average) 방식으로 통합 비율을 산출하였다(<식 1> 참조).

$$P_{\text{통합}} = \frac{(P_{\text{정보요청형}} \times 8,531) + (P_{\text{FAQ}} \times 299)}{8,531 + 299} \quad (1)$$

예를 들어 ‘정보요청형’의 경우, 정보공개청구 데이터에서 해당 유형에 속하는 동사의 출현 비율 합계는 44.7%로 나타났으며, 기록관리 FAQ 데이터에서는 해당 동사가 출현하지 않았다. 이에 따라 두 데이터셋의 규모를 고려하여 가중평균 방식으로 통합비율을 산출, 43.1%로 계산되었다. 같은 방식으로 여섯 가지 질문 유형에 대한 비율을 산출한 결과는 <표 3>의 통합비율에 제시하였다.

<표 3> 유형분류별 정규화 비율 및 평가셋 구성 결과

의도(Intent)	통합비율(%)	√변환 값	정규화 후 비율	평가셋 건수
① 정보요청형	43.1	6.57	63.4	30(-2)
② 사실확인형	1.9	1.38	13.3	7
③ 정보열람형	0.6	0.77	7.4	4
④ 업무처리형	1.1	1.05	10.1	5
⑤ 조건확인형	0.1	0.32	3.1	2(+1)
⑥ 권한설정형	0.08	0.28	2.7	2(+1)
합계			100.0	50

<표 3>의 통합 비율에 나타난 바와 같이 정보요청형이 차지하는 실질적 비율은 43.1%로, 상위 20개 중심 동사로 한정된 질의 집합(전체 46.66%) 내에서 약 92%에 해당한다. 이러한 쏠림 현상을 해소하기 위해 본 연구에서는 통합비율을 기반으로 하되, 각 비율에 제곱근(√)을 적용하여 고빈도 유형의 비중을 완화하고 저빈도 유형의 비중을 확대하는 정규화 방식을 적용했다. 제곱근 변환은 극단값을 완화하면서도 원비율을 유지하는 일반적인 정규화 방법으로, 제곱근 변환 후 각 값을 전체 합으로 나누어 100%로 재정규화한 결과, 최종 평가셋의 구성비율은 <표 3>의 정규화 후 비율과 같다.

평가셋 50건은 PoC 단계의 챗봇 성능을 측정하기 위한 검증용 데이터셋으로, 각 질문 유형별 응답 메커니즘의 작동 가능성을 확인하는 데 목적을 두었다. 또한 저빈도 유형의 사례가 평가 과정에서 사실상 누락되는 문제를 방지하기 위해 각 질문 유형별로 최소 2건 이상이 포함되도록 하한값을 설정하였다. 이를 통해 통계적 비중이 낮은 유형도 평가 과정에서 일정 비율을 확보하여 전체 평가셋의 균형성과 대표성을 강화했다. 평가 문항은 생성형 AI(GPT-5)를 활용하여 총 5회의 반복 생성 과정을 거쳐 초안을 마련함으로써 표현의 다양성과 유형별 적합성을 확보하였다. 이후 모든 문항을 직접 검토하여 의미 중복, 비문, 분류 기준과의 불일치가 확인된 항목을 제거하고, 최종적으로 평가셋 구성 목적에 부합하는 문항만을 선별했다.

4.2.2 인텐트 분류 정확도

정량적 평가 결과, 인텐트 분류는 전체 50문항 중 48문항이 정확하게 분류되어 96%의 정확도를 보였다(<표 4> 참조). 대부분의 문항에서 중심 동사 기반 분류 규칙이 안정적으로 작동하였으며, 질문 유형에 따른 핵심 동사의 식별이 전반적으로 일관되게 수행되었다.

<표 4> 인텐트 분류 정확도 평가 결과 요약

평가 항목	결과 요약	주요 사항
① 인텐트 분류 정확도	96% (48/50)	<ul style="list-style-type: none"> 중심 동사 기반 분류가 대부분 정확하게 작동함 오분류 2건은 복문 구조로 인한 중심 동사 식별 실패(1건)와 보조 규칙 우선순위 적용에 따른 유형 왜곡(1건)에서 발생함

다만 일부 문항에서 인텐트 오분류가 확인되었다. 한 사례에서는 복문 구조로 인해 논항구조 분석 단계에서 사실확인형의 중심 동사가 안정적으로 식별되지 못하였고, 다른 사례에서는 중심 동사가 사실확인형으로 인식되었음에도 보조 규칙 적용 과정에서 ‘공개’와 같은 키워드가 우선 반영되면서 권한설정형으로 분류되는 결과가 나타났다. 이러한 결과는 단일 동사 기반 규칙이 복합 문장이나 다중 단서가 포함된 질의에서 취약할 수 있으며, 키워드 기반 보조 규칙의 우선순위 설정 방식에 따라 의도 분류가 왜곡될 가능성이 있음을 보여준다(<표 5> 참조).

