

# 한의학 분야 문헌 분석을 통한 생물학적 네트워크 분석시스템 개발

## Implementing Biological Network Analysis System through Oriental Medical Literature Analysis

유석종\*, 조용성\*, 이준학\*, 서동민\*, 예상준\*\*, 김철\*\*  
한국과학기술정보연구원 생명의료융합기술연구실\*, 한국한의학연구원 K-herb연구단\*\*

Seok Jong Yu(codegen@kisti.re.kr)\*, Yongseong Cho(yscho@kisti.re.kr)\*,  
Junehawk Lee(juneh@kisti.re.kr)\*, Dongmin Seo(dmseo@kisti.re.kr)\*,  
Sang-Jun Yea(tomita@kiom.re.kr)\*\*, Chul Kim(chulnice@kiom.re.kr)\*\*

### 요약

최근 한의학에 대한 과학적 접근이 진행되면서 한약재 성분의 효능을 검증하고자 하는 다양한 분자 생물학 분야의 연구가 진행되고 있다. 하지만 관련 한약재의 주요 성분과 관련된 생화학적 기작을 손쉽게 검색할 수 있는 시스템이 갖추어져 있지 못한 실정이다. 본 연구는 국내 한약재에 대한 약효 성분과 생물학적 기작에 대한 정보를 수집 및 텍스트마이닝을 수행하여 한약재 정보 데이터베이스를 구축하고자 하였다. 연구자가 손쉽게 분석된 한약재의 화합물, 유전자 그리고 생물학적 상호작용 정보를 검색할 수 있는 웹사이트 원형을 개발하였다. 문헌 분석결과 한의학분야 주요 화합물 및 유전자/단백질 정보를 추출할 수 있었고 현대 한의학 연구 현황의 특징을 보여주었다. 분석된 결과는 웹을 통해 한약재별 PubMed 문헌 정보와 관련된 한약재의 약재 정보 및 생물학적 상호작용 정보를 가시화하여 볼 수 있도록 개발하였다.

■ 중심어 : | 한의학 | 생물학적 네트워크 | 네트워크 분석 | PubMed |

### Abstract

Currently, oriental medicine research is focused with modern research technology and validate it's various biochemical effect by combining with molecular biology technology. But there are few searching system for finding biochemical mechanism which is related to major compounds in oriental medicine. In this research, we aimed developing korean herb database based on text-mining system by analyzing PubMed data. We have developed prototype system for searching chemical, gene and biological relation in oriental medicine. It is characterized by modern oriental medicine research trend with major chemical, gene and protein information. Analysis results can be searched on the prototype system with visualization of the biological interactions.

■ keyword : | Oriental Medicine | Biological Network | Network Analysis | PubMed |

\* 본 연구는한의학연구원의 「ICT기반 한약자원 발굴시스템 구축 (K15433)」 과 한국과학기술정보연구원의 「초고성능컴퓨팅기반 건강한 고령사회 대응 빅데이터 분석기술개발 (K-15-L03-C02-S02)」 사업으로부터 지원을 받아 수행된 연구임.

접수일자 : 2015년 08월 12일

심사완료일 : 2015년 08월 20일

수정일자 : 2015년 08월 20일

교신저자 : 김철, e-mail : chulnice@kiom.re.kr

## I. 서론

한의학에서 다루는 한약재는 다양한 천연물을 기반으로 천연물내의 다양한 화합물의 약효지식 축적을 해 온 학문 분야이다. 최근 한약재를 대상으로 현대의 분자 생물학 기술을 응용한 연구가 활발히 진행되고 있는데, 대표적인 사례로 자가 면역을 억제할 수 있는 물질을 상산(常山, *Dichroa febrifuga*) 나무에서 추출하여 보고한 바 있다[1]. 특히 오믹스 연구를 통해 다양한 세포 내의 생명 현상들이 밝혀지면, 이러한 지식을 기반으로 한약재에 존재하는 다양한 약제의 효능을 현대 과학적으로 검증하고 고부가가치의 신약으로 개발할 수 있는 기회가 열릴 것으로 전망된다.

한의학의 정보화 기술도 온톨로지 기술을 중심으로 다양한 한약재 간의 정보를 체계적으로 관리할 수 있는 기술이 활발히 개발되었다[2][3]. 하지만 한약재의 성분과 이에 대한 생체 내의 신호전달 기작과의 연관성을 파악하는 데는 상당한 어려움이 있는 실정이다. 국제적으로 생명과학 분야 연구 논문정보를 제공하고 있는 PubMed의 경우 한 해 등록되는 논문 수가 지속적으로 증가하여 2014년도에는 1,184,604개의 논문이 보고되고 있다. 이렇게 빠르게 증가하고 있는 논문 정보를 연구자가 모두 읽어서 관련 정보를 추출하는 일은 점점 어려운 일이 되어가고 있다. 최근 그 대안으로 주목받고 있는 기술이 텍스트마이닝이다[4].

