

맥상용어목록을 이용한 맥상표현 자동추출방법 연구 - 『의종손익』을 중심으로 -

¹대구한의대학교 한의과대학 원전외사학교실 대학원생 · ²동국대학교 한의과대학 원전외사학교실 교수
³원광대학교 한의과대학 원전학교실 교수 · ⁴대구한의대학교 한의과대학 원전외사학교실 교수
금유정¹, 이병욱², 엄동명³, 송지청^{4*}

A Research on Automatic Data Extract Method of Pulse Descriptions Using the List of Pulse Terminology - Based on 『Euijongsonik』 -

Keum Yujeong¹, Lee Byungwook², Eom Dongmyung³, Song Jichung^{4*}

¹Graduate Student at Dept. of Medical Classics and History, College of Korean Medicine,
Daegu Haany University

²Professor at Dept. of Medical Classics and History, College of Korean Medicine, Dongguk
University

³Professor at Dept. of Medical Classics, College of Korean Medicine, Wonkwang University

⁴Professor at Dept. of Medical Classics and History, College of Korean Medicine, Daegu Haany
University

Objectives : Pulse descriptions in Korean Medical texts are comprised of combinations of pulse terminology, where various combinations of pulse terminology are used to describe disease symptoms. For Korean Medical doctors and professionals, however, it is impossible to identify the entirety of pulse description combinations, and their understanding is mostly limited to those learned from classical texts studied individually.

Methods : This research was carried out by using Access of Microsoft Office 365 in Windows 10 of Microsoft. Pulse descriptions were extracted from the text, 『Euijongsonik』. In the final stages, the automatically extracted list of pulse descriptions was refined through [excluded terminology of pulse description].

Results : The PC environment of this research was Intel Core i7-1065G7 CPU 1.30GHz, with 8GB of RAM and a Windows 10 64bit operation system. Out of 6,115 verses 6,497 descriptions were primarily extracted, and after a refinement process, the final list contained 5,507 pulse descriptions.

Conclusions : Based on the assumption that classical texts are available in data form to be processed by programs, current research methodology demonstrated that it was more efficient in regards to time and man power to create a pulse description database compared to when the researcher manually created one.

Key words : pulse description, data extract, terminology, 『Euijongsonik』.

* Corresponding author : Song Jichung

Dept. of Medical History, College of Korean Medicine, Daegu Haany University, 1 Haanydae Rd., Kyungsan, Kyungbuk, ROK

Tel : +82-53-819-1826, Fax : +82-53-819-1850, E-mail : prunedias@gamil.com

Received(October 19, 2020), Revised(November 8, 2020), Accepted(November 9, 2020)

Copyright © The Society of Korean Medical Classics. All rights reserved.

© This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. 서론

한의학에서 맥에 관한 지식은 『黃帝內經』부터 기록되었으며, 이후 『脈經』과 같은 전문의서를 통해 확대되었다. 한의학에서 맥은 脈象을 통해 지식화되었으며, 수많은 한의서에는 병증에 대한 주요 정보로 脈象을 제시하였다.

脈象은 주로 諸脈體狀의 28脈으로 표현되고 있으며 이 외에도 脈象을 표현하는 용어들이 있다. 맥상을 표현하는 용어는 많아 30여 가지이지만, 맥상 용어¹⁾의 조합으로 수백, 수천 가지로 표현될 수 있으며 실제 한의서에도 다양한 조합의 맥상용어를 통해 병증을 설명하고 있다. 하지만 한의사 또는 한의학 전문가들은 이러한 다양한 조합의 맥상표현에 대해 모두 파악하는 일은 불가능하며 개인적인 노력에 의해 한의서를 접하는 과정에서 단편적으로 맥상표현에 대한 지식을 접할 수 있을 뿐이다. 따라서 다양한 의서에 기재된 다양한 맥상표현을 개인이 정리한다는 것은 불가능할 것이다.

최근에 IT 기술을 활용하여 방제분류, 본초분류, 방제 및 본초 검색, 방제 분석 등 임상과 직접적인 연관이 있는 시스템들이 만들어져 있으며 이와 관련된 다양한 연구²⁾³⁾⁴⁾⁵⁾⁶⁾들도 진행되었다. 하지만 IT 기술을 활용하여 맥상표현을 추출하는 방법을 제시한 연구는 없었다.

이에 본 연구에서는 IT 기술을 활용하여 고문헌 의서에 기재된 맥상표현을 자동적으로 추출하는 방법에 대해 제시해보고자 한다.

II. 연구방법

1. 방법

본 연구는 Microsoft社의 Window 10 환경에서 Microsoft Office 365의 Access를 사용해 진행하였다. 맥상표현 추출 대상 원문은 『醫宗損益』이고, 디지털 원문은 한국한의학회에서 제공하는 “한의학고전DB”⁷⁾에서 추출하였다. 일차적으로 ‘맥상용어’ 데이터 셋⁸⁾을 이용하여 원문에서 맥상표현을 자동추출하고자 하였으며, 최종적으로 ‘맥상계의용어’ 데이터 셋을 이용하여 자동으로 추출된 맥상표현을 정제하고자 하였다.

2. 원리

본 연구에서 사용한 맥상표현 자동추출 원리는 [맥상용어]를 사용하여 원문에서 맥상표현 관련 용어를 추출하고, 문자열의 위치 관계를 이용하여 맥상표현 정보를 정제하고 생성하는 것이다. 이 과정은 크게 3단계(그림 1과 2)로 구분할 수 있다.

