

한의학 처방 분석 애플리케이션 개발: EveryFormula

오준호 책임연구원

한국한의학연구원

EveryFormula: Unlocking the Knowledge of Traditional East Asian Medicine Prescriptions with Data-Driven Analysis

Junho OH

Korea Institute of Oriental Medicine

Abstract

Herbal prescriptions play a crucial role in traditional East Asian medicine on the aspects of disease treatment and knowledge understanding. However, the vast number of these prescriptions makes it challenging to unlock their hidden knowledge. To address this, we developed EveryFormula, a web-based application designed to facilitate the understanding of the relationships between herbal prescriptions, medicinal herbs, and diseases through data-driven search and analysis. EveryFormula features [SEARCH] for relevance analysis, [COWORD-NET] for term network analysis, and [SEMANTIC-MAP] for semantic analysis. These tools enable users to search for prescriptions, analyze related medicinal herbs and diseases, and trace similar terms, thereby making it easier to derive explicit knowledge from traditional East Asian medicine prescriptions. This application promises to aid anyone interested in understanding and utilizing traditional East Asian medicine prescriptions.

Keywords: prescription of traditional East Asian medicine, prescription of traditional Korean medicine, data analysis, knowledge discovery

Correspondence: 오준호(Junho OH)

1672 Yuseong-daero, Yuseong-gu, Daejeon, 34054, Rep. of Korea

Tel: +82-42-868-9317, E-mail: junho@kiom.re.kr

Received 2024-05-27, revised 2024-06-07, accepted 2024-06-10, available online 2024-06-11
doi:10.22674/KHMI-12-1-7



서론

동아시아 전통의학(이하 한의학)은 동아시아 전통사회에서 이루어진 임상 의학의 총화(總和)이다. 그 가운데 본초를 조합하여 만들어진 처방은 이러한 임상 경험의 결정체이다. 한의학의 오랜 역사 속에서 처방은 환자의 증상이나 의사의 치료 방식과 같은 의학 내적인 요소에 따라 새로운 처방이 만들어지기도 하고 기존의 처방이 변형되기도 하면서 발전해 왔다. 시대에 따라 의학 사상이 변화하였고, 지역 풍토에 따라 발생하는 질병과 생산되는 약재가 달랐다. 이러한 의학 외적 요소도 처방을 다양하게 분화시키는 동인이 되었다. 그러므로 한의학에서 처방은 유구한 역사의 족적(足跡)이자 한의학 지식을 이해하기 위한 최선의 데이터라고 할 수 있다.

처방의 이해는 한의학 교육, 연구, 임상에서 핵심적인 과정이지만 처방을 이해하고 분석하기 위해서는 매우 많은 시간과 노력이 요구된다. 한의학의 처방은 그 수가 많고 매우 다양하기 때문이다. 한의학 처방의 전체 수량은 알 수 없으나 현대 중국에서 간행한 《中医方剂大辞典》에는 대략 10만여 종의 처방이 수록되어 있다(1993년 1판 기준 96,592종, 2판에는 여기에 2,400여 종이 추가되었다)¹⁾. 범위를 좁혀 보더라도 한국의 임상가들이 많이 사용하는 《동의보감(東醫寶鑑)》에 약 4천 종이 실려 있고²⁾, 《방약합편(方藥合編)》만 하더라도 약 467종의 처방이 수록되어 있다³⁾. 하지만 이 처방들은 본초가 가감되거나 용량이 증감되는 등의 방식으로 다양하게 변주되기 때문에 사실상 무한히 확장될 수 있다.

이에 본 연구에서는 한의학 처방을 데이터로 만들어 검색하고 분석함으로써 처방-본초-병증 사이의 관계를 이해하는 데 도움을 줄 목적으로 개발된 웹 기반 애플리케이션 EveryFormula⁴⁾를 소개하고 어떻게 활용할 수 있는지 설명하고자 한다.

본론

EveryFormula는 한의학 처방을 검색하고 분석하는 것을 목적으로 개발되었다. 검색 방법은 단순하면서도 다양한 검색 쿼리를 조합할 수 있도록 설계하였다. 분석은 주로 처방을 매개로 본초와 본초, 본초와 병증, 병증과 병증의 관계를 정량적으로 계측하는 데 주력하였다. 또 사용자가 직관적으로 분석 결과를 이해할 수 있도록 적절한 시각화 방법을 선택하고 이를 적용하여 분석 결과를 표현하였다.

