

지능형 기록정보서비스 챗봇을 위한 논항구조 의미역 분석 기반 응답 논리 구조 설계*

Logical Answer Structure Design based on Semantic Role Labeling (SRL) for an Intelligent Archival Information Service Chatbot*

최유리(You-Ree Choi)¹, 이예은(Ye-eun Lee)², 오효정(Hyo-Jung Oh)³

E-mail: choiulee@jbnu.ac.kr, yenny923@naver.com, ohj@jbnu.ac.kr



- 1 제1저자 전북대학교 기록관리학과 석사과정
- 2 공동저자 전북대학교 기록관리학과 석사과정
- 3 교신저자 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 공동연구원

논문접수 2025.07.16
최초심사 2025.07.25
게재확정 2025.08.28

ORCID

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

Ye-eun Lee
https://orcid.org/0009-0000-5853-0352

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.

- 본 논문은 2025년도 한국연구재단 연구비 지원에 의한 결과의 일부임(과제번호: NRF-2021R111A3047435)

초 록

국가기록원은 우리나라를 대표하는 기록관리 핵심기관으로서 국민의 다양한 기록 열람 요청에 응답하고 기록관리 전반에 관한 총괄적 민원을 처리하는 역할을 담당한다. 본 연구는 최근 AI정부를 표방하는 기조에 맞춰 국가기록원을 대상으로 기록정보서비스에 특화된 지능형 챗봇 개발을 위한 사전 단계로, 논항구조분석을 통한 질문 유형 분류와 응답 시나리오 템플릿 설계를 목적으로 한다. 챗봇 구축을 위한 학습 데이터로는 국가기록원에 실제 접수된 정보공개청구 제목과 기관에서 제공하고 있는 기록관리 FAQ를 수집하였으며, 논항구조분석을 통해 여섯 가지 질문 유형을 도출하고 각 유형에 따라 자연어 이해 요소를 구조화하였다. 이후 이용자 질의에 내재된 의도를 식별하고 유형별 맞춤형 응답을 구성할 수 있도록 질문 유형 분류와 응답 시나리오 템플릿을 설계하였다. 제안한 질문 유형 기반 응답 구조화는 기록정보서비스 챗봇 개발에 있어 체계적인 응답 생성의 모형이 되며, 향후 실제 서비스 환경에서의 챗봇 구현 시 사용자 만족도와 업무 효율성 향상에 기여할 수 있을 것으로 기대된다.

ABSTRACT

The National Archives of Korea (NAK), as the nation's central authority for records management, is responsible for responding to diverse public requests for access to records and addressing general complaints related to records management. As a preliminary step toward developing an intelligent chatbot specialized in archival information services for the NAK, this study aims to classify question types and design response scenario templates through semantic role labeling (SRL). Research data consist of information disclosure request titles submitted to the NAK and the records management FAQs it provides. From these, six question types were derived and structured with natural language understanding elements. Then, question type classification and response scenario templates were developed to identify user intentions and generate customized responses. The proposed question type-based response structuring offers a systematic model for response generation in archival information service chatbot development and is expected to improve user satisfaction and operational efficiency in future service environments.

Keywords: 기록정보서비스, 지능형 챗봇, 논항구조분석, 질문 유형 분류, 응답 시나리오 템플릿
Archival Information Services, Intelligent Chatbot, Semantic Role Labeling (SRL), Question Type Classification, Response Scenario Template

1. 서론

1.1 연구배경 및 필요성

디지털 정보환경의 확산으로 정보 접근과 활용 능력은 현대 시민의 필수 역량으로 자리 잡았으며, 이는 기록정보서비스 분야에서도 예외가 아니다. 최근 이용자의 정보 요구는 점차 정교해지고 있으며 이에 맞춰 기록관리기관 역시 신속하고 정확한 정보 제공 체계를 마련해야 할 필요성이 커지고 있다. 특히 이용자가 원하는 기록정보를 효과적으로 탐색할 수 있도록 검색 기능을 고도화하고, 이용자 중심의 체계적인 서비스 개발이 요구된다(강례림, 2020).

국가기록원은 우리나라를 대표하는 기록관리 핵심기관으로서 국민의 다양한 기록 열람 요청에 응답하는 역할을 담당한다. 디지털 전환 시대를 맞아 24시간 언제든지 접근 가능하고 점차 정교해지는 이용자의 정보 요구에 효과적으로 대응하기 위해서는 질문 맥락을 실시간으로 이해하고 맞춤형 정보를 제공할 수 있는 지능형 서비스의 도입이 필요하다. AI 기반 챗봇은 이용자의 다양한 질문 유형을 분석하여 적합한 기록정보를 즉시 안내하고, 복잡한 검색 과정을 단순화하여 기록정보 접근성을 혁신적으로 향상시킬 수 있다. 이는 AI를 국가 핵심 전략 기술로 설정하고 공공서비스의 구조적 혁신을 추구하는 정부의 디지털 전환 정책(정책브리핑, 2025)과도 부합하는 방향이다.

따라서 본 연구는 국가기록원 기록정보서비스에 특화된 지능형 챗봇 설계를 위해 기록정보 질의의 구조적 특성을 고려하여 논항구조 기반 의미역 분석을 통한 질문 유형 분류와 응답 논리 구조 설계 방법론을 제시하고자 한다. 이를 통해 이용자의 정보 요구를 정밀하게 파악하고 실제 서비스에 적용할 수 있는 맞춤형 응답 체계를 마련함으로써, 기록관리기관의 업무 효율성과 이용자 만족도 향상에 기여하는 것을 목표로 한다.

1.2 선행연구

기록정보서비스를 위한 지능형 챗봇과 관련해 다음 두 가지 측면의 선행연구를 검토하였다. 첫째는 기록정보서비스 이용자의 질문 및 정보 요구 분석과 관련된 연구이고, 둘째는 기록관리기관에서의 챗봇 및 AI 기반 질의응답 시스템의 적용 사례와 기술적 특성에 관한 연구이다.

먼저 이용자 요구에 초점을 맞춘 주요 선행연구는 다음과 같다. 백지연과 오효정(2019)은 국가기록원 웹 포털 서비스에서 수집된 12년간(2007년~2018년)의 질의로그 빅데이터를 분석해 이용자의 정보요구와 검색결과 유형을 파악하였다. 이를 바탕으로 통합검색 고도화, 최신 기록정보 반영, 시각화 및 추천 서비스 도입 등 검색 만족도 향상 방안을 제안하였다. 진주영과 이해영(2018)은 국가기록원 웹사이트의 유입경로와 10년(2007년~2017년) 동안의 상위 검색어 데이터를 분석하여 이용자 요구를 파악하고, 맞춤형 기록정보서비스 개선 방안을 제시하였다. 주요 관심사는 토지조사부, 일제강점기, 6·25 전쟁, 기록관리 등으로 나타났으며, 이를 바탕으로 포털·모바일 연계 강화, 외국어 콘텐츠 제공, 검색 서비스 고도화 등을 제안하였다.

기록관리기관의 챗봇 및 AI 기반 질의응답 기술 적용에 대해 검토한 논문은 아래와 같다. 강윤아와 오효정(2023)은 생성형 AI 기술을 기록관리 실무에 적용하여 전자기록관리 효율화 및 기록정보서비스의 만족도 향상 방안을 제안하였다. 특히 기록물 생산현황 통보, 민원·정보공개청구 처리, 전시 콘텐츠 생성 등에 생성형 AI를 활용할 수 있는 업무 영역별 실무 방안을 구체적으로 제시하였다. 박은비(2021)의 연구에서는 보존 중심에서 이용 중심으로 전환 중인 기록관리 흐름에 맞춰 국가기록원 기록정보서비스의 웹 인터페이스를 개선하고 이용자 친화적인 챗봇을 적용한 사례를 제시하였다. 사용성 평가를 통해 문제점을 도출하고 DIRKS 방법론을 적용해 초보자도 쉽게 활용할 수 있는 챗봇을 설계 및 구축하였으며, 메신저 플랫폼과 연동해 간편한 정보 검색과 안내

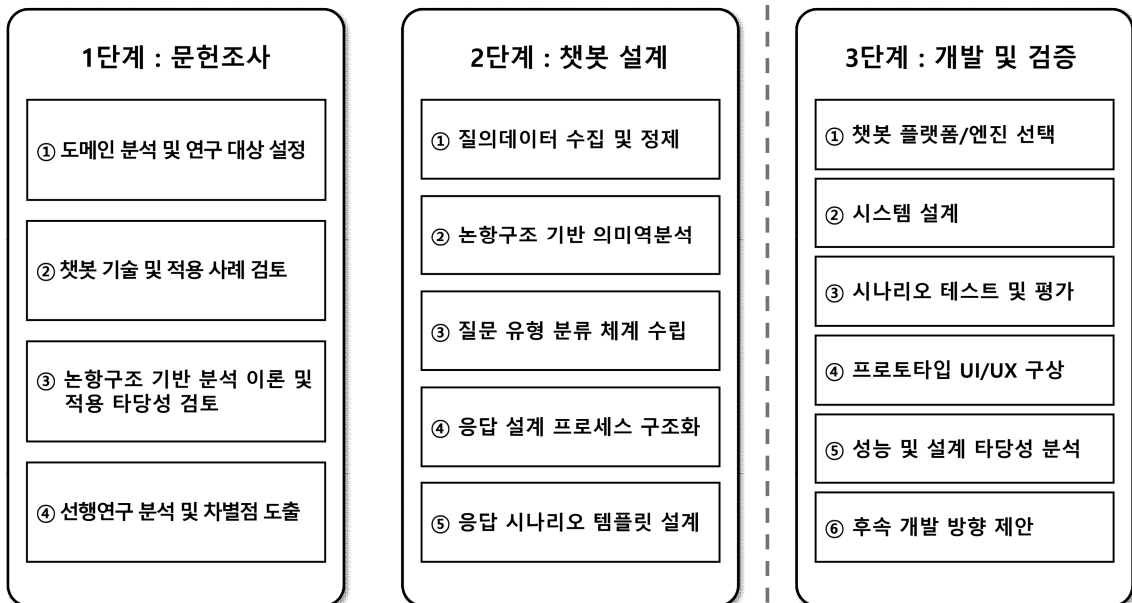
기능을 제공하였다. 김선옥(2025)의 연구와 이창희 외(2018)의 연구에서는 기록정보서비스 강화를 위해 챗봇을 기록관에 도입하는 방안을 모색하고 각각 대학사료실 및 대학기록관을 사례로 챗봇 프로토타입을 설계·개발하였다. 연구 결과, 챗봇은 반복 질의응답, 이용자 맞춤형 정보 제공, 상호작용 확대에 효과적이었으며 도입을 위해 서비스 범위 설정, 요구 분석, 도구 선정, 대화 구성이 핵심 절차로 제시되었다.

