Journal of Knowledge Information Technology and Systems(JKITS), Vol. 15, No. 5, pp. 721-730, October 2020
https://doi.org/10.34163/jkits.2020.15.5.015

Journal of Knowledge Information Technology and Systems

ISSN 1975-7700 (Print), ISSN 2734-0570 (Online)

http://www.kkits.or.kr

Construction of Big Data Center for Medicinal Biological Resources

Yunji Jang, Taehong Kim, Myeong-Ku Lee, Ho Jang, Sang-Jun Yea, Sang-Kyun Kim*

Intellectual Information Team, Korea Institute of Oriental Medicine

ABSTRACT

This study aims to propose a big data center to construct the information on medicinal biological resources and provide the infrastructure to open and share them. The information on the medicinal biological resources consists of basic data such as effects, treatments, and warnings of medicinal biological resources, biochemical data such as chemical compounds of medicinal materials and mappings among compounds and targets, and utilization data such as locations and times of discovery of medicinal resources. In order to ensure the data compatibility, all data was constructed in an open format. The bigdata of medicinal biological resources are opened and searched on our bigdata center (http://kmbigdata.kr) which is constructed based on CKAN. Our bigdata center provides medicinal biological resource data in CSV and RDF file format along with their metadata in DCAT format. Moreover, medicinal biological resource data are provided to the forest bigdata platform using ESB. Therefore, our data can be obtained on the forest big data exchange (http://bigdata-forest.kr). By using the bigdata center and forest big data exchange, it is expected that domain experts and the publics will be able to use the accurate and useful data on medicinal biological resources for the health care and biotechnology companies will be able to search our data to make products such as new drugs and functional foods.

KEYWORDS: Medicinal biological resources, Forest big data, Open format, Shared infrastructure, NiFi, CKAN

ARTICLE INFO: Received 6 August 2020, Revised 21 August 2020, Accepted 13 October 2020.

*Corresponding author is with the Intellectual Information Team, Korea Institute of Oriental Medicine, Korea Institute of Oriental Medicine, 1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054, Korea.
E-mail address: skkim@kiom.re.kr

1. 서론

산림 약용자원은 과거 생약이나 간단한 가공만 거친 형태로 주로 식용 또는 약용으로 이용되어 왔으나, 최근 물질 추출, 성분 연구 등을 통해 의약품, 건강기능식품, 화장품 등의 소재로 활용 가치가 확대되고 있으며 건강기능식품 기술의 특허 출원수도 꾸준히 상승하고 있다[1]. 특히 산림 자원을 이용한 의약품 개발은 고부가가치 품목으로 미래 4차 산업혁명의 주요 분야로 부각되어, IoT와 AI, 빅데이터 등 4차 산업혁명의 핵심 기술을 임업 분야에 접목시켜 기술 집약 산업으로 발전시키기 위한 연구가 각계에서 진행 중이다[2].

삶의 질에 대한 인식 변화와 인구 초고령화 시대가 도래함에 따라 건강한 삶에 대한 국민들의 관심이 나날이 증가하고 있다. 건강한 노후에 대한 욕구 또한 상승하여 약용 생물자원을 이용한 약선 음식이나 민간요법 등 다양한 정보를 누구나 쉽게 찾아볼 수 있게 되었다. 그러나 국가 차원에서 일반인에게 제공되는 자료는 아직 부족한 상태로[3], 정확한 정보와 함께 검증되지 않은 정보들도 무분별하게 혼재되어 각종 미디어 매체를 통해 노출되고 있다[4]. 생물유전자원의 접근 및 이익 공유(ABS, Access to Genetic Resources and Benefit-sharing)에 관한 나고야의정서[5] 발효에 따라 해외 생물자원의 이용이 많은 우리나라에서는 약용 생물자원의 사용 로열티 상승에 따른 수급의 불안정, 해당 국가의 사전 승인 필수 조건으로 인한 연구 개발 지연 등 수입 약용작물에 대한 대응 방안 마련이 시급한 상황이다[6]. 약용 생물자원의 임의 복용 혹은 오용을 방지하기 위해서는 올바르고 정확한 효능 및 복용 정보를 구축하여 공신력 있는 건강 정보 서비스를 제공해야 한다.

이에 본 연구에서는 약용 생물자원에 대한 정보를 수집 및 가공하여 빅데이터를 구축하고, 이를 개방하기 위한 공유 인프라 마련을 통해 4차 산업 서비스에서 활용 가능한 빅데이터 플랫폼과의 연계 방안을 제안하였다.

