

본초 목록을 이용한 방제의 본초 구성 자동 추출 방법

¹동국대학교 한의과대학 원전·의사학교실

金基郁¹ · 金太烈¹ · 李丙旭¹ *

Automatic Extraction Method of Compositional Herb Using Herb List

Kim Ki-Wook¹ · Kim Thae-Yul¹ · Lee Byung-Wook¹ *

¹Dept. of Medicine Classics and History, College of Korean Medicinel, Dong Guk University.

Objectives : Many keyboarding staff are required for compositional herb data entry from ancient documents. It's the most difficult thing to build DB. So, we have studied to solve this problem.

Methods : We have used Windows 7 and Access2013 for automatic extraction compositional herb in the *Taipinghuiminhejijufang*(太平惠民和劑局方). First of all, we have made synonyms list of herbs and list of herb volumes notation. And then we have analyzed order of arrangements between names of herb and notations of herb volume.

Results : We have analyzed a part of prescriptions in the *Taipinghuiminhejijufang*(太平惠民和劑局方) (519 prescriptions). After 91 seconds, we can complete automatic extraction of compositional herb from 519 prescriptions.

Conclusions : If we can analyze order of arrangements between names of herb and notations of herb volume, we can automatically extract compositional herb of prescription from ancient documents.

Key Words : Prescription, compositional herb, order of term arrangement, automatic extraction of compositional herb

I. 서 론

* Corresponding Author : BW Lee, College of Korean Medicinel, Dongguk University, Dongdaero 123, Gyeongju, Gyeongsangbuk-Do, Korea.
E-mail: omis@dongguk.ac.kr Tel: 054-770-2665.
이병욱. 경북 경주시 동대로 123. 동국대학교 한의과대학.
접수일(2014년 07월21일), 수정일(2014년 08월10일),
게재확정일(2014년 08월11일).

한의학에서의 약물과 관련된 지식은 先秦時期的 『山海經』부터 문헌에 기록되었으며, 이후 馬王堆 의 서인 『五十二病方』, 그리고 東漢時期的 『傷寒雜病論』 등의 문헌을 통하여 지속적으로 축적되고 후대로 전달되었다. 이후로도 약물을 이용한 치료 경험에 대한 지식은 더욱더 풍부해지고, 한 두 사람의 지식이 아닌 국가 전체의 약물 지식을 다루는 문헌이 저술되는 수준에 이르렀으며, 그러한 지식의 양적 증

가는 인간의 지력으로 감당하기 어려운 상황까지 이르렀다. 지식의 양적 증가는 질병 치료의 측면에서 매우 긍정적인 현상임에도 불구하고, 유효한 방제와 관련된 정보를 감별하여 선택하는데 어려움을 주는 측면도 있다. 이러한 문제의 해결을 위하여 질병, 병인, 효능을 이용한 방제의 분류체계가 만들어지고 또한 방제 구성의 유사성에 기반한 類方이라는 방제의 분석체계도 등장하였다.

최근에는 IT 기술을 활용하여 질병, 병인, 효능에 따른 방제 분류 및 검색 시스템과 類方 체계 분석 시스템상을 구현하여 임상활용에 편리함을 도모하고자 하는 연구¹⁾²⁾³⁾⁴⁾⁵⁾⁶⁾가 진행되었으며 일정 부분 성과도 있었다. 그러나 이러한 IT 기술을 활용한 방제에 대한 분석 및 검색 시스템은 몇몇 문헌을 중심으로 DB가 구축되어 활용되고 있다. 몇몇 문헌은 법률적 의미를 가지는 문헌을 중심으로 선정되어 DB화가 진행되었다. 이러한 성과는 한의학 전반에 대한 약물 지식을 모두 반영하였다고 보기 힘든 측면이 있다. 때문에 몇몇 문헌에 국한되지 않고 한의학 문헌의 폭 넓은 지식을 온전하게 활용하기 위해서는 더욱 다양한 문헌에 대한 약물 지식의 체계적인 정리가 필요하다.

그러나 언제나 이러한 사업과 연구의 어려운 점은 고문헌에 기록된 내용으로부터 체계화된 방제구

성 및 주치 효능 정보를 얻기 위하여 대량의 인적 자원을 투입해야 한다는 점이다. 한 번에 많은 인적 자원을 투입하여 연구를 하기 위해서는 인적 자원의 체계적인 협업을 지원할 수 있는 시스템을 추가로 구축해야하는 부담도 함께 감당을 해야 한다.

이에 본 연구에서는 컴퓨터 시스템을 이용하여 고문헌에 기록된 원문으로부터 체계적인 지식을 DB화하는 방법을 연구하는 과정에서 얻은 방제의 본초 구성 추출 방법 분야에 대한 연구 결과를 보고하는 바이다.

II. 본 론

1. 연구방법

본 연구는 MS社의 Windows 7 환경에서 Access 2013을 사용하여 진행하였으며, 본초 구성 추출의 원문은 『太平惠民和劑局方』을 이용하였다. 본초구성의 추출은 본초명-용량 정보 추출을 목표로 진행하였으며, 본초의 修治 정보의 추출은 본초명-용량 정보의 검색에 사용하는 용어 목록만 수정하면 동일한 성능을 발휘 할 수 있으므로 본 연구에서는 생략하였다.