EveryFormula는 PC 및 모바일 등 웹브라우저가 탑재되어 있고 인터넷이 연결된 환경이라면 어디에서든 사용할 수 있도록 웹애플리케이션(Web Application)으로 개발되었다.

1. 데이터 수집

EveryFormula에 사용된 처방 데이터는 특허청 한국전통지식포털(www.koreantk.com)에 공개된 전통의료(처방 부분) 데이터이다. 수록된 처방의 개수는 모두 11,344종이고, 처방에 사용된 본초는 모두 1,886종이며, 처방과 관련된 병증은 모두 7,081종이다. 이 처방들은 의서 25종에서 수집되었다⁵⁾.



2. 기능 구현

EveryFormula의 기능은 크게 다음과 같다. 검색과 연관성 분석을 위한 [SEARCH], 용어 네트워크 분석을 위한 [COWORD-NET], 용어 의미 분석을 위한 [SEMANTIC-MAP]. 이하 각각의 기능과 구현 방법 및 도출 결과 해석 등에 관해 설명하고자 한다.

1) [SEARCH] 검색과 연관성 분석

[SEARCH]는 EveryFormula의 기본 기능으로 검색과 연관성 분석 결과를 제공한다. 기본적인 기능이면서도 가장 중요한 기능이다. 본초-병증-처방 분석에서 가장 흔히 사용되는 방법이기 때문이다^{6,7)}.

(1) 검색

검색 기능은 ‘본초’, ‘병증’, ‘처방명’, ‘서적명’을 조합하여 원하는 처방을 찾아볼 수 있는 기능이다. 사용자의 편의를 위해 [포함할 키워드]와 [제외할 키워드] 이렇게 2가지 검색창에 키워드를 입력하여 검색하도록 설계하였다. 각각의 키워드에는 처방의 이름, 처방을 구성하는 본초, 처방의 주치(主治)가 되는 병증, 처방이 수록된 서적 등 4가지 필드 값을 구분 없이 입력할 수 있다. 또한 검색창마다 복수의 키워드가 함께 입력될 수 있다.

2개의 검색창에 키워드를 입력하는 간단한 형식을 따르고 있지만 입력 키워드에 4가지 필드 값을 입력할 수 있고 복수의 키워드를 허용하기 때문에 다양한 방식으로 활용할 수 있다. 예를 들어 포함할 키워드에 “오적산”을 넣으면 처방 이름이 오적산인 처방을 검색해 준다. 이때 오적산과 함께 “동의보감”을 넣으면 동의보감에 실려 있는 오적산을 검색해 준다. 또 포함할 키워드에 “백지”와 “고본”을 넣고, 제외할 키워드에 “천궁”을 넣으면 백지와 고본은 포함되어 있으면서 천궁은 포함되지 않은 처방을 검색해 준다. 이 기능은 병증과 함께 사용할 수도 있다. 포함할 키워드에 “두통”, “백지”를 넣고 제외할 키워드에 “천궁”을 넣으면 “두통”을 주치로 하는 처방 중에 “백지”는 포함되어 있지만 “천궁”은 포함되어 있지 않은 처방만을 찾을 수 있다(그림 1~3).



그림 1. [SEARCH] 검색 예시: 처방 이름으로 검색하기



그림 2. [SEARCH] 검색 예시: 본초 조합으로 검색하기

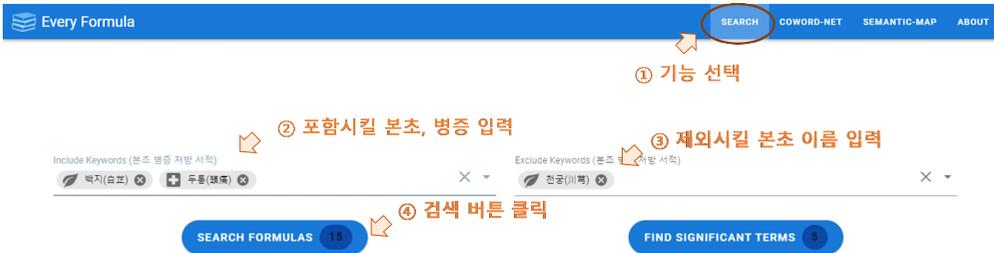


그림 3. [SEARCH] 검색 예시: 본초 및 병증 조합으로 검색하기

(2) 연관성 분석

연관성 분석 기능은 검색된 처방에서 중요하게 사용된 본초와 병증을 알아낼 수 있는 기능이다. 이 기능은 검색 결과를 바탕으로 분석을 진행하기 때문에 검색 기능과 같은 페이지에 구현하였다.

