Journal of Korean Library and Information Science Society
Vol. 55, No. 4

시맨틱 추론 규칙을 이용한 대규모 언어 자원의 품질 고도화 방안*

- 국립중앙도서관 주제명표목표를 중심으로 -

Method for Enhancing the Quality of Large-Scale Lexical Resources Using Semantic Inference Rules: Focusing on the National Library of Korea Subject Headings

정 도 헌 (Do-Heon Jeong)**

〈 목 차 〉

요 약 : 본 연구는 대규모 언어 자원에 관한 연구 동향을 바탕으로, 자동화된 시맨틱 추론 기법을 이용한 거대 언어 자원의 효율적 품질 제고 방안을 제시하고 응용 가능성을 제안하고자 한다. 이를 위해, 우선 언어 자원 내 용어 간의 다양한 관계를 분석하여 도출한 정오 사례를 바탕으로 공통의 시맨틱 추론 규칙을 정의하였다. 정의된 추론 규칙을 바탕으로 거대한 언어 자원의 네트워크를 고속 탐색하고 오류를 검출하는 스프레딩 알고리즘 기반의 시맨틱 추론 엔진을 개발하였다. 국립중앙도서관의 주제명표목표에 대한 실험을 통해, 본 연구에서 제안한 자동화 기법과 웹 기반 관리 시스템을 활용하여 대용량 데이터의 품질 고도화 작업을 효율적으로 수행할 수 있음을 확인하였다. 본 연구는 대규모 언어 자원의 품질 고도화를 위한 시맨틱 추론 기법을 새롭게 제안한 점, 복잡한 용어 네트워크에서 발생하는 논리적 오류 사례를 분석하고 체계화하는 최초의 시도였다는 점에서 의의가 있으며, 일련의 과정을 통해 인공 지능 시대의 인간과 기계의 협업 방식을 논의하였다는 데 의의가 있다.

주제어 : 언어 자원, 대용량 데이터, 시맨틱 추론, 인공 지능, 주제명표목표, 인간 기계 협업

ABSTRACT : The purpose of this study is to propose an efficient quality enhancement method for large-scale lexical resources using automated semantic inference techniques, based on current research trends in large lexical resources, and to suggest practical applications. To achieve this, common semantic inference rules were first defined by analyzing various relational cases among terms within lexical resources and identifying correct and erroneous patterns. Using these defined inference rules, a semantic inference engine based on a spreading algorithm was developed, enabling rapid network traversal and error detection across very large lexical resources. Through experiments on the Subject Headings of the National Library of Korea, it was confirmed that the automated methods and web-based management system proposed in this study enable effective quality enhancement of large-scale data. The study is significant in that it proposes a novel semantic inference approach for enhancing the quality of large-scale lexical resources, as well as the first attempt to analyze and organize logical error cases arising within complex term networks. Furthermore, it is meaningful in discussing methods of human-machine collaboration in the era of artificial intelligence.

KEYWORDS : Lexical Resources, Large-Scale Data, Semantic Inference, Artificial Intelligence, Subject Headings, Human Machine Collaboration

* 본 연구는 2023년도 덕성여자대학교 교내연구비 지원에 의해 이루어졌음(3000008144).
** 덕성여자대학교 글로벌융합대학 문헌정보학전공 부교수
(doheonjeong@duksung.ac.kr / ISNI 0000 0004 6099 1600)

• 논문접수: 2024년 11월 23일 • 최초심사: 2024년 12월 8일 • 게재확정: 2024년 12월 16일
• 한국도서관·정보학회지, 55(4), 245-263, 2024. http://dx.doi.org/10.16981/kliss.55.4.202412.245

Ⅰ. 서 론

대규모 언어 자원(lexical resources)은 현대 정보 사회에서 그 중요성이 더욱 커지고 있다. 시소러스(thesaurus)와 워드넷(Wordnet)은 단어 간의 의미적 관계를 체계적으로 정리한 지식 베이스로 질의 확장과 텍스트 유사성 측정 등 자연어 처리(NLP; natural language processing)와 텍스트 마이닝 연구 분야에서 광범위하게 활용되고 있으며(허고은, 2019; Navigli & Ponzetto, 2012), 위키피디아(Wikipedia)/디비피디아(DBpedia)는 방대한 양의 구조화된 데이터를 기반으로 지식 그래프 구축 및 정보 추출 연구 등에 기여하고 있다(Nothman et al., 2013; Paulheim, 2017). 또한, 다양한 융합적 연구가 진행되면서 다국어 시소러스 구축, 워드넷과 딥러닝 모델의 결합, 위키피디아 기반의 지식 베이스 확장 등 언어 자원의 활용 범위가 더욱 넓어지고 있다. 최근에는 생성형 인공 지능(generative artificial intelligence)의 시대가 열리면서 기반이 되는 거대 언어 모델(LLM; large language model)의 발전까지 촉진하고 있어, 언어 자원을 비롯한 지식 자원의 활용 가치는 더욱 증가하게 되었다(An et al., 2019; Marciniak, 2020).

이렇듯 공공 기관을 비롯한 많은 기관이 보유한 대규모 지식 베이스의 효율적 품질 제고 방안이 필수적임에도 용어 간에 설정된 복잡한 관계 정보를 기계적으로 진단하는 시맨틱 추론 기반의 자동화 방안은 아직 본격적으로 연구된 바가 없다. 이에 본 연구는 기구축된 대규모 언어 자원에 대해 전문가가 직접 수행하기 어려운 용어 간 시맨틱 관계의 오류 존재 여부를 기계적으로 점검하는 새로운 품질 관리 방안을 제안하고자 한다. 이를 위해 다양한 용어 관계를 분석하여 도출된 논리적 오류 사례를 바탕으로 시맨틱 추론 규칙의 기초 체계를 구축하고자 하며, 수립된 규칙 체계는 향후 관련 연구의 발전에 기여할 것으로 기대한다. 마지막으로, 본 연구에서 제안한 일련의 품질 제고 과정을 통해 인간과 기계의 협업 방식에 대해 논의하고자 한다. 향후 후속 연구를 통해 고품질의 국가 지식 자원을 기반으로 공공 기관의 독자적인 인공 지능 모델(on-device LLM)을 구축하는 데 기여할 수 있을 것으로 기대한다.

Ⅱ. 이론적 배경

1. 대규모 언어 자원 활용 연구

언어 자원을 활용한 응용 연구는 주로 시소러스, 워드넷, 위키피디아/디비피디아와 같은 대규모 자원을 이용하여 정보 검색, 텍스트 마이닝, 자연어 처리 등 여러 분야에서 활발하게 진행되고 있다. 다양한 언어 간의 시소러스를 구축하여 다국어 정보 검색 시스템에 활용하는 다중 언어

시소러스 구축 연구가 수행되었으며(Navigli & Ponzetto, 2012), 워드넷을 기반으로 텍스트 마이닝 기법을 활용하여 용어의 의미적 관계성을 파악함으로써 단어의 불확실성을 해소하고자 하는 연구가 수행되기도 하였다(허고은, 2019). 또한, 위키피디아를 기반으로 구조화된 개체 간의 시맨틱 지식 그래프를 구축하여 정보 검색 및 질의 응답 시스템에 활용하는 연구가 수행된 바 있으며(Paulheim, 2017), 방대한 언어 정보를 기반으로 텍스트에서 유용한 정보를 자동 추출하는 NER(named entity recognition) 연구가 수행되기도 하였다(Nothman et al., 2013).