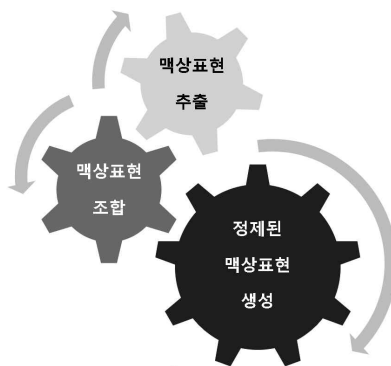


그림 1. 맥상표현 자동추출원리

- 1) 맥상용어와 맥상표현이란 용어를 구별하여 사용하였다. 맥상용어는 맥상에 대해 지식화되어 있는 용어를 의미하는 것으로 대표적인 것이 제맥체상의 28맥이다. 맥상표현은 한의서에 자연어로 기재되어 있는 용어를 의미하는 것으로 맥상용어를 활용하여 기술된 용어이다.
- 2) 박병선, 김은하, 이선아, 이병욱. 방제학에 기재된 방제효능과 본초 구성을 기반으로 도출된 효능의 비교 연구. 대한한의학회전학회지. 21(1). 2008. pp.79-92.
- 3) 김정훈, 이병욱. DB를 활용한 方劑의 類方分析 방법 설계. 대한한의학회전학회지. 21(1). 2008. pp.143-151.
- 4) 김현호, 홍효신, 유제혁, 권오민, 차용석. 객체지향형 처방 데이터베이스의 구축과 처방 검색 프로그램의 설계 및 개발. 한국한의학회연구논문집. 17(2). 2011. pp.73-83.
- 5) 김기욱, 김태열, 이병욱. 본초 목록을 이용한 방제의 본초 구성 자동 추출 방법. 대한한의학회전학회지. 27(3). 2014. pp.155-166.
- 6) 김중현, 배효진, 김창업, 이충열, 신상원. 텍스트마이닝(Text mining)을 활용한 한의학 원전 연구의 가능성 모색. 대한한의학회전학회지. 31(4). 2018. pp.27-46.

- 7) 한의학고전DB: <https://mediclassics.kr/> 2020년 10월 16일 검색
- 8) 향후 데이터 셋의 명칭은 대괄호를 통해 표시하며, 이는 Access의 테이블명이다.

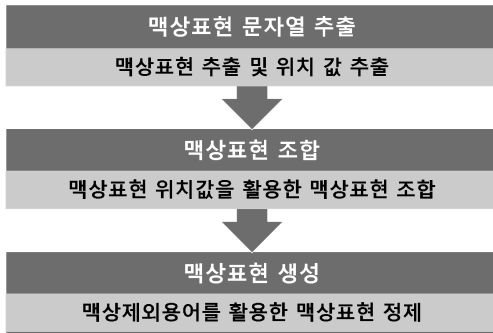


그림 2. 맥상표현 자동추출 과정

3. 원리 구현 개념 설명

본 연구의 대상이 된 『醫宗損益』 중 卷六 「寒門」의 맥상표현이 들어있는 부분은 다음(그림 3)과 같다.

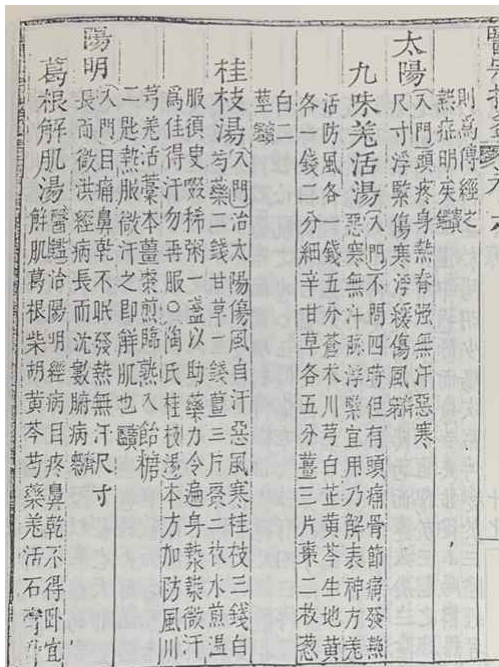


그림 3. 『醫宗損益』卷之六·「寒門」 원문일부

이 중 대상으로 한 원문과 원문에 포함된 맥상표현은 아래와 같다.

[入門]目痛鼻乾不眠發熱無汗尺寸長而微洪經病長而沈數腑病濟衆

한국한의학연구원 한의학고전DB에서 해당문장을 가져오면 아래와 같다.

入門目痛■鼻乾■不眠■發熱■無汗，尺寸長而微洪經病，長而沈數腑病.<濟衆>

이를 통해 맥상표현을 자동추출한 결과는 아래(그림 4)와 같다.



그림 4. [맥상용어]를 이용한 맥상표현 자동추출 결과

1) [맥상용어] 작성

[맥상용어]는 다음과 같은 용어들을 사용하여 작성하였다.

- 諸脈體狀의 28개 맥상용어; 浮 沈 遲 數 滑 澀 大 緩 洪 實 弦 緊 長 芤 微 細 濡 弱 虛 革 動 散 細 伏 絕 促 結 代
- 기타 맥상용어로 사용 가능한 용어; 疾 急 牢
- 脈과 脈의 위치를 나타내는 용어; 寸 關 尺 脈
- 맥상용어 이체자; 脈-脈-脉, 沈-沉, 澀-澀
- 맥상용어 이음동체자; 牢(뇌)-牢(뢰), 數(삭)-數(수)-數(축), 沈(침)-沈(심)
- 한자어로 맥상표현을 하고 있는 경우; 雙弦

이와 같은 방법으로 표제어를 ‘표준용어’로 연관어를 ‘비표준용어’로 규정하여 작성한 [맥상용어]는 다음(표 1)과 같다.