분석 방법은 다음과 같다. 검색 결과를 부분 집합으로 보고, 전체 집합과 비교했을 때 연관성이 높은 본초나 병증 용어를 찾아내 준다. 예를 들어 포함할 키워드에 “동의보감(東醫寶鑑)”을 넣어 검색하면, 다른 의서의 처방과 비교했을 때 동의보감에 더 두드러지게 사용된 본초와 병증이 도출된다. 포함할 키워드에 “반하(半夏)”를 넣어 검색하면, 다른 처방과 비교했을 때 반하가 포함된 처방에 더 두드러지게 사용된 본초와 병증이 도출된다. 또한 포함할 키워드에 “해수(咳嗽)”를 입력하고, 제외할 키워드에 “해역(咳逆)”을 입력하면 해수는 치료하지만 해역과는 무관한 본초 혹은 병증을 찾아 준다(그림 4). 검색 결과는 표 형태뿐만 아니라 워드클라우드(Wordcloud)를 통해 시각화하여 사용자가 직관적으로 이해할 수 있도록 하였다(그림 5~6).

키워드 사이에 연관성을 분석하기 위해서는 이를 정량적으로 계측할 수 있어야 한다. 이를 위해 G-score를 이용하였다. G-score는 Ted Dunning이 제안한 방식으로 두 범주형 변수의 독립성을 비교하는 방법이다. 두 변수가 함께 관찰된 실제 관찰값(observed value)을 O, 두 변수가 함께 관찰될 이론적 기댓값(Expected value)을 E라고 하였을 때, G-score의 근사값은 아래와 같은 공식으로 계산할 수 있다.⁸⁹⁾ G-score가 낮을수록 대상 키워드는 독립적으로 출현한다고 해석할 수 있고, 반대로 값이 크다면 대상 키워드는 검색 키워드와 연관성이 높다고 해석할 수 있다.

$$G \approx 2 \left(O \cdot \log \frac{O}{E} - (O - E) \right)$$



EveryFormula에 키워드를 입력하여 검색 결과가 도출되면 내부적으로 검색 결과에 포함된 본초 및 병증의 개수가 계산된다. 이를 관찰값으로 삼는다. 한편 전체 데이터 포함된 본초 및 병증 개수를 바탕으로 검색 결과에 도출된 본초 및 병증의 개수를 추정할 수 있다. 이를 기댓값으로 삼는다. 이 양자의 차이를 바탕으로 G-score가 계산된다.

워드클라우드(Wordcloud)는 텍스트를 분석하여 텍스트에 포함된 각 키워드의 중요성을 상대적으로 보여주기 위해 자주 사용되는 시각화 방법이다. 워드클라우드에서 하나의 키워드는 키워드의 색깔과 크기 2가지 변수로 표현될 수 있다. 색깔은 키워드가 속한 카테고리 등 범주형 데이터를 표현할 때 사용되며, 크기는 키워드의 출현 빈도와 같은 연속형 데이터를 표현할 때 사용된다. 사용자는 워드클라우드를 통해 어떤 단어가 중요하게 사용되었는지 한눈에 파악할 수 있다.

EveryFormula에서는 범주형 자료는 따로 적용하지 않았고, G-Score의 제공근에 비례하여 폰트의 크기를 조절하도록 하였다. 해당 텍스트가 2차원 평면에 그려지기 때문에 제공근을 취하여 나타내면 폰트의 크기가 아니라 폰트의 면적에 비례하여 그려지기 때문이다. 이를 통해 사용자는 검색 키워드와 가장 연관성이 높은 키워드를 직관적으로 확인할 수 있다.