언어 자원을 활용한 융합 연구 사례로, 워드넷의 의미 관계와 딥러닝 모델을 결합하여 텍스트의 의미를 더 정교하게 파악하는 연구(Saedi et al., 2018) 등이 수행되기도 하였다. 최근에는 언어 자원을 활용해 생성형 AI 기술의 학습 모델인 거대 언어 모델(LLM)의 성능을 향상시키기 위한 다양한 시도가 수행되고 있다. 워드넷, 시소러스와 같은 언어 자원이 용어 간의 의미적 관계를 풍부하게 학습시키는 데 효과적이며, 이를 통해 인공 지능 모델의 문맥 이해, 문장 생성, 질의 응답 능력이 향상될 수 있음을 보여준다(An et al., 2019; Marciniak, 2020).

2. 언어 자원 자동 구축 및 품질 고도화 연구

대량의 정보원으로부터 의미 관계를 추출하여 새로운 지식을 생성하고자 하는 응용 연구가 다수 수행되어 왔다. 논문, 특허 등 기술 문서의 정의문으로부터 용어를 추출하고 사전을 구축하고자 하는 연구(한희정 외, 2017), 학술정보로부터 전문 용어의 시맨틱 네트워크를 자동 생성하고 통제하여 언어 자원의 활용률을 제고하고자 하는 연구(정도헌, 2018) 등이 수행된 바 있다. 또한, 한국어 디비피디아를 기반으로 새로운 트리플 관계 데이터를 생성하여 기존의 지식 베이스를 증강시키는 방안(김선동, 강민서, 이재길, 2014), 자동화된 방법으로 대량의 학술 논문으로부터 기술 용어의 시맨틱 네트워크를 자동 생성하고 두문자어의 의미를 식별하는 WSD(word sense disambiguation) 방안(Jeong, Hwang, & Sung, 2011) 등의 다양한 지식 베이스 구축 및 활용 연구가 수행된 바 있다.

언어 자원의 품질 평가 및 개선 관련 연구와 관련하여, 위키피디아를 기반으로 지식 베이스의 품질과 신뢰성을 높이는 방법과 응용 시스템을 소개한 연구가 있었으며(Mendes et al., 2011), 디비피디아에 온톨로지를 추가하여 데이터의 의미적 정합성과 품질을 개선하고자 하는 연구도 수행된 바 있다(Paulheim & Gangemi, 2015).

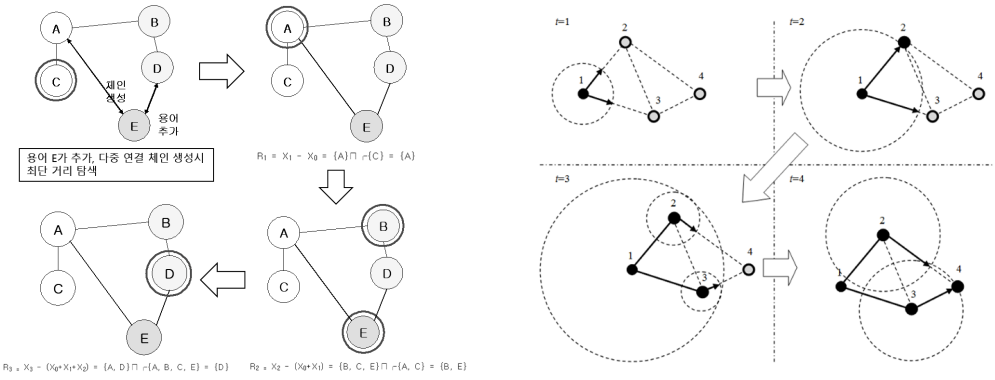
국립중앙도서관 주제명표목표의 활용 및 품질 제고 연구와 관련하여, 주제명표목표의 주제명 구축 및 활용 현황에 대한 심층 분석을 통해 발전 방안을 제시하고자 하는 연구가 있었으며(이혜경, 이용구, 2023), 국립중앙도서관의 주제명표목표를 활용한 검색 시스템의 개선 방안을 제안하거나(백지원, 정연경, 2014), 고품질의 주제명표목표를 구축함으로써 관련 서비스의 향상

을 도모하고자 한 품질 제고 연구가 수행된 바 있다(여지숙 외, 2022; 최윤경, 정연경, 2014). 그러나 언어 자원에 설정된 복잡한 시맨틱 관계 정보를 분석하여 논리적 관계 오류를 검출하는 자동화된 품질 관리 방안은 본격적인 연구가 수행된 바 없다. 이에 본 연구는 전문가가 직접 수행하기 어려운 대규모 언어 자원에 대해 시맨틱 추론 기술 기반의 효율적인 품질 고도화 방안을 제안하고자 한다.

3. 용어의 시맨틱 네트워크 탐색 기법

스프레딩 알고리즘(spreading algorithm)은 그래프 이론에서 노드 간의 영향을 퍼뜨리는 방법이다. 주로 네트워크 상에서 정보, 질병, 영향력 등이 퍼져 나가는 과정을 모방하는 데 유용한 개념으로 네트워크 그래프 간의 최단 경로의 탐색 및 최적화 등에 활용되고 있다. 대표적인 알고리즘 중 하나인 파문 확산 알고리즘(RSA; Ripple Spreading Algorithm)은 물결의 파문을 모방하여 다중 경로 문제(KSPP; K-shortest path problem)를 해결하고자 하였다(Hu et al., 2012). 또한, 다양한 자원으로부터 통합 언어 자원을 구축하고 의미 관계로 형성된 복잡 네트워크의 최단 거리 탐색을 위한 시맨틱 네트워크 탐색 알고리즘(SNSA; semantic network search algorithm)이 제안되기도 하였다(정도헌, 최희윤, 2006).

〈그림 1〉은 스프레딩 알고리즘의 확산 방식에 대해 각 알고리즘의 특징을 도식화하여 소개하고 있다. 시맨틱 네트워크 탐색 알고리즘은 대용량 언어 자원의 고속 처리를 위해 해시테이블 구조를 기반으로 시맨틱 관계를 추론하기 위한 용도로 특화되어 개발되었으므로, 기능의 개선과 확장이 용이하여 본 연구를 위한 추론 방법의 기본 알고리즘으로 활용하고자 한다.



〈그림 1〉 시맨틱 네트워크 탐색 알고리즘(정도헌, 최희윤, 2006, 그림 일부 수정함)과 파문 확산 알고리즘(Hu et al., 2012)의 최단 거리 탐색 방식 비교

시맨틱 네트워크 탐색 알고리즘, 파문 확산 알고리즘과 같은 스프레딩 알고리즘은 특정 지점에서 사방으로 확산해 나가는 방식으로 최단 경로를 찾는 너비우선탐색(BFS; breadth-first search) 방식의 알고리즘이기 때문에 일정한 단계 수를 제한하거나 특정 조건을 부가하여 경로를 탐색하는 것이 가능하다는 장점이 있다. 단, 본 연구를 수행하기 위한 기능을 구현하기 위해서는 경로의 길이나 연결 단계 수를 통제하여 지정된 단계를 거치도록 하는 제약 기능을 구현 일부 알고리즘을 수정해야 한다. 파문 확산 알고리즘은 에지(edge)의 가중치를 고려하는 모델인 반면, 시맨틱 네트워크 탐색 알고리즘은 가중치가 없는 단순 모델이므로 특정 노드를 출발해 되돌아오는 순환 패턴 탐지 기능의 구현에 용이하여, 본 연구에서 제안한 시맨틱 추론 엔진 개발을 위한 기본 알고리즘으로 채택하였다.