2. 원리

본 연구에서 사용한 본초-용량 추출 원리는 본초명과 용량에 해당하는 문자열의 위치 관계를 이용하여 본초의 용량 정보를 생성하는 것이다. 그 과정은 크게 3단계로 구분할 수 있다.



Fig. 1. fundamental of automatic extraction of compositional herb. 본초 조합의 자동 추출 원리

- 1) 박병선, 김은하, 이선아, 이병욱. 『방제학』에 기재된 방제 효능과 본초 구성을 기반으로 도출된 효능의 비교 연구. 대한한의학원전학회. 대한한의학원전학회지21(1). 2008. pp.79-92.
- 2) 이병욱, 엄동명, 김정자, 박지하, 개념중심 용어테이블을 이용한 효능 중심 본초 검색방법 연구. 대한한의학원전학회. 대한한의학원전학회지20(4). 2007. pp.311-318.
- 3) 김정훈, 이병욱. DB를 활용한 方劑의 類方分析 방법 설계. 대한한의학원전학회. 대한한의학원전학회지21(1). 2008. pp.143-151.
- 4) 이병욱, 정훈, 권영규. 인터넷상 동의보감 방제의 약물구성 계보 표현 방법에 관한 연구. 한국의사학회. 한국의사학회지21(2). 2008. pp.41-47.
- 5) 오용택, 김상찬, 이병욱. 본초 구성을 이용한 방제 효능 예측 연구 및 본초의 효능 강도 설정안. 대한한의학원전학회. 대한한의학원전학회지21(2). 2008. pp.49-57.
- 6) 김현호, 홍효신, 유재혁, 권오민, 차용식. 객체지향형 처방 데이터베이스의 구축과 처방 검색 프로그램의 설계 및 개발. 한국한의학연구원. 한국한의학연구원논문집17(2). 2011. pp.73-83.

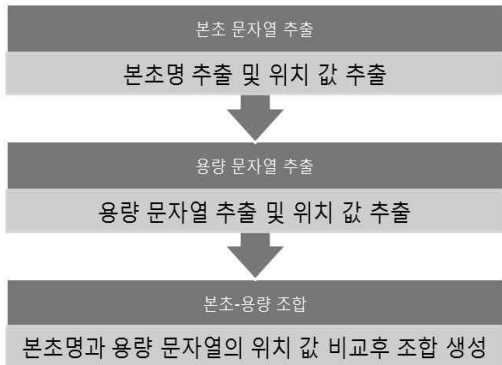


Fig. 2. Process of automatic extraction of compositional herb. 본초 조합의 자동 추출 과정

3. 원리 구현 개념 설명

<원문>龍腦芎犀丸

消風化痰，除心肺邪熱，去頭面諸風。治偏正頭痛，心忪煩郁，面熱目，鼻塞腦昏，痰熱咳嗽，咽隔不利。石膏(細研)川芎(各四兩)生龍腦(別研)生犀角山梔子(去皮，各一兩)朱砂(研飛，四兩，內一兩爲衣)人參(去蘆)茯苓(去皮，用白者)細辛(去苗)甘草(炙，各二兩)阿膠(碎炒，一兩半)麥門冬(去心，三兩)上除別研，後入外，并搗，羅爲細末，煉蜜爲丸。每服一丸至二丸，細嚼，茶，酒任下，食後服。

방제구성 검토

상위목차ID	상위	완료
14	治諸風(附腳氣)	
하위목차ID	하위	
28	龍腦芎犀丸	

원문	방제구성																																										
<div style="border: 1px solid gray; padding: 5px; margin-bottom: 5px;">하위원문</div> <p style="font-size: small;">消風化痰，除心肺邪熱，去頭面諸風。治偏正頭痛，心忪煩郁，面熱目，鼻塞腦昏，痰熱咳嗽，咽隔不利。石膏(細研)川芎(各四兩)生龍腦(別研)生犀角山梔子(去皮，各一兩)朱砂(研飛，四兩，內一兩爲衣)人參(去蘆)茯苓(去皮，用白者)細辛(去苗)甘草(炙，各二兩)阿膠(碎炒，一兩半)麥門冬(去心，三兩)上除別研，後入外，并搗，羅爲細末，煉蜜爲丸。每服一丸至二丸，細嚼，茶，酒任下，食後服。</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 30%;">본초문자열</th> <th style="width: 30%;">용량원문</th> <th style="width: 40%;">본초ID</th> </tr> </thead> <tbody> <tr><td>石膏</td><td>四兩</td><td></td></tr> <tr><td>川芎</td><td>四兩</td><td></td></tr> <tr><td>龍腦</td><td>一兩</td><td></td></tr> <tr><td>犀角</td><td>一兩</td><td></td></tr> <tr><td>山梔子</td><td>一兩</td><td></td></tr> <tr><td>朱砂</td><td>四兩</td><td></td></tr> <tr><td>人參</td><td>二兩</td><td></td></tr> <tr><td>茯苓</td><td>二兩</td><td></td></tr> <tr><td>細辛</td><td>二兩</td><td></td></tr> <tr><td>甘草</td><td>二兩</td><td></td></tr> <tr><td>阿膠</td><td>一兩半</td><td></td></tr> <tr><td>麥門冬</td><td>三兩</td><td></td></tr> <tr><td>*</td><td></td><td></td></tr> </tbody> </table>	본초문자열	용량원문	본초ID	石膏	四兩		川芎	四兩		龍腦	一兩		犀角	一兩		山梔子	一兩		朱砂	四兩		人參	二兩		茯苓	二兩		細辛	二兩		甘草	二兩		阿膠	一兩半		麥門冬	三兩		*		
본초문자열	용량원문	본초ID																																									
石膏	四兩																																										
川芎	四兩																																										
龍腦	一兩																																										
犀角	一兩																																										
山梔子	一兩																																										
朱砂	四兩																																										
人參	二兩																																										
茯苓	二兩																																										
細辛	二兩																																										
甘草	二兩																																										
阿膠	一兩半																																										
麥門冬	三兩																																										
*																																											