이렇게 한약재 약효 성분에 대한 생물학적 네트워크 정보를 추출하기 위한 방법으로써 텍스트마이닝 기술이 부각되고 있는 이유는 기존 연구를 통해 얻어진 지식을 대부분 문헌으로 보고되며, 문헌 분석이 정보 추출을 위한 최상의 원본 소스가 되기 때문이다. 정보기술의 발달로 자연어 처리의 기술의 정확성이 증가하고 이를 기반으로 텍스트마이닝을 통해 생체 내의 개체명 인식과 개체 간의 상호작용 정보를 추출할 수 있는 국제적인 문헌 분석 대회(BioCreative)가 개최되고 있다[5].

텍스트마이닝을 통해 한약 성분과 관련된 생물학적 네트워크 간의 연계가 가능하다면 한약 성분의 효능을 신호 전달 네트워크의 분석을 통해 예측할 수 있다. 네

트워크 약물학은 시스템 생물학을 기반으로 한약 성분의 작용을 이해하고자 하는 시도로써 시스템 생물학의 발전을 통해 세포 내의 다양한 단백질 간의 관계가 단백질 네트워크로 구성될 수 있으며, 한약 성분과 단백질 간의 상호작용 정보를 수집한다면 한약재의 특징을 신호 전달 관점에서 파악할 수 있을 것이다[6].

현재 한약 성분과 단백질 간의 상호작용을 추출하는 텍스트마이닝 연구는 미비한 상황으로 신뢰성이 높은 한약 성분과 단백질의 상호작용에 대한 정보 추출 기술의 연구가 필요한 상황이며, 이를 기반으로 한약재와 세포 내 단백질 간의 상호작용 네트워크를 구축하여 검색할 수 있는 검색 시스템의 확보가 한의학 연구를 체계적으로 진행하는데 중요한 상황이다. 이를 확장하여 신호 전달과 질병과의 관계를 기반으로 한약 성분과 질병과의 관계성도 추출할 수 있다. 또한 분석 결과를 보다 직관적으로 파악할 수 있도록 분석 결과를 가시화할 수 있는 웹 시스템을 개발하고자 한다. 기존 텍스트마이닝 시스템은 대부분 분석된 개체명을 문헌에 표시해주는 수준의 가시화만 제공하고 있는 상황이어서 개별적인 문헌 분석 결과를 읽고 분석 결과 간의 관계를 연구자가 파악해야 하는 문제점이 극복할 수 있을 것이다.

## II. 관련연구

시스템 생물학 연구를 통해 개별적인 유전자 및 단백질 연구에서 세포 내의 생물학적 시스템에 대한 연구가 확대되면서 단백질-단백질 간의 상호작용 그리고 유전자와 질병 간의 연관성 정보를 확보하기 위한 초기 연구가 진행되었다[7][8]. 또한, 단백질/유전자 등 생물학적 작용을 수행하는 주요 물질에 대한 개체명 인식 기술에 대한 연구를 통해 문헌에서 기술하고 있는 다양한 형태의 생체 내 물질을 추출하고자 하는 연구가 진행되었다. 개체명 인식은 미국 국립보건원(NIH)에서 제공하고 있는 PubMed 데이터베이스에 수록된 전 세계 생물학 분야 연구논문의 초록 정보를 통해 분석이 수행되었다. ABNER[9]와 BANNER[10]는 기계학습을 통해

문장 내에 생물학 용어 개체를 추출하는 도구로 사용되고 있으며, 단백질, 유전체, 세포명 등을 주로 추출하도록 개발되었다. 한의학 분야 약재 정보와 같은 화합물 정보는 별도의 개체명 인식도구들이 개발되어졌는데, 대표적으로 OSCAR4[11] 도구가 사용되어지고 있다. 특히 문헌으로부터 텍스트마닝 기법을 이용한 단백질-단백질 간 상호작용에 대한 연구를 통해 PubMed 정보를 활용한 단백질 네트워크의 구축을 시도하였다[12][13].

한의학 분야 문헌의 텍스트마닝 기술은 2000년대 중반 이후 연구가 본격적으로 시작되었다. 2007년 Zhou 등은 한의학 분야의 증상과 유전자 간의 관계를 bubble-bostrapping 방법을 통해 증상과 질병 간의 관계와 질병과 유전자 간의 관계를 추출하여 통합함으로써 약 20만 개의 증상-유전자 간의 관계를 구축하였다[14]. 또 다른 유사 연구로 한의학과 유전자, 질병, 한의학 효과와 한의학 약재 간의 연관 정보를 문헌 분석으로 통해 TCMGenDIT라는 데이터베이스를 구축하였는데, 한의학 약재의 치료 기작을 이해하기 위한 유전자 관계 및 한의학 정보를 같이 제공하고 있다[15].

