

# 학술 정보 기반 한의학 처방을 위한 확장 적응증 데이터베이스 구축 Extended Adaptation Database Construction for Oriental Medicine Prescriptions Based on Academic Information

이소민\*, 백연희\*\*, 송상호\*, RETITI DIOP EMANE CHRISTOPHER\*\*, 함선중\*, 홍성연\*, 김익수\*, 임종태\*, 복경수\*\*,  
MINH NHAT TRAN\*\*\*\*, QUYNH HOANG NGAN NGUYEN\*\*\*\*, 김소영\*\*\*\*, 김안나\*\*\*\*, 이상훈\*\*\*\*  
유재수\*

충북대학교 정보통신공학과\*, 충북대학교 빅데이터 협동과정\*\*, 원광대학교 SW 융합학과\*\*\*,  
한국한의학연구원\*\*\*\*, 과학기술연합대학원대학교 한의융합의학\*\*\*\*

So-Min Lee(somin@cbnu.ac.kr)\*, Yeon-Hee Baek(byh741@naver.com)\*\*,  
Sang-Ho Song(pshtkdgh9@cbnu.ac.kr)\*,  
RETITI DIOP EMANE CHRISTOPHER(retitidiopchristopher@nate.com)\*\*,  
Xuan-Zhong Han(2225608139@qq.com)\*, Seong-Yeon Hong(hsy0521@naver.com)\*,  
Ik-Su Kim(dlrtn654@naver.com)\*, Jong-Tea Lim(jtlim@chungbuk.ac.kr)\*,  
Kyoung-Soo Bok(ksbok@wku.ac.kr)\*\*\*, MINH NHAT TRAN(tnminh@kiom.re.kr)\*\*\*\*,  
QUYNH HOANG NGAN NGUYEN(noel123@kiom.re.kr)\*\*\*\*, So-Young Kim(ssuying1221@kiom.re.kr)\*\*\*\*,  
An-Na Kim(ankim2012@kiom.re.kr)\*\*\*\*, Sang-Hun Lee(ezhani@kiom.re.kr)\*\*\*\*,  
Jae-Soo Yoo(yjs@cbnu.ac.kr)\*

## 요약

의료의 질은 효과, 효율, 적절성, 과학적-기술적 측면 등과 같은 4가지로 정의할 수 있다. 과학적-기술적 측면의 질 관리를 위해 의료기관에서는 매년 보수교육의 형태로 최신 지견을 현장에 보급하고 있다. 하지만 최신 지견이 가장 빠르게 보급되는 연구 결과들을 단발성인 보수교육만으로 임상 현장에 충분히 보급하는 것에는 명백한 한계가 존재한다. 빅데이터, 인공지능과 같은 지능정보처리 기술이 의료 분야에 적용될 경우 기존에 문헌 조사 등으로 연구되어 적은 정보만으로 연구를 수행해야 했던 한계를 극복할 수 있다. 본 논문은 기존 약재 처방 적응증을 확장할 수 있는 근거가 되는 데이터베이스를 구축한다. 이를 위해 한의학 관련 국내외 논문 정보를 수집, 저장 관리, 분석하는 작업을 수행한다. 약재 처방전의 확장 적응증 콘텐츠 구축을 위한 한의학 근거문헌 데이터의 처리 및 분석 기법을 설계한다. 본 연구를 통해 한의학 의사결정지원시스템에서 근거 기반 약재처방 주치 정보의 기본 콘텐츠로 활용할 수 있을 것이라 기대한다.

■ 중심어 : | 학술 정보 | 한의학 | 확장 적응증 | 데이터베이스 구축 |

## Abstract

The quality of medical care can be defined as four types such as effectiveness, efficiency, adequacy, and scientific-technical quality. For the management of scientific-technical aspects, medical institutions annually disseminate the latest knowledge in the form of conservative education. However, there is an obvious limit to the fact that the latest knowledge is distributed quickly enough to the clinical site with only one-time conservative education. If intelligent information processing technologies such as big data and artificial intelligence are applied to the medical field, they can overcome the limitations of having to conduct research with only a small amount of information. In this paper, we construct databases on which the existing medicine prescription adaptations can be extended. To do this, we collect, store, manage, and analyze information related to oriental medicine at domestic and abroad Journals. We design a processing and analysis technique for oriental medicine evidence research data for the construction of a database of oriental medicine prescription extended adaption. Results can be used as a basic content of evidence-based medicine prescription information in the oriental medicine-related decision support services.

■ keyword : | Academic Information | Oriental Medicine | Extended Adaptation | Database Construction |

\* 본 연구는 한국 한의학회 연구주요사업인 AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축 사업(No.KSN201211), 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2019R1A2C2084257), 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업(IITP-2021-2020-0-01462), 2020년 대한민국 교육부와 한국연구재단의 지원(NRF-2020S1A5B8103910), 그리고 중소벤처기업부 '산업전문인력역량강화사업'의 재원으로 한국산학연합회(AURI)의 지원(2021년 기업연계형연구개발인력양성사업, 과제번호 : S3047889)을 받아 수행된 연구임.

