

에너지 분야 한국어 특허의 기능적 분석을 위한 Ko-SAO(Subject-Action-Object)구조 제안

김선혜^{1,2} · 장혜진^{1*}

¹한국외국어대학교 AI데이터융합학부 / ²캠브리지대학교 제조연구소

Ko-SAO Structure for Functional Analysis of Korean Patents in Energy Field

Sunhye Kim^{1,2} · Hyejin Jang¹

¹Division of AI Data Convergence, Hankuk University of Foreign Studies

²Institute for Manufacturing, University of Cambridge

This study proposes an analysis framework based on the SAO (Subject, Action, Object) structure to efficiently extract technical information from Korean patent documents. Traditional patent analysis has required substantial time and effort due to the vast amount of data and complex sentence structures, with existing SAO extraction tools primarily focused on English. However, considering the increasing number of Korean patents filed with the Korean Intellectual Property Office and the leading role of Korean companies in technological advancements, there is a need for an SAO structure tailored specifically to the Korean language. To address the challenges posed by the syntactic flexibility and contextual dependencies of Korean sentences, this study employs a hybrid approach that combines machine learning models with rule-based and statistical methods. Additionally, the effectiveness of the proposed methodology is validated through comparisons with bilingual models that leverage the strengths of English SAO extraction techniques to enhance Korean patent analysis. The findings of this study are expected to contribute to the development of Korean patent analysis tools, supporting domestic researchers and companies in advancing technological innovation and strengthening global competitiveness.

Keywords: SAO Structure, Patent Analysis, Natural Language Processing(NLP), Lexical Information, Text Mining

1. 서론

특허 분석은 기술 동향 파악, 경쟁자 모니터링, 연구 개발 방향 결정에 중요한 역할을 한다. 그러나 전통적인 방법은 방대한 데이터와 복잡한 문장 구조로 인해 시간과 노력이 많이 소요된다(Kim and Yoon, 2021). 이를 극복하기 위해 SAO(subject, action, object) 구조가 도입되었으며, 이는 특허 문장을 주어(S), 동작(A), 객체(O)로 분해해 주요 정보를 자동 추출하는 데 유용하다(Moehrle *et al.*, 2005; Choi *et al.*, 2011; Kim *et al.*,

2018). SAO 구조는 복잡한 문장을 관리하기 쉬운 요소로 분해해 기술 혁신과 관계를 명확히 이해하는 데 도움을 준다(Park *et al.*, 2012; Choi *et al.*, 2012). 그러나 대부분의 SAO 추출 연구는 영어에 집중되어 있으며, 이는 영어 문법의 단순성, 풍부한 데이터, 그리고 많은 NLP 모델이 영어로 개발되었기 때문이다. 한국 특허청(Korean Intellectual Property Office; KIPO)에 출원되는 한글 특허 수의 증가와 한국 기업의 기술 선도를 고려할 때, 한글 특허 분석에 SAO 구조 도입이 필요하다. 기계번역을 통한 우회적 활용 시도가 있었지만 이는 번역 품질 문제

이 연구는 (2025학년도) 한국외국어대학교 교원연구지원사업 지원에 의하여 이루어진 것임.

* 연락저자 : 장혜진 교수, 17035 경기도 용인시 처인구 모현읍 외대로 81 한국외국어대학교, Tel : 031-330-4494, E-mail : hj@hufs.ac.kr
2024년 10월 17일 접수; 2024년 12월 26일 수정본 접수; 2025년 1월 24일 게재 확정.

