

생성형 AI를 활용한 지식그래프 구축: 구술생애사를 중심으로

Constructing a Knowledge Graph Using Generative AI: Focusing on Oral Life Histories

이혜경 (HyeKyung Lee)*
이용구 (Yong-Gu Lee)**

초 록

본 연구의 목적은 구술생애사 자료를 대상으로 생성형 AI를 활용하여 RDF 기반 지식그래프 구축의 가능성을 탐색하는 데 있다. 이를 위하여 실제 파독 간호사 7인의 구술생애사를 대상으로 생성형 AI에 적용하였으며, RDF 트리플을 자동 생성하도록 하였다. 이후 7인에 대한 RDF 트리플을 통합하여 지식그래프를 형성하고자 하였다. 그 결과, 첫째, 생성형 AI는 비정형 데이터인 구술생애사에서 일정 수준의 RDF 트리플을 생성하여 이러한 자료에 대한 지식 구조화 가능성을 확인할 수 있었다. 둘째, 동일하거나 유사한 의미의 술어를 일관성 있게 생성하지 못하는 한계가 드러났으며, RDF 구축을 위해서는 통합화된 술어 체계가 필요함을 시사하였다. 셋째, 통합 지식그래프를 도출함으로써 개별 구술생애사 간 RDF 연계가 가능하고, 이를 통해 시각화하여 각각적 관점에서 데이터를 파악할 수 있음을 확인하였다. 다만, 구술생애사 자료의 구어적 특성, 맥락 의존적 표현, 사투리 등은 AI의 개체명 인식에 어려움을 주었으며, 객체 수가 불필요하게 증가하는 문제점 등이 한계로 나타났다. 그럼에도 불구하고 본 연구는 구술생애사와 지식그래프를 연결한 탐색적 시도라는 점에서 의의가 있으며, 복잡한 절차를 거쳐야 했던 지식그래프 구축 과정을 생성형 AI를 통해 간소화할 수 있는 가능성을 제시하였다.

ABSTRACT

The purpose of this study is to explore the feasibility of constructing RDF-based knowledge graphs by applying generative AI to oral life history data. For this purpose, the oral life histories of seven Korean nurses dispatched to Germany were analyzed and processed through generative AI to automatically produce RDF triples. These triples were subsequently integrated to form a unified knowledge graph. The findings are as follows. First, generative AI was able to produce RDF triples from unstructured oral life history data, thereby demonstrating the possibility of knowledge structuring for such narrative sources. Second, limitations were observed in the AI's ability to consistently generate predicates with identical or semantically similar meanings, indicating the necessity of establishing a consolidated predicate framework for effective RDF construction. Third, by integrating the RDF triples into a unified knowledge graph, it was possible to interlink individual oral life histories and visualize them for multidimensional exploration. However, challenges were observed due to the colloquial expressions, contextual-dependent nuances, and dialectal features inherent in oral history transcripts, which hindered accurate entity recognition and led to unnecessary proliferation of objects. Despite these limitations, this study holds significance as an exploratory attempt to connect oral life history research with knowledge graph construction. Furthermore, it demonstrates the potential to simplify the otherwise complex process of building knowledge graphs through generative AI, thereby suggesting broader applicability of such methods as a data resource in fields such as diaspora studies and migration history.

키워드: RDF, 지식그래프, 구술생애사, 생성형 AI, 한인 디아스포라
RDF, knowledge graph, oral life history, generative AI, Korean diaspora

* 한남대학교 문과대학 문헌정보학과 조교수(keilee@hnu.kr) (제1저자)

** 경북대학교 사회과학대학 문헌정보학과 교수(ygglee@knu.ac.kr) (교신저자)

■ 논문접수일자 : 2025년 8월 22일 ■ 최초심사일자 : 2025년 9월 12일 ■ 게재확정일자 : 2025년 9월 19일

■ 정보관리학회지, 42(3), 237-255, 2025. <http://dx.doi.org/10.3743/KOSIM.2025.42.3.237>

※ 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. 서 론

21세기 지식정보 사회에서 텍스트 데이터는 단순한 기록을 넘어, 의미와 맥락을 담은 지식 자원으로서 재구성되고 활용될 필요가 있다. 특히 인터넷과 월드 와이드 웹의 발전으로 방대한 양의 비정형 데이터가 생산 및 축적되는 사회로 도래하게 하였으며, 이러한 데이터를 보다 효과적으로 처리하고 더 나아가 이를 텍스트를 대상으로 구조화하고 의미적 연결망을 형성할 중요성이 점점 증대되고 있다.

이러한 맥락에서 지식그래프(Knowledge Graph)는 현실 세계의 개체와 관계를 데이터 구조로 표현함으로써, 데이터 간 연관성을 명시적으로 드러내고, 단순한 정보에서 유의미한 지식을 추출할 수 있는 핵심 기술로 주목받고 있다. 다만 어느 영역에서든 지식그래프를 생성하는 것은 다소 노동집약적이어서 많은 시간이 소요되며, 해당 영역의 정확하고 깊이 있는 전문 지식이 요구되어 연구 및 결과물 도출이 용이하지 않다.

최근에 생성형 인공지능(AI)의 급속한 발전으로, 최신 AI 모델이 자연어 텍스트의 이해와 생성에서 박사급 전문가 수준의 성능을 주요 벤치마크 테스트에서 보여주고 있으며, 실제로 존재하지 않는 가상의 정보나 내용을 만들어내는 환각 현상의 발생률도 점차 감소하고 있다. 이들 모델은 문제해결 과정에서도 여러 논리적 단계를 나누어 처리하는 추론을 유도하는 CoT (Chain-of-Thought)와 같은 기법을 활용함으로써, 인간처럼 사고하여 보다 정확하고 신뢰할 수 있는 결과를 도출하고 있다. 이러한 자연어로 된 문장을 이해하고 처리할 수 있는 인공지

능의 발전으로 텍스트로부터 지식그래프의 구축에서 난관을 해결할 수 있을 것으로 판단된다.

한편, 구술생애사는 개인과 집단의 기억을 면담으로 수집하고 기록으로 남기는 방식으로, 개인의 삶의 궤적과 사회적 경험을 동시에 담아내는 중요한 인문·사회적 자료이다. 구술생애사는 주로 인터뷰를 진행하고, 채록하는 방식으로 그 생애를 기록하고 있어서, 구어체적 표현을 비롯한 일반적 텍스트와는 상이한 비정형적 특성을 보인다. 이러한 특성으로 인해 일반적인 자연어를 대상으로 하는 텍스트 처리 기법으로는 보다 수준 높은 분석과 구조화가 어렵다. 반면, 이러한 구술 데이터는 개인의 경험을 사회적·역사적 맥락과 연결하는 데 있어 중요한 가치를 지니고 있어, 이를 체계적으로 구조화할 경우 새로운 학술적 통찰을 제공할 수 있다.

이에 본 연구는 한인 디아스포라로서의 파독 간호사 구술생애사를 대상으로 생성형 AI를 활용하여 RDF 기반 지식그래프 구축의 가능성을 탐색하고자 한다. 보다 구체적으로, 구술생애사를 RDF로 구조화하는 것은 단순히 특정 집단의 삶을 기록하는 데 그치지 않고, 한국 사회사 및 이주사 연구 전반에 기초 자료로 활용될 수 있는 확장성을 지닌다. 또한 본 연구는 구술생애사 연구에서 생성형 AI의 새로운 활용 가능성을 제시함과 동시에, 지식그래프 구축 과정의 복잡성으로 인해 접근이 어려웠던 연구자들에게 방법론적 통찰을 제공할 수 있을 것으로 기대된다. 따라서 본 연구의 궁극적 목적은 구술생애사와 같은 비정형 데이터를 생성형 AI를 활용해 RDF 기반 지식그래프로 구조화함으로써, 구술 데이터의 체계적 활용 가능성과 인문 사회 연구 전반에서의 확장 가능성을 제시하는 데 있다.