레코드: 1/1 레코드: 1/12

Fig. 3. Result of automatic extraction of compositional herb. 본초 조합의 자동 추출 결과

1) 본초명 및 위치 정보 추출

본초명으로 사용된 문자열을 문장에서 찾고 그 문자열이 몇 번째 문자열에서 시작되는지 정보를 추출하는 과정이다. 문자열 검색시에는 동일한 개념을 가진 문자열이 여러 가지 존재할 경우 가장 긴 문자열을 우선하여 본초명을 추출하도록 하였다. 예를 들면 山梔와 山梔子라는 본초 목록을 가지고 있으며, 그 중 원문에 山梔子라고 기록된 경우는 山梔子를 추출하도록 하였고 원문에 山梔라고만 표현되어 있을 경우에는 山梔라는 문자열을 추출하도록 하였다. 또한 본초명의 표기는 簡體와 繁體 모두 동의어 처리를 하여 검색 목록에 포함시켰다.

본초문자열	본초위치
石膏	58
川芎	65
龍腦	74
犀角	82
山梔子	85
朱砂	98
人參	116
茯苓	123
細辛	135
甘草	142
阿膠	153
麥門冬	165

Table 1. Position of herb notation.
 본초명 표기의 위치

2) 용량 표현 문자열 및 위치 정보 추출

방제 구성에서 용량 정보를 표현하는 문자열을 숫자-문자(단위 포함) 조합을 생성하여 문자열을 추출하고 위치 정보를 함께 추출한다. 용량 정보를 나타내는 문자열이 동일한 문자열을 부분적으로 포함하는 경우가 있을 경우에는 가장 긴 문자열을 용량 정보로 추출하도록 하였다. 예를 들면 三十五兩半이라는 용량 정보가 원문에 기록되어 있다면, 검색 조건에 입력된 조건중 이를 만족하는 것은 五兩, 五兩半, 十五兩, 十五兩半, 三十五兩, 三十五兩半 이라는 문자열 모두 검색이 된다. 이 중 가장 긴 문자열인 三十五兩半을 본초의 용량 정보로 활용하는 것이다.

용량 문자열	용량위치
四兩	69
一兩	94
四兩	105
一兩	110
二兩	149
一兩半	160
三兩	173

Table 2. Position of volume notation.
 용량 문자열의 위치

3) 본초명과 용량 문자열의 위치 정보 비교

본초명의 위치와 용량의 위치를 비교하여 본초명의 위치보다 나중에 위치한 용량 문자열 중 가장 앞에 위치한 용량 문자열을 찾아 해당 본초의 용량 정보로 입력한다. 위에 추출된 두 자료를 위치 정보를 중심으로 정리하면 아래와 같다.

본초명	본초위치	용량	용량위치
石膏	58		
川芎	65		
		四兩	69
龍腦	74		
犀角	82		
山梔子	85		
		一兩	94
朱砂	98		
		四兩	105
		一兩	110
人參	116		
茯苓	123		
細辛	135		
甘草	142		
		二兩	149
阿膠	153		
		一兩半	160
麥門冬	165		
		三兩	173

Table 3. Comparison between position of herb name and volume notation. 본초 문자열과 용량 문자열의 위치 비교

Table 3의 정보를 조합하면 石膏(58)와 川芎(65)보다 뒤에 위치한 용량 정보를 나타내는 문자열은

四兩(69), 一兩(94), 一兩(110), 二兩(149), 一兩半(160), 三兩(173)이 있다. 이 중 가장 앞에 위치한 문자열은 四兩(69)이다. 따라서 石膏와 川芎의 용량 정보는 四兩이라는 값을 입력한다. 龍腦(74), 犀角(82), 山梔子(85) 보다 뒤에 위치한 용량 문자열은 一兩(94), 一兩(110), 二兩(149), 一兩半(160), 三兩(173)이 있다. 이 중 가장 앞에 위치한 문자열은 一兩(94)이다. 따라서 용량 정보는 一兩의 값을 입력한다. 이후의 본초의 용량 정보 역시 동일한 방법으로 입력한다.

