

국내학술논문에서 약국관련 연구동향분석

김양우*

한국보건복지인재원

(2023년 4월 28일 접수 · 2023년 5월 3일 수정 · 2023년 5월 19일 승인)

Analysis of Pharmacy Related Research Trends in Domestic Academic Papers

Yang-woo Kim*

Korea Human Resource Development Institute for Health & Welfare

(Received April 28, 2023 · Revised May 3, 2023 · Accepted May 19, 2023)

ABSTRACT

Keywords:

Pharmacy

Data

Morphological analysis

NLP

Natural Language Processing

This research was conducted to investigate trends in domestic research on pharmacies using the open API of the KOREA Citation Index (KCI) based on the keyword “pharmacy”. We collected online big data using the Biblio Data Collector, performed a morphological analysis of the collected data, and visualized high-frequency keywords in a word cloud. The period covered three years of data from March 2019 to November 2022 to investigate pharmacy-related paper information after the COVID-19 pandemic. Results As a result of this study, the words with the highest frequency of pharmacy are “research”, “service”, “business”, “region”, “corporation”, “export”, etc., followed by “impact”, “society”, “Product” was the order. There was only one paper on the COVID-19 pandemic, appearing with the keyword “mask”, and research on pharmacy and the COVID-19 pandemic was minimal compared to the social role and topicality of pharmacy at the time. As of December 2022, studies on 24,389 pharmacies in the last 3 years totaled 26 of the KCI criteria, indicating poor research on pharmacies compared to the social role and weight they contribute.

서론

COVID-19 pandemic 이후 약국의 공적 마스크 판매 등 약국의 사회적 역할과 기능이 재조명 되었다. 이를 계기로 2000년 7월 의약분업 시행이후 처방전에 따라 조제하고, 일반의약품을 판매하는 고착화된 시스템 하에서 지내오던 약국의 새로운 역할과 약사의 직능에 대한 연구와 고민의 필요성이 커져가고 있다. 본 연구에서는 이를 반영하는 하나의 정량적 분석방법으로 ‘약국’ 키워드를 바탕으로 한국 학술지인용색인(KCI; Korea Citation Index)의 open API를 활용하여 약국에 대한 국내 연구의 트렌드를 살펴보았다. 약국관련 논문의 제목과 키워드 초록에서 비정형 텍스트

를 수집하고 형태소분석을 거쳐 워드 클라우드로 시각화 하였다.

워드 클라우드는 비정형 텍스트 데이터를 분석하는 대표적인 기법으로, 사전에 전 처리된 텍스트 데이터로부터 단어(명사 또는 형용사)들을 추출하고 출현 빈도수를 계산하고 워드 클라우드 기법을 사용하여 시각화 분석하는 것이다. 출현 빈도수가 높은 단어는 크게 표시되고 각각의 단어들은 색상이 다르게 표시된다. 여기서 출현 빈도수가 높은 단어는 중요도가 크거나 관심도가 높은 것을 의미한다.¹⁾ 워드 클라우드를 하기 위해서는 전처리로 형태소분석이 필요한데, 형태소란 더 이상 분리하면 본래의 의미를 잃어버리는 말의 최소 단위이다. 한국어에서 문장은 공백으로 구분

*Corresponding author: Yang-woo Kim, Korea Human Resource Development Institute for Health & Welfare, Osongsangmyeong 2-ro, Cheongju-si, Chungcheongbuk-do, Republic of Korea