본 논문의 구성은 다음과 같다. 제 2장에서는 약용 생물자원의 데이터를 다루는 관련 연구의 특징을 살펴본다. 제 3장에서는 본 연구에서 구축한 데이터의 수집과 구축 방법에 대해 설명하고, 제 4장에서는 데이터 개방을 위한 플랫폼의 구축에 대해 설명한다. 제 5장에서는 본 연구가 시사하는 방향과 한계점에 대해 고찰하고 총 결론을 기술한다.

2. 관련 연구

약재의 처방, 병증과 같은 전통의학 관련 정보를 제공하는 한국전통지식포탈[7]은 대부분 전문 용어로 되어 있어 비전문가의 접근성이 떨어지며, 약재의 구성 성분에 최신 연구 정보는 반영되어 있지 않다. 온톨로지를 기반으로 한의학 교과서의 수록 정보 간 연계를 보여주는 한의시맨틱검색[8]은 한의학 지식을 체계화하여 제공하지만 모든 용어가 한자로 되어있어 마찬가지로 일반인이 접근하기에 쉽지 않다. PubMed로부터 한국·중국·일본 천연물 약재들에 대한 최신 구성성분 정보를 구축한 TM-MC[9]는 약재의 성분에 대한 상세 정보가 다소 부족하다.

3. 데이터베이스 구축

약용 생물자원 빅데이터 구축을 위해 총 3개년에 걸친 약용 생물자원 데이터 생산 계획을 수립하였다. 총 33,896,700건(107,818MB) 구축을 목표로 하며 연차별 생산 계획 항목과 세부 내용은 <표 1>과 같다. 특히 1차년도인 2019년에 생산한 총 20종의 데이터 항목들에 대한 설명은 <표 2>에 기술하였다.

표 1. 연도별 데이터 구축 계획
Table 1. Planning of data to be produced by year

데이터셋	데이터명	2019년	2020년	2021년
약용 생물자원 기본 데이터	약용 생물자원 목록	목록 500여건 정제	변경시 업데이트	변경시 업데이트
	약용 생물자원 사진	기존 데이터 정제 (500건)	신규 사진 추가 (1,200건)	신규 사진 추가 (1,000건)
	약용 생물자원 효능 정보	기구축 교과서 데이터 정제 (1,200건)	번역 및 가공 (1,200건)	수집, 번역, 가공 (600건)
	약용 생물자원 치료 정보	기구축 교과서 데이터 정제 (2,800건)	번역 및 가공 (2,800건)	수집, 번역, 가공 (1,000건)
	약용 생물자원 주의사항	데이터 수집 및 정제 (50건)	데이터 수집 및 정제 (150건)	데이터 수집 및 정제 (150건)
	약용 생물자원 온톨로지	기구축 온톨로지 정제 (50,000 트리플)	온톨로지 업데이트 (50,000 트리플)	온톨로지 업데이트 (50,000 트리플)
	고문헌 약용 생물자원 목록	식재료 목록 추출 및 정제 (250건)	변경시 업데이트	변경시 업데이트
	고문헌 약용 생물자원 효능 정보	-	데이터 추출, 정제, 번역, 가공 (300건)	데이터 추출, 정제, 번역, 가공 (300건)
	고문헌 약용 생물자원 치료 정보	-	데이터 추출, 정제, 번역, 가공 (300건)	데이터 추출, 정제, 번역, 가공 (300건)
바이오 연계 데이터	약용 생물자원 구성 성분 목록	기구축 구성성분 목록 정제 (20,000건)	신규 성분 목록 추출 및 정제 (500건)	신규 성분 목록 추출 및 정제 (500건)
	구성 성분 구조식 그림	기구축 구성성분 구조식 정제 (500건)	구성성분 구조식 드로잉 (12,000건)	구성성분 구조식 드로잉 (8,500건)
	구성 성분 동의어 정보	동의어 정제 (2,000,000건)	동의어 정제 (5,000,000건)	동의어 정제 (5,000,000건)
	구성 성분 구조 데이터	SDF 파일 생성 및 정제 (500건)	SDF 파일 생성 및 정제 (12,000건)	SDF 파일 생성 및 정제 (8,500건)
	구성 성분 SMILES 정보	SMILES 생성 및 정제 (500건)	SMILES 생성 및 정제 (12,000건)	SMILES 생성 및 정제 (8,500건)
	구성 성분 PubChem 매핑	기구축 매핑 정보 정제 (12,000건)	신규 성분 매핑 정보 구축 (100건)	신규 성분 매핑 정보 구축 (100건)
	구성 성분 물리 화학 특성	구성성분 특성분석 (3,500건)	구성성분 특성 분석 (84,000건)	구성성분 특성 분석 (59,500건)
	구성 성분 DL (Drug-likeness)	DL 정보 생성 및 정제 (500건)	DL 정보 생성 및 정제 (12,000건)	DL 정보 생성 및 정제 (8,500건)
	구성 성분 OB (Oral-bioavability)	OB 정보 생성 및 정제 (500건)	OB 정보 생성 및 정제 (12,000건)	OB 정보 생성 및 정제 (8,500건)
	단백질 유전체 목록	단백질 유전체 목록 정제 (500건)	단백질 유전체 목록 정제 (3,000,000건)	단백질 유전체 목록 정제 (3,000,000건)
	성분-타겟 매핑 정보	성분-타겟 매핑 정보 정제 (3,000건)	성분-타겟 매핑 정보 정제 (7,500,000건)	성분-타겟 매핑 정보 정제 (7,500,000건)
약용 생물자원 활용 데이터	산림 약용 생물자원 분포 위치 정보	기구축 위치정보 정제 (2,000건)	위치정보 수집 및 정제 (5,000건)	위치정보 수집 및 정제 (4,000건)
	산림 약용 생물자원 발견 시기 정보	기구축 시기정보 정제 (2,000건)	시기정보 수집 및 정제 (5,000건)	시기정보 수집 및 정제 (4,000건)
	약용 생물자원 생산 가공 정보	-	생산 가공 정보 정제 (200건)	생산 가공 정보 정제 (200건)
	약용 생물자원 가격 정보	-	소비자 가격 정보 수집 (210,000건)	소비자 가격 정보 수집 (210,000건)
	웰빙 푸드 레시피	-	레시피 정보 수집, 가공, 구축 (500건)	레시피 정보 수집, 가공, 구축 (500건)
데이터 (건)		2,100,800건	15,921,250건	15,874,650건