4. 본초 구성 조합 추출 데이터 베이스의 작성

1) 테이블 작성

(1) 내용텍스트 테이블

원문이 저장되는 테이블로 ID, 제목, 내용, 길이 필드를 가진다.

(2) 단위목록 테이블

본초 구성을 표현하는데 사용되는 단위 목록이 저장되는 테이블로 용량단위 필드를 가진다.

(3) 방제본초구성 테이블

원문으로부터 추출된 방제의 본초 구성이 최종적으로 저장되는 테이블로 방제ID, 본초문자열, 본초위치, 용량원문 필드를 가진다.

(4) 본초 표준 변환 목록 테이블

원문으로부터 본초명을 추출하기 위한 본초명이 수록된 테이블로 표준본초명, 비표준본초명 필드를 가진다. 원문에서 본초명을 추출할 때는 비표준본초명 필드의 값을 활용하여 추출한다. 비표준본초명 필드에는 본초명의 동의어, 간체와 번체 등의 경우의 수를 고려한 값들이 입력되어 있다.

(5) 용량앞꿈기 테이블

용량 표현에 필요한 숫자 부분과 단위의 조합으로 생성된 문자열에서 가장 긴 문자열을 추출하여 저장하는 테이블이다. 처방ID, 용량위치, 용량문자열

길이의최대값 필드를 가진다. 三十五兩半이라는 용량 정보가 원문에 기록되어 있다면, 검색 조건에 입력된 조건중 이를 만족하는 것은 五兩, 五兩半, 十五兩, 十五兩半, 三十五兩, 三十五兩半 이라는 문자열 모두 검색이 된다. 이 중 시작점이 같은 문자열 중 가장 긴 문자열을 가진 三十五兩半, 十五兩半, 五兩半을 용량 정보로 저장하고 나머지 일부분의 문자열 만을 포함한 것을 제거하기 위하여 임시로 생성된 값을 저장하는 테이블이다.

(6) 용량표기목록 테이블

용량 표현에 사용되는 숫자 부분 목록을 저장한 테이블이다. 六(육)과 六(륙)을 고려하여 목록을 작성한다. 용량 필드를 가진다.

(7) 임시용량 테이블

원문에 표현된 용량 표기와 관련된 내용을 모두 추출하여 임시로 저장하는 테이블이다. 처방ID, 용량원문, 용량문자열길이, 용량위치, 추출, 위치길이합 필드를 가진다.

(8) 임시처방구성 테이블

원문에 표현된 본초명과 용량 표기에 대한 추출 결과를 모두 임시로 저장하는 테이블이다. 방제ID, 본초문자열, 본초문자열길이, 본초위치, 추출, 위치길이합 필드를 가진다.

(9) 최장본초문자추출임시테이블

원문에 표현된 본초명의 여러 표현 중 가장 긴 본초명을 선택하여 임시로 저장하는 테이블이다. 방제ID, 위치길이합, 본초문자열길이의최대값 필드를 가진다. 예를 들면, 원문에 熟地黄이라는 본초명이 표현되어 있으면 熟地和 熟地黄 모두 문자열 검색에서 추출되고 그 중 가장 긴 문자열인 熟地黄을 찾기 위한 정보가 기록된 임시 테이블이다.

(10) 최장용량문자추출 테이블

용량 표현에 필요한 숫자 부분과 단위의 조합으로 생성된 문자열에서 가장 긴 문자열을 추출하여

저장하는 테이블이다. 처방ID, 위치길이합, 용량문자열길이의최대값 필드를 가진다. 三十五兩半이라는 용량 정보가 원문에 기록되어 있다면, 검색 조건에 입력된 조건중 이를 만족하는 것은 五兩, 五兩半, 十五兩, 十五兩半, 三十五兩, 三十五兩半 이라는 문자열 모두 검색이 된다. 이 중 용량앞끊기 테이블에서 수정한 三十五兩半, 十五兩半, 五兩半에서 가장 긴 문자열을 가진 三十五兩半을 용량 정보로 저장하고 나머지 일부분의 문자열 만을 포함한 것을 제거하기 위하여 임시로 생성된 값을 저장하는 테이블이다.

2) 쿼리의 구성

본 연구에서 본초명과 용량과 관련된 문자열을 추출하여 본초명과 용량 문자열의 위치 관계를 고려하여 자동으로 정보를 입력하는 과정은 삭제, 선택, 수정 및 추가 쿼리의 조합을 생성한 뒤 필요한 과정을 순차적으로 실행하는 방식으로 설계되었다. 기존 원문에 포함된 내용으로부터 필요한 정보를 추출하고 비교하는 과정을 모두 쿼리로 표현하였으며, 쿼리문의 내용은 다음과 같다.

(1) 삭제쿼리

① 00최장본초문자추출임시테이블삭제

이전 작업에서 본초명의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE 최장본초문자추출임시테이블.* FROM 최장본초문자추출임시테이블;

② 01임시용량삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE * FROM 임시용량;

③ 02임시처방구성삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리

문장은 아래와 같다.