접수일자 : 2021년 04월 21일  
수정일자 : 2021년 05월 18일

심사완료일 : 2021년 05월 18일  
교신저자 : 유재수, e-mail : yjs@cbnu.ac.kr

## I. 서론

의료의 일차적 목표는 환자의 건강을 위하여 질병을 치료하는데 있는 것으로 의료의 질은 효과, 효율, 적절성, 과학적-기술적 측면 등과 같은 4가지로 정의된다 [1]. 과학적-기술적 측면의 질 관리를 위해 의료기관에서는 매년 보수교육의 형태로 최신 지견을 현장에 보급하고 있으며 의료법에서 “의료인은 보수교육을 받아야 하며, 보수교육을 이수하지 않은 의료인에 대해서는 의료인 실태 신고가 반려되도록 규정”하는 등 의료법에 따라 이를 엄격히 관리하고 있다[2]. 그러나 이러한 최신지견이 가장 빠르게 보급되는 연구결과들을 단발성인 보수교육만으로 임상현장에 충분히 보급하는 것은 명백한 한계가 존재한다. 빅데이터, 인공지능과 같은 지능정보처리 기술을 통해 의료 분야에 적용될 경우 기존에 문헌 조사 등으로 연구되어 적은 정보만으로 연구를 수행해야 했던 한계를 극복할 수 있다. 데이터를 활용한 연구를 통해 과거에는 발견하지 못했던 새로운 처방에 대한 데이터를 통해 많은 연구자들이 이를 공유할 수 있도록 하는 데이터베이스를 구축할 수 있으며 의사결정지원 서비스에 활용될 수 있다.

기존에는 사람이 직접 수작업으로 논문 사이트에서 논문을 수집 및 분석하였다[3]. [3]은 5개년 간의 논문을 수집하여 한의학계에서 항염증 효과와 약제 및 처방의 연구 동향을 분석하였다. 하지만 발행된 지 5개년 이상의 논문까지 수작업으로 수집하여 의료분야에 활용하기에는 한계가 있다. 현재 다양한 출처로부터 생성되는 주치병증 데이터를 분석하여 의사결정지원 서비스에 활용하기 위해서는 빅데이터 분석에 기반을 둔 기술이 필요하다.

본 논문에서는 이를 보완하기 위해 한의 처방과 관련한 최신지견이 포함된 연구논문을 통해 기존의 의서에 국한된 적응증 뿐 아니라 최근 임상 활용으로 인해 확장되고 있는 새로운 적응증에 대한 정보를 제공할 수 있는 데이터베이스를 구축한다. 이를 통해 한의사의 임상 진료에 최신 연구결과를 보다 잘 활용 할 수 있는 과학기술적 콘텐츠를 구축하고자 하였다. 국내외 사이트로부터 크롤링, API를 활용하여 1980년대부터 현재까지의 논문을 자동화하여 수집하고 분석하였으며, 논문

에서 나타난 최신 처방을 파악하고 이를 데이터베이스로 제공한다. 확장 적응증 콘텐츠 구축을 위한 텍스트 분석을 수행하기 위해 근거문헌의 수집 및 근거문헌 데이터의 저장 관리 기법을 설계 및 구현한다. 약재처방연의 확장 적응증 콘텐츠 구축을 위한 빅데이터 분석 기반의 한의학 근거 문헌 데이터의 처리 및 분석 기법을 설계 및 구현한다. 최종적으로 빅데이터 분석을 이용한 확장 적응증 콘텐츠를 구축하여 의사결정 서비스에 활용한다.

## II. 관련 연구

최근 데이터를 수집 및 분석 방법을 이용한 데이터 마이닝에 관한 연구[4-8]가 다양하다. 데이터 마이닝 처리 과정을 데이터 수집, 데이터 정의, 데이터 저장, 데이터 분석, 데이터 시각화 순으로 진행된다. 텍스트 마이닝은 데이터 마이닝 중 하나로 텍스트 데이터를 분석 대상으로 한다. 텍스트 마이닝은 데이터 분석을 위해 텍스트 분석, LDA 기반 토픽 모델링 등을 이용한다.

[4]에서는 데이터 마이닝 기법을 이용하여 유튜브 인기 동영상 콘텐츠 요소들의 상관성을 분석하였다. 인기 동영상 콘텐츠의 키워드, 조회 수, 동영상의 likes와 dislikes의 수, 댓글 수 등을 수집하여 빈도 수를 분석하여 상관 관계를 분석하였다. [5]는 scopus에 게재된 영어 논문 초록을 대상으로 텍스트 분석, 시대별 빈도 분석, 토픽모델링을 이용하여 인공지능과 헬스케어 융복합 분야의 시대별 주요 연구주제와 연구동향을 분석하였다. [6]은 텍스트 마이닝을 이용하여 4차 산업과 관련된 연구 동향을 파악하였다. 2016년부터 2019년까지 “4차 산업”이라는 키워드로 논문을 수집한 후 LDA 알고리즘 기반의 토픽 모델링 기법을 활용하여 자료 분석을 수행하였다. [7]은 승례문과 관련된 신문 기사를 수집한 후 핵심 키워드를 추출하여 특징을 분석하였다. 또한 핵심키워드 간의 연관 관계 분석을 통해 시각화 결과를 제공하였다. [8]은 “DMZ 관광” 키워드에 대한 소셜미디어 데이터를 수집 및 분석한 후 토픽 모델링을 적용하여 핵심 토픽을 도출한 결과를 시각화하여 제공하였다.