로 인해 오류가 누적될 가능성이 크며 자연스러운 한국어 문법 구조를 반영하기 어렵다(Yoon and Kim, 2011; Yoon *et al.*, 2013; Jang *et al.*, 2023). 이는 한국어뿐 아니라 일본, 중국, 대만 등 비영어권 국가의 특허분석에서도 동일하게 나타나는 어려움이다(Liu *et al.*, 2022). 특히 한글 문장은 구문 구조가 유연하고 문맥에 의존적이어서 SAO 요소를 명확히 식별하기 어렵다. 이를 해결하려면 한글의 특성에 맞춘 NLP 도구 개발이 필요하다. 본 연구는 이러한 한계를 극복하기 위해 한글에 적합한 SAO 구조 추출 프레임워크를 제안하고 실험하는 것을 목표로 한다. 자연어 처리의 한국어 품사 태깅 및 문법 구조를 활용한 룰 기반의 한국어 SAO 추출 알고리즘을 개발한다. 한글 특허 분석 도구 개발 및 적용 가능성을 탐구하여 국내 연구자와 기업의 한국어 분석을 통해 기술 R&D 기획, 관리 및 평가 등에 적용되는 것을 기대한다.

2. 배경이론: SAO(subject, action, object)와 자연어 처리

SAO 구조는 소련의 발명가 Genrich Altshuller가 제안한 TRIZ 방법론 중 하나인 기능 분석에서 제안되었다. 기능 분석의 목적은 시스템 기능의 문제를 이상적으로 파악하고, 전체 시스템을 SAO 구조로 매핑하는 것이다. 기능분석을 통해 문장에서 추출된 SAO 구조는 수단과 목적의 관계를 통해 시스템을 표현할 수 있으며, <Figure 1>과 같이 주어-동작-객체 (Subject-Action-Object) 구조이다. 여기서 AO(Action-Object)가 기술 문제를 설명하고 S(Subject)가 해결책을 나타낸다면, 문제와 해결책 사이의 관계로 해석할 수 있다(Moehrle *et al.*, 2005). 따라서 SAO구조는 기술의 구성요소와, 구성 요소 간의 관계, 기능을 잘 표현할 수 있다(Park *et al.*, 2012; Choi *et al.*, 2012). 이러한 특성으로 인해 SAO 구조는 개발된 이후에 특허를 분석하는 연구에서 널리 이용되었다. 기술 기획 탐색, 특허 침해 분석, 특허 유사도 분석 등의 분야에서 SAO구조는 활용되었으며, 유용성이 입증되었다(Choi *et al.*, 2012; Wang *et al.*, 2017; Kim and Yoon, 2021; Park *et al.*, 2012; Wang *et al.*, 2019).

Sentence: Electric vehicle batteries lose efficiency due to temperature fluctuations.
 subject action object

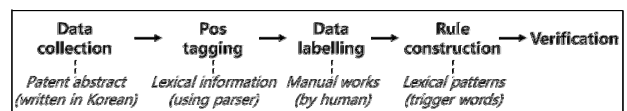
Figure 1. Example of SAO Structure

대부분의 연구는 미국 특허청에 출원된 특허를 대상으로 SAO 구조를 적용했다. 미국 외 특허청 특허를 사용하는 경우에도, 원문에서 SAO 구조를 직접 추출하지 않고 영어로 번역 후 추출하거나, 추출된 영문 SAO 구조를 다시 원문으로 번역해 사용했다(Yoon and Kim, 2011; Yoon *et al.*, 2013; Liu *et al.*, 2022; Jang *et al.*, 2023). 이는 기존 SAO 구조 추출 방법론이 영어 기반으로 구축되었기 때문이다. SAO 구조 추출 방법은 크

게 전문가 기반, 규칙 기반, 학습 기반으로 나뉜다. 전문가 기반 방식은 자원 소모적이고 human error의 가능성이 있으며, 학습 기반 방식은 연구가 적어 정확도가 낮다. 가장 많이 사용되는 규칙 기반 방식은 두 단계로 구성된다. 첫째, Parser를 이용해 영문 pos tag를 추출하며, 종종 dependency parser로 pos tag와 dependency relation을 함께 추출한다. 대부분 Stanford Parser를 사용한다(Danqi Chen and Christopher D Manning, 2014; Anna Rafferty and Christopher D. Manning, 2018). 둘째, 규칙에 따라 S, A, O 구문을 정의하는데, 이때 action word와 관련된 명사 또는 명사구를 주어나 객체로 정의한다. 한국어에서 SAO 구조를 추출하는 방법은 크게 세 가지로 나뉜다. 첫째, 한글 데이터로 훈련된 기계 학습 모델을 개발해 SAO 추출 정확도를 향상시키는 방법. 둘째, 영어 SAO 도구의 강점을 활용한 이중 언어 모델 도구를 사용하는 방법. 셋째, 한글 구문의 복잡성을 다루기 위해 규칙 기반과 통계적 방법을 모두 고려하였다. 마지막으로, 검증 챗터에서 이중 언어 모델 도구와 비교할 예정이다. 하지만 한글의 경우, pos tagging 과정보다 문제가 있다. 영어처럼 대중적으로 사용되는 parser가 없으며, 연구도 오래되지 않았다. 본 연구에서는 KoNLPy parser를 이용하며, 한글 기반 SAO 구조 추출 방법론은 이어지는 3장에서 논의한다.