<표 5> 인텐트 분류 과정에서 발생한 오분류 사례

예상 유형	질문	유형 결과	인텐트 분류
사실 확인형	기록물이 등록되어 있는지 확인할 수 있나요?	업무 처리형	{‘intent’: ‘업무처리형’, ‘verb’: ‘등록’, ‘ARG1’: ‘기록물’, ‘ARG0’: ‘’, ‘ARG2’: ‘’, ‘ARGM-CND’: ‘’}
사실 확인형	해당 기록물이 공개 대상인지 알 수 있나요?	권한 설정형	{‘intent’: ‘권한설정형’, ‘verb’: ‘알’, ‘ARG1’: ‘공개 대상인지’, ‘ARG0’: ‘해당 기록물’, ‘ARG2’: ‘’, ‘ARGM-CND’: ‘’}

본 연구에서는 설계 단계에서 일반 이용자의 질문을 중심으로 한 ‘정보요구형’과 업무 담당자의 질문을 중심으로 한 ‘업무지원형’을 구분하고, 각 범주를 기준으로 처리 우선순위를 설정하였다. 그러나 실제 평가 과정에서는 일반 이용자형 동사와 업무지원형 동사가 하나의 문장에 혼합되어 등장하는 사례가 확인되었다. 이는 실제 질의 상황에서 두 유형의 문장이 엄격하게 분리되지 않으며, 이용자와 실무자의 질의 경계가 생각보다 유연함을 시사한다. 따라서 향후 인텐트 분류 체계는 정보요구형과 업무지원형을 이중적으로 구분하는 방식에서 나아가, 혼합적 문맥에서도 안정적으로 작동하는 통합 우선순위 체계를 갖추도록 확장할 필요가 있다. 이러한 개선은 복문 구조의 처리 정확도를 높이고, 다양한 문장 유형을 포괄하는 견고한 분류 체계를 구축하는 데 기여할 것이다.

4.2.3 템플릿 적용 정확도

템플릿 적용 정확도 평가 결과, <표 6>과 같이 평가 문항 전반에서 템플릿 구조의 오류나 슬롯 누락과 같은 구조적 문제는 확인되지 않았다. 이는 문항별로 도출되는 논항의 구성과 출현 양상이 서로 다름에도 불구하고, 템플릿 구조가 안정적으로 유지되었음을 의미한다. 실제 질의 데이터에서는 주요 논항의 출현 여부가 문항마다 다르게 나타나며, 이러한 특성은 슬롯 기반 응답 생성 과정에서 문장 구성의 안정성을 저해할 수 있다. 그럼에도 본 연구에서 제시한 템플릿은 의미역의 누락이나 조합 차이를 고려한 구조로 설계되어, 각 문항에서 추출된 의미역 정보가 지정된 슬롯에 일관되게 삽입되었다.

<표 6> 템플릿 적용 정확도 결과 요약

평가항목	결과요약	주요사항
② 템플릿 적용 정확도	구조적 오류 없음	<ul style="list-style-type: none"> 템플릿 적용 과정에서 구조적 오류는 확인되지 않음 모든 문항에서 지정된 의미역이 템플릿 구조에 정상 삽입됨

예를 들어 정보요청형 질의 “훈령 개정 이력 자료를 요청합니다”에서는 VERB(요청)과 ARG1(훈령 개정 이력 자료)이 템플릿의 해당 슬롯에 정확히 삽입되었다. 마찬가지로 사실확인형 질의 “문서의 생산연도를 알고싶습니다”의 ARG1(문서의 생산연도), 권한설정형 질의 “개인정보가 포함된 자료의 열람 권한을 설정할 수 있나요?”의 ARG1(자료의 열람 권한) 모두 지정된 슬롯에 정상 반영되었다. 즉, 추출된 의미역이 슬롯에 삽입되지 않거나 잘못된 슬롯에 배치되는 구조적 오류가 본 평가에서는 발견되지 않았다. 이는 질의마다 논항 구성이 상이함에도 불구하고 템플릿이 논항 정보를 안정적으로 수용하여 구조적 완결성을 유지했음을 보여준다.

특히 정보요청형, 사실확인형, 정보열람형과 같은 기록물 중심 질의에서는 ARG1을 중심으로 한 논항구조가 공통적으로 나타났으나, 응답에서 요구되는 정보의 성격과 전달 방식은 유형에 따라 상이하였다. 템플릿은 이러한 특성을 반영하여 기본적인 슬롯 구조의 일관성을 유지하면서도 유형별 응답 요구사항에 적절히 대응할 수 있도록 설계되었으며, 평가 결과 이를 안정적으로 구현할 수 있음을 확인하였다. 이는 논항구조 기반 템플릿 설계 방식이 기록관리 분야의 질의응답 구조와 높은 정합성을 보임을 의미한다.

4.2.4 문장 구조의 적절성

문장 구조의 적절성은 조사 선택, 주술 호응, 어순의 자연스러움, 비문 여부를 기준으로 평가하였다. 평가 결과, <표 7>과 같이 전체 응답의 84%가 문장 구조 측면에서 전반적으로 자연스럽게 생성된 것으로 나타났다. 정보요청형과 정보열람형처럼 단문 구조를 기반으로 하는 템플릿은 논항구조 배치와 고정 문구가 일관되게 작동하며, 문장 구조의 안정성이 비교적 높게 유지되었다.