위와 같이 데이터 마이닝은 다양한 분야에서 연구가

다양하게 진행되고 있지만 의학 분야에서 데이터 마이닝을 적용한 연구는 활발히 이루어지지 않고 있다. 따라서 본 연구는 한의학 분야 데이터를 수집 및 분석하여 연구자들이 활용할 수 있도록 데이터베이스를 구축하고자 한다. 한의 처방의 확장 적응증 콘텐츠를 구축하기 위해 2개의 한의학 관련 논문 사이트에서 논문을 수집하였다. OASIS는 국내 논문에 대한 정보를 제공하며 PubMed는 국외 논문에 대한 정보를 제공한다. 다음은 OASIS와 PubMed에 대하여 조사한 내용이다.

OASIS(Oriental Medicine Advanced Searching Integrated System)[9]는 한의학 연구자와 한의약 기획자에게 전통의학 정보를 제공하기 위한 포털사이트다. 전통의학 연구정보의 통합 서비스를 구축하고 학술 교류 네트워크를 확대하여 한의학 연구지원 기능을 강화하기 위해서 다양한 한의학 분야의 콘텐츠와 관련 서비스를 제공한다. 웹 크롤러(Web Crawler)는 웹상에 존재하는 웹 문서들 중 특정 사이트의 콘텐츠를 자동으로 수집하는 기술이다[10]. 웹 크롤러를 통한 키워드 검색을 수행하여 OASIS 사이트에서 제공하는 논문 정보에 대한 데이터를 수집한다. 본 논문에서는 국내 근거문헌을 수집하기 위한 대표적인 사이트로 [9]에서 근거문헌 DB 구축을 위한 데이터를 수집한다.

PubMed[11]은 생명과학 및 생물의학 주제에 대한 참조 및 요약에 담고 있는 MEDLINE 데이터베이스를 주로 접근할 수 있게 해주는 자유 검색 엔진이다. 미국 국립 보건원의 미국 국립 의학 도서관이 정보 검색의 Entrez 시스템의 일부로서 데이터베이스를 유지 및 관리한다. PubMed에서는 E-utilities라는 API를 제공한다. E-utilities는 9개의 프로그램(EInfo, ESerch, ESummary, EFetch, ELink, EGQuery, ESpell, ECitMatch)으로 구성되며 각각의 프로그램은 다른 기능을 제공한다. E-utilities 구성 요소를 결합하여 애플리케이션 내에서 맞춤형 데이터 파이프라인을 형성하여 데이터를 조작한다. E-utilities는 데이터의 입력 및 출력에 데이터베이스 레코드 식별을 위한 정수 ID인 UID를 사용한다. 본 논문에서는 ESerch를 이용해 질의에 일치하는 UID 목록을 추출한 뒤 Efetch를 이용해 지정된 데이터베이스의 UID 목록에 응답하여 데이터 레코드를 수집한다. 본 논문에서는 국외 근거문헌을

수집하기 위한 대표적인 사이트로 [11]에서 근거문헌 DB 구축을 위한 데이터를 수집한다.

### III. 확장 적응증 데이터베이스 구축

#### 1. 시스템 구조

확장 적응증 콘텐츠 구축을 위해 국내 OASIS, 국외 PubMed 사이트에서 논문을 수집, 저장 관리, 분석을 수행한다. [그림 1]은 전체 시스템 구조를 나타낸다. 먼저 한의학 처방명 온톨로지를 전처리하고 이를 활용하여 OASIS와 PubMed에서 각각 크롤링과 API를 통해 논문을 수집한다. 수집된 논문을 바탕으로 처방과 관련이 있는 질환에 대한 근거문헌 DB를 구축한다. 엑셀 파일로 확장 적응증 콘텐츠를 구축하고 '천식'에 대해 과거의 처방과 최근의 처방을 비교한다. 이를 표로 시각화한다.

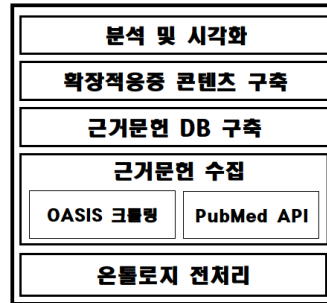


그림 1. 전체 구조

PubMed, OASIS 등의 출처로부터 한의학 관련 처방 연구의 국내외 근거문헌을 수집하였다.