표 1. [맥상용어]

순서	표준용어	비표준용어	순서	표준용어	비표준용어
1	結	結	22	細	細
2	關	關	23	數	數
3	急	急	24	實	實
4	緊	緊	25	沈	沈
5	牢	牢	26	緩	緩
6	短	短	27	濡	濡
7	代	代	28	長	長
8	大	大	29	絶	絶
9	動	動	30	遲	遲
10	牢	牢	31	疾	疾
11	脈	脈	32	尺	尺
12	脈	脉	33	促	促
13	脈	脈	34	數	數
14	微	微	35	寸	寸
15	伏	伏	36	沈	沈
16	浮	浮	37	沈	沉
17	數	數	38	虛	虛
18	散	散	39	革	革
19	瀋	瀋	40	弦	弦
20	澁	澁	41	洪	洪
21	澁	澁	42	滑	滑
			43	雙弦	雙弦

다만 맥상용어로 사용되는 한자라 해도 다음의 용어들은 맥상용어가 아니므로 제외해야 한다. 다음 용어는 맥상용어 테이블에 제외용어로 표기하여 추가한다.

표 2. 맥상표현 검색시 제외되어야 하는 용어

순서	맥상용어	제외용어	순서	맥상용어	제외용어	순서	맥상용어	제외용어
1	結	結梗	61	脈	陰脈	121	弱	虛弱
2	結	結塊	62	脈	任脈	122	緩	苦緩
3	結	結胸	63	脈	冲脈	123	長	長流
4	結	痰結	64	脈	衝脈	124	長	長夏
5	結	便結	65	脈	冲脈	125	疾	痼疾
6	結	熱結	66	脈	血脈	126	疾	宿疾
7	結	鬱結	67	脈虛	氣虛	127	疾	疾病
8	結	凝結	68	微	微熱	128	疾	瘡疾
9	結	燥結	69	微	微妙	129	促	喘促
10	結	便結	70	微	微火	130	沈	沈重

11	結	寒結	71	伏	伏龍肝	131	沈	沉重
12	急	苦急	72	伏	伏於	132	沈	沈香
13	急	窘急	73	伏	伏于	133	沈	沉香
14	急	急食	74	伏	伏飲	134	沈	昏沈
15	急	氣急	75	浮	浮小麥	135	沈	昏沉
16	急	喘急	76	浮	浮腫	136	虛	肝虛
17	短	氣短	77	浮	浮萍	137	虛	氣虛
18	短	短氣	78	散	氣散	138	虛	內虛
19	大	大概	79	散	散血	139	虛	膽虛
20	大	大驚	80	散	生脈散	140	虛	大腸虛
21	大	大恐	81	澁	乾澁	141	虛	兩虛
22	大	大戟	82	澁	乾澁	142	虛	膀胱虛
23	大	大忌	83	澁	尿澁	143	虛	脾虛
24	大	大怒	84	澁	尿澁	144	虛	三焦虛
25	大	大豆	85	澁	淋澁	145	虛	小腸虛
26	大	大麥	86	澁	淋澁	146	虛	乘虛
27	大	大便	87	澁	澁痛	147	虛	腎虛
28	大	大病	88	澁	語澁	148	虛	心虛
29	大	大補	90	澁	語澁	149	虛	兩虛
30	大	大腹	90	澁	血澁	150	虛	陽虛
31	大	大蒜	91	澁	血澁	151	虛	胃虛
32	大	大小便	92	澁	澁痛	152	虛	陰虛
33	大	大碗	93	細	細末	153	虛	太虛
34	大	大人	94	細	細辛	154	虛	肺虛
35	大	大腸	95	細	細剉	155	虛	表虛
36	大	大抵	96	小	小腸	156	虛	下虛
37	大	大調中	97	小	小青龍	157	虛	虛怯
38	大	大青龍	98	小	小青龍	158	虛	虛勞
39	大	大青龍	99	數	頻數	159	虛	虛冷
40	大	大便	100	數	生數	160	虛	虛勞
41	大	大黃	101	數	生數	161	虛	虛煩
42	大	豆大	102	數	成數	162	虛	虛損
43	大	맥상	103	數	成數	163	虛	虛弱
44	大	腹大	104	腎	腎實	164	虛	虛熱
45	大	子大	105	實	肝實	165	虛	虛人
46	動	舉動	106	實	芡實	166	虛	虛風
47	動	跳動	107	實	鷄頭實	167	虛	虛寒
48	動	動搖	108	實	膽實	168	虛	虛火
49	動	動血	109	實	大腸實	169	虛	血虛
50	動	變動	110	實	膀胱實	170	滑	便滑
51	動	移動	111	實	脾實	171	滑	潤滑
52	動	躁動	112	實	三焦實	172	滑	精滑
53	動	潤動	113	實	小腸實	173	滑	便滑
54	脈	經脈	114	實	實熱	174	滑	滑石
55	脈	躄脈	115	實	心實	175	滑	滑泄
56	脈	帶脈	116	實	胃實	176	滑	滑脫
57	脈	督脈	117	實	枳實			
58	脈	脈經	118	實	充實			이 표는 필요에 따라 추가적으로 작성될 수 있음
59	脈	陽脈	119	實	肺實			
60	脈	維脈	120	弱	衰弱			

2) [맥상용어] 추가

‘漚而數’과 같이 맥상용어 사이에 ‘而’를 이용하여 맥상표현을 하고 있는 경우 [맥상용어]에 추가하였다. ‘而’를 이용하여 [맥상용어]를 작성하는 과정은 쿼리를 이용한다. 우선 ‘而’를 포함한 [말이율이]를 작성하고 이를 바탕으로 쿼리를 실행한다. 쿼리문장은 다음과 같다.

```
SELECT [맥상용어].[표준용어] & [말이율이.말이율이] & [맥상용어_1].[표준용어] AS 표준용어,
[맥상용어].[비표준용어] & [말이율이.말이율이] & [맥상용어_1].[비표준용어] AS 비표준용어
FROM 맥상용어, 맥상용어 AS 맥상용어_1, 말이율이;
```

3) 맥상표현 문자열 및 위치 정보 추출

맥상을 표현하는 문자열을 추출하고 각 문자열의 위치정보를 함께 추출한다.