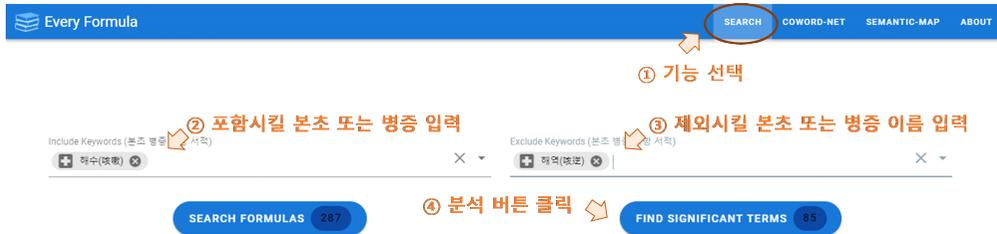


그림 4. [SEARCH] 연관성 분석 예시: 검색창 쿼리 입력

Term	G2-score	P-value	Category
상백피(桑白皮)	160.573	<=0.01	H
향인(杏仁)	131.027	<=0.01	H
파모(枹母)	97.242	<=0.01	H
전호(前胡)	92.211	<=0.01	H
길경(枳椇)	89.331	<=0.01	H
관동화(款冬花)	72.649	<=0.01	H
오미자(五味子)	66.672	<=0.01	H
천문동(天門冬)	64.270	<=0.01	H
담성(痰盛)	55.800	<=0.01	S
역문동(異門冬)	55.702	<=0.01	H

Rows per page: 10 1-10 of 85

DOWNLOAD DATA

그림 5. [SEARCH] 연관성 분석 예시: 표 출력 결과



그림 6 [SEARCH] 연관성 분석 예시: Wordcloud 출력 결과

그림 7. [COWORD-NET] 용어 네트워크 분석 예시: 쿼리 입력

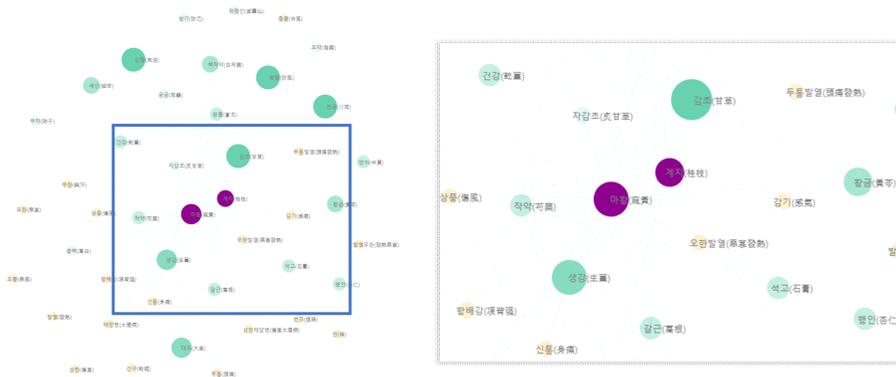


그림 8. [COWORD-NET] 용어 네트워크 분석 예시: 네트워크 출력 결과



2) [COWORD-NET] 용어 네트워크 분석

한의학 처방 분석에서 흔히 사용되는 분석 방법은 네트워크 분석이다¹⁰⁾. [COWORD-NET]은 입력 키워드를 중심으로 자주 함께 사용되는 본초나 병증을 네트워크(Network)로 보여준다. 예를 들어 “마황”, “계지”를 입력하여 분석하면, 처방 데이터를 기반으로 “마황” 그리고 “계지”와 공통으로 연관된 용어를 네트워크로 도출해 준다(그림 7~8).

이렇게 도출된 네트워크는 입력된 검색 키워드에 관한 일종의 에고 네트워크(Ego Network)이다. 에고 네트워크는 네트워크 분석에서 특정 대상과 직접 연결된 대상들로 구성된 하위 네트워크를 말한다. 여기서 기준이 되는 중심 대상을 ‘에고(ego)’, 에고와 연결된 다른 대상들을 ‘알터(alter)’라고 부른다.

EveryFormula에서 키워드가 입력되면 먼저 해당 키워드가 포함된 처방을 모두 검색한다. 검색된 처방에 포함된 본초와 병증은 모두 노드(node)가 되며, 동일한 처방에 존재하는 본초와 병증은 엣지(edge)로 연결된다. 최초로 입력된 검색 키워드가 에고가 되며 나머지 노드는 알터가 된다. 입력된 키워드가 하나가 아니라 복수라면 에고는 복수가 될 수도 있다. 이렇게 만들어진 네트워크는 전체 처방 데이터로 구성할 수 있는 네트워크에서 검색 키워드가 포함된 노드와 엣지만을 간추려낸 에고 네트워크가 된다.