이상의 선행연구 검토 결과를 토대로 본 연구에서는 반복적으로 제기되는 민원 및 질의를 중심으로 문장 구조를 분석하고, 질문 유형을 체계화함으로써 기록정보서비스에 특화된 챗봇 응답 구조를 설계하고자 한다. 기존 선행연구들이 주로 이용자 경험 향상, 인터페이스 편의성, 기술 도입 가능성 등의 측면에 초점을 맞춘 데 비해, 본 연구는 실제 기록정보서비스 현장에서 축적된 질의 데이터를 분석하여 반복 질의의 구조적 특성을 문장 내 논항구조를 통해 살펴보고, 질문 유형 간 차이점을 도출함으로써 질문 유형별 응답 시나리오를 제시했다는 점에서 차별성을 가진다. 특히 실제 이용자들이 접수한 정보공개청구 제목과 같은 축약된 문장 형태의 질의에서도 구조적 반복성을 추출하고 실무자용 FAQ와 일반 이용자 질의 간의 표현 방식 및 논리 구조 차이를 반영한 응답 체계를 수립함으로써, 보다 정밀하고 일관된 정보 제공이 가능한 기반을 마련하였다. 이러한 접근은 기존의 사전 정의된 시나리오를 일방적으로 적용하는 방식을 보완하여, 실데이터 기반의 구조화된 응답 설계로 확장하려는 실천적 시도라 할 수 있다.

1.3 연구방법 및 범위

본 연구의 최종 목적은 지능형 기록정보서비스를 실현하기 위한 AI 챗봇 개발로, 전체 연구는 <그림 1>과 같이 문헌조사, 챗봇 설계, 개발 및 검증의 총 3단계로 나뉜다. 본 논문은 이 중 1단계와 2단계까지의 연구 결과에 해당하는 것으로, 최근 기술 발전이 빠르게 진행되고 있는 AI 챗봇 엔진을 활용해 실제 서비스를 구축하기 위한 사전 설계 및 도메인 특화 지식(domain adaptive knowledge) 구축 단계에 해당한다.

먼저 1단계 문헌조사에서는 기록정보서비스의 개념과 기능을 고찰하고 연구의 도메인과 대상을 설정하였다. 또한 지능형 응답 시스템과 챗봇 기술의 구조를 살펴보고 기록정보서비스에 적용된 사례들을 분석하여 기술적



<그림 1> 지능형 기록정보서비스 챗봇 구축 연구 방법

적용 가능성을 검토하였다. 아울러 논항구조 기반 의미역 분석에 대한 이론적 배경을 정리하고 기존 연구 사례 검토를 통해 해당 분석 기법이 기록정보서비스 질의의 구조를 해석하는 데 적합한 방법임을 확인하였다. 마지막으로 관련 선행연구들을 분석하여 본 연구와의 차별점을 도출하고 이후 챗봇 설계를 위한 분석 관점을 정립하였다.

본 논문의 핵심인 2단계 챗봇 설계에서는 실제 이용자들이 국가기록원에 접수한 정보공개청구 목록과 실무자용 기록관리 FAQ 데이터를 수집, 동사를 기준으로 데이터를 정제하여 질의문 분석 대상을 확정하였다. 이후 심층 자연어 분석 기술을 활용해 각 질의문에서 중심 동사와 의미역을 추출하고, 논항구조 기반 질의문 분석을 수행하였다. 이러한 분석을 통해 도출된 중심 동사 및 의미역 패턴을 기준으로 질의 유형을 분류하였으며, 이후 유형별로 응답 구성 요소를 식별하고 응답 흐름 구조 및 시나리오를 설계한 뒤, 질문 유형별 특성을 반영한 모듈형 응답 템플릿을 구축하였다. 이는 이후 AI 챗봇 개발 시 기록관리 분야에 특화된 지식으로, 질문 분석 및 응답 문장 생성 시 참조할 수 있는 증강 학습 데이터로 활용 예정이다.

2. 이론적 배경

2.1 기록정보서비스를 위한 지능형 응답 서비스

기록정보서비스는 기록관의 핵심 기능 중 하나로 이용자의 정보 요구에 맞춰 기록정보를 제공하고 그 활용을 지원하는 일련의 활동을 의미한다. 미국아키비스트협회(SAA, Society of American Archivists)의 용어사전에 따르면, 기록정보서비스는 ‘고객의 관심사와 관련된 자료를 찾도록 돕는 서비스’로 정의되며(Pearse-Moses, 2005), 『기록학용어사전』(한국기록학회, 2008)에서는 이를 ‘정보 제공 서비스’라는 용어로 제시하고 있다. Pugh(2005)는 기록정보서비스를 ‘이용자가 소장기록을 활용하여 원하는 정보를 찾도록 돕는 것’으로 정의하며, 더 나아가 잠재적 이용자를 포함한 모든 이용자와 상호작용하는 공공서비스 프로그램으로 개념을 확장하였다. 이처럼 다양한 정의를 종합해 보면 기록정보서비스는 이용자의 요구에 맞는 기록을 제공하기 위한 기록관의 전반적인 서비스로서, 이용자 정보를 수집하고 이를 바탕으로 맞춤형 서비스 프로그램을 기획·제공하는 기능까지 포함하는 개념으로 정리할 수 있다.

기록정보서비스의 중요성이 점차 부각되면서 이용자의 정보 요구를 중심으로 한 맞춤형 서비스 기획이 강조되고 있다. Freeman(1984), Pugh(1982)등은 기록관이 기록 중심에서 벗어나 이용자 중심으로 전환되어야 한다고 주장하였으며, 국내에서도 2000년대 이후 이용자 요구 분석에 대한 연구가 활발히 이루어져 왔다. 국가기록원은 이러한 흐름에 따라 기록정보서비스 담당부서를 신설하고 다양한 부가가치 서비스를 제공하고 있다. 이러한 변화 속에서 기록정보서비스는 이용자와 기록을 연결하는 핵심적 기능으로서, 단순한 검색·열람 지원을 넘어 이용자의 다양한 정보 요구에 맞춘 포괄적 서비스 체계로 발전해 왔다. 특히 디지털 환경의 확산과 함께 온라인 기반 서비스의 비중이 증가하면서, 보다 효율적이고 이용자 친화적인 서비스 제공 방안에 대한 관심이 높아지고 있다. 그러나 기존 연구들은 주로 일반적인 이용자 행태나 만족도 조사에 치중되어 있어 실제 질의에 내재된 구조적 특성이나 반복적 요구 패턴에 대한 체계적 분석은 충분히 이루어지지 않은 상황이다. 따라서 본 연구에서는 이용자 질의에 나타난 논항구조를 분석해 질문 유형을 체계화하고, 이를 기반으로 한 맞춤형 응답 설계 방안을 제시하고자 한다. 이러한 접근은 최근 공공분야에서 확산되고 있는 AI 기반 질의응답 시스템의 발전 방향과도 일치한다.

실제로 공공기관에서는 AI 기술을 활용한 지능형 응답 시스템을 적극 도입하며, 이용자 질의에 대한 정보 접근성과 행정 효율성을 동시에 제고하고자 하는 노력을 기울이고 있다. 단순한 키워드 기반 검색에서 벗어나 질의의 의미를 이해하고 문서 내용을 함께 분석하는 방식을 통해 보다 정밀하고 맥락에 맞는 응답을 제공하는 시스템

구축이 확산되는 추세이다. 이러한 흐름의 대표적인 사례로 국회전자도서관은 문서 검색 기능을 넘어 이용자의 질문과 문서 내용을 동시에 분석하는 기계독해 기반 질의응답 시스템을 운영하고 있다. 이러한 시스템을 통해 정보 제공의 정확성을 높이는 데 기여하고 있으며, 응답 정확도 향상과 개별 요구에 특화된 정보 제공을 통해 공공 정보서비스의 질적 개선에 기여하고 있다.

지방자치단체 차원에서도 이러한 흐름에 발맞춰 AI 기반 챗봇 시스템을 행정에 도입하는 사례가 등장하고 있다. 경상북도는 초거대 언어모델을 기반으로 한 ‘챗경북(챗GB)’을 개발해 행정지원 서비스를 운영하고 있으며, 이는 한국어 기반 사전학습 모델에 지역 행정 데이터를 학습시켜 지방행정에 특화된 응답을 제공하는 국내 최초의 시도이다. 향후에는 사업건의서, 보도자료, 이메일 작성, 보고서 요약 등 반복적인 행정업무에 특화된 기능을 추가하여 공무원의 업무 부담을 경감하고 창의적 업무로의 전환을 도모한다는 계획을 발표하였다(양승진, 2024).

공공도서관 분야에서도 이용자 중심의 응답 시스템 도입이 확대되고 있다. 전주시는 2024년 6월부터 카카오톡 기반 ‘도서정보 챗봇 서비스’를 운영 중이며, 이를 통해 이용자는 모바일 환경에서 실시간으로 도서관 서비스를 이용할 수 있다. 해당 시스템은 도서 검색 및 예약, 대출 현황 조회, 회원증 확인, 상호대차 신청, 대출 기간 연장 등의 기능을 제공하며, 단순한 질의응답을 넘어 실질적인 서비스 이용까지 지원하는 편의 기능을 구현하였다(윤난슬, 2024).

이상의 사례들에서 보듯이 공공부문의 AI 기반 질의응답 시스템은 질의 의미 파악, 유형별 분류, 맥락 기반 응답 생성 등의 기능을 통해 고도화되고 있다. 이러한 기술적 발전은 기록정보서비스 분야에서도 이용자의 정보 요구에 효과적으로 대응하고 체계적인 응답 구조를 설계할 수 있는 토대를 제공한다.

본 연구는 이러한 흐름과 궤를 같이하면서도, 기록정보서비스 분야에 특화된 응답 구조 설계를 위해 논항 구조 분석을 적용한 점에서 기술적 차별성을 갖는다. 기존 공공기관 챗봇의 대부분은 키워드 기반 또는 정규 패턴 매칭 방식으로 운영되며, 이는 사전에 정의된 문장 형태에서는 안정적인 응답이 가능하나 질의 표현이 다양해지거나 중심 구조가 달라질 경우 응답 실패 가능성이 높다. 반면 본 연구는 질의의 중심 동사와 논항구조를 분석하여 질문을 유형화하고, 각 유형에 적합한 응답 시나리오를 설계하였다. 이러한 접근은 표현 방식이 다르더라도 의미 구조가 동일하다면 동일한 응답 흐름으로 연결될 수 있는 유연성을 확보하며, 키워드 기반의 접근 방식과의 주요한 차별점이다. 특히 기록정보서비스 분야의 질의는 표면적인 표현은 다양하나 실제로는 기록물 요청, 이관, 열람 등에 해당하는 중심 동사와 요청하는 대상 정보의 구조적 반복성이 높은 특징을 가진다. 따라서 SRL 기반 분석은 질문 유형을 구조화하고 응답을 표준화하는데 효과적인 접근 방식이며 기록관리 업무의 특성과 도메인 전문성을 반영한 설계 논리로 기능할 수 있다.