#### 2. 온톨로지 전처리

처방명에 대한 정보(한글/한자/중국어병음/영문표기/이명)가 정리된 온톨로지 파일로부터 처방명의 이명 목록을 딕셔너리 형태로 생성한다. 질환명에 대해서도 하나의 질병명을 키로 하여 나머지 이명 목록이 값이 되는 딕셔너리 형태로 생성한다. 논문 수집에 활용하기 위해 처방명과 질환명의 이명을 OR 연산으로 묶고 처

방명과 질환명은 AND로 묶어서 검색에 사용한다. PubMed는 외국 사이트이기 때문에 검색 시 영어 단어만 활용한다. OASIS는 국내 사이트이기 때문에 검색 시 한글과 영어 단어를 모두 활용한다. 처방명 id가 기준 key이며 영어 단어가 공식 이름으로 사용된다. 이명들은 other-names라는 키의 값으로 사용되며 검색 시 OR 연산으로 공식 이름과 이명을 연결한다. pub\_dict는 PubMed에서 사용하는 처방명 사전이다. [그림 2]는 온톨로지 전처리를 통해 생성하여 PubMed에서 검색어 입력 시 활용할 pub\_dict의 출력 결과이다.

```
key : F080108001
formular-name : MA HUANG TANG
other-names : ['마황탕', 'MA HUANG TANG', 'mahuang decoction', 'Mahuang Tang', 'Ephedrae Decoction', 'Ephedra Decoction', 'Ephedra Decoction', 'Ma Huang Combination', 'Ephedra Decoction', 'Ephedra Decoction', 'Ma-huang Decoction', 'Mahuang Combination']
```

그림 2. pub\_dict 출력 결과

[표 1]은 검색 키워드에 대한 디셔너리이다. 사이트별로 사용할 검색어가 다르기 때문에 각각 디셔너리를 따로 만들었다. 처방명 온톨로지와 질환명 온톨로지를 이용해 각각 디셔너리를 만들고 AND 연산으로 조인하여 최종 검색에 사용하였다. 논문 사이트에서 검색만 했을 때보다 논문을 다운로드까지 했을 때 시간이 더 소요되므로 실제로 검색 결과가 있는 단어만 크롤링에 사용하기 위해 필터링을 거쳤다.

표 1. 검색 키워드에 대한 디셔너리

구분	필터링 전	필터링 후	조인
OASIS	처방명 사전	oasis_dict_OR_before	oasis_dict_AND
	질환명 사전	oasis_disease_OR_after	
PubMed	처방명 사전	pub_dict_OR_before	pub_dict_AND
	질환명 사전	pub_disease_OR_after	

### 3. 근거문헌 수집 및 DB 구축

OASIS는 연도별로 파일을 제공한다. 파일은 엑셀 파일 형태로, 한글제목, 영문제목, 한글저자, 영문저자, 한글초록, 영문초록, 한글키워드, 영문키워드, 발행처, 저널, 발행일, 권 호, 페이지, ISBN, ISSN, PDF 항목이 제공된다. 파일을 수집하기 위해 OASIS의 검색 정보를 크롤링 하였다. 크롤링 과정은 [그림 3]과 같다. 첫째, 키워드를 검색하여 URL을 통해 검색결과 존재 여부를 확인한다. 검색 결과가 있는 경우 다음 단계로 진행

하고 검색 결과가 없는 경우는 다음 키워드를 검색한다. 둘째, 검색결과에 대한 논문을 다운로드한다. 셋째, 논문 수집이 완료된 리스트를 이용하여 수집된 논문을 처리한 후 근거문헌 DB에 저장 작업을 수행한다.

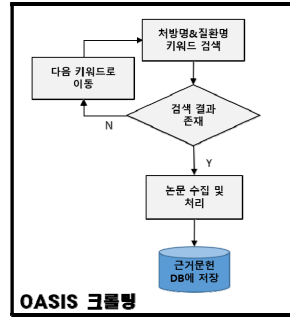


그림 3. OASIS 크롤링 과정

PubMed에서는 E-utilities API를 이용하여 근거문헌 수집을 하기 위해서 총 4단계를 실행한다. [그림 4]는 PubMed API 수집 과정이다. 첫째, ESearch (텍스트 검색)명령어를 이용하여 질환명, 처방명을 검색한다. 둘째, 필터링 함수를 통해 검색 결과가 없는 경우를 제외한다. 셋째, ESearch 결과로 나온 PMID를 List로 저장하여 EFetch (데이터 레코드 다운로드) 명령어로 검색하여 근거문헌 데이터 수집한다. 넷째, EFetch (데이터 레코드 다운로드) 명령어로 나온 XML에서 근거문헌 데이터를 반복하여 수집 및 저장하는 함수를 구현한다. 다섯째, 근거문헌 DB에 저장 작업을 수행한다.

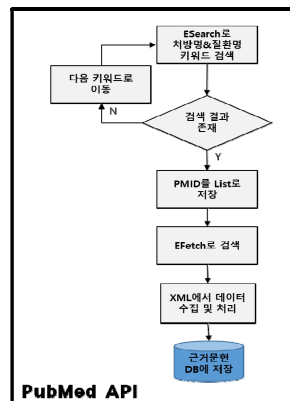


그림 4. PubMed API 수집 과정

4. 확장 적응증 콘텐츠 구축

한국표준질병·사인분류(KDC) 8차 개정 파일과 의학 주제표목(MeSH) 파일을 이용하여 질병 사전을 구축하였다. [그림 5]는 질병 사전의 예시이다. 질환 사전에서 검색 키워드의 질환명과 일치하는 키를 찾아 질병분류 코드, 한글 질환명, 영어 질환명, MeSH 유의어를 추출하고 처리결과 데이터베이스에 저장한다. [그림 6]은 확장 적응증 콘텐츠를 추출하는 과정이다. 근거문헌 DB에 저장된 OASIS, PubMed의 논문 원문에서 필요한 정보를 추출하여 처리결과 DB에 저장한다. 처리결과 DB에서 데이터를 추출하면서 형식에 맞게 전처리를 하고 이를 엑셀 파일로 추출하여 확장 적응증 콘텐츠를 구축한다.