특정 본초와 본초, 본초와 병증, 병증과 병증이 여러 처방에 반복적으로 나타나면 엣지의 강도가 강화된다고 해석할 수 있다. EveryFormula에서는 엣지의 크기를 변경하는 대신 입력 키워드와의 관계에 따라서 노드의 크기를 표시하였다. 즉 네트워크 결과에서 노드(node)의 크기가 클수록 함께 사용된 빈도가 높은 용어이다.

3) [SEMANTIC-MAP] 용어 의미 분석

[SEMANTIC-MAP]은 검색 키워드와 의미가 유사한 본초 또는 병증을 분석하여 도출해 준다. EveryFormula에서는 처방 데이터를 Word2Vec 기법으로 분석한 다음 각 용어를 고차원 벡터로 변환시켰다.

검색 결과는 도표와 그래프로 출력된다. 그래프는 Word2Vec을 통해 고차원 벡터로 대응시킨 뒤 다차원척도법을 이용하여 2차원 평면에 투영한 결과이다. 원의 크기는 키워드의 출현 빈도를 의미하며, 색깔은 검색 키워드와의 거리를 의미한다.

예를 들어 “발열”을 넣어 분석하면, 의미를 기준으로 발열과 가까운 “발열오한(發熱惡寒)”, “자한부지(自汗不止)”, “한열(寒熱)” 등을 찾아 준다. 또 “계지(桂枝)”를 검색하면 “계피(桂皮)”, “관계(官桂)”, “계심(桂心)”, “육계(肉桂)”가 도출된 것을 볼 수 있다(그림 9~10).

분석에 사용된 Word2Vec은 인공신경망(artificial neural network)을 이용한 단어 임베딩(embedding) 기법 가운데 하나이다¹¹⁾. 모든 용어는 사용된 맥락 속에서 의미를 추측할 수 있다. 곧 주변 용어를 통해 대상 용어의 의미를 파악할 수 있는 것이다. 이런 원리를 바탕으로 Word2Vec은 텍스트에서 전후 맥락을 고려하여 대상 단어를 고차원 공간에 벡터 형태로 변환한다. 그 결과, 의미가 유사한 단어는 고차원 공간에서 가까이 위치하며, 의미가 먼 단어는 멀리 위치하게 된다. 이를 통해 단어 사이의 의미 관계를 정량적으로 계산할 수 있다.

한의학 처방 데이터에도 이 방법을 적용할 수 있다. 같은 처방에 등장하는 본초와 병증은 텍스트에서 같은 문맥 속에 나타나는 단어를 비유할 수 있다. 따라서 EveryFormula에서는 처방 데이터를 Word2Vec으로 학습하여 유사한 의미를 가진 본초 혹은 병증을 고차원 공간 위에 벡터로 표현할 수 있었다. 이 벡터 사이의 거리를 연산하면 해당 본초 혹은 병증 사이의 관계를 상대적인 수치로

표현할 수 있다¹²⁾.

시각화에 사용된 다차원척도법(Multidimensional Scaling, MDS)은 고차원 공간에 존재하는 데이터를 저차원 공간에 투영하는 기법 가운데 하나이다¹³⁾. 인간이 직관적으로 인식할 수 있는 공간은 2차원 혹은 3차원에 국한되기 때문에 고차원 공간에 펼쳐진 데이터를 직관적으로 인식할 능력이 없다. 이때 다차원척도법 등을 이용하여 고차원 공간의 데이터를 2차원 혹은 3차원으로 투영하여 인간이 인식할 수 있는 차원으로 축소하는 것이다.

그러나 고차원 벡터를 2차원 평면에 투영하면 정보의 왜곡을 피할 수 없다. 이는 아래 분석 예시에서도 확인할 수 있다. “계지”의 검색 결과를 보면 Word2Vec을 통한 의미 분석 결과에서 “계피”가 “관계”보다 “계지”와 가깝다는 점을 알 수 있다. 그러나 다차원척도법을 이용하여 투영한 그래프에서는 “관계”가 “계피”보다 중심에 있는 “계지”와 더 가깝게 그려져 있다. 이는 차원 축소로 나타난 불가피한 정보의 왜곡 때문이다. 이 때문에 키워드의 색깔을 통해 이를 보완하였다(그림 10).