텍스트마닝을 연구자들이 손쉽게 사용할 수 있는 분석 도구의 개발에 관해서는 PathText가 개발되어졌는데, MEDIE, KLEIO 그리고 FACTA도구를 연계하여 분석 시스템화 하였다[16].

현재 대부분의 한의학 분야 문헌 분석은 분석 도구를 통해 얻어진 관련 유전자와 질병 등의 정보를 데이터베이스로 구축하여 제공하고 있다. 이미 구현된 분석 알고리즘을 대상 데이터에 적용하여 구축되었지만, 새로운 한약재와 신규 문헌 정보에 대한 분석을 손쉽게 수행할 수 있는 분석 시스템이 제공되지 못하고 있는 실정이다.

### III. 제안하는 분석시스템 설계

한의학 약재에 대한 문헌 분석을 수행하기 위해 국내 한약재 학명을 기반으로 PubMed의 문헌 정보를 요청하고 다운로드 하여 텍스트마닝 도구를 활용하여 한

약재의 약효성분과 생물학적 네트워크 추출을 수행하고자 하였다. 이를 구현하기 위해서 본 연구에서는 [그림 1]과 같이 클라이언트 서버 구조의 분석시스템을 설계하였다. 클라이언트 프로그램은 Eclipse RCP를 기반으로 JeUtils, Neo4J manager, NeoClipse 등 관련 오픈소스 모듈을 기반으로 설계하였으며, BioNewtork는 PubMed 데이터 구조와 문헌 분석 결과를 기반으로 생물학적 객체들의 정보를 모델링 하였다. 분석 결과의 가시화를 위해서 분석된 결과는 그래프 데이터베이스인 Neo4J를 활용하여 생물학적 네트워크 데이터를 저장하였으며, 각 노드는 개체명 인식을 통해 추출된 생물학적 개체로 속성 값을 부여하였으며, 엣지는 상호작용의 타입을 설정할 수 있도록 하였으며, 데이터의 저장과 로딩은 Neo4J 에서 제공하는 자바 API를 활용하여 설계하였다. 또한 분석 시스템과 클라이언트와의 연계를 위해서 본 연구에서는 NBCR에서 개발한 OPAL 프레임워크를 활용하여 SOAP프로토콜 방식으로 분석 제출과 결과물의 확인이 가능하도록 설계하였다[17].

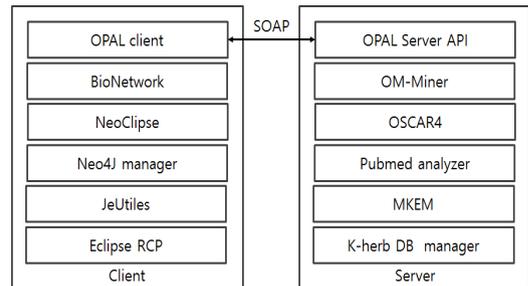


그림 1. 시스템 구성도

한의학 분야 문헌 분석을 위해서 다운로드된 PubMed의 xml데이터를 대상으로 크게 3가지 분석을 진행하였다. 먼저 NCBI로부터 다운로드 받은 xml을 분석하여 연구자가 기술한 관련 화합물 정보를 추출할 수 있는 화합물 추출 도구를 개발하였다. 다음으로 문헌의 제목과 초록을 문장별로 분리하여 OSCAR4[11]를 활용한 화합물 용어를 추출하였으며, ABNER[9]를 이용하여 단백질 및 유전자 개체를 추출하였다. 최종적으로 생물학적 개체 간 상호작용 정보는 MKEM[18] 도구를 활용하여 추출하였다[그림 2].

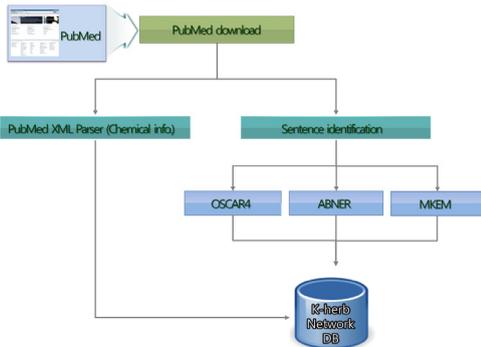


그림 2. 관계 추출 프로세스

본 연구에서는 각 한약제에 대한 기존 연구된 논문에서 추출한 생물학적 네트워크를 저장하고 이를 연구자들이 손쉽게 검색할 수 있도록 그래프 기반의 데이터베이스를 최신의 빅데이터 기술을 기반으로 구축하고자 하였다. 특히 데이터베이스 분야에서도 대용량의 복잡한 네트워크 정보를 저장하고 관리할 수 있는 NOSQL이라는 새로운 형태의 데이터베이스가 등장했으며, 대용량의 네트워크 데이터를 저장 및 관리할 수 있는 NOSQL 데이터베이스로 Neo4J를 활용하였다. Neo4J는 기존의 데이터베이스에 비해 약 1000배의 검색속도 개선이 있으며, 2개의 단백질 간의 최단 거리 검색 등의 기능 등을 포함하고 있어 향후 구축된 데이터베이스를 분석하고 확장하고자 할 때 높은 성능을 제공하였다.