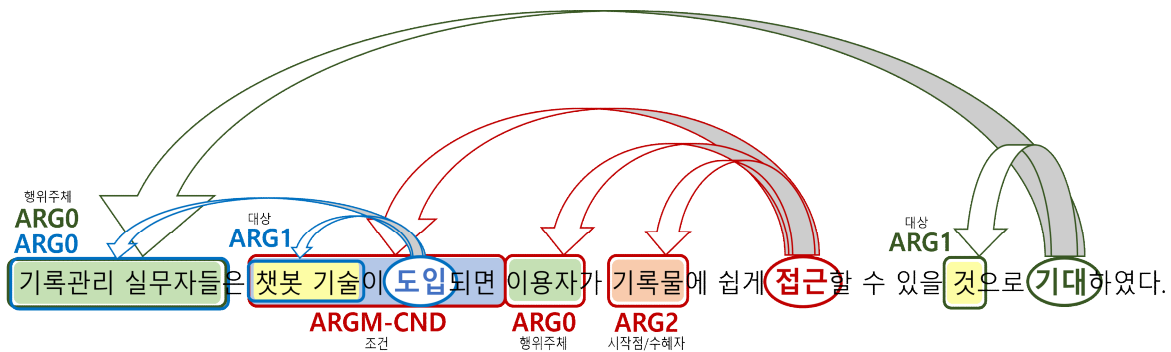
2.2 논항구조 기반 의미역 분석(SRL)

논항구조는 동사나 술어가 요구하는 성분 간의 관계를 체계적으로 기술하는 개념으로, 각 논항은 ‘행위자’, ‘피동자’, ‘수혜자’ 등 특정한 의미역(Semantic role)을 가지며 이를 결정하는 방법을 의미역 분석(SRL: Semantic Role Labeling)이라 한다. 이는 문장의 통사 구조와 의미 해석을 연결하는 핵심 이론으로, 자연어처리(NLP) 분야에서는 문장의 의미 분석과 문장 간 논리 구조 파악을 위한 기반으로 활용된다. 특히 한국어처럼 문법 구조가 유연하고 생략이나 중의성이 잦은 언어의 경우, 자연어를 정밀하게 처리하고 이해하는 인공지능 시스템을 개발하기 위해서는 의미역 정보가 포함된 다층적인 언어 분석 데이터가 필수적이다(이선웅, 2019). 의미역 분석은 단순한 구문 분석(parsing)을 넘어 동사 중심의 사건 구조를 정밀하게 기술하고, 이를 통해 질문 의도 파악 및 응답 설계의 논리 기반을 제공하는 도구로 활용될 수 있다.

의미역	정의
ARG0	서술어의 동작주, 행위자
ARG1	서술어의 피동작주, 대상
ARG2	시작점, 수혜자 등
ARG3	착점

의미역	정의
ARGM-LOC	장소 (locatives)
ARGM-DIR	방향 (directional)
ARGM-CND	조건 (condition)
ARGM-MNR	방법 (manner)
ARGM-TMP	시간 (temporal)
ARGM-EXT	범위 (extent)
ARGM-PRD	보조 서술 (secondary predication)
ARGM-PRP	목적 (purpose clauses)
ARGM-CAU	발생 이유 (cause clauses)
ARGM-DIS	담화 연결 (discourse)
ARGM-NEG	부정 (negation)
ARGM-INS	도구 (instrument)

<그림 2> 논항구조 내 의미역의 필수격과 부가격 종류



<그림 3> 복문구조 논항 기반 의미역 분석 결과 (기록관리 분야 예시 문장)

일반적으로 논항구조 내 의미역은 크게 필수격과 부가격으로 구분된다. <그림 2>는 본 연구에서 차용한 심층언어분석 기술인 Exobrain(임수종 외, 2015)의 의미역 분석 결과로 제시되는 격(格, Case)정보로, 필수격은 ARG0(행위주체), ARG1(대상), ARG2(수혜자 또는 시작점) 등 술어가 반드시 요구하는 핵심 논항 정보로, 부가격은 ARGM-CND(조건), ARGM-LOC(장소) 등 술어의 의미를 보완하는 선택적 정보로 분석된다. 본 연구에서는 이러한 의미역 중 기록관리 분야 특성을 고려하여 ARG0, ARG1, ARG2, ARGM-CND에 중점을 두었다. 예를 들어 필수격인 ARG0, ARG1, ARG2는 기록관리 업무의 핵심인 ‘누가-무엇을-누구에게’ 구조를 드러내며, 부가격인 ARGM-CND는 ‘법적 요건 충족 시’, ‘보존기간 만료 시’ 등 조건부 처리 상황이 효과적으로 표현됨을 파악하였다.

<그림 3>은 “기록관리 실무자들은 챗봇 기술이 도입되면 이용자가 기록물에 쉽게 접근할 수 있을 것으로 기대하였다.”라는 문장을 대상으로, 세 개의 동사(‘도입’, ‘접근’, ‘기대’) 각각에 대해 의미역을 부착한 결과를 시각화한 것이다. 이 문장은 복문 구조로 동사 간의 인과 및 기대 관계가 내포되어 있으며, 각 동사는 고유한 의미역 구조를 통해 문장의 전체 논리 흐름을 형성한다. 먼저 첫 번째 동사 ‘도입’에는 ARG0(행위 주체)로 ‘기록관리 실무자들’, ARG1(대상)으로 ‘챗봇 기술’이 부착되어 기술 도입의 주체와 객체가 명확히 설정된다. ‘접근’에는 ARGM-CND(조건)로 ‘챗봇 기술이 도입되면’, ARG0(행위 주체)로 ‘이용자’, ARG2(수혜자 또는 시작점)로 ‘기록물’이 부착되어 기술 도입이라는 선행 조건 아래, 이용자의 정보 접근 행위를 구조화하고 있다. 마지막으로 주절에 해당하는 ‘기대’에는 ARG0(행위 주체)로 ‘기록관리 실무자들’, ARG1(대상)으로 ‘이용자가 기록물에 접근할 수 있을 것’이 할당되어 문장의 전체 핵심 의미를 완성한다. 이러한 도식화된 분석은 복수의 술어가 독립적인 의미역 구조를 가지면서도 조건, 기대 등의 논리적 연결을 통해 통합된 의미를 형성한다는 점을 보여준다. 특히 조건절, 다양한

행위 주체, 복합 의미 구조 등이 혼재된 문장은 기록정보서비스 질의의 정밀한 해석을 위해 의미역 기반 분석이 효과적임을 시사한다.

3. 이용자 질의데이터 분석

3.1 국가기록원 질의 데이터 수집 및 전처리

본 연구에서는 국가기록원 이용자의 정보 요구 유형이 반영된 챗봇 논리 구조 설계를 위한 기초 데이터로, 이용자들이 실제 국가기록원에 요청한 정보공개청구 내역과 국가기록원 실무자 관점에서 자주 물어보는 정보를 사전에 정리해 제공하는 자료(FAQ)를 분석 대상으로 삼았다. 정보공개청구의 경우, 일반 이용자가 기록물을 청구하는 과정에서 작성한 것으로, 대부분 기록물이라는 실체적 대상을 중심으로 한 특징을 가진다. ‘판결문’, ‘지적원도’, ‘특정 기록물’ 등 구체적인 문서나 자료의 명칭이 중심이 되며, 정보의 성격보다는 형태와 존재 여부, 열람 가능성 등 기록물 자체에 대한 접근이 주된 내용이다. 반면 FAQ 데이터는 국가기록원의 기록물 관리 업무에 대한 실무자 질의를 바탕으로 구성되어 있으며 주로 보존기간, 이관 기준, 폐기 절차, 생산 등록 요건 등 법령 및 행정 지침에 기반한 구조적 질의가 다수를 차지한다. 이러한 질의는 단순한 정보 요청을 넘어 행정적 판단이나 규정 해석을 요구하는 경우가 많으며 복합 조건에 따른 절차 안내가 필요한 특성을 보인다.

<표 1> 국가기록원 정보공개청구 제목 데이터 개수

데이터 종류	정보공개청구 제목 목록 원본데이터	정보공개청구 제목 목록 분석 대상 데이터 (동사)
개수	15,897	8,531

<표 2> 국가기록원 기록관리 FAQ 데이터 개수

데이터 종류	FAQ 원본 데이터	FAQ 분석 대상 데이터 (동사)
개수	161	299

<표 1>과 <표 2>는 각각 본 연구의 분석 대상인 정보공개청구 자료와 FAQ 자료의 통계로, 먼저 <표 1> 정보공개청구 데이터는 2021년부터 2024년까지 국가기록원에 접수된 정보공개청구 목록을 입수, 총 15,897건의 청구 제목 데이터를 확보하였다. 이 중 단어 수준의 명사나 의미 불명확한 제목, 정보 요구를 표명하는 명사나 동사를 포함하지 않은 경우, 의미역 분석이 불가능한 항목 등을 제외하였다. 구체적인 선별 항목으로는 첫째, 문장성이 결여된 단어 수준의 제목으로 단일 명사 혹은 키워드로 구성되어 문장의 구조를 식별할 수 없는 경우이다. ‘연혁’, ‘근거’, ‘자료’, ‘기록물’ 등을 예시로 들 수 있다. 둘째, 중심 동사가 결여되어 의미역 분석이 불가능한 경우로 ‘관련 업무 처리 기준’, ‘등록 서식’, ‘정보 공개’ 등이 있다. 셋째, 비정형 데이터 및 시스템 오류로 특수기호 혹은 공개된 정보 중 마스킹처리가 되어있는 정보이다. 이는 국가기록원의 정보공개 신청 결과 개인정보에 해당하는 부분이 포함되어 있어 ‘.’, ‘##’, ‘***’ 등으로 마스킹(masking) 처리되어 있는 정보를 말한다. 본 연구에서는 이같은 사례(46.3%)를 제외하고 최종적으로 8,531건을 분석 대상으로 선정하였다.

<표 2>의 국가기록원 FAQ 데이터는 기록원 공식 웹사이트의 FAQ 페이지로부터 자체 개발한 크롤링을 통해 총 161건의 항목을 수집하였다. 이 중 상당수는 한 항목 내에 다수의 문장이 복합적으로 포함되어 있었으며, 세분화된 이용자 정보 요구를 파악하기 위해 하나의 질문 내에 여러 개의 동사가 포함된 경우, 각 문장을 단문으로

분리, 최종 분석 대상을 299건으로 확정하였다. 이와 같은 전처리 결과는 데이터 유형 간의 구조적 차이를 반영한 것으로 정보공개청구 제목은 상대적으로 간결하고 단일 목적의 요청이 많은 반면, FAQ는 복문 형태로 구성되어 있어 동사 단위로 분리한 후 의미역 중심 분석이 가능했던 사례가 많았다는 것을 시사한다.

3.2 논항구조 기반 의미역 분석

본 연구에서는 의미역 분석을 위해 한국전자통신연구원(ETRI)이 개발한 Exobrain 언어분석 Open API를 활용하였다. ETRI 발표에 따르면, Exobrain Open API는 기존 대비 형태소 분석과 개체명 인식 정확도를 각각 96.8%, 89.4%까지 향상시켰으며, 구어체 대응 모델에서는 분석 오류가 40% 이상 감소하는 성능 개선을 나타냈다(박주영, 2020). 이처럼 Exobrain은 한국어 기반의 다양한 문장 유형에 대해 안정적인 분석 성능을 제공하며 널리 활용되고 있는 자연어처리 도구로, 특히 구어체분석을 위해 전이학습(transfer learning) 및 데이터 증강(data augmentation) 기법을 적용하고 있다. 본 연구에서는 이 같은 범용성 및 높은 정확도라는 장점을 활용, 기록관리 분야의 질의 의미역 특성을 파악하고자 하였다. 다만 본 연구 도메인에서 나타나는 행정 전문 용어나 축약 표현, ‘기록관리’ 영역의 고유한 용어 체계와 복합적인 맥락을 포함한 질의 구조에 대한 정밀한 대응에는 한계가 있어 이를 위해 사용자 사전 및 API 적용 이후 복합명사 처리 등의 후처리 과정을 수행하여 보완하였다.