2. 이론적 배경

2.1 지식그래프

지식그래프는 현실 세계를 데이터 형태로 구조화하여 표현한 그래프로, 데이터 간의 관계를 명시적으로 연결함으로써 단순한 데이터를 지식으로 전환하고 의미를 전달하는 역할을 한다(Peng et al., 2023). 이러한 지식그래프는 2009년 팀 버너스 리의 W3C가 제시한 시맨틱 웹(Semantic Web)을 구현하기 위한 핵심 기반으로, 온톨로지 어휘의 구성과 데이터 관계의 명확한 표현이 필수적이다.

특히 W3C(World Wide Web Consortium)가 제안한 RDF(Resource Description Framework)는 웹상에서 정보를 표현하고 교환하기 위해 마련된 표준 프레임워크로, 주어(Subject), 술어(Predicate), 목적어(Object)로 연결하는 트리플(Triple) 구조를 통해 데이터를 표현한다 (World Wide Web Consortium, 2014).

이에 조은경과 강승식(2020)은 정서 장애 주제 분야의 지식그래프를 개발하기 위해 웹 데이터, 전문용어집, 신문 기사 등을 기반으로 ‘조현병’ 관련 용어와 유사어를 추출하여 지식 노드를 형성하였고, 이를 통해 지식그래프의 구현 가능성을 탐색하였다. 나아가 이들은 현대사회 의 인문학자들이 지식그래프를 언어학적·언어 지능적 발달을 지원하는 전문가로서 활용할 필요성을 강조하였다.

박하람과 김학래(2021)는 과편화된 일본군 ‘위안부’ 디지털 아카이브들에서 소장하고 있는 기록물 2,248건을 대상으로 메타데이터를 확인하고 지식그래프를 설계하였으며, 디지털 기록

의 연계를 통해 분산된 디지털 기록에 대해 연계 검색 및 의미 조건에서 활용 가능성을 파악하였다. 또한 그래프 구축을 통해 기록의 활용성 증대를 기대할 수 있음을 나타내었다.

한편, 박지연과 김홍중(2023)은 한국전쟁 역사 사료에 기반한 지식그래프 및 온톨로지 구축을 위해서 한국어 기사 말뭉치 중 51,009건의 데이터와 한국전쟁 역사 말뭉치 739건 등을 활용해 딥러닝 모델인 BERT를 기반으로 개체명 인식 예측을 수행하고, NetworkX를 통해 지식그래프를 형성하였다. 이를 통해 기존의 검색어 매칭으로 결과를 도출하는 데이터베이스에 비해 해당 검색어를 포함한 전투명, 인명, 지명 등 폭넓은 검색 결과를 파악함에 있어 지식그래프의 활용성을 강조하였다.

이처럼 지식그래프는 인공지능이 일상화된 현대사회에서 방대한 데이터를 보다 효과적이고 효율적으로 이해할 수 있는 기술로 주목받고 있으며, 선행연구에서 살펴본 것처럼 다양한 분야에서 이를 활용한 연구가 활발히 이루어지고 있었다. 더불어 메타데이터부터 신문기사, 디지털 기록, 역사 사료의 문장 등 정제되지 않은 비정형 데이터 등 다양한 데이터유형을 활용하여 지식그래프 형성의 가능성을 파악하고 있었다.

2.2 구술생애사

구술생애사는 집단적 구술사와 자전적 구술사로 구분할 수 있으며, 자전적 구술사는 자신이 살아온 삶의 궤적에 따라 연대기 형태로 표현하는 방식과 생애 중 자신의 주관적 경험과 느낌을 표현하는 방식으로 다시 나눌 수 있다

(김성례, 2002). 이러한 관점에서 구술생애사는 녹음하거나 이를 채록한 텍스트 등 대표적인 비정형 콘텐츠 및 데이터로(Cherukuri et al., 2025), 인간의 기억과 경험을 심층 면담을 통해 수집하고, 이를 역사적 자료로 보존하는 연구 방법이다(유명상, 2025).

한신갑과 이상직(2017)은 구술생애사 자료에 대하여 생애사 시간표(Life History)를 통해 조직하면 보다 효과적으로 생애 이력을 분석할 수 있다고 하였으며, 생애 과정에서 시기나 연도 혹은 연령에 따른 가족관계 혼인, 학업, 노동 등의 범주와 함께 주요 사건의 발생 시점과 지속 기간 등을 표현한다면, 수집된 사례들의 유사성과 특수성을 파악해 생애 이력의 패턴을 파악하고 집단 생애사 측면에서 그 의미를 유추하는 등의 생애 연구의 새로운 방향을 제시할 수 있을 것이라 언급하였다.

결국 이는 특정 생애사의 대상이었던 한 인물과 다른 인물의 어떠한 특성을 통해 연결되고 서사구조의 겹침이 나타날 수 있다는 것을 의미하고, 이들의 생애사를 시각화 한 형태 즉, 지식그래프 형식으로 구축한다면, 생애사 데이터의 체계적 설계 및 분석의 가능성을 파악할 수 있고 나아가 관련 연구 분야의 깊이와 관점을 확장 시킬 수 있는 기제 될 수 있음을 시사한다.

Cherukuri et al.(2025)은 보편적 구술사 연구 및 분석에서도 적용할 수 있는 재사용 가능한 데이터의 형성과 보존을 위하여, 일본계 미국인의 강제수용 구술사를 대상으로 대규모 언어모델(LLMs)을 구축하고자 하였다. 이들은 실제 전문가들이 태깅한 데이터와 생성형 AI인 GPT4o, Llama, Qwen 등을 활용하여 형성된 데이터를 비교 및 평가를 수행하였다. 그 결과

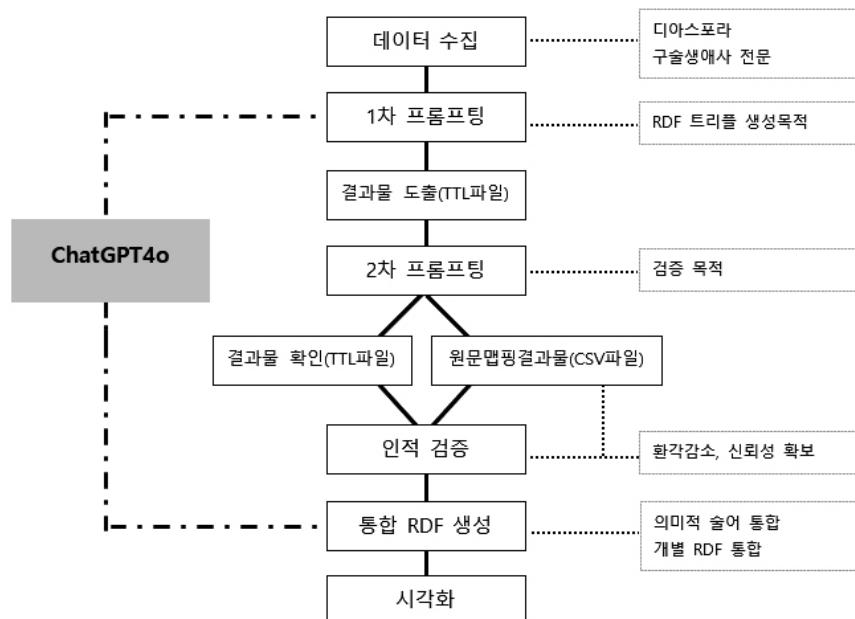
의미분류 부문에서는 ChatGPT가, 감정분석 부문에서는 Llama가 높은 성능을 보인 것으로 확인하였으며, 구술사와 같은 비정형 데이터의 가공이나 활용에서 생성형 AI의 사용 가능성을 밝혔다.

국내에서는 구술생애사 데이터에 대하여 토픽모델링을 적용하거나(박신규, 이준엽, 2023), 텍스트 마이닝 기법을 적용해 구술 아카이브를 분석한 연구(문형진, 2023) 등이 존재하였으나, 아직까지 생성형 AI를 적용한 연구는 찾아보기 어려웠다. 또한 지식그래프 구축에 대한 선행 연구들은 데이터에 기반해 지식그래프를 구축해 왔으나, 역시 생성형 AI를 적용하고자 한 시도는 거의 이루어지지 않았다.