DELETE 임시처방구성.* FROM 임시처방구성;

④ 03 용량앞끊기삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE * FROM 용량앞끊기;

⑤ 04용량문자뒤끊기

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE * FROM 최장용량문자추출;

⑥ 05최장용량문자추출삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE * FROM 최장용량문자추출;

⑦ 15중복본초명삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE 임시처방구성.추출 FROM 임시처방구성 WHERE (((임시처방구성.추출)=No));

⑧ 20-2본초명앞용량삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE 임시용량.추출 FROM 임시용량 WHERE (((임시용량.추출)=No));

⑨ 26-1용량삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE 임시용량.추출 FROM 임시용량
WHERE (((임시용량.추출)=No));

⑩ 27용량수정삭제

이전 작업에서 본초 용량의 자동 추출을 위하여 임시로 생성한 테이블의 값을 삭제하는 쿼리로 쿼리 문장은 아래와 같다.

DELETE 임시용량.추출 FROM 임시용량
WHERE (((임시용량.추출)=No));

(2) 선택쿼리

① 27본초위치기준용량제거

원문에서 본초명을 뜻하는 문자열의 위치보다 문장의 앞에 위치한 용량표기를 삭제하기 위하여 목록을 만드는 쿼리임.

SELECT 임시용량.용량위치, 임시용량.추출, 임시용량.처방ID, 임시용량.용량원문
FROM 본초위치최소값 INNER JOIN 임시용량
ON 본초위치최소값.방제ID = 임시용량.처방ID
WHERE (((임시용량.용량위치)<[본초위치최소값].[본초위치최소값]));

② 건본초문자열

원문으로부터 본초명의 다양한 표기 가운데 가장 긴 문자열을 추출하여 목록을 만드는 쿼리임.. 예를 들면 ‘熟地黃’을 뜻하는 문자열로 ‘熟地’와 ‘熟地黃’의 2 개 문자열을 등록한 경우, 원문에서 ‘熟地’라는 표기 만으로 숙지황을 표현했다면 ‘熟地’라는 문자열을 추출하고, 원문에서 ‘熟地黃’이라고 표기했다면 ‘熟地黃’이라는 문자열을 추출하기 위한 과정이다.

SELECT 임시처방구성.방제ID, 임시처방구성.본초위치, Max(임시처방구성.본초문자열길이) AS 본초문자열길이의최대값
FROM 임시처방구성
GROUP BY 임시처방구성.방제ID, 임시처방구성.본초위치;

③ 본초_용량01단계

원문에서 본초명을 뜻하는 문자열 보다 뒤에 있

는 용량 표기 전체를 추출하는 쿼리임.

SELECT 임시처방구성.방제ID, 임시처방구성.위치길이합, 임시처방구성.본초문자열, 임시용량.용량위치
FROM 임시처방구성 INNER JOIN 임시용량 ON
임시처방구성.방제ID = 임시용량.처방ID
WHERE (((임시용량.용량위치)>[임시처방구성].[위치길이합]))
ORDER BY 임시처방구성.방제ID, 임시처방구성.위치길이합;

④ 본초_용량02단계

원문에서 본초명을 뜻하는 문자열 보다 뒤에 있는 용량 표기에서 가장 앞에 있는 용량 표기의 위치 값을 추출하는 쿼리임.

SELECT 본초_용량01단계.방제ID, 본초_용량01단계.본초문자열, Min(본초_용량01단계.용량위치) AS 용량위치의최소값
FROM 본초_용량01단계
GROUP BY 본초_용량01단계.방제ID, 본초_용량01단계.본초문자열, 본초_용량01단계.위치길이합
ORDER BY 본초_용량01단계.방제ID, 본초_용량01단계.위치길이합, Min(본초_용량01단계.용량위치);

⑤ 본초위치최소값

원문에서 가장 앞에 있는 본초명의 위치 값을 추출하는 쿼리임.

SELECT 임시처방구성.방제ID, Min(임시처방구성.본초위치) AS 본초위치의최소값 FROM 임시처방구성
GROUP BY 임시처방구성.방제ID;

⑥ 용량과단위

용량 표기를 위하여 사용되는 숫자 부분과 도량형을 표현하는 문자부분의 목록을 조합하여 용량 표기에 문자열을 조합하여 생성하는 쿼리. 예를 들면 ‘三十’이라는 숫자를 표기하는 문자열과 ‘錢’, ‘錢半’, ‘兩’, ‘兩半’, ‘斤’이라는 도량형 문자 목록을 이용하여 ‘三十錢’, ‘三十錢半’, ‘三十兩’, ‘三十兩半’, ‘三十

斤'이라는 용량 표기 검색용 문자열을 자동으로 생성하는 쿼리임.

```
SELECT [용량표기목록].[용량] & [단위목록].
[용량단위] AS 용량단위 FROM 용량표기목록, 단
위목록;
```

⑦ 임시용량최대

하나의 원문에 동일한 용량표현이 2 차례 이상 표기된 경우에 각각의 표기의 위치와 본초명의 위치를 여러 차례 비교할 수 있도록 문자열의 위치를 계산하는 시작점을 새롭게 설정하기 위한 쿼리임.

```
SELECT 임시용량.처방ID, 임시용량.용량원문,
임시용량.용량문자열길이, Max(임시용량.용량위치)
AS 용량위치의최대값
FROM 임시용량
GROUP BY 임시용량.처방ID, 임시용량.용량원문,
임시용량.용량문자열길이;
```