<표 7> 문장 구조의 적절성 평가 결과 요약

평가항목	결과요약	주요사항
③ 문장 구조의 적절성	84% (42/50)	<ul style="list-style-type: none"> 전체 응답 중 84%가 조사 선택, 문장 배열, 주술 호응 측면에서 전반적으로 자연스러운 문장 생성이 이루어짐 오류 8건 발생: 조사 선택 오류(5건), 논항 적용 슬롯-고정 문구 중복으로 인한 부자연스러운 표현(2건), 복문 처리 과정에서 중심 동사 식별이 불안정하여 비문이 생성된 사례(1건)

반면 일부 문항에서 문장 생성 과정에서 반복적으로 나타나는 오류 유형이 확인되었으며, 대표적인 사례는 <표 8>에 제시하였다. 첫째, 의미역 슬롯 값과 템플릿 고정 문구가 동시에 반영되면서 ‘관련하신 ○○ 관련 기록물’과 같이 중복된 표현이 생성되는 사례가 2건 확인되었다. 이는 요청 또는 청구 행위를 나타내는 중심 동사 대신 ‘관련’이 의미역 분석 단계에서 중심 동사로 인식되면서, 슬롯 값과 고정 문구가 함께 호출된 데에서 비롯된 것으로 볼 수 있다. 이러한 오류는 동일 유형 내에서도 동사의 사용 빈도와 역할을 고려한 우선순위 규칙이 필요함을 시사한다.

둘째, 사실확인형 응답에서는 괄호 표기가 포함된 명사에 대해 목적격 조사가 적절히 선택되지 않는 사례가 총 5건 확인되었다. 이는 템플릿 내부의 조사 선택 규칙이 괄호나 특수 표기를 포함한 문자열을 충분히 고려하지 못하면서, 조사 부착 단계에서 규칙이 정상적으로 적용되지 않은 결과로 해석할 수 있다.

셋째, 복문 구조가 포함된 문항에서는 중심 동사가 안정적으로 식별되지 않으면서 비문이 생성된 사례가 확인되었다. 논항구조 분석 단계에서 의도했던 ‘확인하다’가 아닌 ‘등록하다’가 중심 동사로 인식되면서 질의문이 업무처리형으로 분류되었고, 이에 따라 업무처리형 템플릿이 우선 적용되었다. 이후 템플릿 적용 단계에서는 문장 내에 남아 있는 ‘확인’ 어근을 토대로 사실확인형 템플릿의 표현인 ‘확인하신’이 부분적으로 호출되면서, 서로 다른 유형의 문형이 혼재된 응답이 생성되었다. 즉, 표면적으로는 조사 선택 규칙의 문제처럼 보이지만, 실제로는

복문 처리 과정에서 중심 동사 식별이 실패한 결과가 이후 단계 전반에 연쇄적으로 영향을 미친 사례로 해석할 수 있다. 따라서 향후 개선은 단순한 조사 규칙 보완보다 복문 구조에서 동사 우선순위를 판단하고 중심 동사를 판별하는 기준을 강화하는 방향이 우선적으로 요구된다. 이러한 분석을 통해 전반적인 문장 생성 성능은 안정적인 수준을 보였으나, 특정 오류 유형이 반복적으로 나타난다는 점에서 조사 선택 규칙, 슬롯 조정 규칙, 복문 처리 기준을 중심으로 템플릿과 문장 생성 규칙의 추가적인 정교화가 필요함을 확인하였다.

〈표 8〉 문장 구조에서 발생한 오류 사례

질문	분류 결과	SRL 분석 결과	답변
조직개편 관련 기록물을 요청합니다.	정보 요청형	슬롯-고정 문구 중복	관련하신 조직개편 관련 기록물에 대한 정보는 현재 총 3건이 검색되었습니다. 해당 자료는 국가기록원(임의)에서 확인하실 수 있으며, 자료 신청은 정보공개 포털 또는 담당부서를 통해 진행하실 수 있습니다.
해당 기록물의 보존연한을 확인할 수 있나요?	사실 확인형	조사 선택 오류	해당 기록물의 보존연한과 관련된 기록은 국가기록원(임의)에서 확인되었습니다. 현재 확인된 건수는 총 3건입니다. 보다 자세한 내용은 국가기록원(임의)을 통해 확인하실 수 있습니다.
기록물이 등록되어 있는지 확인할 수 있나요?	업무 처리형	복문 처리 오류	해당 기록은 에 의해 생산되어, 이관 또는 등록된 절차가 확인되었습니다. 처리 유형은 '확인하신'로 분류됩니다. 본 기록은 현재 대상으로 처리되고 있습니다. 절차에 대한 자세한 사항은 관련 지침을 참조해 주세요.

4.3 정성적 평가

제안한 챗봇의 실무 적용 가능성을 검증하기 위해 기록관리 현업에 종사중인 기록연구사를 대상으로 FGI를 수행하였다. 본 연구에서 개발한 챗봇은 중앙연구기록물관리기관인 국가기록원의 질의 데이터를 기반으로 구축된 점을 고려, 정성적 평가 참여자 역시 국가기록원의 관리 체계 하에 있는 중앙행정기관 및 공공기관 소속 실무자를 대상으로 선정하였다.