Ⅲ. 시맨틱 추론 기반 언어 자원 품질 고도화

1. 시맨틱 네트워크의 관계 탐색 및 자동 검출 방법

본 장에서는 우선 스프레딩 알고리즘을 이용해 다단계(multi-degree)로 연결된 용어의 시맨틱 네트워크를 탐색하고 오류 판정을 위한 단위 네트워크를 검출하는 추론 방법을 설명하고, 오류를 진단하기 위한 추론 규칙을 도출하는 과정을 소개한다. 본 연구에서는 언어 자원에서 가장 중요하다고 할 수 있는 용어 간 관계 오류 요소를 찾아 이를 제거함으로써 고품질을 유지하고자 하였다. 따라서 전문가가 수작업을 통해 파악하기 어려운, 네트워크 내에 존재하는 A-B-C-A 형식으로 연결되는 삼각형 형태의 연결 관계를 우선 도출하고자 하였다(〈그림 3〉 참고). 삼각 연결 관계는 본 연구에서 제안한 기계적인 추론을 하기 위한 가장 작은 단위의 네트워크 구조이다.

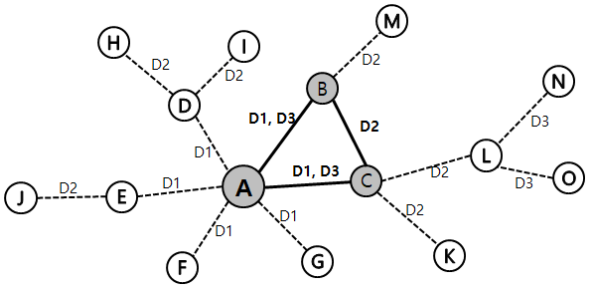
대량의 복잡한 네트워크 내에서 용어의 삼각 연결 관계를 찾아내고 진단하기 위한 고속의 탐색 도구를 개발하기 위해 시맨틱 네트워크 탐색 알고리즘을 기반으로 부가 기능을 추가하였다. 〈공식 1〉은 기본 추론을 위한 이행 함수 관계 “if A → B and B → C, then A → C”를 처리하기 위한 탐색 알고리즘이다. 네트워크 탐색 과정에서 용어 탐색이 중복적으로 일어나므로 반복 발생하는 노드 값을 제어하기 위해 해시 구조 기반의 고속 연산 처리를 실행한다. N단계(degree)의 노드 값인 Rn은 네트워크를 반복 탐색하여 결과 값이 공집합이 될 때 연결 차수를 반환하며 종료된다. 앞서 소개한 바와 같이 이 방식은 가중치가 없는 연결 중심의 모델이므로 시소러스와 같은 용어 간의 관계 데이터 처리에 특화되어 있다.

$$R_n = X_n - S_{n-1} \ (S_n = \sum_{i=0}^{n} X_i) \text{ 또는 } R_n = X_n - \sum_{i=0}^{n-1} X_i$$

X_n: n차 단계에서의 이웃 노드 집합 (단, X_0는 시작 노드 값)
R_n: n차 단계의 최종 노드 집합

〈공식 1〉 시맨틱 네트워크 탐색 알고리즘(정도헌, 최희윤, 2006)

먼저, 기존의 최단 거리 탐색 알고리즘을 〈그림 2〉의 예시를 통해 설명하면 (1) 시작 노드 A에서 1단계 탐색이 진행된 결과는 $R_1 = X_1 - X_0 = \{B, C, D, E, F, G\} \cap \{A\}^C = \{B, C, D, E, F, G\}$ (2) 1단계 탐색 결과 집합에서 2단계 연결 관계로 진행한 결과는 $R_2 = X_2 - \sum_{i=0}^{1} X_i = \{A, M, C, B, L, K, H, I, J\} \cap \{A, B, C, D, E, F, G\}^C = \{M, L, K, H, I, J\}$ 이다. 따라서, A노드로 부터 2단계 진행 결과로 각각 노드 M, L, K, H, I, J로 확장된다. (3) 2단계 결과 노드로 부터 다시 3단계까지 진행한 결과는 $R_3 = X_3 - \sum_{i=0}^{2} X_i = \{B, C, N, O, D, E\} \cap \{A, B, C, D, E, F, G, M, L, K, H, I, J\}^C = \{N, O\}$ 로 계산된다. 출발 노드 A로부터 최종 N, O 노드까지 최단 거리인 3단계를 거쳐 도달하게 된다.



〈그림 2〉 시맨틱 네트워크 내 삼각 연결 관계 탐색 예시
(에지의 속성 값 D는 연결 차수(degree)를 의미하며, 출발 노드인 A를 기준으로 D1, D2, D3로 표기함)

이와 같이, 기존의 알고리즘은 출발 노드로부터 거리가 멀어지는 확산 모델이므로, 아래 〈공식 2〉와 같이 본 연구를 위해 알고리즘을 일부 수정하였다. 3단계에서 출발 노드로 돌아오는 회귀 패턴인 삼각 연결 관계를 검출하기 위해 생성된 해시테이블에서 출발 노드(X_0)를 사전 제거한 후 확산 탐색을 진행하도록 변경되었다.

$$R_n = X_n - \sum_{i=0}^{n-1} X_i \ (n=1 \text{ or } 2\text{일 때}),\ R_n = X_n - (\sum_{i=0}^{n-1} X_i - X_0) \ (n=3\text{일 때})$$

X_n: n차 단계에서의 이웃 노드 집합 (단, X_0는 시작 노드 값)
R_n: n차 단계의 최종 노드 집합

〈공식 2〉 용어의 삼각 연결 관계를 찾기 위한 수정 알고리즘

〈그림 2〉를 예로 들어 네트워크 내 존재하는 삼각 연결 관계를 찾아내는 과정 즉, 노드 A로부터 출발해 B와 C를 거쳐 다시 A로 돌아오는 회귀 사이클을 추출하는 과정을 간략히 설명하면 다음과 같다. N차 단계에서의 연결 용어 집합을 R_n이라 할 때, 출발 노드인 A를 기준으로 $R_1(A) = \{B, C, D, E, F, G\}$ 이다. 각 노드에 대해 다시 2단계 탐색을 진행한 결과 $R_2(A) = \{(B,M), (B,C), (C,B)\ (C,L), (C,K), (D,H), (D,I), (E,J)\}$ 의 2차원 구조로 나타낼 수 있다. 동일한 방식으로 3단계까지 최단 거리 탐색을 한 결과, 2단계에서 진행이 종료되는 경우를 제외하면 $R_3(A) = \{(B,C,A), (B,C,K), (C,B,A), (C,B,M), (C,L,N), (C,L,O)\}$ 만 남게 된다. 이때 각 항목의 세 번째 값이 출발 노드인 A가 되는 경우인 $(B, C, A), (C, B, A)$를 반환하고 중복 제거 과정을 거쳐 최종 데이터는 (A, B, C)가 된다. 이러한 단위 탐색 과정을 네트워크 내의 모든 노드에 대해 반복적으로 실시함으로써 전체 네트워크 내의 모든 관계를 도출할 수 있다. 연구를 수행하기 위해 Python 프로그래밍 언어와 네트워크 분석용 라이브러리인 NetworkX를 이용해 용어의 삼각 연결 관계 탐지가 가능한 고속 탐색 엔진 및 실시간 오류 분석 도구를 모두 직접 개발하여 활용하였다.