현재 텍스트마이닝을 통해 추출된 정보는 아래 [그림 3]에서와 같이 각 노드를 단백질/유전자 정보를 기반으로 구성된다. 이러한 정보를 모델링하기 위해 각 노드의 속성으로는 생체 내의 객체 이름과 엣지의 타입으로

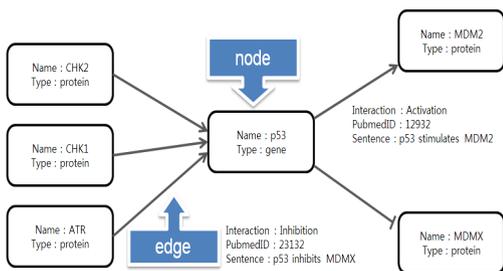


그림 3. 설계된 단백질/유전자의 상호작용 그래프 모델

는 상호작용의 타입을 정의하였고, 상호작용 정보에는 관련된 PubMed의 논문 정보와 문장 정보를 저장하여 원하는 경우 세부 내용을 검토할 수 있도록 하였다. 또한 화합물, 한약 제재 등 다양한 정보를 속성정보에 추가적으로 확장할 수 있도록 설계하였다.

문헌 분석을 통해 수집된 내용을 설계된 한약제에 대한 생물학적 네트워크의 그래프 데이터베이스(K-herb network DB)에 저장하기 위해서, 문헌 분석된 결과물을 K-herb network DB에 저장할 수 있는 입력 도구를 개발하였다. 개발 과정은 Neo4J에서 제공하는 자바 라이브러리를 활용하였으며, 제공된 API를 이용하여 문헌 분석 결과를 불러와 노드 생성과 상호작용 정보를 저장하였다. 그래프 데이터베이스로 저장하기에 비효율적인 PubMed 논문 정보 및 문헌 분석으로 추출된 화합물 정보는 MySQL을 이용하여 별도로 검색할 수 있도록 하였다.

향후 확장적인 웹사이트 설계를 위해서 본 연구에서는 루비 온 레일즈(Ruby on Rails)프레임워크를 활용하였다. 문헌 분석을 통해 구축된 K-herb network 데이터베이스의 구조를 기반으로 관련된 정보를 검색할 수 있는 웹서비스 기능은 크게 448개의 한약제의 검색 및 각각의 한약제별 관련 논문과 화합물 정보를 나타내며 보다 상세하게 생물학적 네트워크를 내비게이션 할 수 있도록 사용자 인터페이스를 설계하였다.

#### IV. 제안하는 분석시스템 설계 개발

##### 1. 한약제 정보 텍스트마이닝 과정

한약제명을 이용한 문헌분석은 [그림 4]의 과정과 같이 한약제를 키워드로 PubMed로부터 논문정보를 다운로드 받아 생물학적 용어 및 네트워크 마이닝 플랫폼을 통해 분석을 수행하였다. 텍스트마이닝은 각각의 문헌 정보를 화합물정보, 유전자/단백질정보, 생물학적 상호작용 정보로 나누어 분석을 수행하여 최종적으로 통합 데이터베이스에 저장하였다.



그림 4. 문헌 분석 과정

## 2. 자료 분석을 위한 PubMed 자료 수집

현재 한의학 연구에서 많이 활용되고 있는 약재는 한국한의학연구원에서 보유하고 있는 다양한 한의학 데이터베이스를 통해 수집된 정보와 대한약전, 생약규격집 등 약재 정보를 통해 총 448개의 약재정보를 수집하였다. 각 한약재의 학명 정보를 활용하여 미국 국립보건원 산하 NCBI에서 보유하고 있는 PubMed 자료를 다운로드 받아 한약재 문헌에 대한 총 17,398건의 관련 논문정보를 수집하여 한약재에 대한 문헌 데이터베이스 구축을 수행하였다.

## 3. 화합물 정보 추출기술 개발

본 연구에서는 현대 문헌에서 보고된 한의학 약재에 대한 정보를 분석하고자 미국 국립보건원 산하 NCBI의 PubMed와 Entrez의 E-Utilities를 이용하여 자료를 수집하였다. 다운로드 모듈은 E-Utilities를 개선한 JeUtils[19]를 활용하였다. 문헌 정보에서 제공하고 있는 화합물 정보를 추출하기 위하여 각각의 논문에서 제공하는 화합물 정보를 추출하는 프로그램을 통해 총 5,716개의 화합물 정보가 수집되었다.