이에 본 연구는 파독 간호사의 구술생애사를 대상으로 생성형 AI를 활용하여 RDF 기반 지식그래프 구축의 가능성을 탐색하고자 하였다. 나아가 생성형 AI를 적용할 수 있는 구체적인 절차와 방법을 시도하고 그 사례를 제시함으로써, 향후 다른 구술생애사 연구에도 활용 또는 적용의 가능성 검토하였다. 이러한 시도는 지식그래프 구현 과정의 복잡성으로 인해 접근이 제한적이었던 연구 영역을 확장시키고, 다양한 연구 주제에 새로운 분석적 통찰과 방법론적 깊이를 제공하는 데 그 의의가 있다.

3. 연구 방법

본 연구의 목적은 생성형 AI를 활용하여 파독간호사의 구술생애사를 기반으로 한 지식그래프의 구축을 시도해 보는 것으로, 연구 절차는 <그림 1>과 같다. 먼저 실험 대상인 텍스트



〈그림 1〉 연구 절차 개요도

로 한인 디아스포라의 구술생애사를 선정하였다. 이는 디아스포라가 단순 구술생애사보다 아주 경험을 포함하여 고유한 문화와 정체성을 유지하며 사회적·역사적 구조를 형성하기 때문에, 더욱 정교하게 구조화할 수 있는 본 연구의 방법론에 적합하다고 판단된다. 두 번째 단계는 생성형 AI를 활용하기 위한 프롬프트 엔지니어링 과정으로 연구 목적 달성을 부합하면서 보편적으로 적용 가능하도록 가장 기본적인 방법을 적용한다. 세 번째 단계는 인공지능의 활각 여부를 확인하고 결과물의 품질에 대한 신뢰성을 확보하기 위한 인적 검증을 수행한다. 마지막으로 동일 내지 유사 술어의 통합화를 수행하고, 그래프 형식을 사용하여 지금까지 도출된 결과에 대해 시각화를 수행한다. 실험 텍스트가 7명의 인물을 포함하므로 이를 통합하여 제공한다.

우선 지식그래프의 대상인 파독간호사의 구술생애사 전문을 수집하였다. 해당 도서(박신규, 이토히로코, 2023)는 파독 간호사 7인과 광부 1인의 생애사를 담고 있으며, 사회학자와의 인터뷰를 바탕으로 채록한 내용을 기록으로 남긴 것이다. 〈그림 2〉는 이러한 생애사 인물 중 한 명에 해당하는 원문의 일부로, 각 인물의 대화 형태와 표현이 그대로 수록되어 있다. 한편, 생애사의 주인공은 총 8인이었으나, 파독 광부의 경우 배우자의 생애를 중심으로 인터뷰가 진행되었으므로, 본 연구에서 파독 간호사를 대상으로 한 지식그래프 작성 목적이 상이하다고 판단하여 제외하였다.

이렇게 수집한 구술생애사는 총 623문단, 3,564문장으로 구성되어 있으며, 개인별 문단 및 문장 수는 〈표 1〉과 같다. 전체 구술생애사의 평균은 1인당 89문단, 509문장으로 나타났

72 독일 퀘른지역 파독간호사·광부의 구술생애사

■ 일본 야마구치현 출생으로 부산에서 성장

난 1940년생이고 동생은 1945년생이에요. 우리 고향은 일본으로, 일본에서 태어났어요. 부모님은 일본 야마구치현으로 1930년 대쯤 들어가신 것 같아요. 아버님은 경남 고성에서 태어나서 거기에서 국민학교도 다니고 중학교 다녔는가 그건 모르겠어요. 모르겠고 직업 없이 그냥 아버지는 부모님 모시고 어부로 있었어요. 그러다가 일본에는 자원으로, 그냥 가서 무조건 무슨 일이 닥치면 일을 하려고 가셨는데 옛날에 직업 없이 일하는 거, 막노동이잖아요. 노동자로 일을 하시고 결혼하고 일본 가셨는가 봐요.

「독일 퀘른지역 파독간호사·광부의 구술생애사」 中 일부 캡처

〈그림 2〉 생애사 전문 데이터(일부)

〈표 1〉 RDF 대상 텍스트의 문단 및 문장 통계

| 대상 인물 | 문단 | 문장 |
|--------|---------|-----------|
| 1번 | 71 | 339 |
| 2번 | 104 | 720 |
| 3번 | 90 | 501 |
| 4번 | 47 | 213 |
| 5번 | 116 | 819 |
| 6번 | 84 | 400 |
| 7번 | 111 | 572 |
| 합계(평균) | 623(89) | 3564(509) |

다. 7명 인물의 구술사 문장과 문단의 수는 통계적으로 다소 큰 편차를 보였다. 4번 인물의 경우, 파독 광부였던 배우자의 생애를 제외하고 수집하였기 때문에 다른 인물들에 비해 텍스트 양이 절반가량 적었다. 3번 인물의 경우 자매가 함께 등장하였으나, 도입부에서 언니가 주체가 되어 두 명의 공통된 부분을 구술하였

으므로 원문 데이터를 분할 하지 않고 텍스트의 원형 모습대로 자매를 하나의 구술 단위로 묶어 RDF를 구현하였다.

일반적으로 자연어 텍스트는 활용 목적에 따라 문단이나 문장을 분리하거나 형태소 분석 등 다양한 전처리 및 분석 과정을 거친다. 또한 RDF 구축을 위해서는 데이터 전처리, 개체명

인식. 그리고 트리플 생성을 위한 술어 추출 과정이 필수적이다. 하지만 이 연구에서는 앞서 수집한 데이터에 대해 별도의 전처리 과정을 거치지 않고, 원문 그대로 생성형 AI가 RDF 트리플을 생성할 수 있도록 지시문과 함께 제시하였다. 즉 생성형 AI가 사전에 가공된 데이터를 활용하는 것이 아니라, 원문을 직접 분석하여 RDF 전체를 구축하도록 하였으며, 이를 통해 구술생애사 연구에서 지식그래프 구축에 있어 생성형 AI의 활용 가능성을 높일 수 있는 방안을 탐색하고자 하였다.

생성형 AI를 활용하여 구술생애사 텍스트로부터 RDF 트리플을 추출하기 위한 프롬프트 엔지니어링 과정은 크게 두 단계로, RDF 트리플의 생성을 목적으로 한 1차 프롬프트와 생성 결과 검증을 목적으로 한 2차 프롬프트로 나누어 수행하였다. <그림 3>과 같이 1차 프롬프트에서는 생성형 AI의 역할과 수행해야 할 사항을 제시한 뒤, 해당 구술생애사를 입력하여 RDF 트리플의 생성을 유도하였다. 1차 프롬프트는 크게 지시문과 안내문으로 나누었으며, 지시문

의 경우 수행해야 할 명령(주요 개체와 개념을 추출하고 RDF 트리플로 변환)과 출력 형식(Turtle 포맷) 지정으로 구성하였다. 특히 CoT 기법과 같은 추론을 유도하기 위해 마지막 문장을 주목할 필요가 있다. 안내문의 경우 프롬프트에 입력할 텍스트에 대한 핵심적인 설명을 명시하였다.