대한민국약전[10]과 대한민국약전외한약(생약)규격집[11]의 수록 약재 중 전문가 자문을 통해 총 541건의 한의 약용 생물자원 목록을 확보하였다. 고문헌 약용 생물자원 정보는 동의보감과 향약집성방에 나오는 약용 생물자원 250건의 목록을 수집하였다. 약용 생물자원 구성성분 정보는 TM-MC(a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine)[9]에 수록된 610건의 구성성분 중 한국약전에 있는 약재 429건에 대해 구축하였다. 단백질 유전체 정보는 STITCH(Search tool for interacting of chemical)[12]데이터베이스에 존재하

표 2. 2019년 생산 데이터
Table 2. Data produced in 2019

데이터셋	데이터명	데이터 설명
약용 생물자원 기본 데이터	약용 생물자원 목록	국내 식약처에서 고시한 약용 생물자원 목록으로 천연물이 아닌 약용으로 활용할 수 있는 목록
	약용 생물자원 사진	약용 생물자원들을 찍은 사진 및 약용으로 판매하는 형태의 사진
	약용 생물자원 효능 정보	한의학 교과서에 명시된 약용 생물자원의 전문 효능 정보와 이를 쉬운 용어로 번역 및 가공한 정보
	약용 생물자원 치료 정보	한의학 교과서에 명시된 약용 생물자원의 전문 치료정보와 이를 쉬운 용어로 번역 및 가공한 정보
	약용 생물자원 주의사항	약용 생물자원 복용시 주의할 사항들을 쉬운 용어로 번역 및 가공한 정보
	약용 생물자원 온톨로지	한의학 교과서 기반으로 구축된 한의 약용 생물자원 온톨로지 데이터
	고문헌 약용 생물자원 목록	고문헌에서 언급된 약용 생물자원 식재료들에 대한 목록을 정리한 데이터
바이오 연계 데이터	약용 생물자원 구성 성분 목록	PubMed 데이터베이스의 연구 논문에서 추출한 약용 생물자원에 대한 구성성분 목록
	구성 성분 구조식 그림	약용 생물자원 구성 성분의 2D 구조식 그림
	구성 성분 동의어 정보	약용 생물자원 구성 성분에 대한 동의어 리스트 정보
	구성 성분 구조 데이터	약용 생물자원 구성 성분에 대한 SDF (Structure data file) 포맷의 데이터 파일
	구성 성분 SMILES 정보	약용 생물자원 구성 성분에 대한 SMILES 문자열 식별자 정보
	구성 성분 PubChem 매핑	약용 생물자원 구성 성분과 PubChem 화합물 데이터베이스와의 연계 정보
	구성 성분 물리 화학 특성	구성 성분의 분자량, 옥탄올-물 분배 계수, 극성표면적값, 회전연결개수 등의 물리화학적 특성 정보
	구성 성분 DL (Drug-likeness)	약용 생물자원 구성 성분의 기존 허가된 약물들과의 유사한 정도 값
	구성 성분 OB (Oral-bioavability)	약용 생물자원 구성 성분의 구강 섭취 후 신체 내 흡수 정도 값
	단백질 유전체 목록	약용 생물자원 구성 성분과 연관된 단백질 및 유전체들의 목록
	성분-타겟 매핑 정보	약용 생물자원 구성 성분과 타겟 단백질 및 유전체와의 매핑 정보
약용 생물자원 활용 데이터	산림 약용 생물자원 분포 위치 정보	자생하는 약용 생물자원에 대한 분포 위치 및 지역 정보
	산림 약용 생물자원 발견 시기 정보	자생하는 약용 생물자원의 위치를 기록한 시기 관련 정보