표 3. 맥상표현의 위치

맥상표현 문자열	맥상표현 위치(이)
尺	19
寸	20
長	21
長而微	21
微	23
洪	24
長	29
長而沈	29
沈	31
數	32

4) 맥상표현 생성

맥상표현 문자열의 위치정보를 활용하여 연속되는 맥상표현을 생성한다.

표 4. 맥상표현 생성

맥상표현 문자열	맥상표현 위치	맥상표현 생성	맥상표현 새 위치
----------	---------	---------	-----------

9) 위치정보는 첫글자만 나타낸다. 왜냐하면 전체용어의 위치 정보는 첫글자의 위치정보에 문자열의 길이를 통해 표현할 수 있기 때문이다.

尺	19	尺寸長	19
寸	20		
長	21		
長而微	21	長而微洪	21
洪	24		
微	23	微	23
長	29	長	29
長而沈	29	長而沈數	29
數	32		
沈	31	沈	31

5) 맥상표현 정제

생성된 맥상표현은 중복되는 것이 있으므로 가장 긴 맥상표현을 기반으로 위치정보를 이용해 중복된 맥상표현을 정제한다.

가장 긴 맥상표현인 ‘長而微洪’의 위치정보는 21부터 24이고 ‘尺寸長’은 19부터 21이므로 두 맥상표현은 위치가 서로 겹친다. 또한 ‘微’의 위치정보도 23이므로 ‘長而微洪’ 내에 있다. ‘長’과 ‘沈’도 ‘長而沈數’ 내에 있다. 이를 정리하면 다음과 같다.

표 5. 맥상표현 생성

맥상표현 생성	맥상표현 새 위치	정제된 맥상표현	정제된 맥상표현 위치
尺寸長	19	尺寸長而微洪	19
長而微洪	21		
微	23		
長	29	長而沈數	29
長而沈數	29		
沈	31		

III. 본문

본 연구에서는 다음과 같이 데이터 베이스를 구성하고 Macro를 작성하여 맥상표현 자동추출과정을 완성하였다.

1. 맥상표현 자동추출 데이터 베이스의 구성

연구방법에서 기술한 내용을 구현하기 위해 다음과 같은 데이터베이스를 작성해야 한다.

1) 테이블 구성¹⁰⁾

(1) 텍스트 테이블: [의중손익_층차원문]

원문 저장을 위한 테이블을 작성해야 하며, 내용 ID, 권수명, 문명, 소재목, 처방명, 원문, 확인 필드를 갖는다.

(2) 문장분할 테이블: [문장분할]

Access 프로그램의 가독성을 위해 255글자 이내의 원문으로 분할하는 테이블을 작성해야 하며, 내용 ID, 문장분할, 순서, 문자시작번호 필드를 갖는다.

(3) 맥상용어 테이블: [용어목록]

맥상을 표현하는데 사용되는 맥상용어 목록 저장을 위한 테이블을 작성해야 하며, (맥상)표준용어, (맥상)비표준용어 필드를 갖는다.

(4) 맥상표현 저장 테이블: [맥상용어_임시저장]

원문에서 추출한 맥상표현을 1차적으로 저장하는 테이블로 내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호, 확인 필드를 갖는다.

(5) 맥상표현 2차 추출 테이블: [용어2단계]

분할된 문장 내 2회 이상 출현하는 맥상표현을 추출하기 위한 테이블로 내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호, 확인 필드를 갖는다.

(6) 맥상표현 위치 표현 테이블: [맥상용어추출]

분할된 문장 내 위치정보를 가지고 있는 맥상표현 문자열을 분할이전 원 문장의 위치정보를 표현하는 테이블로 내용ID, 맥상용어, 용어길이, 문자위치, 확인 필드를 갖는다.

(7) 연속적 맥상용어 정리 테이블: [용어3단계]

문장 내 연속적으로 나타나는 맥상용어를 연속된 맥상표현으로 연결하는 테이블로 내용ID, 맥상용어,

용어길이, 문자위치 필드를 갖는다.

(8) 맥상용어 추출 후 조합 테이블 1단계: [맥상용어추출]

연속된 맥상용어로 생성된 맥상표현의 중복을 삭제하는 테이블로 내용ID, 맥상용어, 용어길이, 문자위치, 확인 필드를 갖는다.

(9) 연속적 맥상표현 정리 테이블: [맥상용어추출_선택]

문장 내 연속적으로 나타나는 맥상표현을 연결하는 테이블로 내용ID, 맥상용어, 용어길이, 문자위치 확인, 필드를 갖는다.

(10) 맥상표현구성 테이블: [맥상용어정제]

원문에서 추출한 맥상표현을 최종적으로 저장하는 테이블로 내용ID, 맥상용어, 용어길이, 문자위치 필드를 갖는다.

(11) 맥상표현 최대길이 추출 테이블: [용어길이 MAX]

자동추출된 맥상표현의 최대길이를 추출하는 테이블로 용어길이의최대값 필드를 갖는다.

(12) 최대길이 맥상표현 추출 테이블: [용어MAX 길이추출]

최대길이 맥상표현을 추출하는 테이블로 내용ID, 맥상용어, 용어길이, 문자위치 필드를 갖는다.