<표 3> 논항구조분석 결과 데이터 개수 및 비율

구분	분석 데이터 (중복 제외 동사 수)	ARG0 (행위 주체)	ARG1 (대상)	ARG2 (수혜자)	ARGM-CND (조건)	총계
정보공개청구 제목	8,531 (433)	631(7.4%)	4,898(57.4%)	2,004(23.5%)	402(4.7%)	7,935(93%)
기록관리 FAQ	299 (144)	52(17.4%)	252(84.3%)	85(28.4%)	25(8.4%)	414(138.5%)

<표 3>은 수집된 질의 두 데이터 집합에 대한 논항구조분석 결과로, 중심 동사 및 주요 논항(ARG0, ARG1, ARG2, ARGM-CND)을 기준으로 집계하였다. 논항별 총계가 100%를 초과하는 것은 하나의 질의문에 동일한 유형의 논항이 복수로 나타나거나, 하나의 동사에 여러 개의 논항을 포함하는 다중 논항 구조를 형성하는 경우가 있기 때문이다.

두 질의 집합은 ARG0(행위주체)의 표현 방식에서 차이를 보였다. 정보공개청구 제목에서는 청구 주체인 주어 생략이 빈번하여 ARG0 비율이 7.4%에 그쳤으나 기록관리 FAQ에서는 17.4%로 상대적으로 높게 나타났다. 이는 정보공개청구는 일반적으로 이용자가 본인의 지칭을 생략하고 간결한 정보 요청 형태로 작성되는 반면, 실무자 질의는 ‘기록물 등록자’, ‘관리기관’, ‘생산부서’ 등 업무 수행의 주체와 책임 소재를 명시적으로 기술하는 경향을 보이기 때문이다. 이러한 특성은 문장의 구조적 완성도와 논항 구성의 복잡성 측면에서도 두 데이터 간의 상이한 성격을 드러낸다.

두 질의 데이터 모두 구체적인 정보 요구 대상 ARG1(대상)의 비율이 높게 나타났으나, 그 활용 방식과 내용에는 명확한 차이가 있었다. 정보공개청구 제목에서 ARG1은 주로 ‘판결문’, ‘기록물’, ‘지적원도’ 등 청구 대상을 구체적으로 명시한 반면, 기록관리 FAQ에서는 ARG1이 업무 행위의 판단 또는 수행 대상으로 기능하였다. 기록관리 FAQ의 경우 ‘보존기간’, ‘이관 기준’, ‘등록 절차’ 등 실무적 판단이 요구되는 요소들로 구성되는 특성을 보였으며, 이는 해당 질의가 단순한 정보 요청을 넘어 기준 해석과 절차적 판단을 포함하는 복합적 성격을 지님을 의미한다.

또한, ARG2(수혜자 또는 시작점)는 정보공개청구 제목에서 23.5%, 기록관리 FAQ에서 28.4%로 나타났다. 예를 들어, “국립현대미술관 과천관 40주년 재개관 연구에 필요한 자료 요청합니다.”라는 질의에서는 ‘국립현대미술관 과천관 40주년 재개관 연구’가 요청 행위의 출발점이 되는 맥락으로 작용하며 ARG2로 분류되었다. 마찬가지로 “특정 기록물군을 특정 업무담당자에게만 열람하게 할 수 있나요?”와 같은 질의에서도 ‘특정 업무담당자’가 열람 행위의 방향이나 시작 대상이 되어 ARG2로 할당되었다. 이처럼 ARG2는 단순한 수혜자를 지칭하는 데 그치지 않고, 행위의 목적지나 출발하는 지점 또는 맥락적 배경을 나타내는 의미역으로 기능한다.

조건 논항인 ARGM-CND는 정보공개청구 제목에서 4.7%, 기록관리 FAQ에서 8.4%로 기록관리 FAQ에서 상대적으로 높은 비율을 보였다. “기관 담당자의 확인이 있어야 청구가 가능한가요?”라는 질의에서처럼, ‘기관 담당자의 확인’은 특정 행위의 전제가 되는 조건으로 기능하며 ARGM-CND로 분류되었다. 이러한 조건 표현은 실무자의 질의가 규정 해석이나 절차 기준 판단 등 조건 기반의 행정적 의사결정을 포함하고 있음을 나타내며, 특히 정보 제공의 가능 여부가 조건 충족 여부에 따라 달라지는 기록정보서비스의 특성과도 밀접한 연관이 있다.

3.3 질문 유형 도출을 위한 중심 동사 분석

<표 4> 이용자 질의 유형에 따른 고빈도 Top 10 동사 및 의미역

구분	동사 종류	사용 횟수	동사별 의미역(Argument Role)	주요 의미역
정보 공개 청구 고빈도 동사	요청	1503(17.6%)	ARG0(요청하는 사람), ARG1(요청하는 자료/정보)	ARG1
	청구	1256(14.7%)	ARG0(청구하는 사람), ARG1(청구하는 정보/자료)	ARG1
	신청	774(9.1%)	ARG0(신청자), ARG1(신청 대상)	ARG1
	알	114(1.3%)	ARG0(알고자 하는 사람), ARG1(알고 싶은 정보)	ARG1
	받	108(1.2%)	ARG0(받는 사람), ARG1(받는 대상)	ARG1
	관련	104(1.2%)	ARG1(관련된 정보/주제)	ARG1
	찾	78(0.9%)	ARG0(찾는 사람), ARG1(찾는 대상)	ARG1
	확인	58(0.7%)	ARG0(확인하는 사람), ARG1(확인 대상)	ARG1
	생산	49(0.6%)	ARG0(생산 주체), ARG1(생산된 자료)	ARG1
	열람	44(0.5%)	ARG0(열람하는 사람), ARG1(열람 대상)	ARG1
FAQ 고빈도 동사	이관	11(3.7%)	ARG0(이관하는 주체), ARG1(이관할 대상), ARG2(이관 장소/기관)	ARG2
	가능	10(3.3%)	ARG0(주체), ARGM-CND(조건)	ARGM-CND
	알리	10(3.3%)	ARG0(정보 제공자), ARG1(정보 내용), ARG2(정보 수신자)	ARG2
	보존	8(2.3%)	ARG0(보존하는 주체), ARG1(보존되는 대상)	ARG1
	생산	7(2.3%)	ARG0(생산 주체), ARG1(생산되는 대상)	ARG1
	포함	7(2.3%)	ARG1(포함하는 대상), ARG2(포함되는 내용)	ARG1
	등록	5(1.7%)	ARG0(등록하는 주체), ARG1(등록 대상)	ARG1
	도입	4(1.3%)	ARG0(도입하는 주체), ARG1(도입 대상)	ARG1
	설치	4(1.3%)	ARG0(설치자), ARG1(설치되는 시스템/장비)	ARG1
	작성	4(1.3%)	ARG0(작성자), ARG1(작성하는 문서)	ARG1

본 절에서는 이용자 질의에 나타난 중심 동사의 유형과 의미역 구성을 파악하기 위해 TF-IDF 가중치 기법을 활용(최유리, 오효정, 2025), 특정 질문 유형에 핵심적으로 작용하는 동사의 중요도를 정량적으로 반영하여 핵심

동사를 선별하였다. <표 4>는 <표 3>의 전체 질의 집합에서 자주 등장한 동사를 기준으로 상위 10위에 해당하는 동사 종류, 사용 횟수, 동사별 의미역, 주요 의미역을 정리한 것이다. 표에서 나타나듯이 지난 4년간 정보공개청구 질의에서 나타난 전체 433개 동사 중 상위 10위 이하 동사는 그 출현 빈도가 0.5% 미만으로, FAQ 질의의 경우 전체 144개 중 상위 10위 이하 동사는 1.3% 미만에 해당하는 것으로 파악되었으며 이는 선별된 중심 동사가 전체 질의 영역에서 대표성을 띠는 의미를 의미한다.

일반이용자가 주로 질의한 정보공개청구에서는 ‘요청’, ‘청구’, ‘신청’ 등 정보 접근을 위한 행위 동사가 상위를 차지하였다. 이들 동사는 모두 ARG1(대상)을 핵심 의미역으로 하며, 구체적인 기록물이나 정보를 지정하는 구조를 보인다. ‘전자문서’, ‘판결문’, ‘회의록’ 등 특정 문서 형태의 기록물을 직접 요구하는 패턴이 주를 이루며, ‘무엇을 요청하는가’에 초점이 맞춰진 특징을 나타낸다. 실무진 기록관리 FAQ에서는 ‘이관’, ‘보존’, ‘등록’, ‘작성’ 등 기록물 관리 절차나 조치를 나타내는 동사가 상위에 위치하였다. 특히 ‘이관’, ‘알리’ 동사의 경우 ARG2(목적지/수신자)가 핵심 의미역으로 나타나며, ‘가능’ 동사는 ARGM-CND(조건)을 중심으로 구성되는 특성을 보인다. 이는 업무 처리 과정에서 조건이나 절차적 판단이 중요하게 작용함을 의미한다.

분석 결과 정보공개청구에서 ARG1은 접근하고자 하는 ‘정보 자체’를 지시하는 반면, 기록관리 FAQ에서는 관리나 처리가 필요한 ‘업무 대상’을 의미한다. 이러한 차이는 일반 이용자가 정보의 ‘이용’을, 실무자가 정보의 ‘관리’를 중심으로 질의를 구성한다는 본질적 특성을 반영한다. 이러한 동사별 의미역 분석 결과는 후속 질문 유형 분류 및 응답 설계에서 핵심적인 기준으로 활용되었다.

4. 질문 유형 기반 챗봇 응답 구조 설계

본 장에서는 앞서 3장에서 도출한 이용자 질의 유형 분석 결과를 바탕으로 이용자 질의 유형 분류체계를 수립하고 이를 토대로 기록정보서비스에 특화된 응답 논리 구조 설계 방안을 제안한다. 논항구조 기반 분석 결과를 실질적인 설계 요소로 전환하여 질문 유형별 응답 시나리오와 템플릿을 구성함으로써 챗봇의 응답 일관성과 정보제공의 정확성을 높이는 것이 본 장의 핵심 목표이다.

4.1 질문 유형 분류 체계 수립

앞서 설명한 바와 같이 실제 이용자 질의문에서 사용된 동사의 종류는 매우 다양하였으나 그중 상당수는 출현 빈도가 극히 낮아 개별 유형으로 분류하거나 독립적으로 분석하기에는 한계가 있었다. 따라서 본 연구에서는 <표 4>의 이용자 질의문의 논항구조분석 고빈도 결과를 기반으로, 중심 동사의 출현 빈도, 의미적 유사성, 그리고 논항 구조 내 기능적 역할을 <표 5>와 같이 여섯 가지의 대표 유형으로 범주화하였다. 기타 동사들은 해당 유형 중 가장 근접한 의미 구조에 따라 통합하는 방향으로 포괄하고자 하며, 이는 유형의 분산을 최소화하고 응답 설계의 실용성을 확보하기 위한 전략적 기준에 따른 것이다.