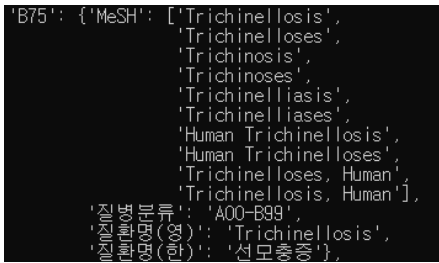


그림 5. 질병 사전

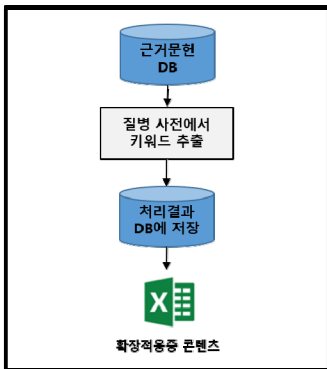


그림 6. 확장 적응증 콘텐츠 추출 과정

[그림 7]은 엑셀 파일로 구축된 확장 적응증 콘텐츠 추출 결과물 형식이다. 논문 서지 정보의 Authors 데이터는 OASIS 논문의 경우 기존 표현 방식에서 PubMed 양식에 맞게 성과 이름의 이니셜을 기재하는

방식으로 수정하였다. 중국인, 일본인, 기타 외국인의 경우 성과 이름의 구분이 어려워 한국인 이름만 추출 적응증 결과 형식에 맞게 처리하였다. (예시 : Hyangsook Lee -> Lee H)

구분	항목	항목 해	내용
지명명	Formula name		[대표어(표준어)] : [요약 용어] KCOM제공 처방용어집 + 자체 작성표기 검색용 set 별도 시트도 링크 제공
	Title		논문 제목
	Authors		논문 저자는 PubMed 양식 , 즉 성과 이름의 이니셜을 기재. 모든 저자 수록. 예시: Hyangsook Lee 는 Lee H
	Journal		PubMed 양식용 이름
	Year		yyyy
	Volume		권 (책) 방식용 이름
	Page		358쪽에서 364쪽까지의 논문이면 358-64 방식으로 기재. 맨 끝에 마침표 기재
	doi		있는 경우만 기재
	URL		•PubMed 에 등재된 경우 : /pubmed/PMID의 형태로 기재 •PubMed 에 등재되지 않은 경우 : URL이 있는 경우만 기재
	MeSH		PubMed 에서 활용되는 경우 기재
키워드	Keyword		해당 논문에서 제시된 핵심어를 세리클론 심으로 구분하여 기재
	Keywords		해당 논문에서 제시된 핵심어를 세리클론 심으로 구분하여 기재
	Korean study		Y, 국내 연구인 경우; 그 외는 표시하지 않음.
			•KOM제공 질환용어집 + 용례드에서 채굴되는 경우 및 자체 검색용 MeSH term 셋을 별도 시트도 링크로 제공
추출적응증	질환명(한글)		•적응증내용이 자동으로 수집되지 않은 경우(값이 없는 경우 수작업반영)
	KCOM 코드		KCOM용 처방용어집 + 자체 작성표기법을 별도 시트도 링크 제공
	KDC 코드		KDC용 코드 기재
	MeSH		해당 질환명의 MeSH term(entry term)
질환계통분류		호출기계 질환	KDC 대분류 또는 현재 건강관리 제공정보 분류

그림 7. 추출 결과물 형식

5. 분석 및 시각화

본 연구에서는 하나의 문헌에서 도출된 키워드에 대하여 텍스트 마이닝을 통해 관련 있는 처방과 질병명을 연결하여 약재처방 적응증 정보 확장 콘텐츠를 구축했다. 본 연구를 통해 구축된 확장 적응증 콘텐츠를 활용하여 국내의 한의학 논문을 기반으로 기존 문헌에는 없는 현대 질병명을 알아낼 수 있다. 최근 연구기반 약재처방 적응증을 활용하여 질환에 대한 과거의 처방과 최근의 처방을 비교 분석한 결과를 표 형태로 시각화하여 성능평가에서 제시한다.