는 약재의 구성성분과 단백질간의 상호 작용 데이터를 활용하여 구축하였다. 산림 약용 생물자원 분포 위치 정보는 식물의 분류와 감별 전문가들이 직접 전국을 탐사하여 채집한 생물자원 정보를 기록하는 모바일 야장 시스템[13]의 누적 데이터를 이용하여 구축하였다.

4. 시스템 구현

4.1. 하둡 테이블 설계

<그림 1>은 2019년에 구축한 약용 생물자원 데이터에 대한 하둡 시스템의 데이터베이스 테이블 스키마이다. 약용 생물자원 기본 데이터인 TB_MM_LIST, TB_MM_ FIGURE, TB_MM_EFFECT, TB_MM_TREAT, TB_MM_WARN, TB_MM_CLASSIC_LIST 테이블과 바이오 연계 데이터에 대한 TB_MM_CC_LIST, TB_MM_CC_IMAGE, TB_MM_CC_SYNONYM, TB_MM_CC_SDF, TB_MM_CC_SMILES, TB_MM_CC_PUBCHEM, TB_MM_CC_PROPS, TB_MM_CC_DL, TB_MM_CC_OB, TB_PN_LIST, TB_CC_PN 테이블 그리고 약용 생물자원 활용 데이터에 대한 TB_PT_LOCN, TB_ PT_TIME 테이블로 구성되어있다.