2. 용어 간 의미 관계 오류 진단을 위한 추론 규칙 정의

추론 기반의 논리적 오류 점검을 위한 최소 단위로 세 용어가 상호 연결된 순환 구조인 삼각 연결 관계를 제안하였다. 본 장에서는 추론 규칙을 작성하기 위해 논리적 오류 유형을 분석하고자 한다. 삼각 연결 관계에서 세 용어 간에 다양한 관계 유형이 발생하며 이를 〈표 1〉과 같이 정리하였다. 국립중앙도서관의 주제명표목표 업무지침(국립중앙도서관, 2021)의 관계를 기반으로 추론에 필요한 관계를 추가하고 사용하지 않는 항목을 제거하여 기초 관계 정의표를 완성하였다. 주제명표목표를 비롯한 모든 언어 자원에서 가장 중요한 기본 관계인 동의어의 경우, 주제명표목표에서는 UF와 USE 관계로 우선어를 설정하고 있으며 이는 방향성이 있는 등위 관계이다. 실제 동의어 관계에서 다양한 형태로 나타나는 포괄적인 동의어 개념을 처리하기 위해 본 연구에서는 무방향성의 Qsyn(quasi-synonym) 관계를 추가하였고, 다시 동의어 그룹을 개념화하여

SYN 그룹으로 설정하였다. SYN 관계의 활용에 대한 설명은 이후 오류 유형 분석 시 자세히 다룬다. 워드넷에서는 전체-부분의 포함 관계(meronymy)를 선언하고 부분(meronym)과 전체(holonym) 관계 등을 세부적으로 정의하고 있으나(Miller, 1995), 본 연구에서는 국립중앙도서관의 주제명표목표를 실험 대상 데이터로 사용하므로, 자료의 관계 설정 범위를 벗어나므로 다루지 않는다.

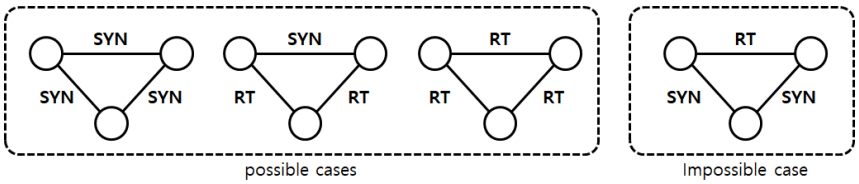
〈표 1〉 시맨틱 추론을 위한 용어 간 관계 설정

그룹 속성	계층속성	관계 표기	방향성 유무	설명	주제명표 목표 관계	추론 엔진 사용 여부
SYN (synonym group)	등위관계	USE	방향성	우선어 지정	○	○
		UF	방향성	비우선어 지정 (used for)	○	○
		QSyn	무방향성	USE, UF 관계 외의 동의어 포괄 (quasi-synonym)	X	○
		PT	방향성	바뀌기 전의 용어(prior term)	○	X
		LT	방향성	바뀐 후의 용어(later term)	○	X
-	계층관계	BT	방향성	상위어(broader term, hypernym)	○	○
		HOL	방향성	전체 관계(holonym)	X	X
		NT	방향성	하위어(narrower term, hyponym)	○	○
		MER	방향성	부분 관계(meronym)	X	X
-	연관관계	RT	무방향성	관련어(related term)	○	○

가. 동의어와 관련어 관계: 무방향성(undirected), 비계층적 등위 관계 검출

우선 언어 자원에서 가장 중요한 기초 관계인 동의어 관계를 중심으로 주요 관계를 설정하고 논리적 오류 유형을 선언하였다. 우선어, 비우선어 관계를 중심으로 하는 동의어 관계의 추론에 앞서 동의어와 관련어로 구성된 용어 네트워크 추론을 우선 진행하였다. 먼저, 동의어 그룹(SYN) 개념을 사용하여 방향성 데이터(USE, UF)를 무방향의 관계 데이터로 전환하여 추출한다. SYN은 RT와의 관계 해석이 용이하도록 동의어 관계를 대표하여 사용하는 추상화된 개념이다. 본 연구에서는 용어 간 시맨틱 관계를 설명하기 위해 트리플 데이터를 표현하는 표준 방식인 RDF(resource description framework) 그래프와 터틀(Turtle) 데이터 포맷으로 기술하였다. RDF 그래프는 주로 정오 추론을 위한 논리의 시각적 관계 표현을 위해 사용하고, 실제 추출된 데이터의 예시를 위해 터틀 형식을 사용한다.

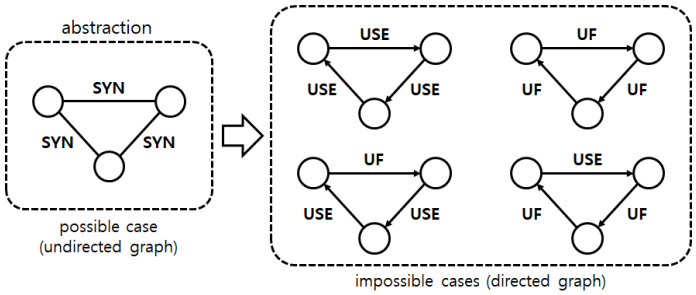
SYN과 RT로 구성되는 모든 네트워크 형태는 〈그림 3〉과 같이 RDF 그래프 형식으로 표현할 수 있다. 이때 불능 관계를 나타내는 경우는 세 용어 간의 관계가 SYN-SYN-RT(순서는 무관)로 선언된 경우이며, 이러한 논리적 오류를 해소하기 위해 오류 관계의 수정 또는 삭제가 필요하다. 전문가의 데이터 검수를 통해 RT를 SYN으로 변경하거나 SYN을 하나 이상 RT로 변경하면 논리적 오류가 해소될 수 있다.



〈그림 3〉 동의어 그룹과 관련어로 구성된 삼각 연결 관계의 시맨틱 관계 유형

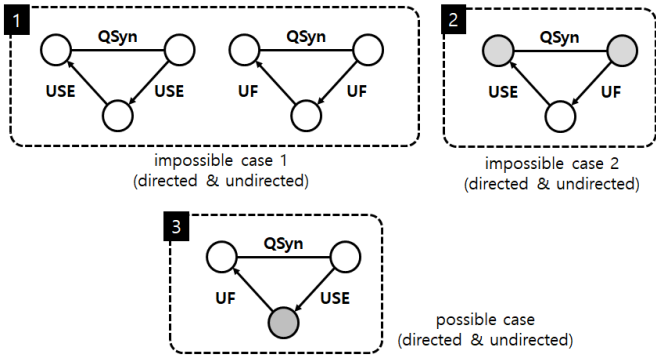
나. 동의어의 우선어, 비우선어 관계: 방향성(directed), 비계층적 등위 관계

우선어, 비우선어로 구성된 용어의 시맨틱 네트워크에서는 한 개의 대표어만 존재하도록 관계를 통제해야 한다. 시소러스 기반의 시스템에서 대표어 설정의 오류는 각종 응용 서비스 품질과 직결되므로 최우선으로 해소해야 하는 중요한 문제라 할 수 있다. 〈그림 4〉는 USE와 UF 관계만으로 구성된 삼각 네트워크에서의 불능 관계를 설명한다. SYN 개념을 활용해 RT와의 가능/불능 여부를 1차 검증한 후, SYN 상태에서 정상으로 판단된 경우를 USE와 UF로 전환하여 2차 검증을 실시한다. 이때 모든 방향성 데이터는 같은 방향으로 정렬하는 전처리를 거친다. 동의어 그룹 SYN으로 연결된 관계(좌측)는 USE, UF만으로 구성된 관계가 되었을 때(우측) 특정 용어로 우선어가 지정될 수 없는 불능 관계가 됨을 알 수 있다.