3. 연구 방법론

제안하는 방법론은 5단계로 구성되어 있다. 먼저 한국어로 작성된 특허 데이터를 수집한다. 수집한 특허의 요약 데이터를 parser를 이용하여 각 구문의 문법적 품사를 태깅한다. 다음으로 수작업으로 특허를 구성하는 문장에 대하여 S, A, O 구조를 라벨링한다. 그 후 품사 태깅 결과와 라벨링 된 결과를 이용하여 구문적 패턴을 도출하여 한국어 SAO구조 추출 규칙을 정의한다. 마지막으로 정의된 규칙에 대한 검증을 수행한다. 각 단계에 대한 상세한 프로세스는 다음 <Figure 2>와 같다.



(Data collection & Pos tagging) 에너지 분야는 국가 기술 경쟁력 및 지속 가능성에서 핵심적인 역할을 차지하는 산업으로, 특허를 통해 기술 동향을 분석하는 연구가 활발히 진행되고 있습니다. 본 연구는 에너지 분야를 대상으로 SAO 구조를 개발하여, 향후 다양한 산업 분야의 기술 분석에도 활용할 수 있는 기반을 마련하고자 한다. 정부 출연 연구기관에서 출원된 특허는 기술적 신뢰성과 데이터의 일관성이 높아 연구 주

제와의 적합성이 크다고 판단하였다. 또한, 정부출연연구원은 국제적인 연구개발뿐 아니라 국내 연구보고서 등 한국어 기반의 보고서를 다수 출간하고 배포하는 역할을 수행한다. 연구분야의 중요성 및 한국어 분석의 필요성을 감안하여 본 연구의 분석 대상으로 적합하다고 보았다. 최종적으로 한국에너지연구원이 한국특허청에 출원한 전체 특허 (NPA=("%한국에너지기술연구원 %")) 4,942건의 특허를 수집하였다 (https://1drv.ms/x/c/7dc2c419eab2f599/ERMSPsa7mBJLoB84--qf6ycBfGeH2b_-5kfcvvpGhUofVQ?e=5HJQ6X). 수집한 특허에서 요약문을 추출하여 분석에 이용하였다. 수집한 특허를 KoNLPy 파서를 이용하여, 문장을 형태소로 분리하고 각 형태소에 품사를 태깅한다. (Data labelling) 다음으로 수작업으로 품사가 태깅된 문장에서 SAO 구조를 추출한다. 자문 대상은 영어 SAO구조를 활용하는 특허 NLP 분석 분야의 박사 학위를 소지하고 있는 전문가 3인이다. 전문가는 한글 특허 문서의 요약을 제공받아, 해당 요약문에 등장하는 기술의 기능을 추출해 내는 작업을 수행하였다. (Rule construction) 다음으로 수작업으로 라벨링한 SAO구조와 품사가 태깅된 문장을 이용하여 구분론적인 규칙을 정의한다. 이 때, 주의해야 하는 점은 한국어는 명사, 동사/형용사, 부사 등으로 구성된 영어와는 다르다는 점이다. 영어의 동사 역할을 한국어는 명사+어미, 어근+어미로 이용한다. 따라서 영어의 SAO 추출 규칙을 한국어에 바로 적용하기는 힘들다. 한국어는 주부와 서술부가 존재하기