IV. 성능 평가

OASIS에서 크롤링을 통해 CSV 형태로 논문 파일을 다운로드 한 후 필요한 키워드를 추출하였다. PubMed에서 API를 통해 XML 형태로 논문 파일을 다운로드 한 후 필요한 키워드를 추출하였다. 추출된 키워드와 논문 서지정보를 바탕으로 논문기반 한의 처방의 확장 적응증 데이터베이스 구축하였고 이를 활용하여 확장 적응증 콘텐츠를 연도별로 구축하였다. 1981년부터 2020년까지 약 40년간의 한의학 관련 논문으로부터 가장 많이 등장한 질환은 ‘천식’이었으며 그 다음으로는 ‘비만’, ‘궤양성 대장염’였다. 성능평가를 위해 천식과 관련하여 가장 많이 나타난 처방명의 순위를 나타내고 연

도별로 새롭게 나타난 처방명을 비교하였다. [표 2]는 성능평가를 수행한 환경이다.

표 2. 성능 평가 환경

이름	값
CPU	Intel(R) Core(TM) i5-4460 CPU @ 3.20GHz
Memory	8GB
Language	python 3.6.8
Database	MySQL

[표 3]은 1981년부터 2020년 까지의 논문 중 ‘천식’과 관련되어 나타난 처방명의 순위이다. 1981년부터 2010년 사이에 논문에서 나타난 처방명으로는 높은 빈도 순으로 ‘소청룡탕’, ‘정천탕’, ‘맥문동탕’, ‘사백산’, ‘육미지황환’이 나타났다. 2011년부터 2020년까지의 논문에서 나타난 처방명으로는 높은 빈도 순으로 ‘소청룡탕’, ‘마황탕’, ‘육병풍산’, ‘정천탕’, ‘신기환’, ‘우귀환’이 나타났다.

표 3. 1981~2020년에 등장한 천식 관련 처방명 순위

순위	1981~2010년	2011~2020년
	처방명(한글)	처방명(한글)
1	소청룡탕(13건)	소청룡탕(11건)
2	정천탕(6건)	마황탕/육병풍산(각 5건)
3	맥문동탕/사백산(각 5건)	정천탕(4건)
4	육미지황환(4건)	신기환/우귀환(각 2건)

[표 4]는 1981년부터 2020년까지의 논문 중 ‘천식’과 관련되어 나타난 처방명을 연도별로 비교하였다. 1981년부터 2000년 사이에 논문에서 나타난 처방명으로는 ‘소청룡탕’, ‘사군자탕’, ‘보중익기탕’, ‘정천탕’이 나타났다. 2001년부터 2010년까지의 논문에서 새롭게 나타난 처방명으로는 ‘맥문동탕’, ‘사백산’, ‘육미지황환’, ‘마황탕’, ‘마행감석탕’, ‘신기환’, ‘교애탕’, ‘좌귀음’, ‘소자강기탕’, ‘윤폐탕’이 나타났다. 2011년부터 2020년까지의 논문에서 새롭게 나타난 처방명으로는 ‘육병풍산’, ‘우귀환’, ‘당귀보혈탕’, ‘진무탕’, ‘이진탕’, ‘패독산’, ‘도적산’, ‘도기환’, ‘소부축어탕’이 나타났다.

표 4. 연도별 천식 관련 새롭게 등장한 처방명

1981~2000년	2001~2010년	2011~2020년
소청룡탕	맥문동탕	육병풍산
사군자탕	사백산	우귀환
보중익기탕	육미지황환	당귀보혈탕
정천탕	마황탕	진무탕
	마행감석탕	이진탕
	신기환	패독산
	교애탕	도적산
	좌귀음	도기환
	소자강기탕	소부축어탕
	윤폐탕	

본 논문에서는 논문기반 한의처방의 확장적응증 데이터베이스 구축을 통해 확장적응증 콘텐츠를 연도별로 구축하였다. 1981년부터 2020년까지 약 40년간의 논문으로부터 가장 많이 등장한 질환은 ‘천식’(83건)으로 총 23개 처방명이 등장하였다. 새로 구축된 콘텐츠의 우수성을 확인하기 위해 한국한의학연구원에서 개발한 “한의진료지원시스템 ALPHA”[12]에서 천식에 사용되는 처방을 검색하여 비교하였다. 한의진료지원시스템 ALPHA에서 “증상”에 ‘천식’을 입력하였을 때 다양한 고문헌 및 교과서 출처를 기반으로 한 총 111개 처방명이 검색되었다. 이중 본 연구에서 새로 구축된 확장 적응증 데이터베이스와 동일한 처방은 9건이었으며, 본 데이터베이스에서 새롭게 구축된 처방은 14건으로 최신 논문 기반 한의 처방의 확장 적응증 데이터베이스가 기존에 알려지지 않았던 새로운 적응증을 제공함을 확인할 수 있었다.

## V. 결론

본 논문은 한의학 관련 국내의 논문 정보를 수집, 저장 관리, 분석하는 작업을 수행하고, 이를 통해 기존 약재처방 적응증을 확장할 수 있는 근거가 되는 데이터베이스를 구축하였다. 약재처방의 확장 적응증 콘텐츠 구축을 위한 빅데이터 분석 기반의 한의학 근거 문헌 데이터의 처리 및 분석 기법을 설계 및 구현하였다.