인터뷰 참여자는 기록물관리 전문요원 자격 보유자로, 현재 중앙행정기관(1인) 및 지방공기업(2인)에서 기록관리 실무를 담당하고 있으며, 실무 경력은 5년 이상(2인), 10년 이상(1인)으로 모두 기록관리 현장에서 이용자의 기록정보 요청에 대한 충분한 경험을 갖춘 실무자들이다. 참여자 A는 소속 기관에서 행정 업무 자동화(RPA) 도입 가능성이 검토되고 있는 상황을 인지하고 있어, 관련 기술에 대한 이해를 바탕으로 현장 관점의 의견을 제시하였다. 참여자 B와 참여자 C는 정보공개 업무를 직접 담당하거나 해당 업무를 수행한 경험이 있어, 정보공개 청구 이전 단계에서 이용자 질의의 특성과 실제 처리 절차에 대한 이해를 바탕으로 평가에 참여하였다. 이들은 국가기록원 기반 기록정보서비스 환경에 대한 이해를 바탕으로 질문 유형 분류 체계와 응답 구조의 실무 적합성을 검토하는 데 적절한 참여자 집단이라 할 수 있다.

〈표 9〉 FGI 평가항목

구분	평가항목
① 도입	이용 경험 기반 맥락 파악
② 질문 유형 분류 타당성	1) 질문 유형 체계의 포괄성
	2) 질문 유형 구분 기준의 명확성
③ 응답 템플릿의 전달 구조 및 이해 용이성	1) 이해 용이성
	2) 구조적 자연스러움
④ 활용 가능성 및 적합성 평가	1) 현장 활용 가능성
	2) 기록관리 업무 특성과의 적합성
⑤ 마무리	추가 의견 및 확장 가능성

인터뷰는 <표 9>와 같이 참여자들의 사전 경험과 시각을 파악하기 위한 도입 질문으로 시작하여 총 다섯 가지 평가항목에 따라 반구조화된 방식으로 진행하였다. FGI는 2025년 11월 대면으로 진행되었으며, 참여자 동의 하에 녹취 후 전사하여 분석하였다.

4.3.1 질문 유형 분류 체계의 타당성

참여자들은 본 연구에서 제시한 6가지 질문 유형이 대민 행정기관에서 반복적으로 나타나는 주요 질의를 상당 부분 포괄하고 있다고 평가하였다. 특히 참여자 B는 신청자들의 요구사항이 정형화되어 있고 틀이 정해져 있기 때문에 ‘정보요청형’, ‘사실확인형’, ‘정보열람형’ 유형이 실제 정보공개 업무에서 빈번하게 등장하는 질문 패턴과 부합한다고 보았다. 나아가 이러한 분류체계가 실무 담당자에게 일관된 응답 흐름을 제공하는 표준형으로 기능할 수 있다는 점을 긍정적으로 언급하였다.

“저희 기관의 경우 신청자의 요구가 상당히 정형화되어 있는 편입니다. 이는 업무가 일정한 범적 기준에 기반하고 있어 질의의 범위와 형식이 비교적 명확하게 한정되어 있기 때문입니다. 이러한 특성으로 인해 정형화된 응답을 제공하는 방식은, 정보공개 업무 담당자가 답변을 수행하는 구조에서 올바른 응답의 표준형을 제시하는 데 도움이 될 수 있으며, 이러한 방식으로 설계된 챗봇은 실무적으로도 유효할 것이라고 생각합니다.” (참여자 B)

한편 참여자 C는 기관별 업무 특성에 따라 분류체계의 적용 범위가 달라질 수 있다는 한계를 지적하였다. 공단, 공사 등 사업 중심 기관의 경우 기록관리와 직접적인 관련이 없는 사업, 운영, 주차 관련 질의가 더 많이 발생하기 때문에 국가기록원 기반의 유형만으로는 모든 질의를 포괄하기 어렵다는 의견을 제시하였다. 이는 챗봇의 실무 적용을 확장하기 위해서는 기관별 도메인 분석과 추가 학습이 병행될 필요가 있음을 시사한다. 다만 참여자들은 대민 행정 중심 기관의 기록정보서비스 범위 내에서는 본 분류체계를 기본 틀로 활용하는 데 무리가 없다는 데 공감하였다.

“저희 공단의 경우 정보공개 질의가 기록관리 업무보다는 사업 운영과 관련된 내용, 예를 들어 주차나 시설 이용과 같은 정보에 집중되는 경우가 많습니다. 이런 점에서 본 분류체계가 그대로 적용되기에는 일부 보완이 필요할 수 있다고 생각합니다. 이러한 한계는 특정 기관에 국한된 문제라기보다는, 기관별 업무 성격이 서로 다르다는 점에서 공통적으로 고려되어야 할 사항이라고 봅니다.” (참여자 C)

또한 참여자들은 질문 유형 간 구분 기준이 전반적으로 명확하게 제시되어 있다고 평가하였다. 참여자 A는 본 챗봇의 분류 기준이 정보공개 청구 이전 단계에서 이용자 질문의 성격을 파악하고 유형을 구분하는 데 실질적인 도움이 될 것이라고 판단했다. 그러나 일부 용어는 의미 범위가 넓게 해석될 가능성이 있다는 지적도 제기되었다. 특히 참여자 C는 ‘업무처리형’이 포괄적인 용어로 인식되어, 이용자에게 6개 유형 전체를 아우르는 상위 범주처럼 보일 수 있다고 언급하였다. 이러한 의견은 분류체계를 정교화하는 과정에서 용어의 범위와 명칭을 보다 명확하게 조정할 필요성을 시사한다. 그럼에도 참여자들은 현재 제시된 분류 기준이 기록관리 실무 환경에서 기본 틀로 활용하기에 충분하다고 의견을 모았다.