<표 5>는 기록정보서비스에 특화된 여섯 가지 질문 유형(intent)을 정리한 것으로, 각 의미역이 기록관리 업무 내에서 어떤 정보 요소와 의미적으로 대응되는지를 예시함으로써 기록정보서비스 도메인 적용에 대한 이해를 높이고자 하였다. 이는 기존의 단순히 특정 키워드의 출현 여부나 명사 중심의 문장 패턴에 기반한 기존 질문 분류 방식과 달리, 기록정보서비스 영역에서 빈번하게 요청되는 중심 동사와 주요 의미역의 결합 패턴을 기준으로 분류함으로써 질의에 내재된 의도와 의미 구조를 중심으로 유형화했다는 점에서 차별성을 갖는다.

<표 5> 동사 범주화 및 분류체계 수립

유형(Intent)		설명	대표동사	주요 의미역	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 발령 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

도출된 여섯 가지의 질문 유형은 이용자의 역할과 정보 요구 목적에 따라 크게 두 그룹으로 구분된다. 일반 이용자 중심의 정보 요구형(정보 요청형, 사실 확인형, 정보 열람형)과 기록관리 실무자 중심의 업무 지원형(업무 처리형, 조건 확인형, 권한 설정형)으로 나뉘며, 각각 서로 다른 응답 구조와 정보 제공 방식을 요구한다.

먼저 정보 요구형의 ① 정보 요청형은 구체적인 기록물이나 정보에 대한 직접적인 요구를 나타내며, ‘요청’, ‘청구’, ‘신청’ 등의 동사와 ARG1(대상)이 핵심을 이룬다. 이 유형의 질의는 명확한 목적성을 가지고 있어 응답 시 해당 정보의 존재 여부와 제공 방법에 초점을 맞춘 구조가 필요하다. ② 사실 확인형은 특정 사실이나 현황에 대한 확인을 목적으로 하며, ‘확인’, ‘알’ 등의 동사를 중심으로 구성된다. 정보 요청형과 유사하게 ARG1을 핵심 의미역으로 하지만, 정보 제공보다는 사실 여부의 확인에 중점을 둔다는 점에서 차이를 보인다. ③ 정보 열람형은 열람 가능성이나 절차에 대한 질의로, ‘열람’, ‘알리’ 동사와 함께 나타난다. 이는 단순한 정보 요청을 넘어 접근 방법이나 절차적 안내가 필요한 유형으로, 응답 설계 시 단계별 안내 정보가 포함되어야 한다.

업무 지원형의 ④ 업무 처리형은 기록관리 실무와 직접 관련된 절차나 조치에 대한 문의로, ‘이관’, ‘보존’, ‘등록’ 등의 전문적 동사가 사용된다. 이 유형은 ARG1과 ARG2를 모두 활용하는 복합적 구조를 보이며 업무 절차의 복잡성을 반영한 다층적 응답이 요구된다. ⑤ 조건 확인형은 특정 조건 하에서의 가능성이나 적용 여부를 묻는 질의로, ‘가능’ 동사와 ARGM-CND(조건)이 핵심을 이룬다. 이 유형은 조건별 분기 응답이 필요하며, 다양한 상황에 따른 시나리오 설계가 중요하다. ⑥ 권한 설정형은 열람이나 접근 권한과 관련된 질의로 ‘설정’, ‘포함’ 등의 동사와 ARG1을 중심으로 구성되며 권한 확인과 관련된 절차 안내가 함께 제공되어야 하는 특성을 가진다.

이후 질문 유형에 따라 실제 이용자 질의문을 분석하고 중심 동사와 의미역의 조합을 통해 질의의 핵심 의미를 도출하는 과정을 수행하였다. 분석은 질의문에 포함된 주요 동사와 그에 수반되는 의미역 간의 연결 관계를 중심으로 핵심 의미를 정제하는 방식으로 이루어졌다. 특히 ARG1은 질문자가 실제로 요구하는 정보, 문서, 결과물 등을 지칭하는 항목으로 질문의 중심 대상 식별에 효과적이었다. 또한 유사한 의미역을 공유하는 질문 유형 간에도, 질의 목적과 응답 흐름의 구조적 차이를 기준으로 세분화하였다. 예컨대 ① 정보 요청형은 구체적인 정보 입수를 전제로 한 수령 의도를 담고 있는 반면, ③ 정보 열람형은 정보 존재 여부와 접근 가능성의 확인이 목적이므로 응답 구조에서 요구하는 정보 유형과 단계가 상이하다. 마찬가지로, ④ 업무 처리형은 기록관리 업무 절차 자체에 대한 문의이며, ⑤ 조건 확인형은 특정 조건 충족 시 절차 수행 가능 여부를 묻는 유형으로, ARGM-CND와 같은 부가적의 유무에 따라 의미적 해석과 응답 방식이 달라진다. 이러한 분류 기준은 응답 시나리오 템플릿에도 반영되어 있으며, 질문자의 내포 의도에 따라 응답 구조를 설계하는 데 있어 핵심 논리로 기능한다. 결과적으로 본 분석은

표면적 질의 형식이 아닌 질의 내 의미역과 핵심 의미의 구조적 연결을 통해 실질적인 이용자 요구를 식별하고, 챗봇 응답 설계의 정밀성을 제고하는 기반 정보를 제공하였다. 이러한 의미 도출 과정은 후속 응답 템플릿 설계 단계에서 질문 유형별 필수 정보 요소를 정의하는 근거로 활용된다.

<표 6> 이용자 질의문 기반 의미역 및 자연어 이해 키워드 도출

유형(Intent)	질의문	주요 동사	주요 의미역	중심 의미	자연어 이해 키워드
① 정보 요청형	(1) 지가증권발급 관련서류 신청합니다.	신청	ARG1	관련서류	신청, 관련서류
	(2) 폐지기관의 이관 기록물 현황 등 요청합니다.	요청	ARG1	이관 기록물 현황	요청, 이관 기록물 현황
	(3) 토지조사부에 관한 정보공개를 청구합니다.	청구	ARG1	정보공개	청구, 정보공개, 토지조사부
② 사실 확인형	(1) 서류의 보존연한을 알고 싶습니다.	알	ARG1	보존연한	알, 보존연한
	(2) 공공기관의 기록물관리전문요원 현황을 알고 싶습니다.	알	ARG1	기록물관리전문요원 현황	알, 공공기관, 기록물관리전문요원 현황
	(3) 지적원도를 확인하고 싶습니다.	확인	ARG1	지적원도	확인, 지적원도
③ 정보 열람형	(1) 기록물 생산현황 통보 절차를 알려주세요.	알리	ARG1	기록물 생산현황 통보 절차	알리, 기록물 생산현황 통보 절차
	(2) 입야사정기록을 열람하고 싶습니다.	열람	ARG1	입야사정기록	열람, 입야사정기록
	(3) 교원발령 기록물을 열람하고 싶습니다.	열람	ARG1	교원발령 기록물	열람, 교원발령 기록물
④ 업무 처리형	(1) 지방자치단체에서 생산한 기록물은 어떤 법령에 의해 관리 및 보존되고 폐기되나요?	보존/폐기	ARG1	관리, 기록물	관리, 보존, 폐기, 기록물, 법령
	(2) 전자문서시스템에서 이관된 기록물의 접근범위 설정은 어떻게 하나요?	이관	ARG1, ARG2	접근범위 설정	이관, 접근범위 설정, 전자문서시스템
	(3) 타 기관으로부터 이관 안내받았습니다.	이관	ARG2	타 기관	이관, 타 기관
⑤ 조건 확인형	(1) CRMS 검색 기능을 통해 소속기관의 기록물 검색이 가능한가요?	가능	ARGM-CND	기록물 검색	가능, 검색, 기록물
	(2) 전자기록생산시스템에서 개인정보가 포함된 문서를 첨부하는 경우 암호화가 가능한가요?	가능	ARG1, ARGM-CND	암호화, 개인정보 포함 문서 첨부	가능, 개인정보, 암호화
⑥ 권한 설정형	(1) 비공개 기록물이 공개로 설정되어 RMS로 이관된 경우는 무엇인가요?	설정	ARG1	공개 설정	설정, 기록물, 공개설정
	(2) 개인정보가 포함된 행정정보 데이터세트는 개인정보 보호법과 공공기록물법 중 어떤 법령에 따라 폐기해야 하나요?	포함	ARG1	개인정보	포함, 개인정보, 행정정보 데이터세트, 법령

<표 6>은 정보공개청구 제목 및 기록관리 FAQ 데이터에 포함된 이용자 질의문을 여섯 가지 질문 유형에 따라 분류하고, 각 질의에서 중심 의미를 도출한 뒤, 이를 기반으로 자연어 이해에 필요한 핵심 키워드를 정리한 결과이다. 예를 들어 ①-③ “토지조사부에 관한 정보공개를 청구합니다.”라는 질의의 중심 동사는 ‘청구’이며, ARG1은 ‘정보공개’로 분석된다. 중심 의미는 ‘정보공개’이며, 자연어 이해 키워드로는 ‘청구’, ‘정보공개’, ‘토지조사부’가 도출된다. 이와 같은 구조는 정보 요청형에서 주로 나타나는 양상으로, 중심 동사와 ARG1 논항이 비교적 명확하게 구성된 단순한 의미 구조를 보인다.

반면 ④-② “전자문서시스템에서 이관된 기록물의 접근범위 설정은 어떻게 하나요?”와 같은 질의는 보다 복잡한 구조를 보이며, 이 질의의 중심 동사는 ‘이관’이고 ARG1은 ‘기록물’, ARG2는 ‘접근범위 설정’으로 분석된다.

중심 의미는 ‘접근범위 설정’이며, ‘이관’과 ‘접근범위 설정’이 자연어 이해를 위한 주요 키워드로 도출된다. 이처럼 복수의 논항이 핵심 정보로 작용하는 구조는 단일 논항 중심의 단순 질의와는 구별되는 특성을 보인다. 이러한 자연어 이해 키워드는 각 질의에서 의미적으로 핵심이 되는 개념어로, 질문의 의도를 이해하고 분류하는 데 중요한 단서로 작용한다.

특히 단순 질의는 하나의 중심 동사와 논항만으로도 질문의 의미가 명확히 드러나지만, 복합 질의는 중심 의미를 구성하는 다수의 논항이 함께 작용하며 더 정교한 해석이 요구된다. 실제로 서비스 환경에서는 ‘열람 요청합니다.’, ‘정보 확인 신청합니다.’ 등과 같이 하나의 문장 내에 두 개 이상의 동사 또는 의미적 표현이 혼재되는 경우가 발생한다. 이처럼 중복되거나 유형 간 경계가 모호한 자연어 표현에 대응하기 위해 본 연구는 중심 동사의 빈도 기반 대표성을 기준으로 질문 유형의 우선 분류 기준을 설정하였다. 예컨대 “열람 요청합니다”와 같은 중의적 표현에서는 ‘열람’과 ‘요청’ 두 동사가 함께 등장하지만, 실제 이용자 질의문에서는 ‘요청’ 동사의 사용 빈도가 상대적으로 높고 의미역 ARG1이 정보 제공 요청의 대상으로 반복 구조화되는 특성이 나타나므로 해당 질의를 정보 요청형으로 분류한다. 이와 같은 기준은 표현 방식이 복잡적이거나 다양하게 변형된 질의에 대해서도 일관된 유형 분류를 가능하게 하며, 향후 자동 분류 시스템 구현을 위한 실용적 기반으로 기능한다.