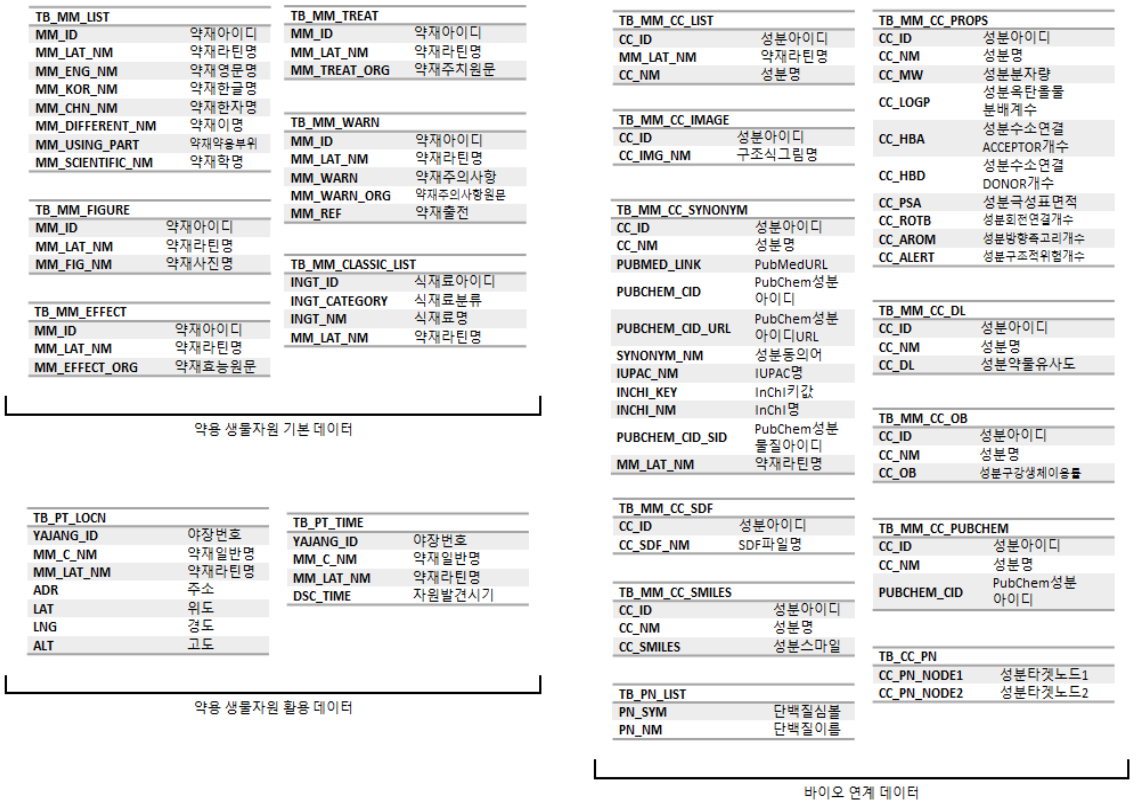


그림 1. 테이블 스키마
Figure 1. Table schema diagram

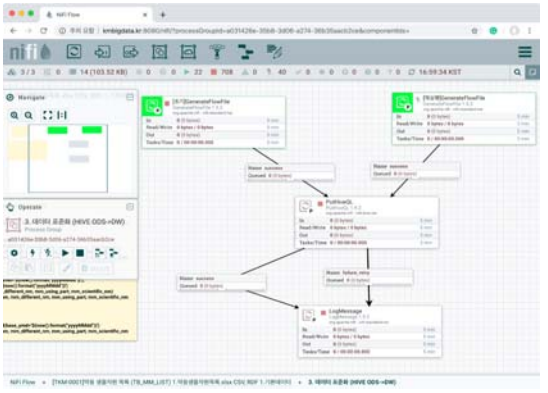


그림 2. NiFi 워크플로우 샘플
Figure 2. Example of NiFi workflow

4.2. 데이터 공유 체계 구축

Apache NiFi는 실시간으로 시스템 간 대량의 데이터를 효율적으로 처리하고 모니터링하기 위한 소프트웨어로서[14] 직관적인 UI로 시각화된 설계가 가능하다. <그림 2>는 약용 생물자원 목록 엑셀 파일을 읽어들여 Hive에 데이터를 적재하고 로그 메시지를 출력하는 NiFi 워크플로우이다.

하둡의 데이터는 오픈 데이터 유통을 위한 데이터 개방 표준 플랫폼인 CKAN[15]을 통해 빅데이터 센터[16]를 구축하여 개방하였다. <그림 3>과 같이 데이터의 상세 정보 조회와 국제표준기반 데이터 카탈로그 형식인 DCAT과 CSV, RDF 형태로 다운로드가 가능하다. <그림 4>는 그 중 RDF 포맷으로 약용 생물자원 목록을 다운로드

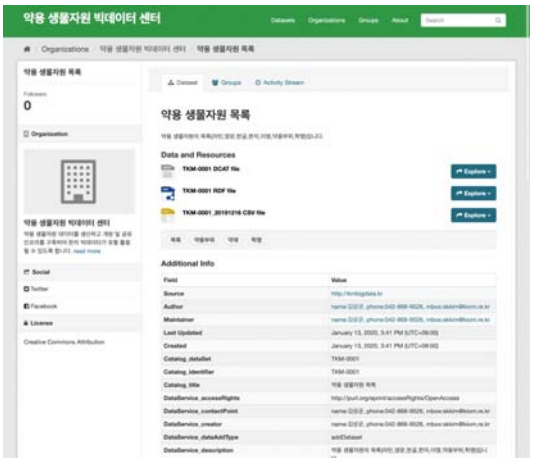


그림 3. 약용 생물자원 빅데이터 센터의 데이터 카탈로그 예제
Figure 3. Example of a data catalog in big data center of medicinal biological resources