4.3.2 응답 템플릿의 전달 구조 및 이해 용이성

참여자들은 본 연구에서 설계한 응답 템플릿이 기존 기록정보서비스의 답변 방식과 크게 어긋나지 않아 이용자가 핵심 내용을 이해하는 데 무리가 없을 것이라고 평가하였다. 참여자 A는 템플릿 구조가 학습 데이터의 답변

사례와 부합하며, 현행 정보 제공 방식과도 일관성을 유지하고 있다는 점을 긍정적으로 언급하였다. 나아가 일부 참여자들은 정보공개 포털 등에서 나타나는 정형화된 답변 패턴을 언급하며, “검색 결과 해당 기록이 없어 부존재 처리합니다”와 같이 기록정보서비스에서 자주 사용되는 문구를 템플릿에 반영할 경우, 이용자의 직관적인 이해를 더욱 높일 수 있다는 구체적인 제안을 덧붙였다.

응답 템플릿의 구조 자체에 대해서는 참여자들로부터 뚜렷한 문제 제기나 부정적 의견이 나타나지 않았다. 참여자들은 구조 자체보다는 응답의 충실도 및 정보 제공 범위에 더 큰 관심을 보였으며, 필요 시 후속 절차 안내, 관련 자료 링크, 참고 정보 등 추가 정보를 포함할 경우 실무 활용도가 높아질 것이라는 의견을 제시하였다. 이는 템플릿이 단순한 정보 제공을 넘어, 업무 흐름과 연계된 안내 요소를 포함하도록 확장될 때 실무 적용성이 강화될 것이라는 점을 시사한다. 따라서 향후 챗봇의 실용성 향상을 위해 안내 요소의 충실도와 기능적 연계 강화가 필요함을 확인할 수 있다.

4.3.3 현장 활용 가능성 및 확장성

참여자들은 본 챗봇이 반복적인 질의 대응이나 기본 절차 안내와 같이 비교적 규칙적인 응답이 요구되는 영역에서 실무자의 업무를 보조하는 도구로 활용될 수 있다는 점에 공통적으로 동의하였다. 참여자 B는 정보공개 업무 처리 시 챗봇을 통해 표준화된 답변 템플릿을 제공함으로써 실무자들이 답변 내용을 신속하게 제시하고, 업무 일관성과 적시성을 확보하는 데 기여할 수 있다는 점을 긍정적으로 평가했다. 다만 챗봇의 실무 도입과 관련하여 고려해야 할 현실적인 제약도 언급되었다. 참여자 B는 챗봇 답변에 대한 이용자 신뢰도는 실제 도입 이후 확인할 필요가 있으며, 답변에 만족하지 못하는 이용자들이 결국 정보공개청구로 이어질 수 있다는 점을 지적하였다. 반면 참여자 C는 즉시 공개 대상 정보에 대한 외부 링크 연계가 답변에 대한 이용자의 수궁 가능성을 높일 수 있다고 언급했다. 이 모든 의견을 종합할 때, 챗봇은 독립적인 응답 시스템이라기보다는 기록정보서비스의 초기 안내 단계에서 청구 전 질의를 보조적으로 지원하는 방식으로 실무 활용 가능성이 더 크다는 점이 확인되었다.

참여자들은 응답 템플릿이 기록정보서비스의 업무 처리 절차와 어떻게 연결되는지가 실무 적합성 판단의 핵심 기준이라고 보았다. 특히 이용자가 궁극적으로 확인하고자 하는 것은 최종적인 처리 결과라는 점이 강조되었다. 참여자 C는 업무가 처리 유형에 따라 진행되기 때문에 공개, 부분공개, 비공개, 타기관이송, 정보부존재 등의 최종 처리 결과가 중요하다고 언급하며, 이 처리 유형이 답변에 반영되어야 이용자가 상황을 더 명확히 이해할 수 있다는 의견을 제시하였다. 이는 응답 템플릿이 실제 기록관리 절차의 흐름과 처리 기준을 명확히 드러낼 때, 기록정보서비스의 업무 특성을 충실하게 반영할 수 있으며 이용자의 이해도를 또한 제고될 수 있음을 시사한다.

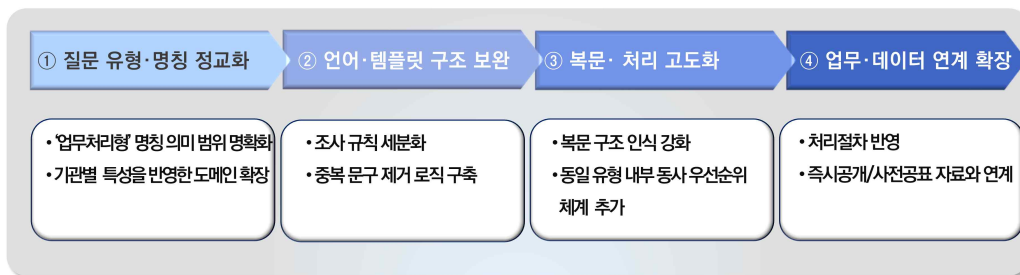
“실제로 이를 처리하는 과정에서 담당자들이 가장 궁금해하는 것은 질문의 형식보다는, 해당 사안을 어떻게 처리할 것인가에 대한 판단입니다. 즉 공개인지, 부분 공개인지, 비공개인지, 다른 기관으로 이송해야 하는 사안인지, 또는 정보가 존재하지 않는 경우인지를 구분하는 것이 핵심입니다. 이러한 처리 유형에 따라 업무가 이루어지기 때문에, 답변 역시 처리 유형을 기준으로 구성된다면 더 적절할 것이라고 생각합니다.” (참여자 C)