Tel: +82-43-710-9205, Fax: +82-43-710-9209, E-mail: line@kohi.or.kr

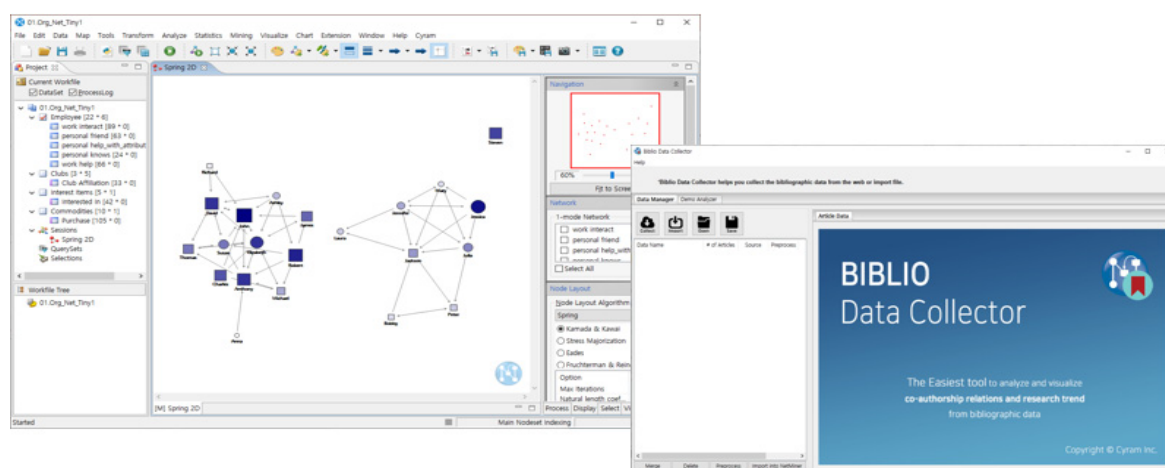


Fig. 1. NetMiner and Data Collector

되는 어절(단어)의 조합이다. 형태소분석은 어절 단위의 문장을 입력 받아서 문맥 정보를 활용하여 형태소 단위로 나누고 각 형태소에 적절한 품사 기호를 부착한 결과를 생성하는 것이다.²⁾ 텍스트는 비정형 데이터로서 주로 정성적 분석의 대상이었지만, 텍스트를 구조적으로 분석하는 방법들이 다양한 분야에서 연구되어왔다. 텍스트에 대한 구조적, 정량적 분석방법은 언어학, 컴퓨터공학 분야를 중심으로 주로 언어연구나 정보 검색을 위해 발전되어 왔고 인문학, 사회과학에서는 내용 분석(Content Analysis)에 사용되어왔으며, 최근에는 거의 모든 학문 분야에서 연구 동향 또는 트렌드 분석 맥락에서 널리 활용되고 있다. 실용적인 측면에서도 뉴스 미디어나 소셜 미디어의 텍스트를 통한 여론 분석, 마케팅 맥락에서의 고객 요구 분석, 특허나 논문 등의 과학기술 트렌드 분석, 녹취록이나 연설문의 의미구조 분석 등 다양한 용도로 쓰이고 있다. 본 연구에서는 텍스트를 정형화된 구조로 모델링하고 분석하는 접근방법에 주목하여 학술문헌의 서지정보에 포함된 텍스트 자료, 문헌의 제목, 초록, 주제 키워드에 이를 적용하였다. 2019년 12월 ~ 2022년 11월 (3년간)의 데이터는 국내 학술 문헌 정보 제공 사이트인 KCL*에 게재된 서지정보(Bibliographic Information)를 수집하였다. 학술지는 시대의 변화와 흐름을 살펴볼 수 있는 유용한 도구이고³⁾, 논문의 초록은 논문의 내용과 주제를 표현하는 가장 적절한 정보로 인식되기⁴⁾ 때문에 KCL에 게재된 '약국' 관련 학술지와 서지정보를 활용하여, 이 시대의 약국관련 연구 트렌드를 알아 보고자 하였다.

* 한국학술지인용색인(KCI; Korea Citation Index)은 국내 학술지 정보, 논문 정보(원문) 및 참고문헌을 DB화하여 논문 간 인용관계를 분석하는 시스템입니다.

연구방법

1. 데이터 수집방법

약국 서지정보의 분석대상 데이터(Biblio data)는 국내 학술 문헌 정보 제공 사이트인 KCL에서 약국 관련 검색 키워드를 통해 수집한 최근 3년 간(2019년 12월 ~ 2022년 11월) 서지 정보이다. 온라인 상에서 자동으로 데이터를 수집하기 위해 넷마이너(NetMiner)의 논문 서지정보 수집기(Biblio Data Collector, 이하 'BDC') 기능을 이용하였다. BDC는 KCL의 Open API를 활용하여 서지정보를 수집할 수 있도록 사용자 UI를 구현한 서지정보 수집 및 전처리 소프트웨어이다.

수집된 서지정보 중 초록, 제목, 키워드 텍스트에 대하여 형태소 분석을 통해 단어를 추출하였다. 이 과정에서 전문 용어와 복합명사, 고유명사 등을 최대한 반영하고 불용어를 제거하기 위해 유의어와 제외어 사전을 작성하여 적용하였으며, 단어의 빈도와 각 문서에서 단어의 중요도 지표인 TF-IDF(Term Frequency-Inverse Document Frequency) 값을 측정하였다. 아울러 연구기관 단위의 분석을 위해 서로 상이하게 표기된 기관명을 동일화 하는 등의 정제작업을 수행하였다.