추출된 화합물 정보 중 등록번호 정보를 통해 향후 PubChem [20] 정보와 연계할 수 있도록 데이터베이스를 구축하였다. 보다 광범위한 화합물 정보를 추출하기 위해서 화합물 정보를 전문적으로 추출할 수 있는 개체명인식 도구인 OSCAR4를 활용하여 총 95,925개의 화합물 정보를 수집하여 통합하였다.

## 4. 유전자/단백질 정보 추출기술 개발

현재 유전자와 단백질의 개체명 인식 도구로는 다양

한 연구가 진행되고 있으며, 이 중에서 ABNER는 단백질, 유전자, RNA, 세포 등의 개체명을 추출할 수 있는 도구로 활용하고자 하였다. PubMed 17,398 건의 논문의 초록정보를 가지고 ABNER를 통한 분석을 수행한 결과 21,952개의 유전자/단백질 정보를 추출하였으며, 이를 분석 결과 데이터베이스에 수록하였다.

## 5. 한약자원 분석 결과

한의학분야의 448개 한약재를 대상으로 분석한 결과, 현대적 연구방법을 적용한 한약재는 196개이며, 한약재당 평균 95개의 논문과 107개의 화합물이 보고되는 것으로 집계되었다. 또한 한의학연구에서 주로 보고된 화합물 분석결과를 살펴보면 Flavonoids, Antioxidants, Glucosides, Alkaloids, Flavonones순으로 빈번하게 보고되고 있을 수 있으며, 식물에서 쉽게 접할 수 있는 화합물들이다[표 1]. 다음으로 한약재별로 다양한 화합물이 보고된 횟수를 분석해 보면, 양파, 노회, 황금, 등황, 작약 순으로 많은 화합물이 보고되고 있으며 이는 해당 한약재의 연구가 활발히 진행되고 있음을 의미한다[표 2].

표 1. 문헌내의 화합물 추출결과

순위	화합물명	보고횟수
1	Flavonoids	781
2	Antioxidants	672
3	Glucosides	470
4	Alkaloids	353
5	Flavonones	326
6	Anthraquinones	308
7	Polysaccharides	307
8	Anti-inflammatory Agents	299
9	Xanthones	285
10	baicalin	261

표 2. 한약재당 보고된 화합물 통계

순위	화합물명	화합물수
1	양파	3186
2	노회	1919
3	황금	1418
4	등황	1259
5	작약	1209

반대로 화합물이 다양한 한약재에 동시에 보고된 경우는 Dinoprostone, Glucosides, Neuroprotective agents, Glycosides, Saponins 등이 다양한 약재에 포함되어진 것으로 연구됨을 알 수 있으며, 특히 Dinoprostone은 prostaglandin E2를 유발하며 약물로 알려져 있다.

표 3. 화합물당 보고된 한약재 통계

순위	화합물명	한약재 수
1	Dinoprostone	44
2	Glucosides	38
3	Neuroprotective Agents	37
4	Glycosides	36
5	Saponins	31
6	Carrageenan	28
7	Triterpenes	27
8	2,2-diphenyl-1-picrylhydrazyl	25
9	Picrates	25
10	Glutathione	24

다음으로 유전자 및 단백질의 출연 빈도를 ABNER를 통해 분석한 결과를 살펴보면, nerve growth factor, catalase, Superoxide dismutase, TNF등이 보고되었다. 특히 growth factor가 많이 보고되는 특징이 있었다[표 4]. 또한 한약재별로 다양한 유전자/단백질 정보를 포함하고 있는 통계[표 5]를 살펴보면 양파, 노회, 동충하초, 등황, 사프란 순으로 나타난다.

표 4. 한약재 연구에 많이 보고된 유전자/단백질

순위	유전자	보고 횟수
1	nerve growth factor	597
2	catalase	107
3	Superoxide dismutase	100
4	TNF-alpha	76
5	cytokines	68
6	Sea	67
7	glutathione peroxidase	66
8	Bcl-2	61
9	alpha-mangostin	53
10	collagen	53

표 5. 한약재별 보고된 유전자 통계

순위	한약재	유전자갯수
1	양파	12041
2	노회	3777
3	동충하초	2383
4	등황	2099
5	사프란	1242

마지막으로 한약재간 많이 연구된 유전자/단백질의 목록을 살펴보면 SOD, cytokine, TNF-alpha, Bax, Bcl-2순으로 나타나며 한약재 연구 시 주로 연구대상이 되는 유전자들임을 확인할 수 있었다[표 6]. 이는 한약재 간 효능연구 시 관련 유전자에 대한 효능연구의 대상을 검색할 때 많은 도움이 될 것으로 보인다.