(3) 수정쿼리

① 11본초위치합길이

본초명을 추출할 때, 본초명의 시작점과 본초명의 길이를 더한 값을 입력하는 쿼리임.

```
UPDATE 임시처방구성 SET 임시처방구성.위치
길이합 = [임시처방구성].[본초위치]+ [임시처방구
성].[본초문자열길이];
```

② 13본초명최종추출

동일한 본초에 대하여 여러 문자열이 추출되었을 경우, 가장 긴 본초명을 자동 추출되도록 설정하는 쿼리임.

```
UPDATE 최장본초문자추출임시테이블 INNER
JOIN 임시처방구성 ON (최장본초문자추출임시테이
블.방제ID = 임시처방구성.방제ID) AND (최장본초
문자추출임시테이블.위치길이합 = 임시처방구성.위
치길이합) AND (최장본초문자추출임시테이블.본초
문자열길이의최대값 = 임시처방구성.본초문자열길
이) SET 임시처방구성.추출 = Yes;
```

③ 14검색오류보정

MS 제품에서 문자열 인식에 오류가 있는 유니코드 문자열을 포함하고 있는 본초명에 대하여 실제로 표기되지 않았음에도 불구하고 위치 정보를 0으로 출력하는 경우를 추출되지 않도록 수정하는 쿼리임.

```
UPDATE 임시처방구성 SET 임시처방구성.추출
= No WHERE (((임시처방구성.본초위치)=0));
```

④ 20-1본초명뒤용량문자추출

원문에서 본초명 뒤에 위치한 용량 표기만 선택하도록 수정하는 쿼리임.

```
UPDATE 임시용량 INNER JOIN 임시처방구성
ON 임시용량.처방ID = 임시처방구성.방제ID SET
임시용량.추출 = Yes
WHERE (((임시용량.용량위치)>[임시처방구성].
[위치길이합]));
```

⑤ 20-3본초용량 초기화

하나의 방제에 여러 번 표기된 본초 용량 문자열에서 추출 대상이었던 본초 용량 문자열을 모두 초기화하여 추출되지 않도록 하는 쿼리임.

```
UPDATE 임시용량 SET 임시용량.추출 = No;
```

⑥ 21용량위치합길이

원문에서 여러 차례 동일 표기의 용량 문자열이 있을 경우, 문자열의 위치를 추출 완료된 이후의 문자열에 대하여 위치를 새롭게 추출할 수 있도록 설정하는 쿼리임.

```
UPDATE 임시용량 SET 임시용량.위치길이합 =
[임시용량].[용량위치]+ [임시용량].[용량문자열길
이];
```

⑦ 26용량최종추출

본초명과 용량 표기 문자열의 길이와 위치 값을 고려하여 생성한 본초-용량 정보를 최종 결정하는 쿼리임.

```
UPDATE (임시용량 INNER JOIN 최장용량문자
추출 ON (임시용량.처방ID = 최장용량문자추출.처
방ID) AND (임시용량.위치길이합 = 최장용량문자추
출.위치길이합) AND (임시용량.용량문자열길이 =
```

최장용량문자추출.용량문자열길이의최대값)) INNER JOIN 용량앞끊기 ON (임시용량.처방ID = 용량앞끊기.처방ID) AND (임시용량.용량문자열길이 = 용량앞끊기.용량문자열길이의최대값) AND (임시용량.용량위치 = 용량앞끊기.용량위치) SET 임시용량.추출 = Yes;

(4) 추가쿼리

① 10본초명추출

원문으로부터 본초명 검색 목록에 포함된 문자열을 기준으로 본초명 문자열과 문자열의 위치 정보를 추출하여 임시처방구성 테이블에 입력하는 쿼리임.

INSERT INTO 임시처방구성 (방제ID, 본초문자열, 본초문자열길이, 본초위치)

SELECT DISTINCT 내용텍스트.ID, [본초 표준 변환 목록].비표준본초명, Len([본초 표준 변환 목록].[비표준본초명]) AS 본초명길이, InStr(1,[내용텍스트].[내용],[본초 표준 변환 목록].[비표준본초명]) AS 본초명위치

FROM 내용텍스트 INNER JOIN [본초 표준 변환 목록] ON 내용텍스트.내용 Like "*" & [본초 표준 변환 목록].비표준본초명 & "*";

② 12최장본초문자추출

임시처방구성 테이블에서 동일 본초를 표현하는 표기 중 가장 긴 문자열을 가진 본초명 표기를 추출하여 그 길이를 저장하는 쿼리임.

INSERT INTO 최장본초문자추출임시테이블 (방제ID, 위치길이합, 본초문자열길이의최대값)

SELECT 임시처방구성.방제ID, 임시처방구성.위치길이합, Max(임시처방구성.본초문자열길이) AS 본초문자열길이의최대값

FROM 임시처방구성

GROUP BY 임시처방구성.방제ID, 임시처방구성.위치길이합;

③ 20용량표기추출

원문으로부터 용량표기를 추출하여 문자열과 위치 값을 임시용량 테이블에 저장하는 쿼리임.