2) 쿼리 구성

본 연구에서 맥상표현 문자열과 위치정보를 자동적으로 추출하는 과정은 삭제, 추가, 업데이트 쿼리를 생성한 뒤 과정순서에 따라 배열하고 이를 실행하는 방식으로 구성하였다. 쿼리 내용은 다음과 같다.

(1) 삭제 쿼리

① 011맥상용어_임시저장삭제

10) 실제 Access에서 사용한 테이블 명칭을 괄호 제목 뒤에 병기하였다.

[맥상용어_입시저장] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 맥상용어_입시저장.* FROM 맥상용어_입시저장;

② 012용어2단계삭제

[용어2단계] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 용어2단계.* FROM 용어2단계;

③ 014맥상용어추출삭제

[맥상용어추출] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 맥상용어추출.* FROM 맥상용어추출;

④ 015용어3단계삭제

[용어3단계] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 용어3단계.* FROM 용어3단계;

⑤ 017맥상용어추출_선택삭제

[맥상용어추출_선택] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 맥상용어추출_선택.* FROM 맥상용어추출_선택;

⑥ 020맥상용어정제삭제

[맥상용어정제삭제] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 맥상용어정제.* FROM 맥상용어정제;

⑦ 021용어길이MAX삭제

[용어길이MAX] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 용어길이MAX.* FROM 용어길이MAX;

⑧ 022용어MAX길이추출삭제

[용어MAX길이추출] 테이블 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 용어MAX길이추출.* FROM 용어MAX길이추출;

⑨ 025맥상용어정제_정리

[맥상용어정제] 테이블에 “삭제”표기로 삭제되어야 할 내용을 삭제하는 쿼리로 쿼리문장은 아래와 같다.

DELETE 맥상용어정제.맥상용어 FROM 맥상용어정제 WHERE (((맥상용어정제.맥상용어)="삭제"));

(2) 업데이트 쿼리

① 쿼리121

기존 [맥상용어_입시저장]에 기록되어 있는 맥상표현과 새로 기록될 맥상표현을 구별하기 위해 기존 맥상표현에 ‘확인’표시를 위한 쿼리로 [맥상용어_입시저장] 확인 필드에 ‘YES’로 표기되며, 쿼리문장은 아래와 같다.

UPDATE 맥상용어_입시저장 SET 맥상용어_입시저장.확인 = Yes;

② 쿼리151

기존 [맥상용어추출]에 기록되어 있는 맥상표현과 새로 기록될 맥상표현을 구별하기 위해 기존 맥상표현에 구별표시를 작성하기 위한 쿼리로 [맥상용어추출] ‘확인’ 필드에 ‘YES’로 표기되며, 쿼리문장은 아래와 같다.

UPDATE 맥상용어추출 SET 맥상용어추출.확인 = Yes;

③ 쿼리23

[용어MAX길이추출]을 이용해 [맥상용어정제] 내 중복되는 맥상표현을 추출하여 이를 삭제하기 위한 준비 쿼리로 [맥상용어정제] ‘맥상용어’ 필드에 ‘삭제’로 표기되며, 쿼리문장은 아래와 같다.

UPDATE 맥상용어정제 INNER JOIN 용어MAX길이추출 ON 맥상용어정제.내용ID = 용어MAX길이추출.내용ID SET 맥상용어정제.맥상용어 = "삭제" WHERE (((맥상용어정제.용어길이)<[용어MAX길이추출].[용어길이]) AND ((맥상용어정제.문자위

치)>=[용어MAX길이추출].[문자위치] AND (맥상용어정제.문자위치)<=[용어MAX길이추출].[용어길이]+[용어MAX길이추출].[문자위치]);

(3) 추가 쿼리

① 쿼리1

[의중손익_층차원문]의 ‘원문’ 필드를 Access가 데이터처리를 할 수 있도록 255글자 이내¹¹⁾로 문장을 분할하고, 원문장의 길이를 파악할 수 있도록 분할 다음 문장은 분할 문장 앞의 문장 길이 ‘250’을 표기하여 [문장분할]에 추가하는 쿼리로 [문장분할]의 내용ID, 문장분할, 문자시작번호 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 문장분할 (내용ID, 문장분할, 문자시작번호 )
```

```
SELECT 의중손익_층차원문.내용ID, Mid([의중손익_층차원문].[원문],1,250) AS 문장분할, 0 AS 문자시작번호
```

```
FROM 의중손익_층차원문
```

```
WHERE (((Mid([의중손익_층차원문].[원문],1,250)) Is Not Null)
```

```
ORDER BY 의중손익_층차원문.내용ID;
```

② 쿼리2, 3, 4, 5, 6

문장의 길이에 따라 문장분할을 250, 500, 750, 1000, 1250 등 250글자 단위로 분할하는 과정으로 쿼리1의 과정을 반복하지만 문장의 원길이를 표기하여 [문장분할]의 내용ID, 문장분할, 문자시작번호 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 문장분할 ( 내용ID, 문장분할, 문자시작번호 )
```

```
SELECT 의중손익_층차원문.내용ID, Mid([의중손익_층차원문].[원문],25112),250) AS 문장분할, 25013) AS 문자시작번호
```

11) 본 연구에서는 250글자로 분할하였다.

12) 쿼리3은 501, 쿼리4는 751, 쿼리5는 1001, 쿼리 6은 1251로 작성되며, 『醫宗損益』의 문장은 1500자가 넘는 내용이 없었기에 쿼리6까지 작성되었다.