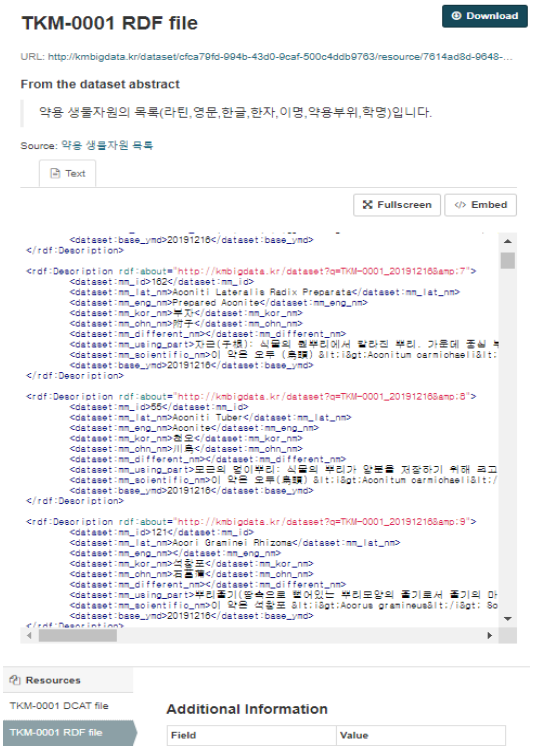


그림 4. 데이터셋 다운로드 화면
Figure 4. Example of downloading datasets

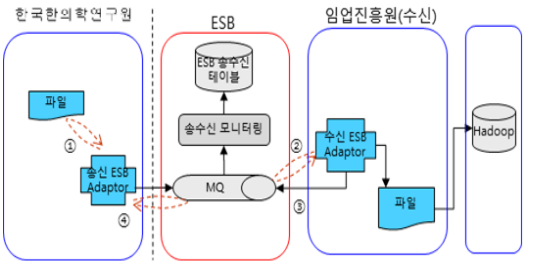


그림 5. ESB 에이전트 연계 구성도
Figure 5. Connection diagram of ESB agent

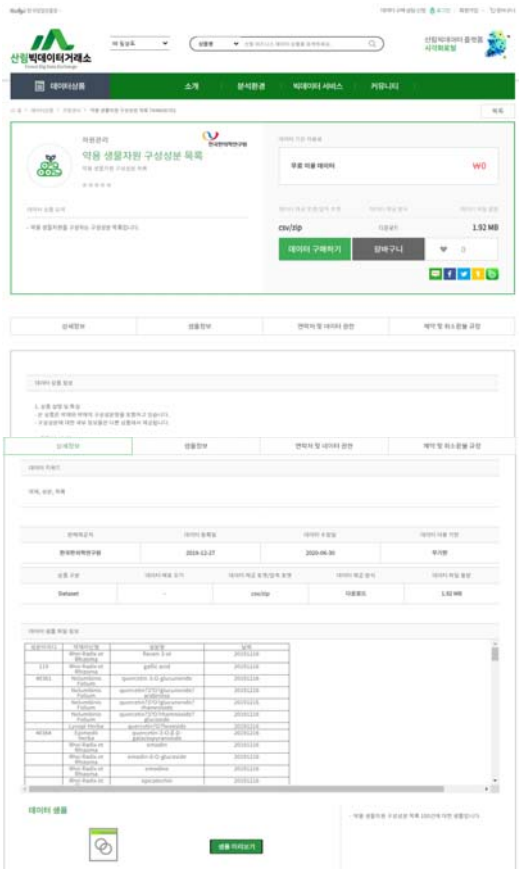


그림 6. 산림 빅데이터 거래소
Figure 6. Example of forest big data exchange

하기 위해 데이터셋을 열람한 화면이다.

또한 ㈜메타빌드의 ESB(Enterprise Service Bus) 에이전트를 통해 한국임업진흥원의 산림 빅데이터 플랫폼과 데이터를 연계하였다<그림 5>. ESB 에이전트는 표준 규약에 맞춰 공급 내역보고 파일을 생성하면 센터와 정보를 자동으로 송수신해주는 프로그램이다[17]. 산림 빅데이터 플랫폼 데이터는 산림 빅데이터 거래소에서 오픈되는데 해당 거래소는 <그림 6>과 같이 서비스 개발이나 데이터 분석 등 관련 사업에 필요한 다양한 포맷의 산림 데이터셋을 유료 또는 무료로 쉽게 판매하거나 구매할 수 있다[18,19]. 본 센터에서 구축한 데이터는 CKAN 기반의 약용 생물자원 빅데이터 센터와 산림 빅데이터 거래소를 통해서 검색 및 다운로드할 수 있다. 단, 약용 생물자원 빅데이터 센터는 본 연구에서 구축한 약용 생물자원 빅데이터만 검색이 가능하지만 산림 빅데이터 거래소에서는 산림 플랫폼의 여러 타 센터들의 데이터들을 ESB 에이전트로 불러와서 함께 제공하고 있다.

3. 고찰 및 결론

산림 약용자원은 과거와 달리 식용 뿐 아니라 의약품과 건강기능식품, 화장품 등의 소재로 활용되며 4차 산업혁명의 핵심 기술과 접목되어 기술집약 산업 요인으로 각광받고 있다. 삶의 질에 대한 인식이 변화하고 인구 초고령화 시대가 도래하여 이제는 많은 이들이 건강한 삶에 관심을 갖게 되고 누구나 다양한 정보를 쉽게 찾아볼 수 있게 되었다.

나고야의정서 발효에 따라 해외 생물자원의 이용이 많은 우리나라는 약용 생물자원의 로열티 부담과 수급 지연 등으로 인해 해외 작물에 대한 대응책 마련이 속히 요구되는 상황이다.

이에 본 연구에서 개방하는 데이터셋은 일반인들도 이해하고 누구나 활용할 수 있는 수준으로 번역 및 가공하여 구축하였다. 비전문가들이 접근하기 어려웠던 기존의 약재 정보 서비스들의 약점을 보완한 것이라 판단되며 이로 인해 약재의 혼·오용이나 부적절한 민간요법의 사용 방지가 가능할 것이라 여겨진다.

또한 천연물 의약품은 안정성과 그에 따른 규제 때문에 상업화 과정에 어려움이 존재해왔는데, 천연물 신약 개발 과정에서 약용 생물자원의 구성성분 정보, 성분의 유전체 및 단백질 타겟들 간 연계를 통해 약용 생물자원에 대한 과학적인 효능 정

보를 사전에 탐색할 수 있도록 하여 각종 업계의 제품 개발에도 활용이 가능하며, 여러 미디어 매체를 통해 노출되는 산업 제품의 과장 광고로 인한 소비자들의 피해도 줄일 수 있다.