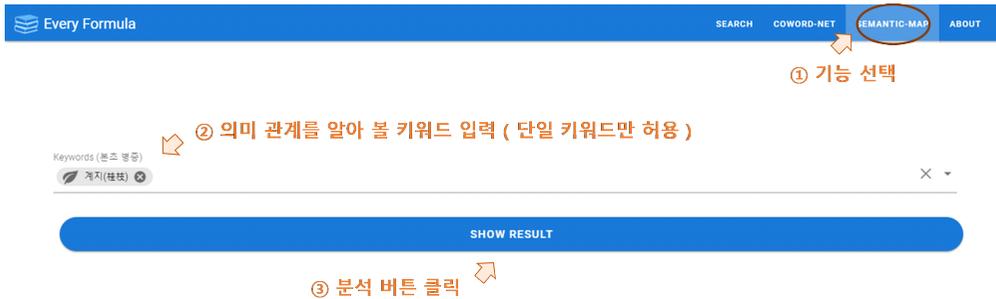


그림 9. [SEMANTIC-MAP] 용어 의미 분석 예시: 쿼리 입력

Name	Freq(BubbleSize)	Distance(BubbleColor)	co-X	co-Y	co-r
계지(桂枝)	651	0	0	0	0.1152
계피(桂皮)	142	0.3363	-0.1997	-0.2211	0.1008
관계(薑桂)	275	0.3761	-0.2519	-0.007	0.1073
계심(桂心)	513	0.3945	-0.305	-0.0139	0.1131
육계(肉桂)	957	0.4393	-0.3007	-0.0807	0.1186
날계(蔞桂)	65	0.5938	-0.3318	-0.0672	0.0925
간지황(乾地黃)	51	0.6574	0.5423	-0.0877	0.0898
지황(地黃)	36	0.6979	0.4231	0.0596	0.0857
조각자(皂角刺)	126	0.7047	0.319	0.0238	0.0996
석류피(石榴皮)	17	0.711	0.1972	-0.1074	0.0762

Rows per page: 10 1-10 of 16



그림 10 [SEMANTIC-MAP] 용어 의미 분석 예시: 표와 그래프 출력 결과



3. 활용 방안

EveryFormula는 한의학 처방을 검색하고 분석하기 위해 개발된 웹애플리케이션이다. 처방 이해는 교육, 연구, 임상을 막론하고 한의학을 이해하고 활용하기 위해 필수적이기 때문에 각 현장에서 두루 활용될 수 있을 것이다.

여기에서는 어떤 연구자 혹은 임상가가 “소갈(消渴)”을 조사한다는 가상의 시나리오를 가지고 설명을 이어 나가보자 한다. 우선 소갈과 관련된 처방에는 어떤 것들이 있는지 스크리닝하기 위해 [SEARCH]의 ‘검색’ 기능을 활용할 수 있다. 대상 범위를 제한하는 일도 수월하다. 특정 의서에 속한 처방만 보고 싶거나 특정 본초가 포함된 처방 혹은 포함되지 않은 처방만 보고 싶은 경우에도 검색 쿼리를 조정하여 쉽게 처리할 수 있다.

다음으로 소갈 치료에 중요하게 사용된 본초는 무엇인가 살펴볼 차례다. [SEARCH]의 ‘연관성 분석’을 통해 소갈에 중요하게 사용된 본초가 맥문동(麥門冬), 천화분(天花粉), 오미자(五味子) 등이라는 사실을 찾을 수 있다. 다시 같은 기능에서 오미자(五味子)를 입력하면 맥문동(麥門冬), 숙지황(熟地黃), 산수유(山茱萸) 등이 오미자와 유의미하게 함께 사용되었음을 알 수 있다.

계속해서 [COWORD-NET]에서 오미자와 맥문동을 함께 검색하여 예고 네트워크를 도출하면, 소갈 이외에 번역(煩熱), 도한(盜汗), 해수(咳嗽) 등의 병증이 네트워크로 연결되어 있다는 사실을 알 수 있다.