본 연구는 기존 한의학 분야의 데이터를 이용해 기존

약재처방 적응증을 확장할 수 있는 데이터베이스를 구축한 것에 의의가 있다. 현재 다양한 출처로부터 생성되는 주치병증 데이터를 분석하여 의사결정지원 서비스에 활용하는 한계를 해결할 수 있다. 국내외 한의학 논문을 기반으로 고문헌에는 없는 현대 질병명 등의 최근 연구기반 약재처방 적응증을 확장 구축함으로써 한의학 의사 결정지원시스템에서 근거기반 약재처방 주치정보의 기본 콘텐츠로 활용할 수 있을 것이다.

본 연구에서는 PubMed 및 OASIS로부터 한의학 문헌을 수집하였으나 이외의 한의학 문헌을 제공하는 사이트로부터 문헌 수집 및 분석을 통해 확장 적응증 콘텐츠를 확장할 수 있을 것이다. 또한 처방명에 대해 한글 처방명과 영어 처방명을 사용하여 구축하였으나 한자 처방명까지 고려한다면 확장 적응증 콘텐츠를 더욱 확장할 수 있을 것이다. 향후 지속적인 연구를 통해 기계학습 및 딥러닝 기술을 적용하여 다양한 학술정보로부터 지능적 확장 적응증 콘텐츠를 구축할 수 있을 것으로 사료된다.

**참 고 문 헌**

- [1] Vuori, Hannu V, Quality assurance of health services, Public health in Europe 16, 1982.
- [2] <https://www.law.go.kr/LSW/lsInfoP.do?efYd=20150105&lsiSeq=166777#0000>
- [3] 유병국, 유재환, “국내 한의학 학술지에 발표된 항염증 한약재 및 한약처방 연구동향-2015 년 이후 발표된 실험논문을 중심으로,” 대한본초학회지 (본초분과학회지), 제36권, 제1호, pp.19-39, 2021.
- [4] 김희숙, “데이터 마이닝을 이용한 유튜브 인기 동영상 콘텐츠 분석,” 한국디지털콘텐츠학회 논문지, 제21권, 제4호, pp.673-681, 2020.
- [5] 윤지은, 서창진, “텍스트 마이닝 기법을 활용한 인공지능과 헬스케어 융· 복합 분야 연구동향 분석,” 한국 IT 서비스학회지, 제18권, 제2호, pp.123-141, 2019.
- [6] 조경원, 우영운, “텍스트 마이닝을 이용한 4 차 산업 연구 동향 토픽 모델링,” 한국정보통신학회논문지, 제23권, 제7호, pp.764-770, 2019.
- [7] 김민정, 김철주, “텍스트마이닝을 활용한 승례문 관련 기사의 트렌드 분석,” 한국콘텐츠학회논문지, 제17호, 제3호, pp.474-485, 2017.

- [8] 박경열, 안희자, “텍스트 마이닝을 활용한 DMZ 관광 이슈의 토픽 모델링 분석,” 관광레저연구, 제31권, 제4호, pp.143-159, 2019.

- [9] <https://oasis.kiom.re.kr/>

- [10] [http://en.wikipedia.org/wiki/Web\\_crawler](http://en.wikipedia.org/wiki/Web_crawler)

- [11] <https://pubmed.ncbi.nlm.nih.gov/>

- [12] <http://cis.kiom.re.kr/diagnosis>

**저 자 소 개**

이 소 민(So-Min Lee)

정회원



- 2019년 2월 : 충북대학교 정보통신공학부(공학사)
- 2019년 9월 ~ 현재 : 충북대학교 정보통신공학과(석사)

<관심분야> : 그래프 처리, 연속 질의 처리, 빅데이터, 기계학습

백 연 희(Yeon-Hee Baek)

정회원



- 2019년 2월 : 충북대학교 경영학부 (복수전공은 빅데이터 연계전공)(학사)
- 2021년 2월 : 충북대학교 빅데이터 협동과정(석사)

<관심분야> : 소셜 네트워크, SIOV, 빅데이터 처리, 데이터베이스

송 상 호(Sang-Ho Song)

정회원



- 2020년 9월 : 충북대학교 정보통신공학과(학사)
- 2020년 9월 ~ 현재 : 충북대학교 정보통신공학 학석사 연계과정(석사)

<관심분야> : 소셜 네트워크, 빅데이터 처리, 데이터베이스, 기계 학습

RETITI DIOP EMANE CHRISTOPHER 정회원



- 2014년 9월 : African Institute of Computer Science (Libreville, GABON), Diploma of Analyst-Programmer
- 2020 9월 ~ 현재 : Chungbuk National University (Cheongju, SOUTH KOREA), Mater's Degree,

Department of Information and Communication Engineering

〈관심분야〉 : 빅데이터 처리, 데이터베이스

함 선 중(Xuan-Zhong Han) 정회원



- 2020년 2월 : 충북대학교 정보통신공학부(공학사)

〈관심분야〉 : 빅데이터, 전자공학, 통신처리

홍 성 연(Seong-Yeon Hong) 정회원



- 2015년 3월 ~ 현재 : 충북대학교 정보통신공학부(공학사)

〈관심분야〉 : 빅데이터, 전자공학, 데이터베이스

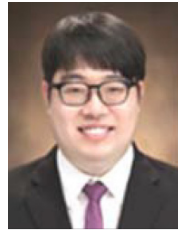
김 익 수(Ik-Su Kim) 정회원



- 2015년 3월 ~ 현재 : 충북대학교 정보통신공학부(공학사)

〈관심분야〉 : 빅데이터, 전자공학, 통신처리

임 중 태(Jong-Tea Lim) 종신회원



- 2009년 2월 : 충북대학교 정보통신학과(공학사)
- 2011년 2월 : 충북대학교 정보통신학과(공학석사)
- 2015년 8월 : 충북대학교 정보통신학과(공학박사)
- 2015년 9월 ~ 2019년 8월 : 충북대학교 정보통신학과 Postdoc.