때문에 이전 연구에서는 '~을/를/이/가' 과 같은 (격)조사를 기준으로 문장을 분리하고자 하였다. 하지만 이러한 방식은 기능을 효과적으로 추출해내지 못한다. 기능이란 기술의 작동 역할, 작용 등을 의미하기 때문이다. 이에 본 연구에서는 (격) 조사 기준이 아닌, NNG와 XSV, ETM 등의 품사를 기준으로 규칙을 정의하였다. 또한 제안하는 규칙에서는 포함구조는 배제하고 기능에 초점을 맞춘 SAO 구조를 추출하는 것을 목표로 하였다. 정의한 추출 단계는 <Table 1>과 같이 크게 3개의 구조로 이루어져 있다. 1 단계: 문장 구분 단계. 2 단계: 품사 전처리 단계. 3 단계: SAO 구조 추출 단계. 각 단계의 규칙은 아래 표 X에 있는 Pseudo code와 같다. 먼저 문장 구분 단계에서는 구분자를 기준으로 긴 문장을 여러 개의 세부 문장으로 구분한다. 이는 Parser의 성능을 높임과 동시에 여러 개의 수식, 병렬 구조로 중첩되어 있는 문장에서 보다 정확하고, 많은 SAO구조를 추출하기 위함이다. 품사 전처리 단계에서는 하나의 단어가 아닌 단어 구 수준의 S와 O를 얻기 위하여 수사 및 조사에 대한 전처리를 수행한다. 주격 조사와 목적격 조사를 이용하여 S와 O에 대한 정의도 수행한다. 마지막 SAO구조 추출단계에서는 정의한 규칙을 이용하여 SAO구조를 추출해 낸다. 정의한 규칙은 <Table 2>와 같다. SAO 추출 규칙은 먼저 등장한 규칙이 더 우선순위를 갖도록 정의하였다.

(Verification) 이를 검증하기 위해서 정량적인 검증과 질적 검증을 수행한다. 정량적인 검증을 위해서는 제안하는 방법론

Table 1. SAO Structure Extraction Rule

Step 1: Sentence separation	
1	For all (word, POS) in doc:
2	If (word, POS) in list_split_del ¹ :
3	Modify (word, POS) into ('EOS', 'DEL') ²
Step 2: POS preprocessing	
1	For all (word, POS) in doc:
2	If (word, POS) in list_NNG_candidate ³ :
3	Modify (word, POS) into (word, 'NNG')
4	Modify seq(word, 'NNG') into (phrase, 'NNGset')
5	If NNGx ⁴ + (('word/은/는', 'JX') or ('word/이/가', 'JKS')):
6	Append NNGx into list_subject_candidate
7	If NNGx + ('word', 'JKO'):
8	Append NNGx into list_object_candidate
Step 3: SAO extraction	
1	For all seq(('EOS', 'DEL') + front + (word, POS) + back + ('EOS', 'DEL')) in doc:
2	For pattern in SAO_extraction_patterns:
3	If (word, POS) == pattern['A']:
4	O = Nearest NNGx in pattern['O'] and in list_object_candidate
5	If pattern['condition'] == NA:
6	S = Nearest NNGx in pattern['S'] and in list_subject_candidate
7	ElseIf pattern['condition'] == 'single NNG can be S':
8	S = Nearest NNGx in pattern['S']

1 list_split_del = [('.', 'SF'), (',', 'SC'), (';', 'SY')]

2 ('EOS', 'DEL') indicates a sentence delimiter

3 list_NNG_candidate = [(word, 'NNBC'), (word, 'NR'), ('화', 'XSN'), ('의', 'JKG'), ('및', 'MAJ')]