〈그림 4〉 방향성 등위 관계(USE와 UF)만으로 구성된 삼각 연결 관계의 불능 유형

복잡한 언어적 관계에서는 USE, UF와 같은 방향성 있는 관계와 함께 대표어의 개념이 없는 무방향성 동의어(Qsyn) 관계가 혼재되는 경우가 흔히 발생하므로 이에 대한 규칙 설정이 필요하다. 〈그림 5〉의 1번 유형은 같은 관계의 연속으로 인해 대표어를 특정할 수 없으므로 불능이다. 2번 유형은 UF→USE의 순서로 인해 대표어 지정이 두 곳에 이루어져 불능 상태이다. 3번의 경우만이 USE→UF 순서로 한 개의 노드를 대표어로 지정할 수 있는 정상적인 경우이다. 〈그림 5〉에 표현되지 않았지만, 두 개의 무방향성 데이터와 한 개의 방향성 데이터로 구성된 경우는 모두 논리적으로 정상 판정이 가능하다.



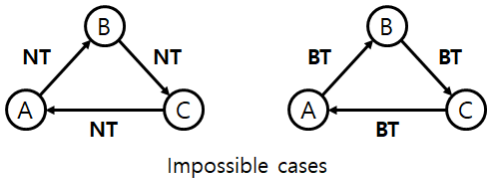
〈그림 5〉 방향성과 무방향성 동의어 관계가 혼재된 삼각 연결 관계의 정오 판정

앞서 논의한 동의어 관련 패턴을 종합하여 공통의 불능 조건을 도출한 최종 추론 규칙은 다음과 같다.

- 방향성 관계가 두 개 이상 포함된 모든 동의어 네트워크에서 USE-USE 또는 UF-UF의 연속적인 패턴이 존재하는 경우는 모두 불능
- UF->USE 순서의 패턴이 존재할 경우는 불능

다. 방향성 있는 계층(hierarchical) 관계 추론

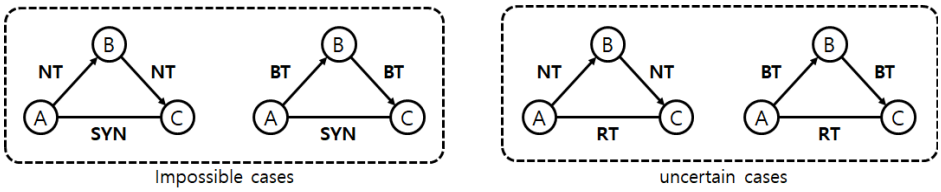
USE, UF는 비계층적이고 방향성이 있는 등위 관계인 반면 상위어(BT)와 하위어(NT)는 방향성이 있는 계층 관계이다. 계층 관계의 추론에는 크게 두 가지의 경우가 존재한다. 방향성 있는 계층 관계로만 구성된 경우와 SYN 또는 RT가 포함되어 방향성과 무방향성 관계가 혼재되어 구성된 경우이다. 첫째, 계층 관계로만 구성된 경우의 불능 상태는 아래 〈그림 6〉과 같이 RDF 형식으로 도식화하여 확인 가능하다. 같은 방향으로 정렬된 용어의 삼각 연결 관계가 NT 또는 BT로만 구성된 경우에는 상위어와 하위어를 특정할 수 없는 상태이므로 모든 관계에 대한 오류 검증 과정이 필요하다.



〈그림 6〉 방향성 계층 관계 BT, NT만으로 구성된 삼각 연결 관계의 불능 상태

둘째, 비계층 무방향 관계인 SYN과 RT가 BT, NT와 함께 나타나는 경우의 해석은 아래 〈그림 7〉과 같다. 이때 RT는 광범위한 관계어를 선언하므로 방향과 계층 개념이 모호하여 명확한 추론

규칙을 적용하기 어렵다. 따라서, SYN-NT-NT와 SYN-BT-BT로 구성된 네트워크는 명확히 불능인 반면, RT가 포함된 경우는 불확실한 상태(uncertain cases)로 판정하였다.



〈그림 7〉 NT, BT 계층 관계와 SYN, RT 비계층 관계가 혼재된 경우의 판정

BT, NT를 중심으로 SYN, RT와의 동시 발생을 고려하여 모든 조합의 경우를 분석하기 위해 〈표 2〉와 같이 종합 정리하여 진단하였다. 앞서 언급한 바와 같이 RT가 한 개 이상 포함된 관계는 모두 불확실한 상태로 간주하므로, 〈표 2〉의 좌측의 NT를 중심으로 표기한 16행(RT는 음영 부분) 이후에는 반복적인 기술이 불필요하다 판단되어 RT 포함 관계들을 표에서 삭제하였다.

분석한 주요 내용을 설명하면, SYN 관계를 포함하는 경우에 BT-NT-SYN으로 혼합 구성된 관계에서 오류가 없으며 BT-BT-SYN 또는 NT-NT-SYN 등 이외의 관계 시 모두 오류로 판정된다. 계층 관계로만 이루어진 경우에 BT-BT-NT 또는 BT-NT-NT로 구성된 관계 이외는 모두 오류로 판정된다.

〈표 2〉 계층 관계 BT 또는 NT를 한 개 이상 포함한 삼각 연결 관계의 추론 규칙 종합표

용어 간 의미 관계			가능(○)/불능(X)/불확실(△) 구분
A→B	B→C	C→A	
NT	NT	NT	X
		BT	○
		SYN	X
		RT	△
	BT	NT	○
		BT	○
		SYN	○
		RT	△
	SYN	NT	X
		BT	○
		SYN	X
		RT	△
	RT	NT	△
		BT	△
		SYN	△
		RT	△

용어 간 의미 관계			가능(○)/불능(X)/불확실(△) 구분
A→B	B→C	C→A	
BT	NT	NT	○
		BT	○
		SYN	○
	BT	NT	○
		BT	X
		SYN	X
	SYN	NT	○
		BT	X
		SYN	X
SYN	NT	NT	X
		BT	○
		SYN	X
	BT	NT	○
		BT	X
		SYN	X
	SYN	NT	X
		BT	X

결론적으로, “방향성 있는 계층 관계인 BT 또는 NT가 한 개 이상 존재하는 삼각 연결 관계”라는 조건 하에서의 시맨틱 추론 규칙은 다음과 같이 요약 정리할 수 있다.