2. 데이터 분석 프로세스

최초의 분석용 데이터는 비정형 텍스트 데이터로서 수집된 문서 수만큼의 텍스트가 존재한다. 형태소 분석과 같은 자연어 처리 과정(Natural Language Processing, NLP)은 비정형 데이터를 정형화된 형태로 변환하기 위한 과정이라고 볼 수 있다. 자연어 처리를 위해서는 형태소 분석이 필수적

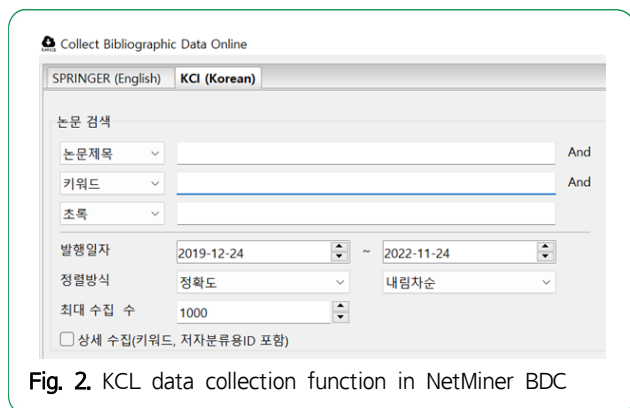


Fig. 2. KCL data collection function in NetMiner BDC

으로 요구되는데, 텍스트에서 형태소를 추출하고 추출한 형태소의 품사를 판별하는 품사 태깅(POS Tagging) 과정을 의미한다. 품사 태깅 시 한 단어가 여러 품사로 사용될 수 있기 때문에 단어의 위치, 앞뒤에 등장하는 단어의 품사 등을 종합적으로 고려하여 판별하게 된다. 자자연어 처리의 최종 산출물은 문서·단어 등장 관계를 정의한 2-모드 행렬과 단어의 품사, 빈도와 같은 속성들이다. 아울러 자연어 처리 단계에서 추출할 단어의 품질을 향상하기 위해 유사어, 제외어 사전을 적용하였는데, 유사어 사전은 동의어와 고유명사 및 복합명사의 처리를 위해, 제외어는 분석에서 제외할 단어 처리를 위해 사용하였다. 데이터 분석에서 입력 데이터의 품질은 분석결과의 품질에 영향을 미칠 수 있으므로 텍스트 분석에서 사전의 구축은 매우 중요한 단

계에 해당하며, 시간이 많이 소요되는 과정이다. 텍스트 분석의 절차는 보통 문서 수집, 파싱(parsing) 및 필터링(filtering), 구조화, 빈도 분석 및 유사도 분석으로 나뉘고, 텍스트 분석 관련 기술은 크게 비정형 텍스트 구조화까지의 단계와 구조화된 문서의 분석과 활용 단계로 나누어 파악한다(Fig. 3).⁵⁾

○ 데이터 수집 및 분석결과

BDC의 KCL 데이터 수집 기능은 제목, 학술지, 주제 키워드에 대한 검색, 데이터의 발행 연도(year) 지정, 질의(query)를 통한 데이터 수집이다. 약국 키워드로 등장한 단어의 논문 빈도와 분포 특성을 살펴보기 위한 데이터수집 결과, 약국데이터는 논문제목 기준 26건, 키워드 기준 16건으로 총 42건이었다. 이들을 동일한 제목, 초록, 주제 키워드, 저자명인 경우 하나로 머지(Merge)하자 34건으로 정리되었고, 그 중에서 관련이 없는 데이터(e.g 박경리소설의 김 약국의 딸들(중략) 등, Exclusive words: 박경리, 사신단, 식민지, 원고적격, 조선, 취약국가론, 한말)를 필터링한 결과 최종적으로 26건이 분석대상 건으로 확정되었다. 수집된 정보 항목은 제목, 초록, 주제 키워드, 저자명 및 소속 기관, 발행연도, 발행기관 등이었다. 위의 표는 수집 결과물의 예시이다(Table 1).