실험에 사용한 생성형 AI인 ChatGPT-4o의 언어모델 GPT-4o는 API 활용 시 최대 컨텍스트 토큰이 약 128,000개에 이르지만(Bergmann, 2024), ChatGPT 환경에서 사용하는 경우 입력·출력 합산 최대 8,192 토큰까지 활용이 가능한 것으로 보고되고 있다(Liang, 2024). 따라서 입력 지시문이 길어지면 결과 또한 방대해질 수 있으나, 토큰 한계에 따라 결과의 일관성 문제 및 오류가 발생할 가능성이 있다. 이에 본 연구에서는 7인 전체의 구술생애사를 한 번에 입력하지 않고, 개인별 구술생애사를 개별적으로 입력하여 RDF 트리플을 생성하고, 그 결과를 확보하였다. 이후 필요에 따라 추가적인 RDF 확장을 요청하였다. 그 결과, 획득한 Turtle 형식

| |
|---|
| 지시문: 다음에 제시되는 텍스트는 구술생애사이며 특히 한인 디아스포라에 해당합니다. 이 구술생애사 텍스트에서 인물, 장소, 사건, 조직 등의 주요 개체(Entity)와 개념을 추출하고, 이를 RDF 트리플(subject, predicate, object) 형식으로 변환하여 주세요. 1) 결과는 Turtle(TTL) 형식의 파일로 주세요. 2) 엑셀파일로 RDF 트리플의 근거 문장을 함께 제시해주세요, 시간을 갖고 천천히 작업하고 결과는 매우 다양하고 깊이있게 처리되어야 합니다. 안내문: 다음은 독일로 파견된 간호사의 일대기를 담은 구술사입니다. |
|---|

<그림 3> 1차 프롬프트 지시문

의 RDF 트리플 파일은 <그림 4>와 같으며, 이 사례는 1번 인물의 구술생애사를 근거로 생성형 AI가 구성한 RDF 트리플의 일부이다. 네임 스페이스로 'ex', 'foaf', 그리고 'schema'를 적용한 것을 볼 수 있으며, 구술생애사 텍스트로부터 추출한 다양한 객체와 술어 표현을 볼 수 있다.

2차 프롬프트는 생성형 AI가 1차로 생성한 RDF 트리플이 원문에 근거한 사실인지를 확인 및 검증하기 위해 진행하였다. 이는 생성형 AI의 대표적인 한계인 환각(hallucination) 가능성을 최소화하기 위한 절차로, <그림 5>와 같이 RDF 트리플 생성 과정에서 원문 근거를 함께 제시하도록 하여, 임의의 가상 RDF가 구축

```

ex:Grandfather_DM a foaf:Person ;
|rdfs:label "담양의 조부(부동)"@ko ;
| schema:additionalType "부동"@ko .

ex:KODCO a schema:Organization ;
|rdfs:label "해외개발공사"@ko .

ex:ParkJaeSoon a foaf:Person ;
| ex:visitedKoreaAfterGwangju true ;
| schema:affiliation ex:Family_Business_Cloth ;
| schema:alternateName "박재순"@ko ;
| schema:alumniOf ex:JeonnamGirlsMiddleSchool,
|   ex:JeonnamNursingSchool ;
| schema:award "쾰른 시장 표창(근속·업적)"@ko ;
| schema:birthDate "1943"^^xsd:gYear ;
| schema:birthPlace ex:DamYang ;
| schema:children ex:Son1,
|   ex:Son2 ;
| schema:educationalCredentialAwarded "조산원 교육 수료"@ko ;
| schema:endDate "1963"^^xsd:gYear ;
| schema:event ex:AbuDhabi_Performance,
|   ex:Account_Scandal ;
| schema:familyName "박"@ko ;
| schema:graduationYear "1961"^^xsd:gYear ;
| schema:grewUpIn ex:GwangJu ;
| schema:hasOccupation "간호사"@ko,
|   "양호교사"@ko ;

```

<그림 4> ChatGPT가 생성한 RDF 트리플 사례

| Subject | Predicate | Object | SourceText |
|-------------------------------|-----------------------------------|----------------------------|---------------------------|
| http://example.org/KimJeongJa | http://xmlns.com/foaf/0.1/name | 김정자 | 난 1940년생이고 동생은 1945년생이에요. |
| http://example.org/KimJeongJa | http://example.org/birthPlace | 일본 아마구치현 | 일본에서 태어났어요. |
| http://example.org/KimJeongJa | http://example.org/birthYear | 1940 | 난 1940년생이고... |
| http://example.org/KimHoJa | http://xmlns.com/foaf/0.1/name | 김호자 | 동생은 1945년생이에요. |
| http://example.org/KimHoJa | http://example.org/birthPlace | 일본 아마구치현 | 일본에서 태어났어요. |
| http://example.org/KimHoJa | http://example.org/birthYear | 1945 | 동생은 1945년생이에요. |
| http://example.org/KimJeongJa | http://example.org/sibling | http://example.org/KimHoJa | 김정자와 김호자 자매 |
| http://example.org/KimJeongJa | http://example.org/parent | http://example.org/Father | 아버님은 경남 고성에서 태어나서... |
| http://example.org/KimJeongJa | http://example.org/parent | http://example.org/Mother | 어머니는 아버지 찾아서 연락선을 타고... |
| http://example.org/baptism | http://example.org/year | 1962 | 1962년도에 영세를 받았어요. |
| http://example.org/KimJeongJa | http://example.org/participatedIn | http://example.org/baptism | 영세를 1962년도에 받았어요. |
| http://example.org/KimJeongJa | http://example.org/workPlace | 쾰른 시립양로원 | 쾰른 시립양로원으로 가게 되었어요. |

<그림 5> RDF 트리플과 추출에 사용된 근거자료 원문(CSV 파일 일부)

되지 않도록 하였다. <그림 5>는 인물 3번에 해당하는 RDF 트리플의 근거자료를 생성형 AI가 작성한 것으로, 실제로 3번 인물은 여동생과 함께 독일로 파견되어 간호사 생활을 하였고, 이에 대한 가족관계를 RDF로 구성한 사례이다. <그림 5>의 첫 행의 트리플을 추출한 근거자료(SourceText 칼럼)와 <그림 2> 원문의 앞부분이 일치하는 것을 알 수 있으며, 생성형 AI가 주변 문맥에 대해 자연언어 처리의 주요 기법인 상호참조 해결(coreference resolution) 기법을 통해 원문에서 직접적인 문장으로 표현하지 않아도 구술 주체의 동생이 누구이며 생년과 출생지를 정확히 추론한 것을 알 수 있다.

다음으로, 생성형 AI가 두 차례에 걸쳐 최종적으로 생성한 RDF 결과물과 이에 기반한 근거자료는 2인의 인적 검증을 거쳤다. <그림 5>와 같이 제시된 결과 파일에서 각 트리플과 근거자료(SourceText)의 문장을 실제 원문에서 찾아 앞뒤 맥락을 확인하고 트리플 생성을 검증하였다. 검증 과정에서 의견이 나뉠 경우에는 주어·술어·목적어의 사실성, 시간·장소 등 핵심 개체의 일치 여부, 해석적 표현 개입 가능성을 기준으로 재검토하고, 2인의 검증자가 근거 표시와 주석을 다시 확인하며 조정하였다. 그럼에도 최종 합의가 어려울 때는 원문 전체 문맥을 재확인하여 보다 보수적인 방향으로 결정하고자 하였다. 이러한 절차는 생성형 AI가 구성한 RDF 트리플의 환각 여부와 사실성을 재확인하여 최종 결과물의 신뢰성을 확보하기 위함이다.

다양한 형태로 나타난 술어들을 의미적으로 유사한 하나의 대표 술어로 정리하는 과정인 술어 통합 절차를 수행한 후, 생성형 AI를 활용

하여 개별적으로 구축된 RDF 데이터를 통합함으로써 일관된 구조의 지식그래프를 구축하였다. 마지막으로 Turtle 형식으로 표현된 최종적인 RDF 트리플을 Python 환경에서 RDFlib 및 NetworkX 패키지를 이용하여 그래프 구조로 변환하였으며, 이를 기반으로 시각화 및 분석을 수행하였다.

4. RDF 지식그래프 구현

4.1 RDF 트리플 구성

앞서 기술된 연구 방법에 따라 1, 2차에 걸쳐 제공한 프롬프트에 따라 생성형 AI가 추출한 RDF 트리플의 각 요소별 통계분포는 <표 2>와 같다. 생성된 RDF 트리플은 총 645건이며, 인물당 평균 92건의 트리플이 도출되었다. 이 트리플들은 총 217개의 고유 개체와 186개의 술어로 구성되어 있었으며, 인물 1인당 평균적으로 31개의 고유 개체와 26개의 고유 술어가 도출되었다. 전체 인물을 통합하여 분석한 결과 중복을 제거한 고유 개체 수는 185개로 확인되었다.