- 모든 경우에 RT가 한 개 이상 존재하면 불확실한 상태로 정오 판정 보류
- NT 또는 BT 단일 관계로만 구성된 삼각 네트워크는 불능 판정
- BT 2개와 SYN 1개, 또는 NT 2개와 SYN 1개로 구성된 경우는 불능 판정
- 계층 관계와 함께 SYN이 2개 포함된 경우는 모두 불능 판정

Ⅳ. 실험 및 데이터 검증

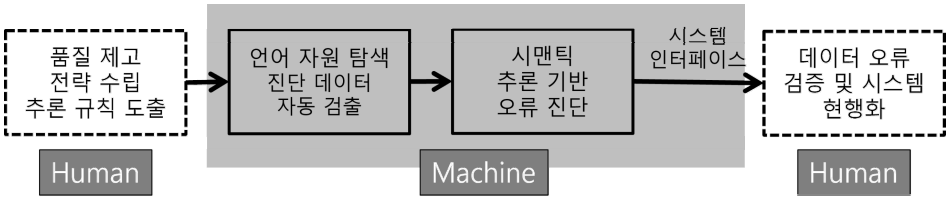
1. 실험 데이터

본 연구에서 제안한 시맨틱 추론에 기반한 대규모 언어 자원의 품질 제고 방안을 실제 적용하기 위한 실험 데이터 셋으로 국립중앙도서관으로부터 주제명표목표의 현행 서비스용 데이터 전량을 제공받아 연구에 활용하였다. 국립중앙도서관 주제명표목표는 국가 장서의 주제 접근이 용이하도록 통제 어휘를 활용하여 표준화된 주제 목록 작성을 지원하는 도구이며 1992년 주제명 검색요어집을 바탕으로 2002년 국립중앙도서관 주제명표목표로 명명된 대규모 시소러스 자원이다(국립중앙도서관, 2021). 주제명표목표의 용어 참조 관계는 동의어 관계(등위 관계), 계층 관계, 연관 관계, 이전/이후 관계 등으로 구분하고 있다. 실험을 위해 구축된 주제명표목표의 용어는 총 488,314건이며, 관계가 정의된 용어는 491,689건으로 나타났다. 이 중 RT 관계가 406,882쌍, BT와 NT가 각각 191,573쌍과 191,570쌍으로 약간의 차이를 보였다. 우선어는 232,657건, 비우선어는 55,176건이며, 이때 모든 비우선어에 대한 우선어가 정의되어 있으므로 USE/UF의 관계는 총 55,176쌍이다. 등위 관계인 이전/이후 관계는 0쌍으로 실험에 포함되지 않았으며, 외국어로 표기된 용어는 148,879건이지만 국립중앙도서관과의 협의 하에 외국어 데이터를 제외한 한국어 용어만을 실험 대상으로 삼았음을 밝힌다.

2. 시맨틱 추론 결과 분석 및 검증

본 장에서는 기계적인 시맨틱 추론 방법과 시스템 환경을 활용하여 전문가가 직접 수행하기 어려운 품질 관리 작업을 효율적으로 수행하는 일련의 프로세스를 종합하여 시스템의 최종 모델로 제안하고자 한다. 〈그림 8〉은 전문가의 지적 작업에 의해 오랜 기간에 걸쳐 구축된 대규모의 언어

자원을 자동화된 방법으로 품질을 진단한 후 다시 전문가가 검증하는 인간과 기계의 협업 과정을 주요 단계별로 설명한다. 우선, 전문가에 의해 품질 점검에 대한 전략을 수립하고 다양한 용어 관계 분석을 통해 추론 규칙 집합을 정의하여 시스템을 개발이 이루어진다. 다음 단계인 품질 점검 과정은 기계화, 자동화되어 대량의 언어 자원을 신속하고 일관성 있게 점검하고 오류로 추정되는 시맨틱 관계를 검출한다. 마지막으로 오류 추정 데이터에 대한 전문가의 검증 및 교정, 수정 데이터 반입을 통한 시스템 현행화의 승인 과정을 거쳐 품질 고도화 과정이 종료된다. 본 연구는 이러한 일련의 과정에서 인간은 품질 제고 전략과 최종 데이터의 검증 및 승인의 역할을 수행하고 기계는 추론 규칙에 따라 직접 데이터의 오류 진단을 수행하는 인간-기계 협업(HMC; human-machine collaboration) 시스템 모델을 제안하고자 하였다.



〈그림 8〉 시맨틱 추론 기반 시스템을 활용한 인간-기계의 협업 시스템 모델

앞서 소개한 바와 같이 스프레딩 알고리즘을 기반으로, 본 연구를 위해 새롭게 제안한 시맨틱 추론 기법을 활용하여 대규모 언어 자원을 자동 진단한 결과, 주제명표목표 내의 삼각 연결 관계는 총 45,842건이 존재하는 것으로 나타났다. 검출한 모든 관계 네트워크에 대해 사전 정의된 추론 규칙에 따라 가능/불능/불확실 패턴을 분석한 상세 결과는 다음과 같다. 이하의 오류 사례는 시맨틱 웹의 데이터 기술 표준인 터틀 형식으로 표현하였다.

가. 계층 관계에서의 가능/불능 관계 분석 결과

NT-NT-NT 또는 BT-BT-BT 유형의 관계 오류는 3건이었으며, BT-NT-NT 또는 BT-BT-NT 형식의 가능 관계는 2,943건으로 주제명표목표의 계층 관계로만 구성된 데이터는 약 99.9% (2,943/2,946)의 정확률(precision)을 보였다. 시스템에 의한 오류 진단 결과에 대해 〈그림 9〉와 같이 전문가에 의한 검증 절차를 수행할 수 있다. 국립국어원 표준국어대사전(https://stdict.korean.go.kr/)에 따르면, 구제비젓을 생선의 내장으로 담근 젓으로 정의하고 있다. 따라서 첫째 행 "젓갈은 구제비젓을 하위어로 간주한다"는 정상 관계라 할 수 있다. 둘째 행 역시 창난젓은 구제비젓의 일종이므로 NT는 정상 관계이다. 셋째 행에서 창난젓은 젓갈의 일종이므로 NT가 아닌 "nl:창난젓@503602 nl:BT nl:젓갈@4280"로 관계를 수정해야 함을 알 수 있다. 'nl:'은 '〈https://www.nl.go.kr/thesaurus#〉'

를 지정하는 네임 스페이스를 의미하며, 각 용어 개체는 '용어@개체식별번호'의 구조로 표현하였다.

```
@prefix nl:<https://www.nl.go.kr/thesaurus#> .
nl:젓갈@4280          nl:NT      nl:구제비젓@503601 . ⇒ True
nl:구제비젓@503601    nl:NT      nl:창난젓@503602 . ⇒ True
nl:창난젓@503602      nl:NT      nl:젓갈@4280 . ⇒ False
```

〈그림 9〉 계층 관계로만 구성된 네트워크의 오류 사례

나. SYN 관계가 포함된 오류와 RT 관계 분석 결과

앞서 〈그림 4〉와 〈그림 7〉을 통해 RDF 그래프 형식으로 도식화한 추론 규칙에 따라 검출된 오류 사례로, 동의어 그룹인 SYN 관계가 포함되어 진단된 오류는 총 8건이었다. 〈그림 10〉과 같이 계층 관계와 SYN(USE/UF) 관계 2종으로 구성된 오류가 1건이었으며, 계층 관계 또는 동의어 관계에서의 오류 여부를 진단해 수정해야 한다. 또한 〈그림 11〉과 같이 USE와 UF 관계로만 구성된 경우가 총 7건이었으며, 불필요한 관계를 삭제해 정확한 대표어를 지정하는 추가 작업이 필요하다. 그 밖에, RT 관계가 포함된 경우는 총 42,920건/45,842건으로 모두 불확실한 관계로 판정하였다. 구축된 언어 자원의 시맨틱 네트워크 전반에 걸쳐 RT 관계가 매우 다수 존재함을 확인할 수 있었다.

```
nl:임신 테스트@500054      nl:USE     nl:임신 진단법@56871 .
nl:임신 진단법@56871       nl:NT      nl:임신 조기 진단법@8809 .
nl:임신 조기 진단법@8809   nl:UF      nl:임신 테스트@500054 .
```