13) 쿼리3은 500, 쿼리4는 750, 쿼리5는 1000, 쿼리 6은 1250로 작성되며, 『醫宗損益』의 문장은 1500자가 넘는

```
FROM 의중손익_층차원문
```

```
WHERE (((Len([의중손익_층차원문].[원문]))>25014)))
```

```
ORDER BY 의중손익_층차원문.내용ID;
```

③ 쿼리11

[용어목록]을 이용하여 [문장분할]에서 [용어목록]과 일치하는 맥상용어를 추출하여 [맥상용어_임시저장]에 추가하는 쿼리로 [맥상용어_임시저장]의 내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호, 확인 필드에 추가된다. 다만 [용어목록]‘확인’에 표시된 맥상용어제외 용어들은 추출에서 제외한다. 쿼리문장은 아래와 같다.

```
INSERT INTO 맥상용어_임시저장 (내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호 )
```

```
SELECT DISTINCT 문장분할.내용ID, 용어목록.비표준용어 AS 맥상용어, Len([용어목록].[비표준용어]) AS 용어길이, InStr(1,[문장분할].[문장분할],[용어목록].[비표준용어]) AS 문자위치, 문장분할.순서, 문장분할.문자시작번호
```

```
FROM 문장분할 INNER JOIN 용어목록 ON 문장분할.문장분할 like"*"& 용어목록.비표준용어&"*" WHERE ((용어목록.삭제표시)=No)
```

```
ORDER BY 문장분할.내용ID, InStr(1,[문장분할].[문장분할],[용어목록].[비표준용어]);
```

④ 쿼리12

[맥상용어_임시저장] 내용 중 ‘확인’ 필드에 ‘YES’가 표기되지 않은 것만 [용어2단계]에 추가하는 쿼리로 [용어2단계]의 내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호, 확인 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 용어2단계 (내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호, 확인 )
```

```
SELECT DISTINCT 맥상용어_임시저장.내용ID,
```

내용이 없었기에 쿼리6까지 작성되었다.

14) 쿼리3은 500, 쿼리4는 750, 쿼리5는 1000, 쿼리 6은 1250로 작성되며, 『醫宗損益』의 문장은 1500자가 넘는 내용이 없었기에 쿼리6까지 작성되었다.

맥상용어_임시저장.맥상용어, 맥상용어_임시저장.용어길이, 맥상용어_임시저장.문자위치, 맥상용어_임시저장.순서, 맥상용어_임시저장.문자시작번호, 맥상용어_임시저장.확인

```
FROM 맥상용어_임시저장
WHERE (((맥상용어_임시저장.확인)=No))
ORDER BY 맥상용어_임시저장.내용ID, 맥상용어_임시저장.문자위치;
```

⑤ 쿼리13

[용어2단계]를 이용하여 [문장분할] 동일ID 문장 내에 존재하는 2번째 이상 출현하는 맥상표현을 추출해서 [맥상용어_임시저장]에 추가하는 쿼리로 [용어2단계]의 내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호, 확인 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 맥상용어_임시저장 (내용ID, 맥상용어, 용어길이, 문자위치, 순서, 문자시작번호 )
```

```
SELECT DISTINCT 문장분할.내용ID, 용어2단계.맥상용어, 용어2단계.용어길이, InStr([용어2단계].문자위치+1,문장분할.문장분할,[용어2단계].맥상용어) AS 문자위치, 문장분할.순서, 문장분할.문자시작번호
```

```
FROM 문장분할 INNER JOIN 용어2단계 ON (문장분할.문장분할 like"*"& 용어2단계.맥상용어&"*") AND (문장분할.문장분할 = 용어2단계.맥상용어)
```

```
WHERE (((InStr([용어2단계].[문자위치]+1,[문장분할].[문장분할],[용어2단계].[맥상용어]))>0))
```

```
ORDER BY 문장분할.내용ID, InStr([용어2단계].문자위치+1,문장분할.문장분할,[용어2단계].맥상용어);
```

⑥ 쿼리14

[맥상용어_임시저장]내 기재된 맥상표현의 문자열 '문자위치'와 '문자시작번호' 정보를 활용하여 문장에서 실제 표현된 문자열의 위치값을 계산하고 이를 [맥상용어추출]에 추가하는 쿼리로 [맥상용어추출]의 내용ID, 맥상용어, 용어길이, 문자위치 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 맥상용어추출 (내용ID, 맥상용어, 용어길이, 문자위치 )
```

```
SELECT DISTINCT 맥상용어_임시저장.내용ID, 맥상용어_임시저장.맥상용어, 맥상용어_임시저장.용어길이, [맥상용어_임시저장].[문자위치]+ [맥상용어_임시저장].[문자시작번호] AS 문자위치
```

```
FROM 맥상용어_임시저장
ORDER BY 맥상용어_임시저장.내용ID, [맥상용어_임시저장].[문자위치]+ [맥상용어_임시저장].[문자시작번호];
```

⑦ 쿼리15

[맥상용어추출]내 '용어길이'와 '문자위치'정보를 활용해서 연속되는 맥상표현을 생성하고 이를 [용어3단계]에 추가하는 쿼리로 [용어3단계]의 내용ID, 맥상용어, 용어길이, 문자위치 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 용어3단계 ( 내용ID, 맥상용어, 용어길이, 문자위치 )
```

```
SELECT 맥상용어추출.내용ID, [맥상용어추출].[맥상용어] & [맥상용어추출_1].[맥상용어] AS 맥상용어, [맥상용어추출].[용어길이]+ [맥상용어추출_1].[용어길이] AS 용어길이, 맥상용어추출.문자위치
```

```
FROM 맥상용어추출 INNER JOIN 맥상용어추출 AS 맥상용어추출_1 ON 맥상용어추출.내용ID = 맥상용어추출_1.내용ID
```

```
WHERE (((맥상용어추출.문자위치)=[맥상용어추출_1].[문자위치]-[맥상용어추출].[용어길이]))
```

```
ORDER BY 맥상용어추출.내용ID, 맥상용어추출.문자위치;
```