인물별 RDF 트리플 구성 현황을 살펴보면, 4번 인물의 경우 구술 인터뷰가 부부를 함께 대상으로 진행되었으나, 본 연구에서는 파독 간호사에 해당하는 부분만을 추려 분석하였기 때문에 다른 인물들에 비해 개체나 술어 수가 상대적으로 적었다. 그리고 2번과 5번 인물은 구술생애사 기록이 많은 편임에도 불구하고(<표 1> 참조), 생성된 RDF 트리플 수는 다른 인물에 비하여 적은 것으로 나타났다. 이는 원 데이

〈표 2〉 RDF 트리플의 구성요소 통계

| 인물 | 트리플 (Triple) | 고유 개체 (Entity) | 고유 URI | | 값 (Literal) | 술어 |
|--------|-----------------|-------------------|-----------|-----------|----------------|-----------|
| | | | 주어 | 객체 | | |
| 1번 | 130 | 44 | 35 | 41 | 62 | 38 |
| 2번 | 94 | 39 | 30 | 31 | 42 | 24 |
| 3번 | 121 | 34 | 29 | 22 | 65 | 25 |
| 4번 | 47 | 5 | 4 | 4 | 39 | 23 |
| 5번 | 69 | 31 | 16 | 29 | 23 | 25 |
| 6번 | 83 | 30 | 22 | 29 | 35 | 27 |
| 7번 | 101 | 34 | 27 | 26 | 40 | 24 |
| 합계(평균) | 645(92.1) | 217(31.0) | 163(23.3) | 182(26.0) | 306(43.7) | 184(26.6) |

터에 기인한 것으로 판단되는데, 실제 2번과 5번 인물은 자신의 생애를 서술함과 동시에 그의 주변 인물의 서술이 풍부하였고, 생성형 AI는 이런 주변적 서술을 RDF 트리플 생성 과정에서 구술 주체에 대해 상대적으로 중요성이 낮다고 판단하고 제외했을 가능성이 있다.

요약하면 인터뷰 내용과 비중에 따라 트리플 구성에도 차이가 나타났으며, 주로 개인에 대하여 구체적이고 풍부한 생애사가 제공될수록 트리플 수와 개체·술어 역시 다양하게 생성되는 경향이 확인되었다.

한편, 생성형 AI는 RDF 트리플 생성에 있어 URI로 활용하기보다는 주로 값(literal)의 형태로 객체를 추출하였는데, 객체 URI는 평균 26개였으며, 값은 평균 43건으로 집계되었다. 본 연구에서 활용한 데이터가 실제 구술을 기록·출판한 데이터라는 점을 고려하면, 구어체 표현이나 맥락적 생략으로 대화가 진행된 것은 생성형 AI가 개체로 구분하기 어렵다고 판단한 경우 이를 값으로 변환한 것으로 추측된다.

구술생애사 인터뷰나 연구 방법에서는 대상 인물의 삶을 체계적으로 분석하고, 구술생애사의 주요 요소를 분류하는 방식으로 ‘생애사적 구조’

또는 생애사적 범주(life course categories)를 적용할 수 있다(한신갑, 이상직, 2017; Atkinson, 1998; Hatch, 2002). 구체적인 범주로는 인물의 기본 속성, 가족관계, 교육과 학력, 직업과 소속, 지리와 장소, 사건과 활동, 시간과 기간 등이 있다. 이 연구에서는 생애사에서 추출된 술어를 이러한 범주별로 군집화하였으며, 결과는 〈표 3〉과 같다. 기본 속성에는 생몰, 종교, 국적 등이 포함되며, 가족관계는 배우자를 비롯한 형제·자매·자녀를 연결하였다. 교육과 학력은 개인별 차이가 존재하였으나 유년기부터 대학 졸업 까지의 학력이 기술되었고, 직업과 소속은 노년기에 이르기까지의 직업 이력을 포괄하였다. 지리와 장소는 생애 전반의 거주지와 언급된 지명을 중심으로 기술되었다. 사건과 활동은 생애사 속 각 역할과 임무 수행을 중심으로 구성되었으며, 시간과 기간은 주요 사건 및 술어와 관련된 시기를 기록하였다. 아울러 생성형 AI는 생애사에서 명시되지 않은 시간 및 장소에 관해서는 표기하지 않았으나, 관련 맥락을 파악한 경우에는 이를 기입하기도 하였다.

본 연구에서 사용한 실험 데이터는 파독 간호사의 생애사로 한인 디아스포라적 요소가 포함

〈표 3〉 술어의 생애사적 범주 굽집화

| 인물 | 인물의 기본속성 | 기초관계 | 교육과 학력 | 직업과 소속 | 자리 및 장소 | 이동과 이주 | 사건과 활동 | 시간과 기간 |
|----|--|--|--|--|---|--|----------------------------------|--|
| 1번 | birthDate birthPlace knowsLanguage nationality religion | children familyName numberOfChildren parent siblingCount spouse | educationalCreden tialAwarded graduationYear | affiliation hasOccupation jobTitle memberOf worksFor | homeLocation studyLocation | conscribedTo | award event participatedIn | endDate startDate hasBeginning |
| 2번 | agent birthDate birthPlace religion | parentOf spouse | | affiliation member memberOf | address addressCountry fromLocation toLocation | wasDerivedFrom arrivalTime departureTime | | |
| 3번 | birthPlace birthYear deathYear nationality religion | child parent sibling spouse | education | memberOf occupation workYearEnd workYearStart workedAs | relatedTo eventYear participatedIn | relatedTo eventYear participatedIn | eventYear participatedIn | baptizedYear year |
| 4번 | birthDate birthPlace hasReligion lifeStage nationality officialBirthPlace | hasChild spouse | attendedSchool | memberOf occupation workedAt | livedIn currentCity | movedTo | event | attendedInstitution |
| 5번 | birthDate birthPlace knowsLanguage nationality | children parent sibling spouse | | employer hasOccupation jobTitle memberOf worksFor | hometown location | | | endDate startDate |
| 6번 | birthPlace birthYear causeOfDeath durationUntilDeath religion | child reason relatedPerson spouse | | jobRole organization | currentResidence country place | emigratedTo | hasEvent | baptismYear date professionStartYear |
| 7번 | agent birthDate birthPlace deathCause deathDate marriageDate nationality | hasSibling children parent IsSpouse | | employmentHistory hasOccupation jobTitle memberOf worksFor | addressCountry location | | | startDate |

된다. 디아스포라는 하나의 민족집단이 보이는 세계 이주 과정을 넘어서 그들이 형성한 공동체 문화나, 사회적 현상을 아울러 표현하고 있으므로(서봉언, 이채문, 2014), 이동과 이주 군집을 추가하여 개인의 지역 내 이동뿐 아니라 디아스포라에서 주요하게 다루어지는 이주 과정도 반영할 수 있도록 하였다.

이에 <표 3>에서 보듯이 구체적으로 살펴보면, 인물마다 채택한 술어는 다르지만 그 의미가 유사한 경우가 존재하였다. 예를 들어, 가족 관계의 경우 1번 · 5번 · 7번 인물에서 사용된 ‘children’은 3번 · 6번 인물의 ‘child’, 4번 인물의 ‘hasChild’와 동일한 의미로 사용되었다.

특히 디아스포라 생애사에서는 중요하게 고려되는 이주 기간과 이주 시기에 있어, 원문 데이터에서는 독일로 이주한 간호사들의 시기와 경로가 확인되지만, 생성형 AI가 구축한 RDF 술어에서는 ‘이주와 이사’나 ‘지리 및 장소’가 혼재되어 나타났다. 원문에서 독일 이주의 계기나 시점이 명확히 서술되지 않은 경우, 사람이 읽었을 때 맥락적으로 파악할 수 있는 요소를 생성형 AI는 구분하지 못하고 ‘hometown’이나 ‘homeLocation’으로 표현하는 경향을 보였다. 반면 한국전쟁 당시의 ‘징용’이나 원문에 ‘이주라는 용어가 직접적으로 등장한 경우에는 ‘emigratedTo’를 사용하기도 했으며, 때로는 ‘movedTo’를 적용하기도 하였다. 다만 국내 이동이 아님을 인식한 경우에는 ‘country’나 ‘addressCountry’를 사용하여 해외 이주를 표현하는 것으로 나타났다.