표 6. 한약재간 많이 연구된 유전자 목록

순위	화합물명	보고 횟수
1	SOD	20
2	cytokines	10
3	TNF-alpha	10
4	Bax	9
5	Bcl-2	9
6	CAT	9
7	PARP	9
8	alkaline phosphatase	8
9	BALB	8
10	cox-2	8

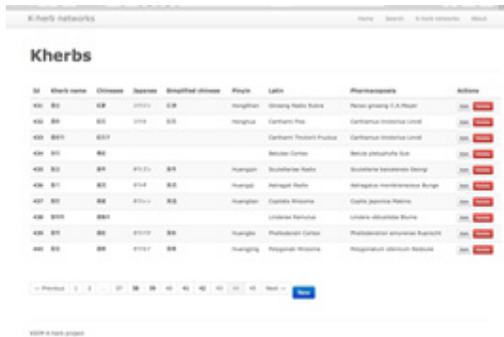
## 6. 한약자원 분석 결과 검색시스템 원형 개발

한의학 분야 연구자가 직접 관심 있는 한약재의 이름을 검색하거나 목록을 검색하여 해당 한약재의 다양한 약재 성분 및 관련 생물학적 네트워크 정보를 열람할 수 있는 웹사이트를 구현하였다. 한약재 정보를 통해 상세 정보로 들어가면 관련 논문과 주요 화합물 정보 및 관련 유전자 정보를 볼 수 있으며, 이와 관련된 생물학적 네트워크를 팝업 화면을 통해 직관적으로 관련 정보를 파악할 수 있도록 구현하였다. 루비 온 레일즈는 HTML5기반의 웹사이트 개발을 위한 고속 개발 환경을 제공하며, Bootstrap과 연계하여 손쉽게 웹사이트 원형을 개발할 수 있는 프레임워크로 활용되었다. 또한

REST서비스를 제공함으로써 향후 모바일 앱 개발과 연동할 수 있도록 확장성을 고려한 시스템을 구현하였다[그림 5].



(a) 메인 화면



(b) 검색 화면

그림 5. 한의학 네트워크 웹사이트 화면

### 7. 네트워크 가시화 설계 및 구현

최근 개발되고 있는 HTML5기반의 웹 프레임워크는 다양한 자바스크립트 기술을 제공하고 있다. 이 중에서 Processing.js[21] 기술은 웹 캔버스를 활용하여 다양한 가시화 기능을 제공할 수 있는 자바스크립트 개발 도구로 활용되고 있다. 최근 빅데이터 기술이 진보함에 따라 다양한 웹 가시화기술이 공개되고 있으나, 생물학적 네트워크를 내비게이션 할 수 있는 도구를 개발하는데 핵심적인 객체의 가시화 및 컨트롤 메커니즘을 고려했을 때 Processing.js는 확장성 있는 내비게이션을 구현하는데 적합한 모델로 최근 많이 활용되고 있다. 본 연구에서는 한의학 약제에 대한 생물학적 네트워크의 가시화를 Neovigator라는 가시화 라이브러리를 활용하여

구현하였다. 한의학 약제의 성분과 관련된 유전자 및 단백질간의 상호작용 정보를 가시화하기 위해서는 사용자가 약제, 유전자 혹은 단백질 정보를 키워드로 검색하면 가시화 모듈은 그래프 데이터베이스를 검색하여 해당 키워드의 노드를 검색하여 화면에 나타내며, 이와 상호작용하는 다양한 노드정보를 노드외곽에 다이얼 형태로 나타낸다.

[그림 6]은 실제 구현된 한약제의 생물학적 네트워크의 한 예로써 사용자가 직접 관련된 상호작용을 클릭하여 확장할 수 있어 원하는 형태의 생물학적 네트워크의 가시화를 구현하였다. 또한 복잡한 네트워크는 전체를 표시할 경우 직관적으로 단백질 및 유전자간의 관계를 파악할 수 없는 단점이 존재하며, 이를 극복하기 위해서 각 노드와 연결된 엣지의 정보를 노드 가장자리에 다이얼형태로 표시하고 연구자가 해당 연결의 내용을 확인한 이후 해당 엣지를 클릭함으로써 네트워크를 확장할 수 있는 능동적 네트워크 확장 기능을 제공하였다.



그림 6. 생물학적 상호작용 가시화 화면

## V. 결론 및 향후 연구

한의학 분야는 최근 현대적인 과학방법론을 도입하여 현대적인 연구가 빠르게 진행되고 있는 분야로 다양한 한약제의 정보와 관련된 약제 및 생물학적 기작에 대한 정보의 중요성이 증가하고 있다. 이 같은 다양한 정보를 수집하고 제공하기 위해서는 과거 연구된 한의학 연구의 문헌 정보를 분석하고 이를 체계적으로 데이

터베이스화하여 지식 기반의 연구로 쉽게 전환할 수 있도록 한의학 분야 지식 기반 데이터베이스의 구축이 필수적인 요소가 되었다.