바이오와 의료 빅데이터 플랫폼뿐만 아니라 기존의 Linked Data 및 국내외 데이터베이스들과의 연계를 위해, 본 연구에서 완성한 플랫폼은 향후 LOD 포맷과 Json, xml 포맷 오픈 지원까지를 최종 목표로 하고 있다.

다만 데이터가 연속성을 띄기 위해서는 빅데이터의 가공과 정제가 가능한 전문 인력과 천연 산림자원 및 약용 생물자원 등 도메인 전문가의 지원이 지속되어야 한다는 제약이 있기에 데이터의 식별과 정제, 번역, 매핑 등 생산 프로세스의 안정적인 정립이 필요하다. 또한 플랫폼이 건재하게 실재하기 위해서는 빅데이터셋과 플랫폼을 꾸준히 활용하는 연구 생태계가 조성될 수 있도록 적극적인 홍보를 통해 다양한 수요자들을 발굴해야 할 것이다.

제안한 플랫폼을 이용하여 약용 생물자원의 정보 공유와 활용이 가능할 것이며 나고야의정서 발효 대비를 위한 국내 약재 데이터베이스를 확보함으로써 국가 경제 이익에 이바지하고 나아가 국민건강 증진에 공헌할 수 있을 것으로 기대된다.

References

[1] J-H. Song, S-G. Yang, G-Y. Choi, B-C. Moon, *Analysis on the trends of Korean health functional food patent based on the medicinal plant resources*, Korean Herbal Medicine Informatics, Vol. 8, No. 1, pp. 25-44, 2020.

[2] H-D. Seok, *Application and development direction of fourth industrial revolution technology in forestry sector*, Korea Rural Economic Institute, 2018.

[3] Ministry of Health and Welfare, *2017 survey on getting oriental medical treatment and herbal medicine consumption*, National Development Institute of Korean Medicine, 2017.

[4] Y-J. Jang, S-J. Yea, B-S. Seong, and C. Kim, *The implementation of medicinal herbs information system*, Journal of Knowledge Information Technology and Systems, Vol. 12, No. 6, pp. 827-836, 2017.

[5] National Institute of Biological Resources, Nagoya Protocol. 2010.

[6] G-M. Lee, D-U. Kim, and Y-M. Kang, *New insight on development plan of herbal medicine resources production through the network of medicinal resources-related institutes in Korea*, Korean Herbal Medicine Informatics, Vol. 5, No. 1, pp. 19-26, 2018.

[7] Korean Traditional Knowledge Portal, https://www.koreantk.com, Aug. 2020.

[8] S-K. Kim, D-H. Park, A-N. Kim, Y-T. Oh, J-Y. Kim, S-J. Yea, C. Kim, and H-C. Jang, *Semantic search system based on Korean medicine ontology*, The Journal of the Korea Contents Association, Vol. 12, No. 12, pp. 533-543, 2012.

[9] S-K. Kim, S-J. Nam, H-C. Jang, A-N. Kim, and J-J. Lee, *TM-MC: a database of medicinal materials and chemical compounds in Northeast Asian traditional medicine*, BMC COMPLEMENTARY AND ALTERNATIVE MEDICINE, Vol. 15, No. 218, pp. 2-8, 2015.

[10] Korea Ministry of Government Legislation, http://www.law.go.kr/%ED%96%89%EC%A0%95%EA%B7%9C%EC%B9%99/%EB

%8C%80%ED%95%9C%EB%AF%BC%EA%B5%AD%EC%95%BD%EC%A0%84/(2019-102,20191106), Aug. 2020.

[11] Korea Ministry of Government Legislation, http://www.law.go.kr/admRulLsInfoP.do?admRulSeq=2100000187066, Aug. 2020.

[12] STITCH, http://stitch.embl.de, Aug. 2020.

[13] Y-J. Jang, B-S. Seong, and Chul. Kim, *Implementation of mobile app for supporting*, Korean Herbal Medicine Informatics, Vol. 6, No. 2, pp. 299-310, 2018.

[14] S-H. Park, and Y-G. Kim, *Application of open data framework for real-time data processing*, Journal of the Korea Institute of Information and Communication Engineering, Vol. 23, No. 10, pp. 1179-1187, 2019.

[15] K-S. Park, T-Y. Kim, N-R. AMIN, H-J. Seo, H-W. Kim, H-S. Won, and Y-K. Lee, *Design and implementation of framework based on DCAT-AP for CKAN*, Database Research, Vol. 33, No. 3, pp. 40-51, 2017,

[16] Big data center of medicinal biological resources, http://kmbigdata.kr, Aug. 2020.