4.2 의미적 술어 통합

텍스트 데이터에서 RDF 트리플을 추출할 때,

동일하거나 유사한 의미를 갖는 술어(predicate)가 다양한 형태나 용어로 사용될 수 있다. 또한, 텍스트에 같은 술어가 사용되더라도 이를 활용하는 생성형 AI는 문맥이나 상황에 따라 다른 술어를 생성할 수 있으며, 세션에 따라 결과가 달라질 수 있다. 이는 생성형 AI의 특성 중 하나인 휘발성에서 기인한다고 볼 수 있다. 유사한 술어들을 하나의 대표 술어로 정리하는 과정인 ‘술어 정규화’(predicate normalization) 또는 ‘술어 통합’(predicate consolidation)이 필요하다. 따라서, 본 연구에서는 앞서 제시한 술어의 차이를 해결하여 ChatGPT 프롬프트에 명시할 술어를 선정하기 위하여 <표 2>에서 출현한 술어의 빈도를 분석하고, 유사 술어를 하나의 표준 술어로 통합하는 작업을 수행하였다. 이에 따라 <표 4>는 7인의 RDF에서 등장한 술어의 빈도를 정리한 결과이다. 구술생애사는 개인의 삶을 반영하기 때문에 특수한 상황이나 경험이 나타난다. 예를 들어, ‘educationalCredentialAwarded’의 1회 등장은 1번 인물이 ‘조산원 교육 수료’를 언급한 경우에 해당하며, ‘durationUntilDeath’와 ‘causeOfDeath’는 6번 인물의 배우자가 간암으로 2달 2주간 입원한 뒤 사망한 사실을 표현한 것이다. 이처럼 특정 인물의 독특한 경험을 반영하는 술어는 비록 저빈도로 나타나더라도 삭제하기보다 그 특수성을 인정할 필요가 있다.

반면, ‘birthDate’와 ‘birthYear’처럼 동일한 의미를 지니는 생년 표기 술어나, 앞서 언급한 ‘children’, ‘child’, ‘hasChild’와 같이 자녀 유무를 나타내는 술어는 특정 인물의 고유한 경험을 반영하지 않는다. 따라서 이러한 일반적 술어는 의미가 중복되지 않도록 통합화 과정을

〈표 4〉 범주별 술어와 빈도

| 범주 구분 | 술어 | 빈도 | 범주 구분 | 술어 | 빈도 |
|----------|------------------------------|----|--------|-------------------|----|
| 인물의 기본속성 | birthDate | 10 | 직업과 소속 | organization | 3 |
| | religion | 7 | | affiliation | 2 |
| | birthYear | 5 | | workedAt | 2 |
| | nationality | 5 | | employmentHistory | 2 |
| | knowsLanguage | 4 | | workedAs | 2 |
| | officialBirthPlace | 1 | | member | 1 |
| | deathDate | 1 | | jobRole | 1 |
| | deathCause | 1 | | location | 18 |
| | causeOfDeath | 1 | | addressCountry | 9 |
| | durationUntilDeath | 1 | | place | 6 |
| | lifeStage | 1 | | address | 4 |
| | hasReligion | 1 | | homeLocation | 2 |
| | | | | currentResidence | 1 |
| 가족관계 | parent | 10 | | currentCity | 1 |
| | spouse | 9 | | hometown | 1 |
| | child | 6 | 이동과 이주 | toLocation | 1 |
| | children | 5 | | departureTime | 1 |
| | relatedPerson | 3 | | conscriptedTo | 1 |
| | sibling | 3 | | arrivalTime | 1 |
| | hasChild | 2 | | movedTo | 1 |
| | parentOf | 2 | | emigratedTo | 1 |
| | numberOfChildren | 1 | | fromLocation | 1 |
| | hasSibling | 1 | 사건과 활동 | hasEvent | 10 |
| | familyName | 1 | | participatedIn | 10 |
| 교육과 학업 | attendedSchool | 3 | | event | 9 |
| | attendedInstitution | 2 | | agent | 9 |
| | studyLocation | 1 | | relatedTo | 5 |
| | education | 1 | | eventYear | 5 |
| | educationalCredentialAwarded | 1 | | award | 1 |
| | graduationYear | 1 | | | |
| 직업과 소속 | memberOf | 14 | 시간과 기간 | startDate | 16 |
| | worksFor | 10 | | date | 9 |
| | jobTitle | 6 | | endDate | 4 |
| | hasOccupation | 6 | | baptizedYear | 2 |
| | occupation | 5 | | baptismYear | 1 |
| | employer | 3 | | marriageDate | 1 |

거쳤다. 또한 디아스포라 구술생애사에서 핵심적으로 다루어지는 ‘이주’와 ‘이동’ 관련 술어 역시 생성형 AI가 혼재하여 사용하는 한계를 고려해 통합화하였다. 이에 따라 최종 선정된

표준 술어는 〈표 5〉에 제시하였다.

각 범주 구분에 대해 유사한 술어들은 생년, 가족, 자녀, 교육, 종교, 직업, 소속, 이주·이동, 거주 등의 범주로 분류하였다. 술어를 선정할

〈표 5〉 유사 술어의 통합화 내역

| 술어 정의 | 술어 | 선정된 술어 |
|-------|--|----------------|
| 생년 | birthDate / birthYear | birthDate |
| 가족 | parent / parentOf | parent |
| | sibling / hasSibling | hasSibling |
| 자녀 | child / children / hasChild / numberOfchildren | hasChild |
| 교육 | attendedSchool / attendedInstitution / education | attendedSchool |
| 종교 | religion / hasReligion | hasReligion |
| 직업 | jobTitle / hasOccupation / occupation / employer / jobRole | hasOccupation |
| 소속 | worksFor / workedAs / affiliation / workedAt | worksFor |
| 이주 이동 | toLocation / movedTo / emigratedTo | moveTo |
| 거주 | homeLocation / currentResidence / currentCity / hometown | homeLocation |

때에는 has 계열이나 동사형 표현을 우선적으로 채택하였다. 다만 디아스포라 구술생애사에서는 ‘이주’와 ‘이동’이 서로 다른 의미로 사용될 수 있으나, 생성형 AI가 이 둘을 명확히 구분하는 데 한계가 있음을 확인하였다. 이에 따라 본 연구에서는 ‘emigratedTo(이주)’ 대신 보다 포괄적인 개념인 ‘moveTo(이동)’를 표준 술어로 채택하였다.