지난 3년간 약국 데이터에 등장한 단어의 논문 빈도와 분포 특성을 살펴보았다. 많이 사용된 단어는 ‘연구’, ‘서

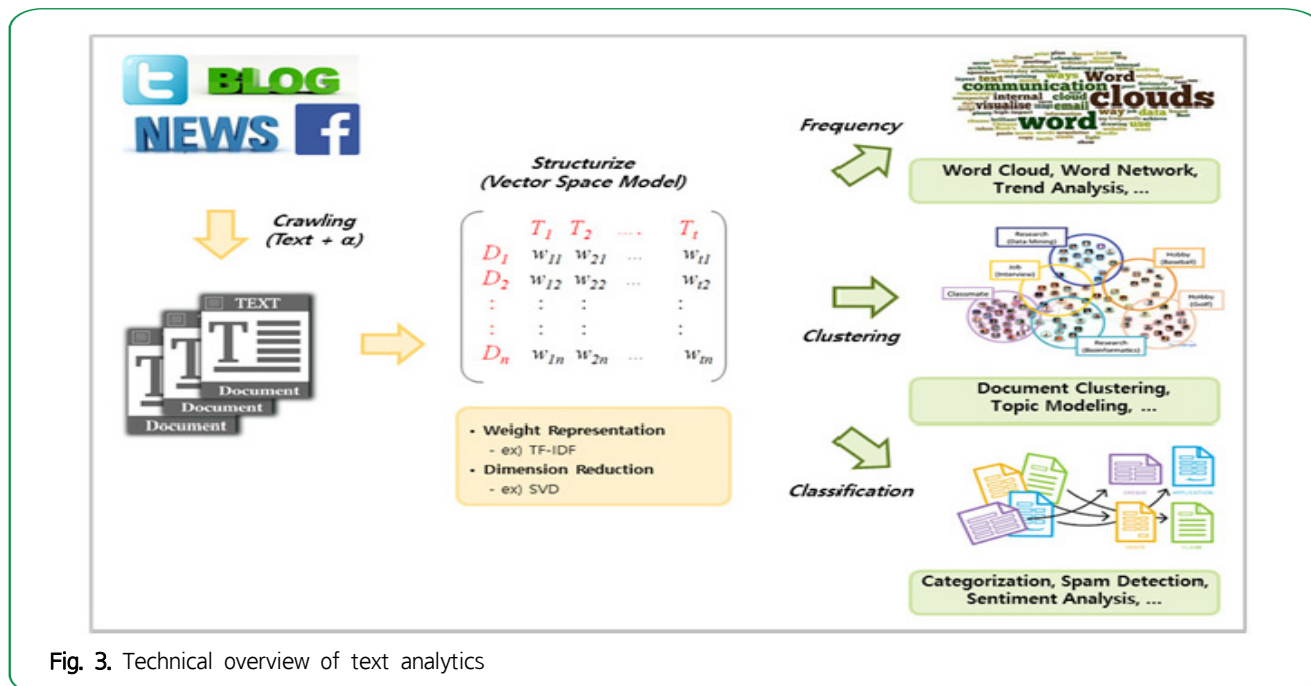


Fig. 3. Technical overview of text analytics

Table 1. Example of BDC_KCL data collection result

	Article_ID	Title_KR	Title_EN
1	ART0026	서울시 세이프약국 사업의 현황 진단과 정책 과제 : 사업 참여 약국 대상 평가 연구를 중심으로	Current status and future directions o
2	ART0027	스포츠 약국 가이드라인의 국제현황	Investigation of International Status o
3	ART0024	지역 약국의 노인 전문 약사 도입 논의	Discussion on the Introduction of Ger
4	ART0026	19세기말~20세기초 전라도 강진 병영지역의 환(煥)사용 실태와 메커니즘 -박약국(朴藥局) 장	A Study on the Use of Hwan in the Bi
5	ART0024	지역 약국으로 회수된 폐의약품의 종류와 약가 분석	Current Status Analysis of Unused anc
6	ART0026	지역약국 약사의 복약지도 현황 및 상담공간에 대한 인식	Community Pharmacists' Medication C
7	ART0025	19세기말 朴藥局의 고객관리와 경영전략	The Client Management and Manage
8	ART0027	코비드-19 마스크 방역 정책과 약사, 약국의 역할	COVID-19 Mask Policy, Role of Pharm
9	ART0024	AHP를 이용한 고령사회 약국약사의 역할에 대한 우선순위 연구	A Study on the Priorities in the Roles
10	ART0025	박경리의 소설에 나타난 비극적 세계관 연구 -『김약국의 딸들』을 중심으로-	A Study on the Tragic View of the Wo
11	ART0025	약국 자동화 관리 시스템에서 변위 센서를 이용한 약품 수량 정확도 개선	Improving Drug Quantity Accuracy us
12	ART0024	의약분업 이후 약국약사의 전문주의 - 전문직 체계를 활용하여 -	Where is the pharmacy professionalis
13	ART0027	일반의약품 약국 외 판매가 소비자의 건강관리에 미치는 영향	The Effect of Sales of Non-prescriptic
14	ART0025	성인을 대상으로 한 지역약국의 산제 조제 현황 및 약사들의 인식 조사 연구	The Status and Perception on Compo
15	ART0026	해외여행자들의 여행상비약 구비 경험 및 약국서비스 이용 의향	Overseas Travelers' Experiences in Pre
16	ART0028	반려동물애호가 대상 동물의약품 사용 및 동물약국 이용 현황 조사	A Survey on the use of veterinary me
17	ART0026	지역약국 기반 환자안전활동: 대한약사회 지역환자안전센터 활동을 중심으로	Community Pharmacy-based Patient S
18	ART0026	우리나라와 일본의 지역약국 약사서비스 관련 공적보험급여의 범위 및 수가 비교	A Comparative Study on Health Insur
19	ART0025	영동지방 의약분업 예외지역 약국들에서의 감기 증상 모의 환자에 대한 약물 처방 현황	Prescription Pattern for a Simulated P
20	ART0026	약국약사의 커뮤니티 재구성: 의약분업 20년, 약사 전문주의 변화	Capturing the Patient-centeredness ar
21	ART0026	법인의 직업의 자유에 대한 소고- 법인약국 개설 금지와 법인 안경업소 개설 금지의 비교를 중	A Study on the Freedom of the Occu