〈그림 10〉 계층 관계와 SYN 관계가 혼재된 네트워크의 관계 오류 사례

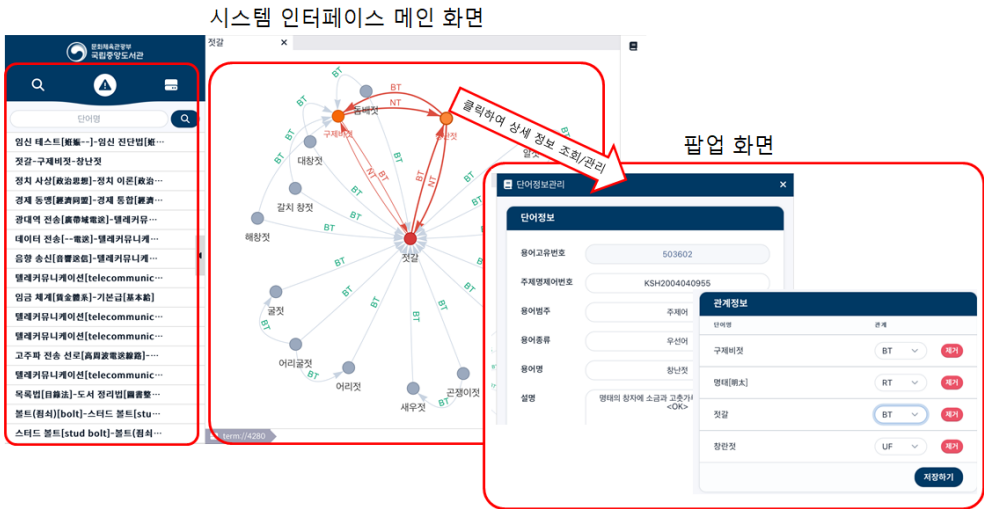
```
nl:네거리@116841              nl:USE     nl:사거리(네거리)@116842 .
nl:사거리(네거리)@116842      nl:UF      nl:십자로@155795 .
nl:십자로@155795              nl:USE     nl:네거리@116841 .
```

〈그림 11〉 USE와 UF 관계로만 구성된 네트워크의 관계 오류 사례

3. 웹 기반 언어 자원 관리 시스템

추론 시스템에 의해 자동 진단된 데이터를 관리하기 위해 전문가의 작업을 지원하는 시각화 기반의 인터페이스 시스템이 필요하다. 본 장에서는 ㈜시즌(https://season.co.kr/)과 공동 개발한 웹 기반 지능형 시소러스 구축 시스템을 간략히 소개한다. 본 연구를 통해 개발한 추론 엔진을

탑재하여 데이터 진단을 수행하고, 결과 데이터의 품질 관리를 위한 몇 가지 필수 기능을 포함하고 있다. 〈그림 12〉는 오류 예시를 시각화하고 이를 확인 및 수정 과정을 간략히 설명한 것으로, 좌측 창은 용어 데이터베이스 검색, 오류 진단 결과 조회, 데이터 반출 기능을 담고 있으며, 우측 창은 네트워크 시각화를 통해 실제 구축된 용어 간의 복잡한 위상 구조를 보여준다. 특정 용어를 클릭하면 팝업 화면이 뜨며 기본 정보 조회 및 관계 수정/삭제 기능을 제공한다. 본 프로토타입 시스템은 후속 연구가 진행됨에 따라 기능 고도화가 이루어질 예정이다.



〈그림 12〉 웹 기반 지능형 시소러스 구축 시스템 인터페이스

V. 결 론

인공 지능 시대가 본격화됨에 따라 시소러스, 워드넷, 위키피디아 등의 대규모 언어 자원은 그 중요성이 점점 증대되고 있으며, 전통적인 정보 검색, 자연어 처리, 텍스트 마이닝 등 다양한 응용 분야에서 핵심적인 역할을 수행할 뿐 아니라, 최근 생성성 AI 기술을 발전시키는 데 핵심적인 자원으로 활용되고 있다.

본 연구는 대규모 언어 자원의 연구 동향을 바탕으로, 자동화된 시맨틱 추론 기법을 이용한 지식 자원의 효율적 품질 제고 방안을 제시하고 응용 가능성을 제안하고자 하였다. 우선 용어 간의 다양한 관계를 추론할 수 있도록 수많은 정오 사례를 바탕으로 공통의 규칙 집합을 정의하였다. 정의된 시맨틱 추론 규칙을 바탕으로 거대한 언어 자원의 네트워크를 고속 탐색하고 오류를 검출하는 스프레딩 알고리즘 기반의 시맨틱 추론 엔진을 개발하였다. 국립중앙도서관의 주제명표

목표에 대한 실험을 통해 데이터의 품질 제고를 위한 일련의 과정을 효율적으로 수행할 수 있음을 확인하였다. 더불어 전문가의 품질 관리 업무를 지원하는 도구인 웹 기반의 시소러스 관리 시스템을 소개하였다.

본 연구는 대규모 언어 자원의 품질 고도화를 위한 시맨틱 추론 기법을 새롭게 제안하였으며, 이를 위해 용어 관계에서 발생하는 다양한 논리적 오류 사례를 분석하고 체계화하는 최초의 시도였다는 점에서 의의가 있다. 또한, 전문가가 직접 수행하기 어려운 복잡한 시맨틱 관계의 오류 진단을 자동화된 추론 기법을 이용하여 효과적으로 수행하는 방안을 제안하였다. 마지막으로, 이 과정에서 인공 지능 시대의 인간과 기계의 협업 방식을 논의하였다는 데 의의가 있다.

본 연구는 한국어 정보를 중심으로 이루어졌으며, 관계를 판단할 수 있는 최소 단위인 삼각 연결 관계에서의 추론을 연구 범위로 한정하였으므로, 향후 연구를 통해 다국어를 기반으로 용어 간의 관계를 더욱 확장함과 동시에 높은 단계의 추론이 가능하도록 발전시킬 계획이다. 본 연구를 바탕으로 국가 지식 자원의 고품질화를 통해 공공 기관의 독자적인 인공 지능 모델(on-device LLM)을 구축하는 데 기여할 수 있을 것으로 기대한다. 또한, 용어 관계의 오류 검증 과정에 외부의 지식 베이스 및 생성형 AI 시스템을 활용하여 전문가를 적극적으로 지원하는 발전된 인간-기계의 협업 방안 등에 대해서도 지속적인 연구를 수행할 계획이다.

참 고 문 헌

국립중앙도서관 (2021). 국립중앙도서관 주제명표목 업무지침. 국립중앙도서관 국가서지과. 출처: https://librarian.nl.go.kr/LI/contents/L20201000000.do

김선동, 강민서, 이재길 (2014). 한국어 디비피디아의 자동 스키마 진화를 위한 방법. 한국정보처리학회 학술대회논문집, 21(1), 741-744.