4 NNGx = ((word, 'NNG') or (phrase, 'NNGset'))

Table 2. Action Rules

Priority	A	S	O	Condition
1	('하', 'XSV'), ('는', 'ETM')	back	forth	single NNG can be S
2	('하', 'XSV'), ('고', 'EC')	forth	forth	
3	('하', 'XSV'), ('여', 'EC')	forth	forth	
4	('할', 'XSV+ETM'), ('수', 'NNB'), ('있', 'VV'), ('는', 'ETM')	back	forth	single NNG can be S, S is last NNG
5	('시키', 'XSV'), ('는', 'ETM')	back	forth	single NNG can be S, S is last NNG
6	('되', 'XSV'), ('는', 'ETM')	-	back	
7	('된', 'XSV+ETM')	-	back	
8	('한', 'XSA+ETM')	back	forth	
9	('한', 'XSV+ETM')	back	forth	single NNG can be S
10	('함', 'XSV+ETN')	forth	forth	
11	('할', 'VV+ETM')	forth	forth	
12	('할', 'XSV+ETM')	forth	forth	
13	('시킬', 'XSV+ETM')	forth	forth	
14	('시켜', 'XSV+EC')	forth	forth	
15	('하', 'XSV')	forth	forth	

을 통해 추출된 SAO 구조가 전체 특허 문서의 텍스트를 얼마나 설명할 수 있는지 수치적으로 검증할 수 있다. 질적 검증을 위해서는 기존의 다른 SAO 구조 방법론인 이중 언어 모델 도구 방법론을 통해 도출된 SAO 구조와 본 연구에서 제안한 방식을 통해 도출된 SAO 구조를 비교한다.

4. 실험 및 검증

4.1. 분석 결과

본 연구에서 제안한 Action rule 기반의 KO-SAO 방식을 통해 추출된 SAO 예시는 <Figure 3>과 같다. 먼저 단계 1에서 길이가 긴 형태의 특허 문장을 분할하고, 단계 2에서 한국어 품사를 태깅하였다. 이를 기반으로 단계 3에서는 <Table 2>의 규

칙을 식별하여 최종적인 SAO 구조를 추출하였다. 길이가 긴 특허 문장에서 복수 개 이상의 유의미한 SAO 구조를 추출하였다.

앞서 정의한 규칙을 이용하여 43,996건의 SAO 구조를 추출하였으며, 결과 형태 및 예시는 <Table 3>과 같다. SAO 구조는 기술적 기능에 해당하는 A를 추출한 후, A 어휘와 문법적으로 연결된 S 및 O 어휘를 추출하는 방식으로 진행하였다. 가장 많이 추출된 A 어휘는 3,433건이 도출된 ‘포함’으로 전체 7.80%를 차지하였다. ‘포함’은 기술 간의 구성요소를 설명하며, 예를 들어 ‘건조물 건조장치’의 구성요소로 ‘열풍 공급부’를 포함하는 기술 관계를 설명한다. 이외에도 ‘이용’, ‘형성’, ‘공급’ 등의 주요 A 어휘가 도출되었다. ‘None’의 경우, A 어휘가 별도의 구체적인 의미 부분 없이 ‘~하다’로 추출된 경우이다. 예를 들어, O 어휘가 ‘장기간 촉매역할’로 추출되어 역할을 하다는 의미로 해석할 수 있다. 한국어 및 특허 작성의 특성상,

Step 1. Sentence separation

means sentence segement

본 발명에 따른 효소를 이용한 이산화탄소의 광물화 반응 장치 및 방법은 탄산무수화 효소를 사용함으로써, 이산화탄소를 흡수제에 흡수 시 흡수 속도 및 흡수량이 현저히 높고, 탄산무수화 효소가 흡수제에 안정적으로 존재하여 장기간 촉매 역할을 할 수 있으며, 물을 계 내에서 지속적으로 순환시킬 수 있음에 따라 전체 공정에서 발생할 수 있는 폐수의 발생을 최소화할 수 있는 효과가 있다.