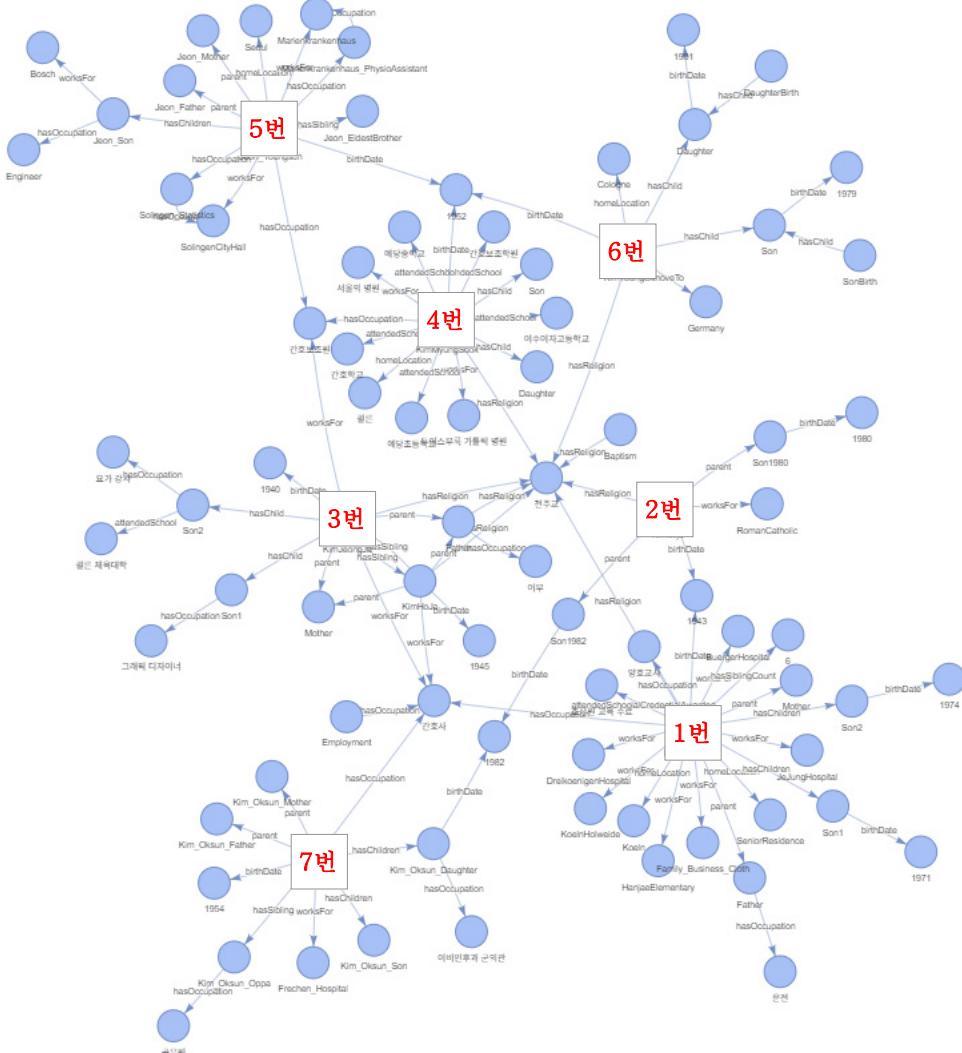
4.3 RDF 트리플 통합

이렇게 개인별로 구축된 RDF는 술어 통합화 작업을 거친 뒤, 해당 술어에 한정하여 생성형 AI에 다시 RDF 생성을 지시하였다. 그 결과 각 인물의 RDF 그래프를 통합하면 〈그림 6〉과 같다. 개별 인물들은 모두 파독 간호사이자 퀼른 지역과 관련된 공통적 특성을 공유하지만, 동시에 〈표 6〉과 같이 서로 다른 관계망 속에서 연결되어 있음을 확인할 수 있었다.

예를 들어, 1번과 2번 인물은 1943년에 출생하였고, 4번 · 5번 · 6번 인물은 1952년에 출생한 것으로 나타났다. 또한 2번 인물의 자녀와

7번 인물의 자녀는 모두 1982년에 출생하여 서로 연결되었다. 직업 측면에서 1번, 2번, 3번 인물과 그녀의 동생, 그리고 7번 인물의 동생은 간호사로 근무하였으며, 3번 · 4번 · 5번 인물은 간호보조원으로서 직업 경력을 가진 것으로 확인되었다. 종교의 경우 5번과 7번 인물을 제외한 대상자가 천주교를 종교로 지니고 있었다. 특히 3번 인물의 동생과 아버지 또한 천주교 신앙을 유지한 것으로 나타났다.

이처럼 개별 인물의 RDF는 한 사람의 생애사를 구체적으로 살펴볼 수 있다는 특성을 지니지만, 통합 RDF를 통해서는 각 인물이 서로 연관되어 있음을 확인할 수 있었다. 즉, 그래프를 통해, 한 인물의 개인적 삶을 넘어 인물 간 관계망 속에서의 생애까지 파악이 가능하였다. 또한 생성형 AI에 통합화된 술어를 기반으로 RDF 생성을 지시했을 때, 보다 효율적으로 RDF 트리플을 형성할 수 있었으며, 이를 통해 생애사 연구에서 텍스트 데이터를 시각화하여 인물들의 연관성을 직관적으로 나타낼 가능성도 확인하였다.



〈그림 6〉 7명의 통합 RDF 지식그래프

〈표 6〉 인물의 연관관계

| 범주 구분 | 객체 | 인물 |
|-------|-------|-------------------------------------|
| 생년 | 1943 | 1번, 2번 |
| | 1952 | 4번, 5번, 6번 |
| | 1982 | 2번의 자녀, 7번의 자녀 |
| 직업 | 간호사 | 1번, 2번, 3번, 3번의 동생, 7번 |
| | 간호보조원 | 3번, 4번, 5번 |
| 종교 | 천주교 | 1번, 2번, 3번, 3번의 아버지, 3번의 동생, 4번, 6번 |

5. 결 론

본 연구는 구술생애사를 대상으로 생성형 AI를 활용해 RDF 형식을 기반으로 하는 지식그래프의 구축 가능성을 확인 및 평가하고자 하였다. 구체적으로 파독간호사 7인의 출판된 구술생애사를 분석 대상으로 삼고, 텍스트의 원문 데이터에 대하여 생성형 AI가 RDF 트리플을 자동 생성하도록 하였으며, 그 결과는 다음과 같다.

첫째, 생성형 AI는 구술생애사의 서술적 특징을 일정 수준 반영하여 RDF 트리플을 생성하였다. 이를 통해 구술생애사와 같이 비정형 적이고 서사적인 자료에 대해서도 지식구조화를 시도할 수 있음을 확인하였다.

둘째, 생성형 AI는 RDF 트리플 추출에서 동일하거나 유사한 의미를 표현하는 술어를 일관되게 생성하지 못하였으며, 상황에 따라 다르게 표기하는 한계가 나타났다. 이러한 특성은 자연어 텍스트에서 표현의 다양성과 생성형 AI가 지닌 휘발성에 기인한 것으로 보이며, RDF 구축 과정에서 통합된 술어 집합 구성의 필요성을 시사한다.

셋째, 이에 본 연구에서는 생성된 술어들을 분석하여 생년, 가족, 자녀, 교육, 종교, 직업, 소속, 이동·이주, 거주 등 구술생애사의 핵심 범주 영역을 중심으로 유사 술어를 통합하고 표준화하였다. 이를 통해 개별 구술생애사의 RDF를 상호 연계할 수 있었으며, 다차원적 관점에서 생애사를 시각화하고 탐색할 수 있는 기반을 마련할 수 있음을 확인하였다.

한편, 본 연구에서 사용한 데이터는 인터뷰 내용을 채록하여 정리한 것으로, 문어체나 문

법적으로 구조화된 구어체에 비해 비정형 텍스트의 특성을 지닌다. 특히 구술 자료 특유의 비표준적 구어체, 사투리, 맥락 의존적 표현 등은 생성형 AI가 개체명을 정확히 인식하고 추출하는 데 잠재적인 장애 요인이 될 수 있다. 또한 생성형 AI는 스스로 인지하지 못하는 객체를 단순한 값으로 처리하였고, 한편 필요 이상으로 객체의 수가 늘어나는 등의 현상은 온톨로지 및 RDF를 구축하는 과정 전반에서도 나타나는 문제로, 생성형 AI를 통한 작업 과정에서도 확인할 수 있었다. 이는 생성형 AI가 전처리, 개체명 인식, RDF 변환을 동시에 수행하는데 있어 완전한 신뢰성을 확보하기 어렵다는 한계를 보여준다. 더불어 본 연구는 RDF 추출 과정에서 원문에 대한 근거 확인과 검증을 수행하였으나, 성과에 대한 체계적인 평가까지는 시도하지 못했다는 점 또한 한계로 지적할 수 있다.