<표 6>은 질문 유형 분류 체계를 기반으로 실제 이용자 질의문을 구조적으로 분석한 결과를 통해 각 질의의 중심 동사와 의미역 구성, 중심 의미, 자연어 이해 키워드를 통합적으로 제시함으로써 질의 구조에 대한 해석과 비교를 가능하게 한다는 점에서 분석적 의의가 있다. 특히 유형별로 중심 동사와 주요 의미역의 조합이 어떻게 형성되는지 문장 단위에서 확인할 수 있어 유형 분류 기준의 적용 타당성을 검증하는 데 유용하다. 또한 동일 유형 내에서도 다양한 동사나 명사 표현이 어떻게 중심 의미로 수렴되는지를 보여주며 질문 해석에서 발생할 수 있는 인식 오류나 유형 간 경계 모호성에 대한 판별 기준을 제공한다. 도출된 키워드는 실제 서비스에서 입수된 질의를 기반으로 수집되었기 때문에 향후 자연어 처리 시스템 학습이나 시나리오 설계 시 실제 어휘 사용 패턴에 관한 기초자료로서도 활용 가능하다.

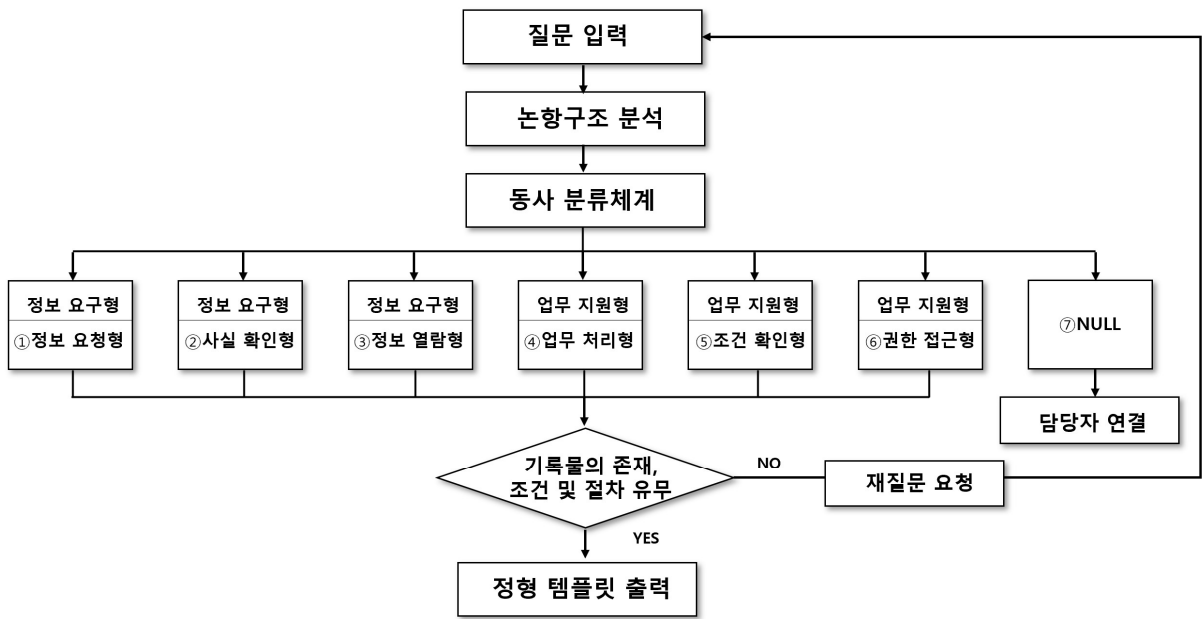
4.2 응답 논리 구조 설계 프로세스

본 절에서는 기록정보서비스 챗봇의 응답 구조를 설계하기 위한 전체 프로세스를 설명한다. 이 과정은 단순히 질문에 대응하는 정보를 검색, 반환하는 것이 아니라, 질의에 내포된 구조적 의미를 해석하고 이를 바탕으로 정형화된 응답을 반환하는 절차로 구성된다. <그림 4>는 이용자의 자연어 질의가 시스템을 거쳐 최종 응답으로 전환되기까지의 과정을 단계별로 시각화한 것이다.

첫 번째 단계에서는 이용자가 자유롭게 입력한 질의문이 시스템에 전달된다. 이후 두 번째 단계에서 입력 문장은 논항구조분석을 거치며, 중심 동사와 함께 ARG0, ARG1, ARG2, ARGM-CND 등 의미역 정보가 자동으로 추출된다. 이 분석은 질문의 구조를 해석하고 의미를 파악하는 기반이 되며 이후 유형 분류의 핵심 입력값으로 작용한다.

세 번째 단계에서는 추출된 동사와 의미역 조합을 바탕으로 질문 유형을 분류한다. 본 연구에서는 여섯 가지 질문 유형(정보 요청형, 사실 확인형, 정보 열람형, 업무 처리형, 조건 확인형, 권한 설정형)을 사전에 정의하고, 이에 따라 분류 작업을 수행하였다. 질의가 구조적으로 불분명하거나 자동 분류가 어려운 경우는 예외적으로 NULL로 처리되며, 이 경우 담당자 연결 방식의 응답이 제공된다. 이는 시스템 응답의 오류를 줄이고 사용자 만족도를 높이기 위한 보완 경로를 제공하는 장치이다.

네 번째 단계는 요청된 기록물의 존재 여부 및 관련 조건·절차의 충족 여부를 확인하는 절차이다. 특히 정보



<그림 4> 응답 설계 프로세스

요청형이나 사실 확인형, 정보 열람형과 같은 정보 요청형 질문 유형의 경우, 요청 대상 기록의 실존 여부에 따라 응답 내용이 달라지기 때문에 이 절차는 응답 정확도를 결정짓는 중요한 요소로 작용한다. 만약 해당 기록이 존재하지 않거나 접근이 불가능한 경우에는 단순히 부존재 사실을 통보하는 데 그치지 않고, 대체 정보 제공, 유사 자료 안내, 재질문 유도 등의 방식으로 응답 흐름을 전환함으로써 서비스의 연속성과 이용자 만족도를 함께 제고할 수 있다.

한편, 업무 지원형 질문 유형의 경우에도 응답 결정에 앞서 관련 조건이나 절차의 충족 여부를 확인하는 과정이 수반된다. 이들 유형은 기록물의 실존 여부보다는 해당 행위의 처리 가능성, 접근 기준, 권한 부여 요건 등의 판단에 따라 응답 흐름이 결정된다. 이에 따라 조건 또는 절차가 충족되는 경우에는 정형 템플릿을 기반으로 응답을 제시하고, 그렇지 않은 경우에는 유사 사례 안내, 대체 절차 제공, 재질문 유도 등의 방식으로 흐름을 전환하도록 설계하였다. 이는 업무 지원형 질문의 맥락적 특성을 반영한 설계로, 유형별 응답의 정합성과 서비스의 실효성을 높이는 데 기여한다.

마지막 단계에서는 질문 유형에 따라 사전에 정의된 응답 템플릿이 선택되어 최종 응답이 자동으로 생성된다. 각 템플릿은 중심 동사와 의미역 정보를 반영하여 구성되며, 정보 제공, 절차 안내, 조건 확인 등 목적에 따라 응답 양식이 달라지도록 설계하였다. 이와 같은 응답 설계 절차는 이용자 질의를 구조적으로 분석하고, 유형별 응답 시나리오를 바탕으로 맞춤형 정보를 제공할 수 있는 기반을 마련한다. 특히 조건 확인, 권한 접근 등 복합적인 해석이 필요한 질문에 대해 일관된 응답 흐름을 확보함으로써, 기록정보서비스 챗봇의 신뢰성과 실효성을 높이는 핵심 요소로 기능한다. 본 연구에서는 템플릿 기반의 응답 구조 설계 중점을 두었으나, 향후 시스템 구현 단계에서는 이용자 피드백을 반영한 동적 응답 흐름 설계도 함께 고려할 계획이다. 예를 들어 사용자가 제공된 응답에 대해 ‘불충분’ 또는 ‘재질문’과 같은 반응을 보일 경우, 질문 유형을 재판단하거나 보완 질의를 유도하는 방식의 유형 전환 흐름을 설계할 수 있다. 이러한 동적 구조는 후속 구현 및 평가 단계에서 실용성과 확장 가능성을 함께 검증해 나갈 예정이다.

4.3 응답 유형별 시나리오 구성

본 절에서는 질문 유형에 따라 어떻게 응답 시나리오가 구성되는지를 구체적으로 설명한다. <표 7>은 본 연구에서 도출한 여섯 가지 질문 유형을 기반으로, 질문의 구조적 특성과 응답 목적에 따른 시나리오 구성 방식을 템플릿으로 정리한 것이다. 각 시나리오는 논항구조분석 결과로 도출된 중심 동사(Verb)와 주요 의미역으로 구성된 자연어 이해 키워드를 바탕으로 설계되며, 질문의 기능적 특성과 응답의 목적에 따라 포함되는 구성 요소가 달라진다. 정보 요청형은 주로 ARG1(대상)의 존재 여부를 중심으로 응답을 구성하며, 사실 확인형은 특정 사건이나 기록의 진위를 판단하기 위해 ARG1과 부가적인 상황 설명이 포함되는 구조를 취한다. 업무 처리형의 경우에는 ARG0(행위 주체)와 ARG2(수혜자 또는 시작점)를 중심으로, 처리 경로 안내나 담당 부서 연결과 같은 응답을 유도한다. 조건 확인형은 ARG1-CND(조건)의 충족 여부에 따라 대응 가능성을 안내하며, 권한 설정형은 특정 정보에 접근하기 위한 권한 조건을 기준으로 답변을 제공한다.

<표 7> 유형별 시나리오 템플릿

	기록 존재	기록 미존재
① 정보 요청형	[Verb]하신 [ARG1]에 대한 정보는 현재 총 [n]건이 검색되었습니다. 해당 자료는 [검색 페이지]에서 확인하실 수 있으며, 자료 신청은 [정보공개 포털] 또는 [담당 부서명]을 통해 진행하실 수 있습니다.	요청하신 [ARG1]은 본 기관에서 보유하고 있지 않거나, 해당 기록이 생성되지 않아 제공이 어렵습니다.
② 사실 확인형	[ARG1]과/와 관련된 기록은 [확인된 출처/문서명]에서 확인되었습니다. 현재 확인된 건수는 총 [n]건입니다. 보다 자세한 내용은 [출처]을/를 통해 확인하실 수 있습니다.	[ARG1]과 관련된 명확한 기록은 현재 확인되지 않았습니다. 관련 정보는 [기관명] 또는 [부서]를 통해 재확인하실 수 있습니다.
③ 정보 열람형	[ARG1]은/는 현재 열람이 가능한 자료로 확인되었습니다. 열람 방법은 [장소/시스템] 또는 [담당자/열람안내지침]을/를 통해 확인하실 수 있습니다.	[ARG1]은 현재 열람 가능한 자료가 확인되지 않았습니다.
④ 업무 처리형	해당 기록은 [ARG0]에 의해 생산되어, [ARG2]로 이관 또는 등록된 절차가 확인되었습니다. 처리 유형은 '[Verb]'로 분류됩니다. 본 기록은 현재 [ARG2]을/를 대상으로 처리되고 있습니다. 절차에 대한 자세한 사항은 [관련 지침/부서]을/를 참조해 주세요.	[ARG1]과 관련된 처리 이력 또는 절차 정보가 현재 확인되지 않았습니다. 관련 업무 여부는 [부서/지침]을 통해 확인 바랍니다.
⑤ 조건 확인형	해당 요청은 [ARG1-CND]을/를 충족할 경우 가능합니다. 조건이 만족될 경우, 관련 절차에 따라 처리가 가능합니다. 보다 정확한 내용은 [담당 부서/기관]에 문의해 주세요.	조건 확인 결과, [ARG1]과 관련된 자료는 현재 확인되지 않았습니다. 관련 처리 가능 여부는 [부서]를 통해 검토해 주세요.
⑥ 권한 설정형	[ARG1]은/는 접근 권한이 설정된 자료로, 해당 정보의 열람 가능 여부는 「관련 법령/지침」에 따라 결정됩니다. 접근 권한이 있는 경우에 한해 열람이 가능하며, 구체적인 열람 조건은 [기관/부서]의 안내를 확인해 주세요.	[ARG1]은 현재 접근 권한이 필요한 자료 중 열람 가능한 항목이 확인되지 않았습니다. 권한 여부 및 제한 기준은 [담당 기관]에 문의해 주세요.
NULL	입력하신 질문은 현재 시스템에 학습된 기준에 포함되지 않아 정확한 응답이 어렵습니다. 관련 용어나 문서명을 보다 구체적으로 입력해 주세요. 필요한 경우 [담당자] 연결을 통해 직접 안내를 받으실 수 있습니다.	