Step2. POS preprocessing

(('물', 'NNG'), ('효', 'JKO'), ('계', 'NNG'), ('내', 'NNB'), ('에서', 'JKB'), ('지속', 'NNG'), ('적', 'XSN'), ('으로', 'JKB'), ('순환', 'NNG'), ('시', 'XSV+ETM'), ('수', 'NNB'), ('있', 'VV'), ('을', 'ETN'), ('에', 'JKB'), ('따라', 'VV+EC'), ('전체', 'NNG'), ('공정', 'NNG'), ('에서', 'JKB'), ('발생', 'NNG'), ('할', 'XSV+ETM'), ('수', 'NNB'), ('있', 'VV'), ('는', 'ETM'), ('폐수', 'NNG'), ('의', 'NNG'), ('발생', 'NNG'), ('을', 'JKO'), ('최소', 'NNG'), ('화', 'XSV+ETM'), ('할', 'XSV+ETM'), ('수', 'NNB'), ('있', 'VV'), ('는', 'ETM'), ('효과', 'NNG'), ('가', 'JKS'), ('있', 'VA'), ('다', 'EF'))

Step3. SAO extraction

Subject	Action	Object	Extraction rule
-	순환	물	[('시키', 'XSV+ETM')]
폐수의 발생	발생	-	[('할', 'XSV+ETM'), ('수', 'NNB'), ('있', 'VV'), ('는', 'ETM')]
효과	최소화	폐수의 발생	[('할', 'XSV+ETM'), ('수', 'NNB'), ('있', 'VV'), ('는', 'ETM')]

Figure 3. Result of SAO Extraction

Table 3. Example of SAO Extraction Using Proposing Method

Rank	Action word	Frequency	Ratio	Example	
				Subject	Object
1	포함	3,433	7.80%	건조물 건조장치	열풍 공급부
2	이용	1,797	4.08%	유동층 건조장치	시스템의 폐열
3	형성	1,736	3.95%	유도 용융부	실리콘 용탕
4	공급	1,437	3.27%	양극용액	전자
5	None(-하다)	1,207	2.74%		장기간축매역할
6	제조	1,106	2.51%	아르곤	수소혼합가스
7	사용	970	2.20%	열병합발전시스템	개질기
8	제공	870	1.98%	전이금속산화물	산점
9	설치	710	1.61%		급수전열관
10	발생	686	1.56%	폴리머재료	마찰전기

A의 주요 어휘가 모두 명사형 어근이 도출된 점이 특징적으로 나타났으며, A 어휘 목록을 통해 기술의 핵심 기능을 파악할 수 있다.

4.2 정성, 정량적 검증

제안한 문법 구조 기반의 KO-SAO 추출 방법론의 타당성을 정량적으로 검증하기 위해 전체 문서에서 등장한 S,A,O와 연관 있는 전체 어휘에 대하여 제안한 규칙을 통해 실제로 추출된 커버리지를 확인한다. 이를 위해 주어, 서술어, 목적어의 핵심 품사 유형을 JX(은, 는), JKS(이, 가), XSV(동사 파생 접미사), JKO(을, 를)로 선정하여 커버리지를 <Table 4>와 같이

도출하였다. 제안하는 방법론을 통해서 도출된 SAO구조는 S, A, O 각각 10.02%, 65.36%, 62.62%로 도출 되었다. 유형 S의 경우, 제안한 문법적 구조(<Table 1>)에서 누락된 SAO 구조를 포함하고 있었으나, 주격 조사로 추출하지 않고 수식어 구 형태로 S를 도출하는 방안을 제안하여 이를 해결하고자 하였다. 도출된 결과는 전체 추출 건수 (17,417) 대비 주격 조사로 추출된 경우 (2,874건)가 16.50% 밖에 차지하지 않았다. 한국어 특성 상 직접적 주어는 생략한 채로 서술되는 경우가 많기 때문에 본 연구는 이러한 특성을 적절하게 반영한 방안을 제안하였음을 알 수 있다.