- 2019년 10월 ~ 현재 : 충북대학교 전자정보대학 정보통신공학부 초빙 조교수

〈관심분야〉 : 소셜 미디어, 빅데이터, 시공간 데이터베이스, 위치기반 서비스 등

복 경 수(Kyoung-Soo Bok) 종신회원



- 1998년 2월 : 충북대학교 수학과(이학사)
- 2000년 2월 : 충북대학교 정보통신학과(공학석사)
- 2005년 8월 : 충북대학교 정보통신학과(공학박사)
- 2005년 3월 ~ 2008년 2월 : 한국

과학기술원 정보전자연구소 Postdoc

- 2008년 3월 ~ 2011년 2월 : 가인정보기술 연구소 차장
  - 2011년 3월 ~ 2019년 8월 : 충북대학교 전자정보대학 정보통신공학부 초빙교수
  - 2019년 9월 ~ 현재 : 원광대학교 SW 융합학과 조교수
- 〈관심분야〉 : 데이터베이스 시스템, 이동 객체 데이터베이스, 이동 P2P 네트워크, 소셜 네트워크 서비스, 빅데이터 처리 등

MINH NHAT TRAN 정회원



- 2014년 7월 : University of Medicine and Pharmacy, Hue University, (Doctor of Medicine in Traditional Medicine)
- 2019년 12월 : University of Medicine and Pharmacy, Hue University(Master in Traditional

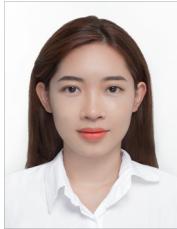
Medicine)

- 2020년 8월 ~ 현재 : 과학기술연합대학원대학교, 한의융합의학(박사)



〈관심분야〉 : 한의의료정보학, 한의데이터베이스, 한의처방 네트워크분석 등

**QUYNH HOANG NGAN NGUYEN**      **정회원**



- 2016년 8월 : Da Nang Traditional Medicine Hospital (Da Nang, VIET NAM), Traditional Medicine Doctor
- 2020년 3월 ~ 현재 : 과학기술연합대학원대학교, 한의융합의학(석박사 통합)

〈관심분야〉 : Network Pharmacology and Traditional Medicine, AI-based healthcare systems, Data-based disease prediction and intervention;

**김 소 영(Soyoung Kim)**      **정회원**



- 2010년 8월 : 북경중의약대학교 침구추나학(학사)
- 2018년 8월 : 과학기술연합대학원대학교, 한의생명과학(석사)
- 2018년 9월 ~ 현재 : 과학기술연합대학원대학교, 한의융합의학(박사)

〈관심분야〉 : 한의학, 한의임상빅데이터, 한의진료지원시스템

**김 안 나(An-Na Kim)**      **정회원**

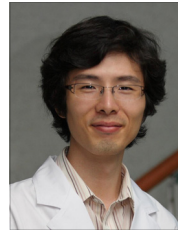


- 2009년 2월 : 우석대학교 한의학과 (한의학사)
- 2011년 8월 : 우석대학교 한의학과 (한의학석사)
- 2017년 2월 : 우석대학교 한의학과 (한의학박사)
- 2012년 4월 ~ 현재 : 한국한의학

연구원 미래의학부 선임연구원

〈관심분야〉 : 한의의료정보학, 한의진단학, 한의진료 지원시스템, 한의처방네트워크분석 등

**이 상 훈(Sang-Hun Lee)**      **정회원**



- 2003년 2월 : 원광대학교 한의학과 (한의학사)
- 2007년 2월 : 원광대학교 한의학과 (한의학석사)
- 2011년 2월 : 원광대학교 한의학과 (한의학박사)
- 2009년 5월 ~ 현재 : 한국한의학연

구원 한의약 데이터부 책임연구원

〈관심분야〉 : 한의의료정보학, 한방의료기기, 의료인공지능, 빅데이터 등

**유 재 수(Jae-Soo Yoo)**      **종신회원**



- 1995년 2월 : KAIST 전산학과(공학박사)
- 1995년 2월 ~ 1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월 ~ 현재 : 충북대학교 전자정보대학 정보통신공학부 정교수
- 2009년 2월 ~ 2010년 2월 : 캘리

포니아 대학교 방문교수

〈관심분야〉 : 데이터베이스 시스템, 멀티미디어 데이터베이스, 센서 네트워크, 바이오 인포메틱스, 빅데이터 등