⑧ 쿼리16

[용어3단계]의 결과값을 [맥상용어추출]에 추가하는 쿼리로 [맥상용어추출]의 (내용ID, 맥상용어, 용어길이, 문자위치 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 맥상용어추출 (내용ID, 맥상용어, 용어길이, 문자위치 )
```

```
SELECT DISTINCT 용어3단계.내용ID, 용어3단
```

계.맥상용어, 용어3단계.용어길이, 용어3단계.문자위치
 FROM 용어3단계;

⑨ 쿼리17

[맥상용어추출] 내 '확인'에 'YES'가 표기되지 않은 데이터를 [맥상용어추출_선택]에 추가하는 쿼리로 [맥상용어추출_선택]의 내용ID, 맥상용어, 용어길이, 문자위치, 확인 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 맥상용어추출_선택 ( 내용ID, 맥상용어, 용어길이, 문자위치, 확인 )
SELECT DISTINCT 맥상용어추출.내용ID, 맥상용어추출.맥상용어, 맥상용어추출.용어길이, 맥상용어추출.문자위치, 맥상용어추출.확인
FROM 맥상용어추출
WHERE (((맥상용어추출.확인)=No));
```

⑩ 쿼리18

[맥상용어추출]과 [맥상용어추출_선택]을 활용하여 연속되는 맥상표현을 생성하고 [용어3단계]에 추가하는 쿼리로 [용어3단계]의 내용ID, 맥상용어, 용어길이, 문자위치 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 용어3단계 (내용ID, 맥상용어, 용어길이, 문자위치 )
SELECT 맥상용어추출.내용ID, [맥상용어추출].[맥상용어] & [맥상용어추출_선택].[맥상용어] AS 맥상용어, [맥상용어추출].[용어길이]+ [맥상용어추출_선택].[용어길이] AS 용어길이, 맥상용어추출.문자위치
FROM 맥상용어추출 INNER JOIN 맥상용어추출_선택 ON 맥상용어추출.내용ID = 맥상용어추출_선택.내용ID
WHERE (((맥상용어추출.문자위치)=[맥상용어추출_선택].[문자위치]-[맥상용어추출].[용어길이]));
```

⑪ 쿼리20

[맥상용어추출] 데이터 중 중복되는 데이터를 제외 한 후 [맥상용어정제]에 추가하는 쿼리로 [맥상용어정제]의 내용ID, 맥상용어, 용어길이, 문자위치 필드에 추가된다. 쿼리문장은 아래와 같다.

INSERT INTO 맥상용어정제 (내용ID, 맥상용어, 용어길이, 문자위치)
SELECT DISTINCT 맥상용어추출.내용ID, 맥상용어추출.맥상용어, 맥상용어추출.용어길이, 맥상용어추출.문자위치
FROM 맥상용어추출
ORDER BY 맥상용어추출.내용ID, 맥상용어추출.문자위치;

⑫ 쿼리21

[맥상용어정제]내 용어길이 최대값을 추출하여 [용어길이MAX]에 추가하는 쿼리로 [용어길이MAX]의 용어길이의최대값 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 용어길이MAX (용어길이의최대값)
SELECT Max(맥상용어정제.용어길이) AS 용어길이의최대값
FROM 맥상용어정제;
```

⑬ 쿼리22

[용어길이MAX] 결과값과 [맥상용어추출]을 이용하여 맥상표현 중 [용어길이MAX] '용어길이의최대값'에 해당되는 맥상표현을 추출하여 [용어MAX길이추출]에 추가하는 쿼리로 [용어MAX길이추출]의 내용ID, 맥상용어, 용어길이, 문자위치 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 용어MAX길이추출 (내용ID, 맥상용어, 용어길이, 문자위치 )
SELECT DISTINCT 맥상용어정제.내용ID, 맥상용어정제.맥상용어, 맥상용어정제.용어길이, 맥상용어정제.문자위치
FROM 맥상용어정제 INNER JOIN 용어길이MAX ON 맥상용어정제.용어길이 = 용어길이MAX.용어길이의최대값;
```

⑭ 쿼리24

[용어길이MAX]의 '용어길이의최대값'을 한 단계

낮추는 것을 목표로하는 쿼리로 [용어길이MAX]의 용어길이의최대값 필드에 추가된다. 쿼리문장은 아래와 같다.