이어서 [SEMANTIC-MAP]에서 “소갈(消渴)”을 입력하고 결과를 도출하면 유사한 문맥에 사용된 키워드로 “신수부족(腎水不足)”, “소증(消中)”, “강증(強中)” 등이 더 있다는 점을 알 수 있다. ‘소갈’이라는 키워드만으로 부족하다면 신수부족, 소증, 강증 등의 키워드로 검색어를 확장해 나갈 수 있다.

이렇게 검색과 분석을 반복해 나간다면 원하는 목적에 맞는 지식을 한층 수월하게 찾을 수 있을 것이다.

4. 한계와 발전 방향

본 애플리케이션은 몇 가지 한계를 가지고 있다. 첫째, 담겨 있는 처방 데이터가 충분히 크지 않다. EveryFormula에서는 공신력 있는 데이터를 사용하기 위해 특허청에서 공개한 처방 데이터를 사용하였다. 하지만 이 데이터는 대상 의서가 25종, 처방의 수가 11,344종에 불과하다. 《中医方剂大辞典》에 수록된 처방이 10만여 종이라는 점을 고려했을 때 적은 분량이다. 더 많은 한의학 처방이 데이터로 구축된다면 분석 결과는 더 신뢰할 수 있을 것이며 그에 따라 활용 범위도 함께 넓어질 수 있을 것이다.

둘째, 처방에서 본초 용량을 고려하지 않았다. 처방 이해에서 본초의 용량은 중요한 요소이다. 동일한 본초 구성을 가지고 있더라도 본초 용량에 따라 처방의 성격은 달라질 수 있다. 그러나 EveryFormula에서는 본초 용량을 데이터로 활용하지 못하였다. 이는 처방 속의 본초 용량이 명시적으로 표시되어 있지 않거나 서로 다른 도량형으로 기록되어 있어 이를 동일한 기준에서 정리할 수 없었기 때문이다. 구성 본초 용량에 관한 연구는 다수 수행되었지만¹⁴⁾¹⁵⁾¹⁶⁾ 이를 고려하여 만들어진 처방 데이터는 많지 않은 형편이다.

셋째, 용어 확장 검색을 지원하지 않고 있다. 비슷한 의미를 담고 있는 키워드를 동일한 키워드로

보고 검색해 준다면 사용자에게 편의를 줄 수 있다. 그러나 이 문제에 대해서는 사용자에게 따라 다양한 요구가 있을 수 있다. 예를 들어 감초(甘草)가 자감초(炙甘草)와 함께 검색된다면 편리할 수 있다. 그렇다면 생감초(生甘草)는 어떨까. 검색과 분석을 할 때 ‘감초’와 ‘자감초’를 결합하고, ‘감초’와 ‘생감초’를 결합하면 논리적으로 ‘자감초’와 ‘생감초’도 결합하게 된다. 이런 문제는 지황(地黃)-생지황(生地黃)-숙지황(熟地黃), 복령(茯苓)-백복령(白茯苓)-적복령(赤茯苓)-복신(茯神) 등 매우 다양하다. 이를 기계적으로 결합하면 자칫 사용자의 의도와 다른 분석 결과를 도출할 염려가 있다. 반대로 사용자에게 판단을 맡긴다면 애플리케이션 사용이 필요 이상으로 어려워질 수 있다. 현재로서 실현 가능한 방식 가운데 하나로는 데이터 규모를 더 확장한 뒤 Word2Vec을 통한 키워드 임베딩 기법을 활용하여 사용자에게 의미가 유사한 단어를 제안한 다음 통합하여 검색하고 분석할 것인지 확인 받는 방식이다.¹²⁾¹⁷⁾

현재 자연어처리(Natural Language Process, NLP) 분야에서는 GPT(Generative pre-trained transformer)나 BERT(Bidirectional Encoder Representations from Transformers)와 같은 대규모 언어 모델(Large language model, LLM)이 새로운 혁신을 만들어 가고 있다. 데이터 크기의 한계 때문에 한의학 처방을 대상으로 이런 모델을 적용할 수 있는지 미지수이다. 그러나 이런 선형 기술을 한의학 처방 더 나아가 고문헌 데이터에 적용한다면 온고지신(溫故知新)이라는 말에 걸맞은 좋은 사례가 될 것이다.