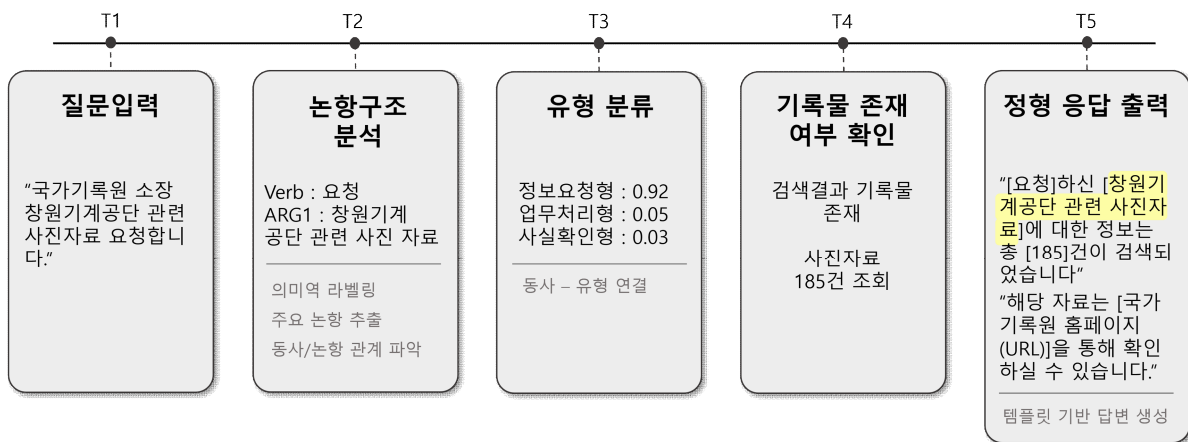
특히 각 시나리오는 기록의 존재 여부에 따라 ‘기록 존재’와 ‘기록 미존재’로 나누어 응답 경로가 달라지도록 설계하였다. 동일한 질문 유형이라 하더라도, 관련 기록이 존재하는 경우에는 정보 제공이나 절차 안내로 이어지며, 미존재 시에는 대체 정보 제공, 관련 부서 안내, 추가 질의 유도 등으로 후속 응답을 구성한다. 이러한 이중 구조는 실제 서비스 환경에서 자주 발생하는 다양한 상황을 반영하며, 이용자의 정보 요구를 보다 유연하고 효과적으로 지원하기 위한 전략으로 기능한다.

이와 별도로 ‘NULL’ 유형은 질문이 구조적으로 불분명하거나 중심 동사 및 의미역의 자동 분석이 어려운 경우에 해당된다. 이 경우에는 질문 재구성 요청이나 담당자 연결 안내와 같은 예외적 응답 시나리오가 제공되며 이는 시스템의 과잉 응답을 방지하고, 자동 응답의 한계를 보완하기 위한 장치로 활용된다. 이와 같은 응답 유형별 시나리오 구성 방식은 질문의 구조와 의도를 바탕으로 일관된 응답을 제공함으로써, 기록정보서비스의 전문성과 신뢰도를 높이는 데 기여한다. 특히 조건 확인형이나 권한 설정형처럼 해석이 복잡한 질문에도 안정적인 응답 흐름을 유지할 수 있어, 실제 서비스에서의 활용도와 시스템 이용자 이해도를 충족시킬 수 있다.

4.4 응답 생성 절차 적용 예시

본 절에서는 앞서 설계한 응답 시나리오가 실제 질의에 어떻게 적용되는지를 구체적으로 살펴보고자 한다. 이를 위해 “국가기록원 소장 창원기계공단 관련 사진자료 요청합니다.”라는 이용자 질의를 예시로, 자연어 형태의 질문이 각 처리 단계를 거쳐 정형 응답으로 생성되는 과정을 단계별로 설명한다. 이는 기록정보서비스 챗봇이 처리할 수 있는 정보 요청형 질의 중 하나로, <그림 5>는 해당 처리 과정을 도식화하여 정리한 것이다. 본 질의는 분석 대상 질의문에서 가장 높은 출현 빈도를 보인 중심 동사 ‘요청’을 포함하고 있어, 정보 요청형 시나리오 구조의 전형을 보여주는 사례로 선정하였다.

첫 번째 단계(T1)에서는 사용자가 자연어로 입력한 질의문이 시스템에 전달된다. 두 번째 단계(T2)에서는 논항구조분석이 수행되어 중심 동사로는 “요청”이, 의미역 ARG1(대상)으로는 “창원기계공단 관련 사진자료”가 자동 추출되며, 이를 통해 시스템은 질의의 핵심 행위와 정보 요구 대상을 구조적으로 인식하게 된다. 세 번째 단계(T3)에서는 분석된 중심 동사와 의미역의 조합을 바탕으로 질문 유형 분류가 이루어진다. 본 예시에서는 정보 요청형이 가장 높은 유사도를 보여 해당 유형으로 분류되었으며 이는 추후 응답 흐름 결정의 핵심 기준이 된다. 네 번째 단계(T4)에서는 정보 요청형 질문에 따라 요청한 정보의 기록물 존재 여부가 확인된다. 실시간



<그림 5> 질문 유형 분류 및 응답 자동화 과정 시각화

검색 결과, “창원기계공단 관련 사진자료”에 해당하는 기록이 총 185건 존재하는 것으로 조회되었으며, 이 결과는 응답 문장을 구성하는 데 핵심 정보로 활용된다.

다섯 번째 단계(T5)에서는 이전 단계에서 확인된 질문 유형과 기록물 존재 여부를 기반으로, 해당 유형에 적합한 정형화된 응답 템플릿이 선택된다. 본 예시에서는 관련 기록물이 185건 존재하므로 시스템은 이에 맞는 응답 문장을 자동 생성한다. 최종 응답에는 “[요청]하신 [창원기계공단 관련 사진자료]에 대한 정보는 총 [185]건이 검색되었습니다.”와 같은 요약 문장 외에도 사용자가 실제 자료를 열람하거나 청구할 수 있도록 사진자료 검색 페이지(URL)에 대한 안내 문구가 함께 포함된다. 이는 단순한 응답 제공을 넘어서 기록 검색 경로와 청구 절차 안내까지 통합함으로써, 이용자가 필요한 정보를 보다 정확하고 신속하게 확인할 수 있도록 설계된 것이다.

이와 같은 적용 예시는 단일 질의에 대한 자동화된 처리 과정을 거쳐 본 연구에서 제안한 논항구조 기반 응답 생성 방식의 실제 구현 가능성과 적용 효과를 보여준다. 이러한 사례를 바탕으로 의미 기반 응답 설계가 실제 서비스 환경에서 충분히 실현 가능하며, 자동 응답의 정밀도와 일관성을 높이는 데 기여할 수 있음을 확인할 수 있다. 아울러 본문에서는 여섯 가지 질문 유형 각각에 대해 <표 6>의 질의문 예시와 <표 7>의 응답 시나리오 템플릿을 통해 사전 구조화된 설계를 제시하고 있다. 향후 연구에서는 유형별 매핑 정확도, 응답 누락률 등의 정량적 지표를 활용하여, SRL 기반 응답 구조 설계의 효과성을 보다 체계적으로 검증해 나가고자 한다.

4.5 기록정보서비스 챗봇 설계의 전략적 시사점

본 연구에서는 지능형 챗봇의 적용 대상 기관으로 국가기록원을 선정, 실제 이용자로부터 접수된 정보공개청구 목록과 기록관리 FAQ 리스트를 수집·정제하여 실증적 연구 데이터셋을 구축하였다. 그 이유는 서두에서 밝힌 바와 같이 국가기록원은 국내에서 가장 다양하고 광범위한 질의가 집중되는 대표 기록관리기관으로, 일반 시민의 정보공개청구부터 전문 연구자의 학술 목적 열람, 행정기관 실무자의 업무 문의에 이르기까지 질의의 스펙트럼이 매우 넓고 복잡적이기 때문이다. 이러한 특성은 다른 기록관리기관에서도 접수될 수 있는 대부분의 질의 유형을 포괄할 수 있음을 의미하며, 본 연구에서 도출한 질문 유형 분류 체계와 응답 템플릿은 지방기록물관리기관, 대학 기록관, 기업 아카이브 등에서 선별적으로 적용하여 각 기관의 특성에 맞게 발전시킬 수 있는 확장성을 지님을 함의한다.

기술적 측면에서 본 연구는 논항구조분석의 적용을 통해 기록정보서비스 질의가 지닌 구조적 특성과 반복 패턴을 정밀하게 식별할 수 있음을 실증하였다. 중심 동사와 논항을 기반으로 질문의 핵심 의미를 유형별로 분류하고 그에 상응하는 응답 구조를 설계함으로써, 챗봇 시스템에서 요구되는 의도 분류, 응답 생성, 대화 흐름 설계 등 핵심 모듈 개발에 기초 자료를 제공하였다. 특히 반복적 질의 유형을 기반으로 설계된 응답 시나리오 템플릿은 챗봇 응답의 일관성을 확보하고 시스템의 유지보수와 기능 확장 측면에서도 높은 활용 가능성을 보여준다. 이상의 분석을 종합하여 본 연구는 기록정보서비스 분야에서 이용자의 정보 요구를 구조적으로 분석하고 이에 기반한 지능형 응답 체계를 설계하였다는 점에서 실무적 활용 가능성과 기술적 확장 가능성 모두를 확보한 사례로 평가할 수 있다. 다만 제한된 유형을 넘어서는 질문에 대한 유연한 대응을 위해 유사 정보서비스 기관의 사례로 확장, 부적절한 요구 및 기록원 업무를 벗어나는 요청에 대한 거절 등의 보완이 필요하다. 또한 이용자가 원하는 정보를 찾은 경우라 할지라도 그 결과의 신뢰성을 검증하기 위한 검색 증강 생성(RAG: Retrieval-Augmented Generation) 기법의 도입도 고려할 예정이다.