INSERT INTO 임시용량 (처방ID, 용량원문, 용량위치, 용량문자열길이)

SELECT 내용텍스트.ID, 용량과단위.용량단위, InStr(1,[내용텍스트].[내용],[용량과단위].[용량단위]) AS 용량위치, Len([용량과단위].[용량단위]) AS 용량단위길이

FROM 내용텍스트 INNER JOIN 용량과단위 ON 내용텍스트.내용 Like "*" & 용량과단위.용량단위 & "*";

④ 20용량표기추출2

하나의 원문에 동일한 용량 표기가 있을 경우 2 번째 이후에 나타나는 표기를 입력하기 위한 쿼리로 사용자가 적절한 수준에서 반복 실행하여 빠짐없이 문자열을 추출하기 위한 쿼리임.

INSERT INTO 임시용량 (처방ID, 용량원문, 용량위치, 용량문자열길이)

SELECT 임시용량최대.처방ID, 임시용량최대.용량원문, InStr([임시용량최대].[용량위치의최대값]+1,[내용텍스트].[내용],[임시용량최대].[용량원문]) AS 용량위치2, 임시용량최대.용량문자열길이

FROM 임시용량최대 INNER JOIN 내용텍스트 ON (내용텍스트.내용 Like "*" & 임시용량최대.용량원문 & "*") AND (임시용량최대.처방ID=내용텍스트.ID)

WHERE (((InStr([임시용량최대].[용량위치의최대값]+1,[임시용량최대].[용량문자열길이]),[내용텍스트].[내용],[임시용량최대].[용량원문]))>0);

⑤ 23최장용량문자추출

원문에 표기된 여러 문자로 이루어진 용량 표기 문자열 중 가장 긴 문자열 정보를 저장하는 쿼리임.

INSERT INTO 최장용량문자추출 (처방ID, 위치길이합, 용량문자열길이의최대값)

SELECT 임시용량.처방ID, 임시용량.위치길이합, Max(임시용량.용량문자열길이) AS 용량문자열길이의최대값

FROM 임시용량

GROUP BY 임시용량.처방ID, 임시용량.위치길이

합;

⑥ 25용량앞끊기

원문에서 같은 위치에서 시작된 용량 표기 중 가장 긴 문자열의 길이 정보를 저장하는 쿼리임.

INSERT INTO 용량앞끊기 (처방ID, 용량위치, 용량문자열길이의최대값)

SELECT 입시용량.처방ID, 입시용량.용량위치, Max(입시용량.용량문자열길이) AS 용량문자열길이의최대값

FROM 입시용량

GROUP BY 입시용량.처방ID, 입시용량.용량위치;

⑦ 28본초_용량최종

본 연구의 최종 단계로 입시로 생성한 본초-용량 원문 정보를 방제의 구성 정보를 저장하는 방제본초 구성 테이블에 입력하는 쿼리임.

INSERT INTO 방제본초구성 (방제ID, 본초문자열, 용량원문, 본초위치)

SELECT 본초_용량02단계.방제ID, 본초_용량02단계.본초문자열, 입시용량.용량원문, 입시처방구성.본초위치

FROM 긴본초문자열 INNER JOIN (입시처방구성 INNER JOIN (본초_용량02단계 INNER JOIN 입시용량 ON (본초_용량02단계.방제ID = 입시용량.처방ID) AND (본초_용량02단계.용량위치의최소값 = 입시용량.용량위치)) ON (입시처방구성.본초문자열 = 본초_용량02단계.본초문자열) AND (입시처방구성.방제ID = 본초_용량02단계.방제ID)) ON (긴본초문자열.본초문자열길이의최대값 = 입시처방구성.본초문자열길이의최대값) AND (긴본초문자열.본초위치 = 입시처방구성.본초위치) AND (긴본초문자열.방제ID = 입시처방구성.방제ID)

ORDER BY 본초_용량02단계.방제ID, 입시처방구성.본초위치, 입시용량.용량위치;

3) 본초 구성 조합 자동 추출을 위한 쿼리의 실행 순서

Access에서 작성한 쿼리를 순서대로 실행하기

위하여 매크로를 작성하였으며, 그 순서는 아래와 같다.

(1) 이전 작업의 입시 자료 삭제 단계

① '00최장본초문자추출입시테이블삭제' 실행

② '01입시용량삭제' 실행

③ '02입시처방구성삭제' 실행

④ '03용량앞끊기삭제' 실행

⑤ '04용량문자뒤끊기' 실행

⑥ '05최장용량문자추출삭제' 실행

(2) 원문에서 본초명과 위치 값을 추출하는 단계

① '10본초명추출' 실행

② '11본초위치합길이' 실행

③ '12최장본초문자추출' 실행

④ '13본초명최종추출' 실행

⑤ '14검색오류보정' 실행

⑥ '15중복본초명삭제' 실행

(3) 원문에서 용량 표기와 위치 값을 추출하는 단계

① '20용량표기추출' 실행

② '20용량표기추출2' 필요한 만큼 반복 실행

③ '20-1본초명뒤용량문자추출' 실행

④ '20-2본초명앞용량삭제' 실행

⑤ '20-3본초용량 초기화' 실행

⑥ '21용량위치합길이' 실행

⑦ '23최장용량문자추출' 실행

⑧ '25용량앞끊기' 실행

⑨ '26용량최종추출' 실행

⑩ '26-1용량삭제' 실행

⑪ '27용량수정삭제' 실행

(4) 방제의 본초명-용량표기 조합 단계

'28본초_용량최종' 실행

5. 결과

본 연구에서는 Access2013 버전을 기준으로 연구가 진행되었기 때문에 Access 프로그램에서 지원하는 문자열 필드의 최대 길이가 255자를 넘는 경