결론

지금까지 한의학 교육, 연구, 임상 현장에서 한의학 처방 검색과 분석에 활용할 목적으로 만들어진 웹 기반 애플리케이션 EveryFormula를 소개하고 사용 방법과 활용 방안을 서술하였다. 본 애플리케이션은 검색과 연관성 분석을 위한 [SEARCH], 용어 네트워크 분석을 위한 [COWORD-NET], 용어 의미 분석을 위한 [SEMANTIC-MAP]의 기능을 담고 있다. 이들 기능을 통해 사용자는 한의학 처방을 처방명, 구성 본초, 적응 병증, 출전 문헌 등을 조합하여 검색할 수 있다. 나아가 치료 대상으로서의 병증, 혹은 치료 수단으로서의 본초를 키워드로 하여 관련된 본초와 병증을 찾아내고, 연관성을 분석하며, 유사한 의미의 키워드를 추적해 나갈 수 있다. 이런 과정을 반복함으로써 정량적인 데이터 분석을 기반으로 방대한 한의학 처방에 내재된 지식을 밖으로 끌어낼 수 있다. 이런 점에서 본 애플리케이션은 한의학 처방을 이해하고 활용하려는 모든 이들에게 도움을 줄 수 있을 것이다.

감사의 글

본 연구는 한국한의학연구원 주요사업 “AI 한의사 개발을 위한 임상 빅 데이터 수집 및 서비스 플랫폼 구축(KSN1923111)”의 지원을 받아 수행되었습니다.

참고문헌

1. 彭怀仁 外. 中医方剂大辞典. 1 版. 北京:人民卫生出版社. 1993.
2. 許浚. 東醫寶鑑. 朝鮮, 1610. In: 한국한의학연구원 (ed). 내 손안에 동의보감(원문강독편). 수퍼노바. 2017.



3. 黃度淵. 證脈方藥合編. 朝鮮, 1884. In: 남산당 (ed). 증맥방약합편. 남산당. 2000.
4. KMCResearch. EveryFormula. Available from: <https://every-formula.netlify.app> (accessed 2024-04-25.)
5. 특허청. 한국전통지식포털. Available from: <http://www.koreantk.com> (accessed 2019-01-12).
6. 김대욱, 박정환, 강영민. 처방 중심으로 본 한약재 시체(柿蒂)의 한의학적 소고(小考). 한약정보학회지. 2018;6(1):91-103.
7. 유희수, 이태훈, 김영은, 이부균. 《聖濟總錄》에서의 麻仁의 주치증별 기본방 및 약대 연구. 한약정보학회지. 2023;11(1):9-29.
8. Stefan Bordag et al. Computational Linguistics and Intelligent Text Processing. Berlin:Springer. 2008:52-63.
9. 오준호. 기계 학습을 이용한 한의학 처방 분석 방안. 대한한의학원전학회지. 2023;36(2):23-34.
10. 김안나 외 5 인. 불면 처방 활용 본초의 네트워크 분석. 대한한의학원전학회지. 2018;31(4):68-78.
11. Mikolov, T., Chen, K., Corrado, G.S., & Dean, J. Efficient Estimation of Word Representations in Vector Space. International Conference on Learning Representations. 2013.
12. 오준호. 기계 학습을 이용한 한의학 용어 유의어 사전 구축 방안. 대한한의학원전학회지. 2022;35(1):93-102.
13. Mead, A. Review of the Development of Multidimensional Scaling Methods. Journal of the Royal Statistical Society. Series D (The Statistician). 1992;41(1):27-39.
14. 김인락, 김광덕. 傷寒論 方寸匕의 현대용량추정. 대한본초학회. 2004;19(4):195-201.
15. 한천희, 이수연, 김지훈, 김윤경. 기존한약서의 현대적 용량 환산을 위한 도량형의 시대적 변천 고찰. 대한한의학방제학회. 2017;25(1):103-13.
16. 김유정, 이병욱, 엄동명, 송지청. 본초 및 용량 용어를 이용한 방제구성 자동추출방법에 대한 연구 : 『의종손익』을 중심으로. 대한한의학원전학회지. 2020;33(4):67-81.
17. 김우주, 김동희, 장희원. Word2vec 을 활용한 문서의 의미 확장 검색방법. 한국콘텐츠학회논문지. 2016;16(10):687-92.