본 연구에서는 최신의 텍스트마이닝 기술을 통합하여 한의학 분야 문헌 분석을 손쉽게 수행할 수 있는 분석 시스템을 개발하였으며, 한의학 약재 448개를 대상으로 관련된 약재 정보와 생물학적 상호작용 정보를 추출하여 통합 데이터베이스를 구축하였다. 분석결과를 통해 한약재별로 주요 화합물과 유전자 및 단백질정보를 추출할 수 있었으며, 다양한 약재에 분포하고 있는 화합물과 유전자/단백질 정보를 분석하였으며, 이를 통해 향후 한약재별 특성을 분석하는데 유용한 정보로 활용될 것으로 보인다. 구축된 데이터베이스를 연구자가 손쉽게 검색할 수 있도록 루비 온 레일즈 기반의 웹사이트의 원형을 개발하였다. 또한 연구자가 직관적으로 생물학적 네트워크 상황을 파악할 수 있도록 생물학적 네트워크 가시화 도구를 최신 웹 기술을 활용하여 구축하였다.

문헌분석결과를 통해 다양한 한약재의 주요성분은 물론 약재성분별로 어떤 한약재에 존재하는지를 네트워크 상에서 직관적으로 이해할 수 있음으로써 주요 한약재의 성분을 파악하는데 활용될 수 있음을 확인하였다. 현재 신약의 개발 속도가 정체되고 있어서 한의학 약재를 대상으로 신규 약효 성분의 중요성이 증가하고 있는 실정이다. 본 한의학 분야 문헌 분석 기술은 향후 다양한 한의학 분야 약재 성분의 효능 및 생물학적 기작을 파악하는데 중요한 도구로 활용될 수 있다.

**참 고 문 헌**

[1] M. S. Sundrud, S. B. Koralov, M. Feuerer, D. P. Calado, A. E. Kozhaya, A. Rhule-Smith, R. E. Lefebvre, D. Unutmaz, R. Mazitschek, H. Waldner, M. Whitman, T. Keller, and A. Rao, "Halofuginone inhibits TH17 cell differentiation by activating the amino acid starvation response," *Science*, Vol.324, No.5932,

pp.1334-1338, Jun. 2009.

[2] 김상균, 박동훈, 김안나, 오용택, 김지영, 예상준, 김철, 장현철, "한의 온톨로지 기반 시맨틱 검색 시스템", *한국콘텐츠학회논문지*, 제12권, 제12호, pp.533-543, 2012.

[3] 예상준, 양창섭, 김철, "헬스 2.0 기반의 한의정보 프레임워크 모델 설계", *한국콘텐츠학회논문지*, 제13권, 제11호, pp.807-814, 2013.

[4] Y. Feng, Z. Wu, X. Zhou, Z. Zhou, and W. Fan, "Knowledge discovery in traditional Chinese medicine: state of the art and perspectives," *Artif Intell Med*, Vol.38, No.3, pp.219-236, Nov. 2006.

[5] L. Hirschman, A. Yeh, C. Blaschke, and A. Valencia, "Overview of BioCreAtIvE: critical assessment of information extraction for biology," presented at the BMC bioinformatics, Vol.6, No.1, p.S1, 2005.

[6] D. C. Hao and P. G. Xiao, "Network pharmacology: a Rosetta Stone for traditional Chinese medicine," *Drug Dev. Res*, Vol.75, No.5, pp.299-312, Aug. 2014.

[7] H. Kitano, "Systems biology: a brief overview," *Science*, Vol.295, No.5560, pp.1662-1664, Mar. 2002.

[8] J. Fluck and M. Hofmann-Apitius, "Text mining for systems biology," *Drug Discov. Today*, Sep. 2013.

[9] B. Settles, "ABNER: an open source tool for automatically tagging genes, proteins and other entity names in text," *Bioinformatics*, Vol.21, No.14, pp.3191-3192, Jul. 2005.

[10] R. Leaman and G. Gonzalez, "BANNER: an executable survey of advances in biomedical named entity recognition," *Pac Symp Biocomput*, pp.652-663, 2008.