기존 접근 방법과의 비교 또한 이중 언어 모델 도구와 제안한 방법론을 이용하여 도출된 SAO 구조를 비교한다. 이중 언

Table 4. Coverage of Proposed method

Type	Frequency	POS	Coverage
S	17,417	JX (은,는) 또는 JKS (이,가)	10.02% (2,874/28,694)
A	43,996	XSV (하, 한, 할, 되, 된 등)	65.36% (40,681/62,243)
O	28,521	JKO (을,를)	62.62% (28,521/45,545)

Table 5. Example of SAO Extraction using a Bilingual Model

Rank	Action word	Frequency	Ratio	Example	
				Subject	Object
1	characterize	1598	5.02%	heat treatment	-
2	provide	1502	4.72%	cooler	refrigerator
3	use	1368	4.30%	transformation strain	carbon comprehrons
4	form	1362	4.28%	thin film	substrate
5	instal	977	3.07%	-	heating tube
6	supply	965	3.03%	thermal paper	thermal energy
7	include	864	2.71%	cylinder	transparent pipe
8	have	815	2.56%	it	effect
9	connect	662	2.08%	conductor	flounder
10	produce	641	2.01%	pyramid resolution	pyrolysis oil
Coverage			33.78%		

어 모델 도구는 기존 영문 SAO 구조 추출 알고리즘을 이용하는 방안이다. 한국어로 작성된 특허를 영문으로 기계 번역 한 후에, 해당 영문 번역문을 SAO 추출 알고리즘에 입력하여 영문 SAO 구조를 추출해 낸다. 다음으로 영문 SAO 구조를 기계 번역하여 한국어 SAO 구조를 도출하게 된다. 이런 단계를 통해 도출된 SAO 구조는 아래 표와 같다. 분석 결과는 총 31,839 개의 SAO구조가 추출되었다. 그 중 상위 10개의 action word는 <Table 5>와 같으며, 커버리지는 약 34%이다. 이 중 have와 같은 action word는 분사를 구성하는 경우와 포함의 뜻 등을 가지고 있는데, 대부분의 경우 분사를 구성하는데 이용되었다. 전반적으로 KO-SAO를 통해 도출된 action word와 많이 차이가 났으며, 한글 원문의 의미를 잘 고려하지 못했다. 또한 it, this, that과 같은 대명사를 잘 고려하지 못했다.

5. 결 론

본 연구는 한국어 특허 문서에서 기술 정보를 효율적으로 추출하기 위해 SAO 구조에 기반한 한국어 SAO구조 추출 프레임워크를 제안했다. 제안된 KO-SAO 방법론이 한국어의 문법적 특성을 잘 반영하여 한국어 특허 문서에서 효과적인 기술 정보 추출을 가능하게 했음을 확인했다. 특허, '포함', '이용', '형성' 등 기술적 기능을 나타내는 주요 어휘를 중심으로 SAO 구조를 추출함으로써, 기술의 핵심 기능을 파악할 수 있었다. 또한, 제안된 방법론은 기존의 이중 언어 모델 기반 SAO 추출 방법과 비교했을 때 한국어의 문법적 특성을 보다 잘 반영하여 더 정확한 SAO 구조를 추출할 수 있다. 특히, 영어가 아닌 한국어 기반의 텍스트 마이닝 접근을 제시하여 기업 및 연구소 내부에 보유한 한국어로 작성된 기술 관련 문건을 분석할 수 있는 체계적인 방법을 제안하였다. 제안한 프레임워크를 기반으로 분석자가 세부적인 Ko-SAO 규칙을 보완할 수 있어 실무자가 직관적으로 적용할 수 있는 실용적인 방법을 제안하였다.

그러나 이 연구에는 몇 가지 한계점이 존재한다. 먼저, 본 연구는 제안한 Ko-SAO 방법론을 에너지 분야에만 한정하여 적용하였다는 한계점이 존재한다. 또한, 한국어 특허 문서의 다양성과 복잡성을 충분히 반영하지 못한 부분이 있다. 일부 특수한 기술 분야에서 사용되는 전문 용어나 비정형화된 문장 구조에 대해 제안된 방법론의 성능이 저하될 수 있다. 또한, SAO 구조 추출 과정에서 문맥을 고려하지 않은 단순한 규칙 기반 접근법이 적용되어, 다의어와 같은 의미적 차이를 완전히 반영하지 못하는 경우가 발생할 수 있다.

따라서 향후 연구에서는 다음과 같은 개선 방안을 고려할 필요가 있다. 첫째, 다양한 기술 분야와 문서 유형에 대한 테스트를 통해 Ko-SAO 방법론의 일반화를 시도해야 한다. 둘째, 최신 자연어 처리 기법을 도입하여 문맥과 의미를 보다 정교하게 고려한 SAO 구조 추출 방법을 개발하는 것이 중요하다.