```
INSERT INTO 용어길이MAX ( 용어길이의최대값)
SELECT Max([용어MAX길이추출].[용어길이]- 1)
AS 용어길이
FROM 용어MAX길이추출
HAVING (((Max([용어MAX길이추출].[용어길이]-1)))>0));
```

2. 맥상표현 자동추출 과정

Access에 작성한 쿼리를 순차적으로 진행하기 위해 매크로를 작성하였고, 그 순서는 다음과 같다.

1) Macro1: 문장분할 단계

- (1) [쿼리1] 실행
- (2) [쿼리2] 실행
- (3) [쿼리3] 실행
- (4) [쿼리4] 실행
- (5) [쿼리5] 실행
- (6) [쿼리6] 실행

2) Macro2: 맥상용어추출 단계

2-1) 맥상용어단순형추출 단계

- 1) [011맥상용어_임시저장삭제] 실행
- 2) [쿼리11] 실행

2-2) 맥상용어반복추출 단계(반복20회)

- (1) [012용어2단계삭제] 실행
- (2) [쿼리12] 실행
- (3) [쿼리121] 실행
- (4) [쿼리13] 실행

3) Macro3: 맥상용어복합형추출 단계

3-1) 맥상연속용어추출준비 단계

- (1) [014맥상용어추출삭제] 실행
- (2) [쿼리014] 실행
- (3) [015용어3단계삭제] 실행
- (4) [쿼리15] 실행

3-2) 맥상연속용어추출반복 단계(반복10회)

- (1) [쿼리151] 실행
- (2) [쿼리16] 실행
- (3) [017맥상용어추출_선택삭제] 실행
- (4) [쿼리17] 실행
- (5) [쿼리151] 실행
- (6) [015용어3단계삭제] 실행
- (7) [쿼리18] 실행

4) Macro4: 맥상용어자동추출 단계

4-1) 맥상연속용어정제준비 단계

- (1) [020맥상용어정제삭제] 실행
- (2) [쿼리20] 실행
- (3) [용어길이MAX삭제] 실행
- (4) [쿼리21] 실행

4-2) 맥상연속용어정제반복 단계(반복10회)

- (1) [022용어MAX길이추출삭제] 실행
- (2) [쿼리22] 실행
- (3) [쿼리23] 실행
- (4) 021용어길이MAX삭제] 실행
- (5) [쿼리24] 실행

4-3) 맥상용어정제최종 단계

- (1) [쿼리25] 실행

최종적으로 RunMacro 명령어를 활용하여 맥상표현자동추출 Macro를 구성하였다.

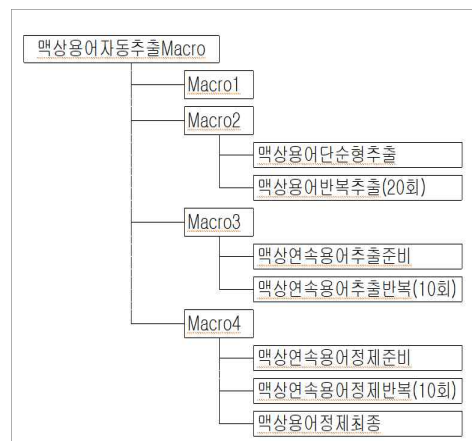


그림 5. 맥상표현 자동추출 Macro 개요도

4. 결과

본 연구를 진행할 때 PC환경은 Intel Core i7-1065G7 CPU 1.30GHz를 사용하였으며, RAM은 8GB를 사용하였고, Window 10 64비트 운영 체제를 사용하였다. 6,115개 문장에서 1차적으로 6,497개 맥상표현을 추출하였으며 정제과정을 거쳐 최종적으로 5,507개 맥상표현을 추출하였다. 이 과정을 추출하는데 걸린 시간은 1분16초였다.

IV. 결론

이와 같은 결과는 고문헌 의서가 프로그램이 파악할 수 있는 데이터파일로 존재한다는 가정하에서 본 연구방법을 통해 맥상을 나타내는 맥상표현 지식을 자동적으로 추출할 수 있음을 보여준다. 물론 맥상용어 데이터의 정확성에 따라 그 결과값이 더욱 정제될 수 있는바, 본 연구과정에서도 비정제 용어를 제외용어로 추가하고, 추출되지 않은 맥상용어를 다시 맥상용어 데이터셋에 추가하는 과정을 반복하였다.

프로그램을 이용한 한의서 지식의 자동추출방법론은 그 자체로 완벽한 것은 아니어서 최종적으로는 전문가가 판단을 해야 하는 영역이 남아있다. 하지만 이와 같이 DB로 작성하는 방법이 연구자가 직접 작성하는 경우보다 시간적, 인적 자원의 절약할 수 있다는 가능성을 보여준다. 왜냐하면 프로그램을 이용하지 않고 정확한 데이터를 얻기 위해서 전문가에 의해 수기로 작성된 데이터라 하더라도 오류가 있을 수 있으며 이 또한 검증 및 정제과정을 거쳐야 하는데, 이러한 과정은 1차적으로 맥상용어를 자동추출하고 이후 전문가에 의해 정제하는 과정에 비해 더 많은 시간과 노력이 드는 것이 자명하기 때문이다.

이 방법론은 이 자체로 완벽한 상태는 아니지만 향후 병증용어 추출과정으로 확대하여 병증-맥상 결합 데이터셋을 추출할 수 있는 선행연구가 될 수도 있다고 판단된다.

감사의 글

이 논문은 2020년도 정부(교육부)의 재원으로 한

국연구재단의 지원을 받아 수행된 기초연구사업임 (2017R1D1A2B05031884)

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (2017R1D1A2B05031884)

References

1. 박병선, 김은하, 이선아, 이병욱. 『방제학』에 기재된 방제효능과 본초 구성을 기반으로 도출된 효능의 비교 연구. 대한한의학회지 21(1). 2008.
2. 김정훈, 이병욱. DB를 활용한 方劑의 類方分析 방법 설계. 대한한의학회지 21(1). 2008.
3. 김현호, 홍효신, 유제혁, 권오민, 차웅석. 객체 지향형 처방데이터베이스의 구축과 처방 검색 프로그램의 설계 및 개발. 한국한의학회지 32(2). 2011.
4. 김기욱, 김태열, 이병욱. 본초 목록을 이용한 방제의 본초 구성 자동 추출 방법. 대한한의학회지 27(3). 2014.
<https://doi.org/10.14369/skmc.2014.27.3.155>
5. 김종현, 배효진, 김창업, 이충열, 신상원. 텍스트마이닝(Text mining)을 활용한 한의학 원전 연구의 가능성 모색. 대한한의학회지 31(4). 2018.
<https://doi.org/10.14369/jkmc.2018.31.4.027>
6. 한의학교전DB: <https://mediclassics.kr/>
2020년 10월 16일 검색