그럼에도 불구하고 본 연구는 구술생애사 연구와 지식그래프 구축을 연결하는 탐색적 시도라는 점에서 의미가 크다. 구술생애사는 개인적 경험과 집단적 기억을 아우르는 중요한 인문 사회적 정보자원이다. 전통적으로 비정형 콘텐츠로 존재해왔다. 본 연구는 이러한 자료를 생성형 AI와 RDF 기술을 접목해 구조화하고, 나아가 디지털 인문학적 활용 가능성을 탐색했다는 점에서 의미가 있다. 특히 디아스포라 연구, 여성사 연구, 이주사 연구 등 다양한 맥락에서 본 연구의 구술생애사의 RDF 구축 시도는 학제적 연구를 위한 데이터 자원으로 확장될 수 있을 것이다.

또한 본 연구는 복잡하고 기계적인 절차를 거쳐야만 구축할 수 있던 지식그래프를 생성형

AI를 통해 보다 간소화된 방식으로 제시함으로써, 그동안 접근성이 낮았던 분야에서도 지식그래프를 쉽게 활용할 수 있는 가능성을 보여주었다. 이는 곧 지식그래프의 구축 필요성은 인식하고 있으나 복잡한 과정 때문에 접근하지 못했던 연구 분야에 새로운 기제가 될 수 있을 것이다.

추후 연구에서는 온톨로지 어휘를 활용한 술어와 객체의 통합화나 표준화 체계 정교화를 통하여 생성형 AI가 일관성 있게 RDF를 구축 할 수 있는 기반을 마련하고, 구술생애사 자료의 언어적 특성을 고려한 맞춤형 전처리 및 개체명 인식 모델을 적용한 RDF 구축의 정확성

향상을 위한 연구를 수행하고자 한다. 더불어 ChatGPT 이외에도 Claude나 Gemini 등 다양한 생성형 AI를 대상으로 지식그래프 구축을 시연하고, 그 결과를 비교 및 성과지표를 통해 평가함으로써 데이터 유형별로 적합한 생성형 AI를 규명하는 연구도 진행되어야 할 것이다.

궁극적으로 본 연구는 생성형 AI를 통한 RDF 구축이 완결된 성과를 제시하기보다, 이러한 방법의 가능성과 약한 탐색적 연구로서 의미가 있다. 이를 통해 구술생애사 연구의 방법론적 확장을 도모하고, 디지털 인문학적 연구의 새로운 방향을 제시하는 기초 자료로서의 가치를 지닐 것으로 기대한다.

참 고 문 헌

- 김성례 (2002). 여성주의 구술사의 방법론적 성찰. *한국문화인류학*, 35(2), 31-64.
- 문형진 (2023). 텍스트 마이닝을 활용한 조선족 생활문화 구술아카이브 분석. *기록과 정보·문화 연구*, 17, 125-156. <https://doi.org/10.23035/KAICS.2023.01.17.125>
- 박신규, 이준엽 (2023). 조선족 여성의 구술생애사에 나타난 삶의 특성과 변화: 토픽모델링 분석을 중심으로. *디아스포라연구*, 17(2), 7-37.
- 박신규, 이토히로코 (2023). 독일 웨른지역 파독간호사·광부의 구술생애사. 대구: 도서출판 책과 세계.
- 박지연, 김홍중 (2024). 한국전쟁 역사 자료에서 지식그래프와 온톨로지를 구축하기 위한 지식 추출 및 표현 기법. *史叢(사총)*, 111, 197-221. <https://doi.org/10.16957/sa..111.202401.197>
- 박하람, 김학래 (2021). 일본군 ‘위안부’ 지식그래프: 패편화된 디지털 기록의 연결. *한국기록관리학회지*, 21(3), 61-78. <https://doi.org/10.14404/JKSARM.2021.21.3.061>
- 서봉언, 이채문 (2014). 키워드 분석을 통해서 본 한국의 디아스포라 연구 동향. *디아스포라연구*, 8(1), 43-69.
- 유명상 (2025). 구술사와 공공역사의 통합적 접근 기록과 기록의 공론장. *기록과 정보·문화 연구*, 165-180. <https://doi.org/10.23035/KAICS.2025.1.특별호.165>
- 조은경, 강승식 (2020). 웹 자료 수집을 활용한 언어 지식그래프 개발: 정서 장애 주제 분야의 사례

- 연구. 한국어 의미학, 67, 203-228. <https://doi.org/10.19033/sks.2020.3.67.203>
- 한신갑, 이상직 (2017). 생애사 연구의 자료기반 확장: 생애사 시간표를 활용한 구술 생애사 자료의 조직 및 분석. 조사연구, 18(2), 99-136. <https://doi.org/10.20997/SR.18.2.4>
- Atkinson, R. (1998). *The Life Story Interview*. Los Angeles: SAGE Publications.
<https://doi.org/10.4135/9781412986205>
- Bergmann, D. (2024, November 7). What is a context window? Available:
<https://www.ibm.com/think/topics/context-window>
- Cherukuri, K. S., Moses, P. A., Sakata, A., Chen, J., & Chen, H. (2025). Large language models for oral history understanding with text classification and sentiment analysis. arXiv preprint arXiv:2508.06729. <https://doi.org/10.48550/arXiv.2508.06729>
- Hatch, J. A. (2002). *Doing Qualitative Research in Education Settings*. State University of New York Press.
- Liang, Z. (2024, July 8). Claude 3.5 Sonnet vs GPT-4o: Context Window and Token Limit. Available:
<https://prompt.16x.engineer/blog/clause-sonnet-gpt4-context-window-token-limit>
- Peng, C., Xia, F., Naseriparsa, M., & Osborne, F. (2023). Knowledge graphs: Opportunities and challenges. *Artificial Intelligence Review*, 56(11), 13071-13102.
<https://doi.org/10.1007/s10462-023-10465-9>
- World Wide Web Consortium (2014, February 25). RDF 1.1 Concepts and Abstract Syntax. Available: <https://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/Overview.html>

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

- Han, Shin-kap & Lee, Sang-jic (2017). Scaling up life histories: Life history calendar as a tool to repurpose oral history archive. *Survey Research*, 18(2), 99-136.
<https://doi.org/10.20997/SR.18.2.4>
- Jo, Eun-kyoung & Kang, Seung-sik (2020). Development of linguistic knowledge graph with collecting language materials in the web: A case study of the subject field of affective disorder. *Korean Semantics*, 67, 203-228. <https://doi.org/10.19033/sks.2020.3.67.203>
- Kim, Seong Nae (2002). Methodology of feminist oral history. *Korean Cultural Anthropology*, 35(2), 31-64.
- Moon, Hyoung-jin (2023). The analysis of the oral archives of Joseonjok living culture using text

- mining. *The Korean Journal of Archival, Information and Cultural Studies*, 17, <http://doi.org/10.23035/KAICS.2023.01.17.125>
- Park, Haram & Kim, Haklae (2021). A knowledge graph on Japanese “Comfort Women”: Interlinking fragmented digital archival resources. *Journal of Korean Society of Archives and Records Management*, 21(3), 61-78. <http://doi.org/10.14404/JKSARM.2021.21.3.061>
- Park, Jiyeun & Kim, Hongjoong (2024). Knowledge extraction and representation techniques for building knowledge graphs and ontologies from Korean War history data. *Sa Chong (The Historical Journal)*, 111(0), 197-221. <https://doi.org/10.16957/sa..111.202401.197>
- Park, Shin-Kyu & Ito, Hiroko (2023) Oral Life Histories of Korean Nurses and Miners Dispatched to Cologne, Germany. Daegu: Books & World Publishing.
- Park, Shin-Kyu & Lee, June-yeop (2023). Characteristics and changes in life revealed in the oral life history of Korean-Chinese women: Focusing on topic modeling analysis. *Journal of Diaspora Studies*, 17(2), 7-37.
- Seo, Bongeon & Lee, Chaimun (2014). Keyword analyses of diaspora research and its trends in Korea. *Journal of Diaspora Studies*, 8(1), 43-69.
- Yoo, Myengsang (2025). An integrated approach of oral history and public history: A public forum of memory and records. *The Korean Journal of Archival, Information and Cultural Studies*, 165-180. <https://doi.org/10.23035/KAICS.2025.1.특별호.165>