셋째, 규칙 기반 접근법에서 축적된 데이터를 활용하여 딥러닝 기반 SAO 추출 모델로 확장함으로써 자동화된 분석을 가능하게 하고, 성능을 더욱 향상시킬 수 있다. 이를 통해 제안된 Ko-SAO 프레임워크가 R&D 기획, 기술 예측 등 다양한 실무와 연구 환경에서 활용될 수 있을 것으로 기대된다.

참고문헌

- Anna, R. and Christopher, D. M. (2008), Parsing Three German Treebanks: Lexicalized and Unlexicalized Baselines, In ACL Workshop on Parsing German.
- Choi, S., Park, H., Kang, D., Lee, J. Y., and Kim, K. (2012), An SAO-based text mining approach to building a technology tree for technology planning, *Expert Systems with Applications*, **39**(13), 11443-11455.
- Choi, S., Yoon, J., Kim, K., Lee, J. Y., and Kim, C.-H. (2011), SAO network analysis of patents for technology trends identification: A case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells, *Scientometrics*, **88**(3), 863, 2011.
- Danqi, C. and Christopher, D. M. (2014), A Fast and Accurate Dependency Parser using Neural Networks, *Proceedings of EMNLP 2014*.
- Jang, H., Park, S., and Yoon, B. (2023), Exploring technology opportunities based on user needs: application of opinion mining and SAO analysis, *Engineering Management Journal*, **35**(3), 209-222.
- Kim, S. and Yoon, B. (2021), Patent infringement analysis using a text mining technique based on SAO structure, *Computers in Industry*, **125**, 103379.
- Kim, Y., Suh, J., and Park, S. (2008), Visualization of patent analysis for emerging technology, *Expert Systems with Applications*, **34**(3), 1804-1812.
- Liu, Z., Zhang, J., Qin, T., Qu, Y., and Li, Y. (2022), One-to-many comparative summarization for patents, *Scientometrics*, **127**(4), 1969-1993.
- Moehrle, M. G. (2005), How combinations of TRIZ tools are used in companies-results of a cluster analysis, *R&D Management*, **35**(3), 285-296.
- Park, H., Yoon, J., and Kim, K. (2012), Identifying patent infringement using SAO based semantic technological similarities, *Scientometrics*, **90**(2), 515-529.
- Wang, X., Ma, P., Huang, Y., Guo, J., Zhu, D., Porter, A. L., and Wang, Z. (2017), Combining SAO semantic analysis and morphology analysis to identify technology opportunities, *Scientometrics*, **111**, 3-24.
- Wang, X., Ren, H., Chen, Y., Liu, Y., Qiao, Y., and Huang, Y. (2019), Measuring patent similarity with SAO semantic analysis, *Scientometrics*, **121**, 1-23.
- Ying, C., Shuyu, Y., Jing, L., Yoon, J., and Kim, K. (2011), Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks, *Scientometrics*, **88**(1), 213-228.
- Yoon, J., Park, H., and Kim, K. (2013), Identifying technological competition trends for R&D planning using dynamic patent maps: SAO-based content analysis, *Scientometrics*, **94**, 313-331.

저자소개

김선혜 : 동국대학교 산업시스템공학과에서 2018 학사, 2020 석사, 2024 박사학위를 취득하고, 한국외국어 대학교, AI데이터융합학부, 동국대학교 산업시스템공학과에서 전문연구원으로, 영국 Cambridge University에서 Research fellow로 재직 중이다. 연구분야는 특허 분석, 텍스트 마이닝, 자연어 처리, 딥러닝, 생

성형 모델이다.

장혜진 : 동국대학교 산업시스템공학과에서 2014 학사, 2016 석사, 2022 박사학위를 취득하였다. 2024년부터 한국외국어대학교 AI데이터융합학부 조교수로 재직하고 있다. 연구분야는 기술경영, 특허분석, 자연어처리이다.