비스’, ‘사업’, ‘지역’, ‘법인’, ‘수출’ 등이었으며 다음으로 ‘영향’, ‘사회’, ‘제품’ 등의 순이었다. 고빈도 단어 500개를 워드 클라우드(Word Cloud)로 시각화 하면 아래(Fig. 4)와 같다.

고찰 및 결론

약국에 대한 사회적인식과 역할론에 대한 고민은 일선 약사들에게 큰 숙제이다. 이를 위해서는 우리사회에서 약국관련 연구가 더욱 활발하게 실행되어야 한다는 데는 이견이 없을

것이다. 본 연구에서는 약국에 대한 연구를 비정형데이터 수집과 자연어처리 그리고 형태소분석을 활용하여 약국에 대한 국내 연구의 트렌드를 정량적으로 살펴보고자 하였다. 이러한 분석방법은 데이터트랜스포메이션(Data Transformation) 시대가 열리면서 거의 모든 분야에서 연구 동향이나 트렌드 분석에 두루 활용되고 있지만 약학분야에서는 아직 시도조차 되고 있지 않은 실정이다. 본 연구에서 COVID-19 pandemic이 시작된 이후 2019년 3월부터 2022년 11월 까지 3년간의 약국 관련 전체논문을 분석한 결과, KCL기준 총 26건으로 약국이 기여하는 사회적 역할과 비중에 비해 약국에 대한 연구가



Fig. 4. Pharmacy -word cloud

빈약한 것으로 나타났다.

COVID-19 pandemic 관련한 논문은 단 1건으로, ‘마스크’라는 키워드로 출현하고 있어, 약국과 COVID-19 pandemic에 대한 연구가 당시 약국의 사회적 역할과 화제성에 비해 매우 미미한 것으로 나타났다. 2022년 12월 현재 24,389개인 약국⁶⁾에 대한 최근 3년간의 연구의 총량과 트렌드를 살펴볼 때, 우리사회에서 약국에 대한 진지한 고찰과 연구에 있어 아직은 아쉬운 점이 많았다. 앞으로 약국이 우리사회 변화에 제대로 기여하고 포지셔닝 할 수 있도록 관련 연구에 대한 지속적인 관심과 지원이 필요하다.

참고문헌

- 1) Lee WJ. A Study on Word Cloud Techniques for Analysis of Unstructured Text Data. The Journal of the Convergence on Culture Technology (JCCT) Vol. 6, No. 4, pp.715-720. November 30, 2020. pISSN 2384-0358, eISSN 2384-0366
- 2) Lee CJ. Korean Morphological Analysis Method Based on BERT-Fused Transformer Model. KIPS Transactions on Software and Data Engineering (KTSDE) Vol. 11, No.4, pp 169-178
- 3) Yoon OH. Analysis of Research Trends in the Journal of Learning and Performance (2000~2014). Journal of Learning and Performance, vol. 18, no. 1, pp. 159-177, June 2016.
- 4) Choi SH, Lee JY. Usability Analysis of Structured Abstracts in Journal Articles for Document Clustering. KOSIM vol. 29, no. 1, pp. 331-349, Mar. 2012. DOI: <https://doi.org/10.3743/kosim.2012.29.1.331>
- 5) Kim NG, Lee DH, Choi HC, William Xiu Shun Wong. Investigations on Techniques and Applications of Text Analytics. The Journal of Korean Institute of Communications and Information Sciences '17-02 Vol.42 No.02 <https://doi.org/10.7840/kics.2017.42.2.471>
- 6) Healthcare Bigdata Hub. <https://opendata.hira.or.kr/op/opc/selectOpenData.mo?sno=11925>