FGI의 마지막 단계에서 참여자들은 질문 유형과 응답 구조 외에도, 기록정보서비스 챗봇이 실무 환경에서 보다 안정적으로 활용되기 위해 고려해야 할 추가적인 확장 요소를 제시하였다. 먼저 민원성 질의와의 연결 가능성이 언급되었다. 참여자 B는 조건확인형과 권한설정형처럼 특정 행동을 요구하거나 가능 여부를 판단하는 유형은 단순한 정보 제공을 넘어 민원 성격의 절차 안내와도 이어질 수 있다고 보았다. 이에 따라 향후 유형 확장 또는 보조 유형 표기를 통해 민원 처리 맥락과의 관계를 보다 명확히 구분할 필요가 있다는 의견이 제시되었다. 또한 일부 참여자들은 정보공개 포털의 즉시공개 자료나 사전공표 자료와 같은 외부 공개 정보와의 연계 기능을 강화

할 경우, 이용자의 이해도와 응답 신뢰도를 높일 수 있을 것이라고 언급하였다. 즉, 챗봇이 단순히 답변을 제공하는 수준을 넘어 공표 자료 및 선행 정보와의 직접적 연결을 제공하는 형태로 확장될 때 실무 활용성이 더욱 높아질 수 있다는 점이 강조되었다. 이러한 제안들은 챗봇의 초기 안내 기능을 고도화하기 위한 실질적 시사점을 제공한다.

4.4 평가 종합

정량적 평가와 정성적 평가를 통합적으로 검토한 결과, 본 챗봇은 기본 응답 구조를 구현하는 데 있어 일정 수준의 안정성을 확보한 것으로 판단된다. 동시에 평가 과정에서 드러난 오류 양상과 실무자 의견을 바탕으로, 향후 고도화를 위한 네 가지 개선 방향이 도출되었다(<그림 5> 참조).



<그림 5> 평가 종합 개선방향

첫째, 질문 유형과 명칭의 정교화가 필요하다. FGI 결과 ‘업무처리형’과 같이 의미 범위가 넓게 인식되는 유형은 이용자에게 상위 범주처럼 보일 수 있어, 용어를 보다 구체화하거나 실제 기관의 업무 특성을 반영한 세분화가 요구되었다. 특히 실무자들은 기관의 성격에 따라 주요 질의 유형 자체가 크게 달라진다는 점을 지적하였다. 예를 들어, 사업·운영 중심 기관에서는 공사 진행, 시설 이용, 주차·안내 등 운영·민원성 질의가 빈번한 반면, 행정·정보공개 중심 기관에서는 기록물 요청·열람·확인 등 전형적인 기록정보서비스 질의가 주로 발생한다. 이러한 차이를 고려할 때 현재 유형 체계는 행정기관 중심의 질의에는 적합하지만 운영기관의 질의를 충분히 포괄하지 못하는 한계가 있으며, 기관별 질의 특성과 업무 범위를 반영하는 방향으로 유형 체계를 확장할 필요가 있다.

둘째, 언어 규칙과 템플릿 구조의 정교화가 요구된다. 평가 과정에서 ‘국가기록원(임의)을’과 같이 괄호 표기가 포함된 단어에 대해 조사 선택 규칙이 정확히 적용되지 않는 사례가 반복되었으며, ‘관련하신 ○○ 관련 자료’처럼 템플릿 고정 문구와 의미역 값이 중복되어 부자연스러운 문장이 생성되는 문제가 확인되었다. 이러한 오류는 괄호·특수문자와 같은 예외적 형태를 충분히 고려하지 못하거나, 의미역 값과 템플릿 고정 표현이 충돌하는 상황에서 후처리 규칙이 정상적으로 작동하지 않은 데에서 비롯된 것으로 보인다. 따라서 조사 선택 규칙을 세분화하고, 중복 표현을 자동으로 조정하는 후처리 로직을 보완하며, 문장 다듬기 규칙을 강화하여 응답의 자연스러움과 구조적 일관성을 확보할 필요가 있다.

셋째, 복문 구조 인식의 고도화와 동일 유형 내부의 동사 우선순위 체계 보완이 요구된다. 정량적 평가에서 확인된 사례처럼, 복합 문장에서 중심 동사가 불안정하게 추출될 경우 인텐트 분류 오류가 템플릿 매핑과 문장 생성 단계까지 연쇄적으로 확산되는 문제가 발생하였다. 더불어 동일 유형에 속하는 여러 동사가 한 문장에 함께 등장하면서 중심 동사 판별이 어려워지는 한계가 드러났다. 이러한 문제를 해결하기 위해서는 복문 구조에서 동사 후보를 안정적으로 식별할 수 있도록 규칙 기반 분석을 강화하고, 동일 유형 내부에서도 세부적인 동사

우선순위 기준을 마련하여 복수 동사 출현 시 중심 동사를 보다 정밀하게 판별할 수 있도록 보완할 필요가 있다.