백지원, 정연경 (2014). 국립중앙도서관 주제명표목표 검색 시스템 개선 방안에 관한 연구. 정보관리학회지, 31(1), 31-51. https://doi.org/10.3743/KOSIM.2014.31.1.031

여지숙, 양기덕, 이토히로코, 이혜경 (2022). 한인디아스포라 관련 주제명표목 개선 방안 연구 - 국립중앙도서관 주제명표목표의 한인 관련 용어를 중심으로-. 한국도서관·정보학회지, 53(1), 103-124. https://doi.org/10.16981/kliss.53.1.202203.103

이혜경, 이용구 (2023). 주제명 활용 분석을 통한 국립중앙도서관 주제명표목표의 현황 연구. 정보관리학회지, 40(2), 157-182. https://doi.org/10.3743/KOSIM.2023.40.2.157

정도헌 (2018). 데이터 활용률 제고를 위한 기술 용어의 상호 네트워크 생성과 통제. 정보관리학회지, 35(1), 157-182. https://doi.org/10.3743/KOSIM.2018.35.1.157

정도헌, 최희윤 (2006). 과학기술 전문용어의 다국어 의미망 생성과 분석. 정보관리연구, 37(4), 25-47. https://doi.org/10.1633/JIM.2006.37.4.025

최윤경, 정연경 (2014). 국립중앙도서관 주제명표목표의 고품질화 방안에 관한 연구. 한국문헌정보학회지, 48(1), 75-95. https://doi.org/10.4275/KSLIS.2014.48.1.075

한희정, 김태영, 두효철, 오효정 (2017). 기술문서 정의문 패턴을 이용한 전문용어사전 자동추출 및 활용방안. 정보관리학회지, 34(4), 81-99. https://doi.org/10.3743/KOSIM.2017.34.4.081

허고은 (2019). Word2Vec과 WordNet 기반 불확실성 단어 간의 네트워크 분석에 관한 연구. 한국문헌정보학회지, 53(3), 247-271. https://doi.org/10.4275/KSLIS.2019.53.3.247

An, Y., Wang, Q., Liu, J., Liu, K., Lyu, Y., Wu, H., She, Q., & Li, S. (2019). Enhancing pre-trained language representations with rich knowledge for machine reading comprehension. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2346-2357. https://doi.org/10.18653/v1/P19-1226

Hu, X.-B., Wang, M., Hu, D., Leeson, M. S., Hines, E. L., & Di Paolo, E. (2012). A ripple-spreading algorithm for the k shortest paths problem. In Proceedings of the 2012 Third Global Congress on Intelligent Systems, 202-208. https://doi.org/10.1109/GCIS.2012.96

Jeong, D. H., Hwang, M., & Sung, W. K. (2011). Generating knowledge map for acronym-expansion recognition. In the Proceedings on U- and E-Service Science and Technology (UNESST 2011), 287-293. https://doi.org/10.1007/978-3-642-27210-3_38

Marciniak, J. (2020). Wordnet as a backbone of domain and application conceptualizations in systems with multimodal data. Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), 1-6. Available: https://aclanthology.org/2020.mmw-1.5

Mendes, P. N., Jakob, M., Garcia-Silva, A., & Bizer, C. (2011). DBpedia spotlight: shedding light on the web of documents. Proceedings of the 7th International Conference on Semantic Systems, 1-8. https://doi.org/10.1145/2063518.2063519

Miller, G. A. (1995). WordNet: a lexical database for english. Communications of the ACM, 38(11), 39-41. https://doi.org/10.1145/219717.219748

Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. Artificial Intelligence, 193, 217-250. https://doi.org/10.1016/j.artint.2012.07.001

Nothman, J., Ringland, N., Radford, W., Murphy, T., & Curran, J. R. (2013). Learning multilingual named entity recognition from Wikipedia. Artificial Intelligence, 194, 151-175. https://doi.org/10.1016/j.artint.2012.03.006

Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web, 8(3), 489-508. https://doi.org/10.3233/SW-160218

Paulheim, H., & Gangemi, A. (2015). Serving DBpedia with DOLCE - more than Just adding a cherry on top. The Semantic Web - ISWC 2015 (Lecture Notes in Computer Science, vol. 9366), 180-196. https://doi.org/10.1007/978-3-319-25007-6_11

Saedi, C., Branco, A., Rodrigues, J. A., & Silva, J. (2018). WordNet embeddings. In Proceedings of the Third Workshop on Representation Learning for NLP, 122-131. https://doi.org/10.18653/v1/W18-3016

• 국한문 참고문헌의 영문 표기
(English translation / Romanization of references originally written in Korean)

Baek, Ji-Won & Chung, Yeon-Kyoung (2014). A study on improving access & retrieval system of the National Library of Korea subject headings. Journal of the Korean Society for Information Management, 31(1), 31-51. https://doi.org/10.3743/KOSIM.2014.31.1.031

Choi, Woon Kyung & Chung, Yeon-Kyoung (2014). A study on improvements for high quality in National Library of Korea subject headings List. Journal of the Korean Society for Library and Information Science, 48(1), 75-95. https://doi.org/10.4275/KSLIS.2014.48.1.075

Han, Hui-Jeong, Kim, Tae-Young, Doo, Hyo-Chul, & Oh, Hyo-Jung (2017). Automatic extraction and utilization of technical term dictionaries using definition patterns in technical documents. Journal of the Korean Society for Information Management, 34(4), 81-99. https://doi.org/10.3743/KOSIM.2017.34.4.081

Heo, Go Eun (2019). Network analysis between uncertainty words based on Word2Vec and WordNet. Journal of the Korean Society for Library and Information Science, 53(3), 247-271. https://doi.org/10.4275/KSLIS.2019.53.3.247

Jeong, Do-Heon (2018). Generating and controlling an interlinking network of technical terms to enhance data utilization. Journal of the Korean Society for Information

Management, 35(1), 157-182. https://doi.org/10.3743/KOSIM.2018.35.1.157

Jeong, Do-Heon & Choi, Hee-Yoon (2006). Building and analysis of semantic network on S&T multilingual terminology. Journal of Information Management. 37(4), 25-47. https://doi.org/10.1633/JIM.2006.37.4.025

Kim, Sundong, Kang, Minseo, & Lee, Jae-Gil (2014). A method of automatic schema evolution on DBpedia Korea. Proceedings of the Korea Information Processing Society Conference, 21(1), 741-744.

Lee, HyeKyung & Lee, Yong-Gu (2023). A study on the current status of National Library of Korea subject headings list through utilization analysis of subject headings. Journal of the Korean Society for Information Management, 40(2), 157-182. https://doi.org/10.3743/KOSIM.2023.40.2.157

National Library of Korea (2021). National Library of Korea subject headings guidelines. National Bibliography Department, National Library of Korea. Available: https://librarian.nl.go.kr/LI/contents/L20201000000.do

Yeo, Ji-Suk, Yang, Kiduk, Ito, Hiroko, & Lee, HyeKyung (2022). A study on the enhancement of korean diaspora-related subject headings: focusing on korean-related terminology in the National Library of Korea subject headings. Journal of Korean Library and Information Science Society, 53(1), 103-124. https://doi.org/10.16981/kliss.53.1.202203.103