우에 대해서는 테이블간의 연결을 이용한 검색이 불가능한 한계가 있었다. 그래서 『太平惠民和劑局方』의 모든 방제 구성에 대한 작업을 진행할 수 없었으며, 방제의 주치와 본초 구성 그리고 복약법을 포함한 내용이 255자 미만인 방제에 519종에 대해서만 연구를 진행하였다. PC는 Intel Core 2 Duo E7500 2.93GHz CPU를 사용하였으며, RAM은 3GB를 사용하였고, Windows 7 32비트 운영체제를 사용하였다. 519개 처방의 3,472개의 본초-용량 정보를 입력하는데 걸린 시간은 1분 31초였다.

이와 같은 결과는 고문헌이 텍스트 파일로 존재하는 환경에서는 방제에 대한 지식과 본초 용례에 대한 지식을 DB로 작성하는데 본초 목록과 용량 표현에 관한 목록을 만들어 문자열의 위치 정보를 서로 비교할 경우 인적자원을 절약할 수 있음을 나타내는 것이다. 이는 본초 구성 뿐만 아니라 효능과 주치 등의 정보에서도 동일한 방식으로 활용할 수 있다.

Ⅲ. 결 론

고문헌에서 방제의 본초 구성을 DB화 하는데 본초 목록과 용량 표기 문자열 목록을 확보하고 문자열 간의 위치 관계를 고려한 시스템을 구축한다면, 시스템을 통하여 효율적인 초고를 입력할 수 있어서 방제 및 본초 용례 연구에 필요한 인적 자원을 줄여 다른 영역의 연구에 투자할 자원을 마련할 수 있을 것으로 판단된다.

References

1. 박병선, 김은하, 이선아, 이병욱. 『방제학』에 기재된 방제 효능과 본초 구성을 기반으로 도출된 효능의 비교 연구. 대한한의학회지. 대한한의학회지21(1). 2008. pp.79-92. Park BS, Kim EH, Lee SA, Lee BW. The Comparative Study between the Effect of Herbal Formula in Schoolbook and the Effect deduced from Compositional Herbal Effects. The Society of Korean Medical

- Classics. Journal of Korean Medical Classics21(1). 2008. pp.79-92.
2. 이병욱, 엄동명, 김정자, 박지하, 개념중심 용어테이블을 이용한 효능 중심 본초 검색방법 연구. 대한한의학회지. 대한한의학회지20(4). 2007. pp.311-318. Lee BW, Eom DM, Kim JJ, Park JH. Searching Methods of Studying Herb Efficacy by The use of Terminology Table. The Society of Korean Medical Classics. Journal of Korean Medical Classics20(4). 2007. pp.311-318.
3. 김정훈, 이병욱. DB를 활용한 方劑의 類方分析 방법 설계. 대한한의학회지. 대한한의학회지21(1). 2008. pp.143-151. Kim JH, Lee BW. Plans on the Methodology for the Classifications of Herbal Formulas with Similar Composition using the DB Systems. The Society of Korean Medical Classics. Journal of Korean Medical Classics21(1). 2008. pp.143-151.
4. 이병욱, 정훈, 권영규. 인터넷상 동의보감 방제의 약물구성 계보 표현 방법에 관한 연구. 한국의사학회. 한국의사학회지21(2). 2008. pp.41-47. Lee BW, Jung H, Kwon YK. Study on the Method of Expressing the Genealogy of the formation of herbal treatments from Treasure Mirror of Eastern Medicines. Korean Society of Medical History. The Journal of Korean Medical History21(2). 2008. pp.41-47.
5. 오용택, 김상찬, 이병욱. 본초 구성을 이용한 방제 효능 예측 연구 및 본초의 효능 강도 설정안. 대한한의학회지. 대한한의학회지21(2). 2008. pp.49-57. Oh YT, Kim SC, Lee BW. Estimation Study of the herbal formula's effects by the compositional herbal effects (Guideline of the herbal effects intensity). The Society

of Korean Medical Classics. Journal of Korean Medical Classics21(2). 2008. pp.49-57.

6. 김현호, 홍효신, 유계혁, 권오민, 차웅석. 객체 지향형 처방 데이터베이스의 구축과 처방 검색 프로그램의 설계 및 개발. 한국한의학연구원. 한국한의학연구원논문집17(2). 2011. pp.73-83. Kim HH, Hong HS, Yoo JH, Kwon OM, Cha WS. Construction of Object-oriented Prescription Database and Design/Development of Prescription Search Program. Korea Institute of Oriental Medicine. Korea Journal of Oriental Medicine17(2). 2011. pp.73-83.