넷째, 업무·데이터 연계 기능의 확장이 요구된다. 실무자들은 챗봇이 단순한 응답 생성에 머무르지 않고, 기록물의 공개 여부나 이송 필요성 등 처리 결과를 명확히 안내하며, 즉시공개 자료나 사전공표 정보처럼 실제로 제공 가능한 외부 데이터를 함께 연계할 때 실무 활용성이 크게 향상된다고 평가하였다. 이러한 의견은 템플릿 구조에 처리 유형 안내, 참고 자료 링크, 후속 조치 설명 등 실무 중심의 정보 요소를 체계적으로 통합하는 방향으로 기능을 확대할 필요가 있음을 시사한다. 이와 같은 개선 방향은 챗봇 구현의 고도화 전략을 구체화하며, 실제 기록정보서비스 환경에서 요구되는 실무적 필요와 기술적 정교화를 균형 있게 반영한 개발 과제로 이어진다. 이러한 보완을 통해 본 시스템은 향후 더 높은 정확성·안정성·실용성을 갖춘 기록정보서비스 챗봇으로 발전할 수 있을 것으로 기대된다.

5. 결론

본 연구는 사전 연구에서 제안된 논항구조 기반 질문 유형 분류 체계와 응답 논리 구조를 토대로, 지능형 기록정보서비스 챗봇을 실제로 구현하고 그 작동 가능성과 성능을 검증하는 데 목적을 두었다. 이를 위해 국가기록원의 실제 질의 데이터를 입력값으로 활용하여 질문 유형 자동 분류와 템플릿 기반 응답 생성이 연계되는 처리 흐름을 구현하였으며, 설계 단계에서 정의한 응답 구조가 시스템 환경에서도 동일하게 작동하는지를 단계별 검증을 통해 확인하였다.

구현 결과, 챗봇은 논항구조 분석 결과를 기반으로 질문 유형을 안정적으로 분류하고, 유형별 응답 템플릿을 적용하여 기록관리 절차와 조건을 비교적 일관된 방식으로 안내하는 것으로 나타났다. 특히 중심 동사와 주요 논항을 기준으로 한 질문 유형 분류 로직이 실제 질의 환경에서도 유효하게 작동함을 확인하였으며, 템플릿 기반 응답 구조가 기록관리 업무의 절차적 특성을 전달하는 데 실질적인 기능을 수행함을 검증하였다. 또한 본 연구에서는 구축된 챗봇 PoC를 대상으로 정량적 평가와 정성적 평가를 병행하여 수행함으로써, 응답 구조의 이해 용이성, 질문 유형 분류의 적합성, 응답 결과의 안정성을 종합적으로 검토하였다. 이를 통해 설계 단계에서 예측한 처리 논리와 실제 구현 결과 간의 일치성을 확인하였으며, 동시에 응답 오류 유형과 개선 필요 지점을 도출하여 향후 고도화를 위한 실증적 근거를 마련하였다.

다만 본 연구는 국가기록원의 정보공개청구 및 기록관리 FAQ로 데이터 범위가 한정되어 있어, 기관별 다양한 질의 유형을 충분히 반영하지 못한 한계가 있다. 템플릿 기반 구조는 절차 안내에는 강점을 가지지만 복합 질의나 예외적 문장 구조에서는 유연성이 낮을 수 있으며, 복문 구조에서 중심 동사 식별이 불안정하게 나타난 사례 역시 규칙 기반 의미역 분석이 복잡한 문장을 처리하는 데 추가적 보완이 필요함을 시사한다.

향후 연구에서는 기관별 기록관리 도메인 특성을 반영하여 질문 유형 체계를 확장하고, 복합 질의에 대한 동사 우선순위 규칙 및 의미역 해석 기준을 보다 정교화할 필요가 있다. 더 나아가 템플릿 기반 구조와 생성형 모델을 결합한 하이브리드 응답 방식을 도입한다면, 구조적 안정성과 응답의 유연성을 동시에 확보할 수 있을 것으로 기대된다. 이러한 후속 연구를 통해 본 연구에서 구현·검증한 챗봇 구조는 기록관리 현장에서 활용 가능한 지능형 기록정보서비스 모델로 발전할 수 있으며, 다양한 공공기관 기록관리 서비스로의 확장 가능성도 함께 제시할 수 있을 것이다.