[17] IBM Cloud, https://www.ibm.com/cloud/learn/esb?mhsrc=ibmsearch_a&mhq=esb, Aug. 2020.

[18] The environmental newspaper, http://www.hkbs.co.kr/news/articleView.html?idxno=554306, Aug. 2020.

[19] Forest Big Data Exchange, https://www.bigdata-forest.kr/service/aboutus, Aug. 2020.

장윤지[1], 김태홍[2], 이명구[1], 장호[2], 예상준[3], 김상균[3]

[1] *한국한의학연구원 지능화추진팀 기술연구원*
[2] *한국한의학연구원 지능화추진팀 선임연구원*
[3] *한국한의학연구원 지능화추진팀 책임연구원*

요 약

본 연구에서는 약용 생물자원 데이터를 생산하고 개방 및 공유 인프라를 제공하는 약용 생물자원 빅데이터 센터를 구축하였다. 약용 생물자원 데이터는 약용 생물자원의 효능, 치료, 주의사항 등의 기본 데이터, 약재의 성분과 성분 타겟 매핑 등의 바이오 데이터, 약용 생물자원의 발견 위치와 시간 등의 활용 데이터들로 구성되어 있다. 데이터의 호환성을 확보하기 위해 모든 데이터는 오픈 포맷으로 구축하였다. 구축한 약용 생물자원 빅데이터는 CKAN 기반의 빅데이터 센터(http:// kmbigdata.kr)에 공개하여 검색되고 있다. 빅데이터 센터에서는 약용 생물자원 데이터를 DCAT 포맷의 메타데이터와 같이 CSV와 RDF 파일 포맷으로 제공한다. 또한 약용 생물자원 데이터는 ESB를 이용하여 산림 빅데이터 플랫폼에 제공하고 있다. 따라서 약용 생물자원 데이터는 산림 빅데이터 거래소(http://bigdata-forest. kr)에서도 얻을 수 있다. 본 연구에서 구축한 빅데이터 센터와 산림 빅데이터 거래소를 이용함으로써 도메인 전문가와 일반인들은 건강 관리를 위한 정확하고 유용한 약용 생물자원 데이터를 얻을 수 있으며, 바이오 분야 업체들은 신약이나 기능성 식품을 개발하는데 활용할 수 있을 것으로 기대한다.

감사의 글

본 연구는 한국한의학연구원 'AI 한의사 개발을 위한 임상 빅데이터 수집 및 서비스 플랫폼 구축(KSN2012110)'과 한국정보화진흥원의 '산림빅데이터 플랫폼 및 센터 구축(ESN1912060)' 과제의 지원을 받아 수행되었습니다.

Yunji Jang received the Master degree in the Department of Computer Engineering from the Hanbat National University. Since 2010, she is a senior researcher at Korea Institute of Oriental Medicine. Her current research interests Korean medicine information and Fuzzy system.

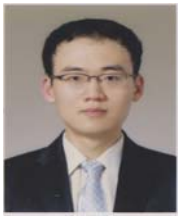
E-mail address: jangbing@kiom.re.kr

Taehong Kim He received his M.S. degree and his Ph.D. degree from the Department of Applied Information Science at University of Science and Technology in 2010 and 2014, respectively. After graduation, he has been researched on Semantic Web, Big Data, and IoT in Korea Institute of Science and Technology Information until 2018. After that, he has been expanded his research area to Korean medicine informatics in Korea Institute of Oriental Medicine.

E-mail address: thkim@kiom.re.kr

Myung-ku Lee received the bachelor's degree in the Department of Oriental Pharmacy from Wonkwang University since 2011. After the graduation, he has been researched on the Korean medicine informatics in the Korea Institute of Oriental Medicine.

E-mail address: 2mj9@kiom.re.kr



Ho Jang received his Ph.D. degree from School of Electrical Engineering and Computer Science at Gwangju Institute of Science and Technology in 2017. He has been working for Korea Institute of Oriental Medicine since 2018. His research interests include computational biology and machine learning.

E-mail address: jh@kiom.re.kr

Sang-Jun Yea received the Ph. D in knowledge service engineering from KAIST, Korea in 2018. Since 2008, he is a principal researcher at Korea Institute of Oriental Medicine. The current research interests is biomedical data science.

E-mail address: tomita@kiom.re.kr

Sang-Kyun Kim received the M.S. degree and the Ph.D. degree in the Department of Computer Engineering from Chungnam National University in 2001 and 2008, respectively. After the graduation, he has been researched on the Korean medicine informatics in the Korea Institute of Oriental Medicine.

E-mail address: skkim@kiom.re.kr