5. 결론

본 연구는 기록정보서비스 챗봇의 정밀한 응답 설계를 위한 기초 연구로서, 논항구조 기반 의미역 분석을 통해 질문 구조를 체계화하고 유형별 응답 템플릿을 설계하는 것을 목적으로 수행하였다. 연구 대상으로는 우리나라 기록관리 대표 기관인 국가기록원을 선정, 실제 이용자들로부터 요청된 질의문들을 입수함으로써 실증적인 연구 데이터셋을 구축하였다. 본 연구를 통해 기록정보서비스 질의를 분류하고 유형별로 요구되는 응답 구성 요소를 식별하여 기록 유무에 따른 이중 템플릿 구조를 설계하였으며, 이러한 구조는 이용자가 원하는 기록이 존재하지 않는 경우에도 대안적 정보나 서비스를 제공할 수 있어 이용자 만족도 향상에 기여할 수 있는 설계이다.

연구의 의의는 다음과 같다. 첫째, 짧고 압축된 정보공개청구 제목에서도 논리 구조를 추출하고, 실무자 질의와 일반 이용자 질의 간 표현 방식의 차이를 반영함으로써 구조 기반 응답 설계의 실증 가능성을 입증하였다. 둘째, 질문 유형에 따라 응답에 요구되는 구성 요소와 의미역 조합의 차이를 반영한 템플릿을 설계함으로써, 실제 적용 가능성이 높은 응답 설계 모델을 제시하였다. 셋째, 본 연구는 기록정보서비스에 특화된 대화형 챗봇 시스템 개발을 위한 구조적 설계 프레임워크를 마련함으로써, 향후 기술적 구현과 실무적 확산의 기반을 제공하였다.

다만 본 연구는 응답 구조 설계에 초점을 둔 기초 연구로, 실제 챗봇 구현 및 이용자 평가까지는 포함하지 못하였다는 한계를 지닌다. 향후 연구에서는 본 논문에서 설계한 질문 유형 분류 체계와 응답 시나리오 템플릿을 기반으로, 기록정보서비스 특화 AI 챗봇 시스템을 구현할 예정이다. 응답 생성 과정에서는 본 연구에서 도출한 중심 동사-의미역 조합 기반의 규칙을 적용하여, 입력 질의의 의미 구조를 분석하고 사전 정의된 템플릿에 따라 응답 항목을 자동 호출하는 방식으로 동작하도록 설계할 계획이다. 구현된 챗봇 시스템에 대해서는 정량적·정성적 평가를 병행할 예정이며, 정량 평가에서는 설계된 시나리오가 실제 질의와 얼마나 정확히 매핑되는지를 확인하기 위해 Coverage 기반 F1-score 등의 성능 지표를 활용하고자 한다. 정성 평가로는 기록관리 실무자 및 일반 이용자 집단을 대상으로 시나리오 기반 사용자 테스트를 실시하고 피드백을 수집하여 응답 로직을 지속적으로 보완해 나갈 계획이다.

참고문헌

- 강레림 (2020). 국가기록원 기록정보서비스 개선방안에 관한 연구. 석사학위논문, 이화여자대학교.
- 강윤아, 오효정 (2023). 전자기록관리 업무 및 기록정보서비스에서의 생성형 AI 기술 활용. 한국기록관리학회지, 23(4), 179-200. <http://doi.org/10.14404/JKSARM.2023.23.4.179>
- 국가기록원 (2024. 10. 17.). 2021-2023 국가기록원 정보공개청구 제목 및 공개결정 구분 (접수번호: 13182230). 대한민국 정보공개포털. 출처: <https://www.open.go.kr/>
- 국가기록원 (2025. 1. 22.). 2024 국가기록원 정보공개청구 제목 및 공개결정 구분 (접수번호: 13632801). 대한민국 정보공개포털. 출처: <https://www.open.go.kr/>
- 국가기록원 (발행년불명). 기록관리 FAQ. 출처: <https://www.archives.go.kr/next/newnews/innovationTFFAQList.do>
- 국회도서관 (발행년불명a). 지능형 법률 검색. 아르고스. 출처: <https://argos.nanet.go.kr/main/mobile/lawqnaMainGuest.do>
- 국회도서관 (발행년불명b). 지능형 질의응답. 국회전자도서관. 출처: <https://dl.nanet.go.kr/qas/qasMain.do>
- 김선옥 (2025). 대학기록관 챗봇 서비스 개발을 위한 RAG 적용: K대학교를 중심으로. 석사학위논문, 경북대학교.
- 대한민국 정책브리핑 (2025. 6. 20.). 이 대통령 "과감한 세계 혜택·규제혁신... AI시대 고속도로 구축". 출처:

<https://www.korea.kr/news/policyNewsView.do?newsId=148944726>

박은비 (2021). 국가기록원 모바일 기록정보서비스를 위한 챗봇 연구. 석사학위논문, 한남대학교.

박주영 (2020. 8. 6.). 국산 인공지능 '엑소브레인' 이제 구어체도 알아듣는다. 연합뉴스.

<https://www.yna.co.kr/view/AKR20200806036400063>

백지연, 오효정 (2019). 국가기록원 질의로그 빅데이터 기반 사용자 정보요구 유형 분석. 정보관리학회지, 36(4), 183-205. <https://doi.org/10.3743/KOSIM.2019.36.4.183>

양승진 (2024. 5. 15.). '경북도정 디지털 대전환' 결실...단순 반복업무 덜어주는 AI '챗경북'. 매일신문.

<https://www.imaeil.com/page/view/2024051215572453222>

윤난슬 (2024. 6. 3.). 전주시, 도서관 이용 더 편리하게 '챗봇 서비스' 시작. 뉴시스.

https://www.newsis.com/view/NISX20240603_0002758522

이선웅 (2019). 의미역 기술 모형 연구 및 시범 구축. (2019-01-11). 국립국어원.

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>

임수중, 권민정, 김준수, 김현기 (2015). ExoBrain을 위한 한국어 의미역 가이드라인 및 말뭉치 구축. 제27회 한글 및 한국어 정보처리 학술대회 논문집, 250-254.

전주도서관 (2024. 6. 3.). 전주시 도서관 챗봇 서비스 시행 안내 (24. 06. 03.~). 출처:

https://lib.jeonju.go.kr/board/view.jeonju?boardId=BBS_0000005&menuCd=DOM_000000106001001000&paging=ok&startPage=1&dataSid=228842

진주영, 이해영 (2018). 국가기록원 웹사이트 유입경로와 이용자 검색어 분석. 정보관리학회지, 35(1), 183-203.

<https://doi.org/10.3743/KOSIM.2018.35.1.183>

최유리, 오효정 (2025). 기록정보서비스 챗봇을 위한 이용자 질문 유형 분석 및 주요 의미 도출. 2025 한국기록관리학회 춘계학술대회 자료집.

한국기록학회 (2008). 기록학 용어 사전. 서울: 역사비평사.

한국전자통신연구원 (2024. 12. 6.). Exobrain 언어분석 Open API. ETRI 기술 사전 체험 플랫폼. 출처:

<https://epretx.etri.re.kr/apiDetail?id=2>

Freeman, E. T. (1984). In the Eye of the Beholder: Archives administration from the Users point of view. American Archivist, 47(2), 111-123.

Pearse-Moses, R. (2005). A Glossary of Archival and Records Terminology. Chicago: Society of American Archivists.

Pugh, M. J. (1982). The Illusion of Omniscience: Subject Access and the Reference Archivist. American Archivist, 45(1), 33-44.

Pugh, M. J. (2005). Providing reference services for archives & manuscripts. Chicago: SAA.

• 국문 참고자료의 영어 표기

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

Baek, Ji-yeon & Oh, Hyo-jung (2019). User Information Needs Analysis based on Query Log Big Data of the National Archives of Korea. Journal of the Korean Society for Information Management, 36(4), 183-205.

<https://doi.org/10.3743/KOSIM.2019.36.4.183>

Choi, You-ree & Oh, Hyo-jung (2025) A Study on User Question Types and Key Meaning Extraction for an Archival Reference Service Chatbot. Proceedings of the Spring Conference of the Korean Society of Archives and Records

Management.

- Electronics and Telecommunications Research Institute (2024, December 6). Exobrain Language analysis technology Open API. e-PreTX. Available: <https://epretx.etri.re.kr/apiDetail?id=2>
- Jeonju Libraries Hub (2024, June 3). Jeonju City Library Chatbot Service Implementation Guide (2024, June 6 ~). Available: https://lib.jeonju.go.kr/board/view.jeonju?boardId=BBS_0000005&menuCd=DOM_000000106001001000&paging=ok&startPage=1&dataSid=228842
- Jin, Ju yeong & Rieh, Hae-young (2018). Analysis of Users' Inflow Route and Search Terms of the Korea National Archives' Web Site. *Journal of the Korean Society for Information Management*, 35(1), 183-203. <https://doi.org/10.3743/KOSIM.2018.35.1.183>
- Kang, Yerin (2020). A Study on the Improvement of Archival Reference Service in the National Archives of Korea. Master's thesis, Ewha Womans University, Korea.
- 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>
- Korean Society of Archival Studies (2008). *Dictionary of records and archival terminology*. Seoul: yukbi.
- 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, Soojong, Kwon, Minjung, Kim, Junsu, & Kim, Hyunki (2015). Korean Proposition Bank Guidelines for ExoBrain. *Proceedings of the 27th Conference on Human and Language Technology*, 250-254.
- National Archives of Korea (2024, October 17). 2021-2023 The title of the request for information disclosure by the National Archives and the decision to disclose it are classified (Reception Number 13182230). Republic of Korea Information Disclosure. Available: <https://www.open.go.kr/>
- National Archives of Korea (2025, January 22). 2024 The title of the request for information disclosure by the National Archives and the decision to disclose it are distinguished (Reception Number 13632801). Republic of Korea Information Disclosure. Available: <https://www.open.go.kr/>
- National Archives of Korea (n.d.). Records Management FAQ. Available: <https://www.archives.go.kr/next/newnews/innovationTFFAQList.do>
- National Assembly Library (n.d.a). Intelligent Legal Search. Argos. Available: <https://argos.nanet.go.kr/main/mobile/lawqnaMainGuest.do>
- National Assembly Library (n.d.b). Intelligent Question Answering. National Assembly Electronic Library. Available: <https://dl.nanet.go.kr/qas/qasMain.do>
- Park, Eun-bi (2021). A Case study on ChatBot for mobile record and information service of the National archives of Korea. Master's thesis, Hannam University, Korea.

Park, Juyoung (2020, August 6). Korea's AI 'Exobrain' can now understand colloquial speech. Yonhap News.

<https://www.yna.co.kr/view/AKR20200806036400063>

Yang, Seung-jin (2024, May 15). Results of 'Gyeongbuk Provincial Government Digital Transformation'... AI 'Chat Gyeongbuk' that Reduces Simple Repetitive Work. Daily Newspaper.

<https://www.imaeil.com/page/view/2024051215572453222>

Yi Seon-ung (2019). A Study on the Description Model of Theta-role and Pilot Buildup (2019-01-11). National Institute of Korean Language.

Yoon, Nanseul (2024, June 3). Jeonju City Launches 'Chatbot Service' to Make Library Use More Convenient. Newsis.

https://www.newsis.com/view/NISX20240603_0002758522