참고문헌

- 강윤아, 오효정 (2023). 전자기록관리 업무 및 기록정보서비스에서의 생성형 AI기술 활용. 한국기록관리학회지, 23(4), 179-200.
<http://doi.org/10.14404/JKSARM.2023.23.4.179>
- 국가기록원 (2025). 국가기록원_나라기록물정보 서비스 출처: <https://www.archives.go.kr/next/newsearch/openAPI01.do>
- 김선욱 (2025). 대학기록관 챗봇 서비스 개발을 위한 RAG 적용: K대학교를 중심으로 석사학위논문, 경북대학교.
- 박주영 (2020. 8. 6.). 국산 인공지능 ‘엑소브레인’ 이제 구어체도 알아듣는다. 연합뉴스.
<https://www.yna.co.kr/view/AKR20200806036400063>
- 배경만 (2024a). 법률 QA API. ETRI 기술사전 체험 플랫폼. 출처: <https://epretx.etri.re.kr/apiDetail?id=70>
- 배경만 (2024b). 언어 분석 기술. ETRI 기술사전 체험 플랫폼. 출처: <https://epretx.etri.re.kr/apiDetail?id=2>
- 설문원 (2022). 국가기록원의 기록제공서비스에 관한 탐색적 연구. 한국기록관리학회지, 22(3), 103-124.
<http://dx.doi.org/10.14404/JKSARM.2022.22.3.103>
- 이창희, 이해영, 김인택 (2018). 기록정보서비스를 위한 메신저 기반의 챗봇 프로토타입 개발 연구: 명지대학교 대학사료실을 중심으로. 정보관리학회지, 35(3), 215-244. <https://doi.org/10.3743/KOSIM.2018.35.3.215>
- 임상빈, 고영우 (2017). AI기반 지능형 지방세브레인시스템 구축 방안. 지방세논집, 4(2), 5-28.
- 정부기록보존소조직제. 대통령령 제4029호.
- 정부기록보존소조직제. 대통령령 제11205호.
- 정책브리핑 (2025. 6. 20.). 이 대통령 “과감한 세계 혜택·규제혁신… AI시대 고속도로로 구축”.
출처: <https://www.korea.kr/news/policyNewsView.do?newsId=148944726>
- 최유리, 이예은, 오효정 (2025). 지능형 기록정보서비스 챗봇을 위한 논항구조 의미역 분석 기반 응답 논리 구조 설계. 한국기록관리학회지, 25(3), 259-279. <http://dx.doi.org/10.14404/JKSARM.2025.25.3.259>
- Google (n.d.). Dialogflow ES [Software]. Available: <https://cloud.google.com/dialogflow/es>
- Gupta, S., Ranjan, R., & Singh, S. N. (2025). Comprehensive Framework for Evaluating Conversational AI Chatbots. arXiv, <https://doi.org/10.48550/arXiv.2502.06105>
- IBM (n.d.). Watson Assistant [Software]. Available: <https://www.ibm.com/watson/assistant>
- Kakao (n.d.). Kakao i Web Chatbot SDK [Software]. Available: <https://docs.kakaoui.ai>
- Microsoft (n.d.). Azure AI Bot Service [Software]. Available: <https://azure.microsoft.com/ko-kr/products/ai-services/ai-bot-service>
- Rasa (n.d.). Rasa Open Source [Software]. Available: <https://rasa.com>

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

- Bae, Kyeongman (2024a). Legal QA API. ETRI Technology Dictionary Experience Platform. Available: <https://epretx.etri.re.kr/apiDetail?id=70>
- Bae, Kyeongman (2024b). Language Analysis Technology. ETRI Technology Dictionary Experience Platform. Available: <https://epretx.etri.re.kr/apiDetail?id=2>
- Choi, You-Ree, Lee, Ye-eun, & Oh, Hyo-Jung (2025). Logical answer structure design based on semantic role labeling (SRL) for an intelligent archival information service chatbot. Journal of Korean Society of Archives and Records Management, 25(3), 259-279. <http://dx.doi.org/10.14404/JKSARM.2025.25.3.259>
- Government Records Preservation Office Regulations. Presidential Decree No. 11205.
- Government Records Preservation Office Regulations. Presidential Decree No. 4029.
- Kang, Yoona & Oh, Hyo-jung (2023). The use of generative AI technologies in electronic records management and archival

- information service. *Journal of Korean Society of Archives and Records Management*, 23(4), 179-200.
<http://doi.org/10.14404/JKSARM.2023.23.4.179>
- Kim, Seonwook (2025). *The Application of RAG for Developing a Chatbot Service in University Archives: Focusing on K University*. Master's thesis, Kyungpook National University, Korea.
- Korea Policy Briefing (2025, June 20). President Lee: "Bold tax benefits, regulatory innovation... Building a highway for the AI era". Available: <https://www.korea.kr/news/policyNewsView.do?newsId=148944726>
- Lee, Chang hee, Rieh, Hae-young, & Kim, Intaek (2018). Development of prototype chatbot based on messenger app for archival reference services: with focus on the archives & records center of Myongji University. *Journal of the Korean Society for Information Management*, 35(3), 215-244. <https://doi.org/10.3743/KOSIM.2018.35.3.215>
- Lim, Sang-Bin & Ko, Young Woo (2017). AI-based intelligent local tax brain system. *Local Tax Review*, 4(2), 5-28.
- National Archives of Korea (2025). *National Archives of Korea_National Records Information Service*. Available: <https://www.archives.go.kr/next/newsearch/openAPI01.do>
- Park, Juyoung (2020, August 6). Korea's AI 'Exobrain' Can Now Understand Colloquial Speech. *Yonhap News*.
<https://www.yna.co.kr/view/AKR20200806036400063>
- Seol, Moon-won (2022). An exploratory investigation of archival reference services in the national archives of Korea. *Journal of Korean Society of Archives and Records Management*, 22(3), 103-124.
<http://dx.doi.org/10.14404/JKSARM.2022.22.3.103>