[11] D. M. Jessop, S. E. Adams, E. L. Willighagen, L. Hawizy, and P. Murray-Rust, "OSCAR4: a

- flexible architecture for chemical text-mining,” J Cheminform, Vol.3, No.1, p.41, 2011.
- [12] T. Ono, H. Hishigaki, A. Tanigami, and T. Takagi, “Automated extraction of information on protein-protein interactions from the biological literature,” Bioinformatics, Vol.17, No.2, pp.155-161, Feb. 2001.
- [13] Y. Hao, X. Zhu, M. Huang, and M. Li, “Discovering patterns to extract protein-protein interactions from the literature: Part II,” Bioinformatics, Vol.21, No.15, pp.3294-3300, Aug. 2005.
- [14] X. Zhou, B. Liu, Z. Wu, and Y. Feng, “Integrative mining of traditional Chinese medicine literature and MEDLINE for functional gene networks,” Artif Intell Med, Vol.41, No.2, pp.87-104, Oct. 2007.
- [15] Y. C. Fang, H. C. Huang, H. H. Chen, and H. F. Juan, “TCMGeneDIT: a database for associated traditional Chinese medicine, gene and disease information using text mining,” BMC Complement Altern Med, Vol.8, p.58, 2008.
- [16] B. Kemper, T. Matsuzaki, Y. Matsuoka, Y. Tsuruoka, H. Kitano, S. Ananiadou, and J. Tsujii, “PathText: a text mining integrator for biological pathway visualizations,” Bioinformatics, Vol.26, No.12, pp.i374-81, Jun. 2010.
- [17] J. Ren, N. Williams, L. Clementi, S. Krishnan, and W. W. Li, “Opal web services for biomedical applications,” Vol.38, No.Web Server, pp.W724-W731, Jun. 2010.
- [18] M. S. D. L. Ali Z Ijaz, “MKEM: a Multi-level Knowledge Emergence Model for mining undiscovered public knowledge,” BMC Bioinformatics, Vol.11, No.2, p.S3, 2010.
- [19] <http://sourceforge.net/projects/jeuutils>
- [20] Y. Wang, T. Suzek, J. Zhang, J. Wang, S. He,

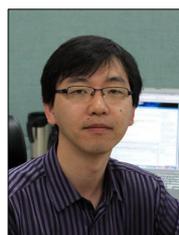
T. Cheng, B. A. Shoemaker, A. Gindulyte, and S. H. Bryant, “PubChem BioAssay: 2014 update,” Vol.42, No.Database issue, pp.D1075-82, Jan. 2014.

[21] <http://processingjs.org/>

## 참고 문헌

### 유 석 종(Seok Jong Yu)

정회원



- 1995년 2월 : 충북대학교 생화학  
과(이학사)
- 1997년 2월 : 충북대학교 생화학  
과(이학석사)
- 2013년 2월 : 충북대학교 정보통신공학과 생물정보협동과정(공학박사)
- 2003년 11월 ~ 현재 : KISTI 책임연구원  
<관심분야> : 생물정보학, 시스템생물학

### 조 용 성(Yongseong Cho)

정회원



- 1996년 2월 : 포항공과대학교 화  
학과(이학사)
- 1998년 8월 : 포항공과대학교 화  
학과(이학석사)
- 2003년 10월 ~ 현재 : KISTI 선  
임연구원  
<관심분야> : 생물정보학, 시스템생물학

### 이 준 학(Junehawk Lee)

정회원



- 2003년 2월 : 한국과학기술원 전  
산학과(공학사)
- 2005년 2월 : 한국과학기술원 바  
이오시스템학과(공학석사)
- 2005년 2월 ~ 현재 : 한국과학  
기술정보연구원 선임연구원  
<관심분야> : 생물정보학, 시스템생물학

서 동 민(Dongmin Seo)

정회원



- 2002년 : 충북대학교 정보통신공학과(공학사)
  - 2004년 : 충북대학교 정보통신공학과(공학석사)
  - 2008년 : 충북대학교 정보통신공학과(공학박사)
  - 2008년 ~ 2010년 : 한국과학기술원 전산학과 연수연구원
  - 2010년 ~ 현재 : 한국과학기술정보연구원 생명의료융합기술연구실 선임연구원
- <관심분야> : 빅데이터 분석, XML, 시맨틱웹, 이동객체 데이터베이스, 센서 네트워크, 그래프 데이터베이스

예 상 준(Sang-Jun Yea)

정회원



- 2002년 2월 : 한국과학기술원 전산학과(공학사)
  - 2004년 2월 : 한국과학기술원 전산학과(공학석사)
  - 2008년 5월 ~ 현재 : 한국한의학연구원 선임연구원
- <관심분야> : 영상처리, 한의약콘텐츠

김 철(Chul Kim)

정회원



- 1998년 2월 : 한국과학기술원 산업공학과(공학사)
  - 2000년 2월 : 한국과학기술원 산업공학과(공학석사)
  - 2009년 8월 : 원광대학교 한의정보학과(공학박사)
  - 2006년 7월 ~ 현재 : 한국한의학연구원 책임연구원
- <관심분야> : RFID/USN, 한의약정보화