용어분해 지원도구를 활용한 약재·처방 주치 분해 연구

김안나 · 오용택 · 김상균 · 서진순 · 장현철*

한국한의학연구원 정보개발운영그룹

Disassembling Indication Terms of Medicinal Treatment Using a Cooperation Tool

An Na Kim, Yong Taek Oh, Sang Kyun Kim, Jin Soon Seo, Hyun Chul Jang*

Informatics Development & Management Group, Korea Institute of Oriental Medicine

We are building an Korean Medicine ontology and disassembling indication terms of medicinal materials and formulas. When we disassemble these terms, an expert can disagree about another expert's result, so participants have to consider another opinion and verify each other concurrently. To achieve this, we have developed a web-based tool that support users in refining terms, disassembling them into the minimum meaning efficiently. With this tool, plural participants refer to each other and make use of this as a verification method. This would have enabled participants to achieve consensus. Due to the nature of the traditional Korean Medicine knowledge which have rich implications and is complicated,E there would be a various result of refinement, disassembly, analysis and another processing. This tool assists users in minimizing the differences and maintaining objectivity by agreement. We have refined and analyzed 4,756 terms from materials and formulas indication in Korean Medicine ontology using this tool. And we concluded that disassembled data increases the percentage of linkage between diseases and medicinal treatments and established an infrastructure for the extensible knowledge.

Key words : traditional korean medicine ontology, disassembling indication terms, medicinal treatment term, cooperation tool

서 론

주치는 한의학적 치료에 있어서 주로 효능을 발휘하는 치료 범위 및 대상을 지칭하는 용어로, 약재 및 처방, 경혈 등은 각기 해당하는 주치를 가지고 있는데, 문헌에서는 주로 "治***"의 형태로 기술되어 있으며, 예를 들면 治心肺氣不足 咳嗽喘促 肢體痿弱 脚軟眼黑 津液枯固 또는 治少陽病 半表半裏 往來寒熱 能和其內熱 解其外邪 傷寒方之 王道也[1] 등이다.

이러한 주치정보는 咳嗽喘促, 肢體痿弱, 脚軟眼黑 등과 같이 많은 경우 주치병증을 의미하지만 心肺氣不足, 津液枯固, 少陽病 半表半裏, 和其內熱, 解其外邪 등과 같이 주치정보 내에 병, 증상, 주치대상, 병인, 병기, 효능, 치법 용어 등을 다양하게 포함하고 있는 복합구조로서, 한의사가 문헌내의 주치정보를 활용할 때에는 적용하고자 하는 상황에 따라 전문지식을 활용하여 다양한 복합적인 주치구조의 내용을 정제 및 재분류한 후 병, 증상, 주치대상, 병인, 병기, 효능, 치법 등 원하는 정보를 얻고 관련 병증과 연계하는 일련의 과정을 직관적으로 종합 판단하게 된다. 그런데 이러한 주치정보를 데이터베이스로 구축하고 활용하기 위해서는 서적으로부터 추출된 지식정보에 대해서 의미단위로 정제 및 분해, 분석이 이루어져야 한다. 즉, 한의사는 직관적으로 판단할 수 있었던 문헌지식들을 IT기반의 시스템에서 활용하기 위해서는 각 용어들이 의미단위로 정제, 분해 되어야 하는데, 정제, 분해된 데이터들은 단순 문자열 일치 검색의 한계나 유사문 검색으로 인한 garbage 데이터 검색의 문제점 등을 해결하여, 실용화 시스템 내에서 검색 정확도를 높이는 등 정보의 활용성 및 효율성을 높일 수 있다.

온톨로지는 용어와 용어 사이의 관계를 명세하고 컴퓨터가 이해할 수 있는 형태로 정의한 것으로 한의학 분야의 경우 병증과 약재, 처방, 경혈, 침구를 중심으로 한 한의학 온톨로지 구축 연구가 한국한의학연구원에서 수행되고 있다[2]. 주치정보는 약재, 처방, 경혈, 침구온톨로지를 병온톨로지와 직,간접적으로 연결할 수 있

* 교신저자 : 장현철, 대전시 유성구 유성대로 1672 한국한의학연구원

· E-mail : hcjang@kiom.re.kr, · Tel : 042-868-9364

· 접수 : 2013/05/03 · 수정 : 2013/06/26 · 채택 : 2013/06/26

는 프로퍼티 중 하나이며, 병증을 기반으로 한 통합온톨로지 내 온톨로지간 연계의 핵심 요소 중 하나이다. 이러한 주치는 앞서 설명한 것과 같이 실제 병인, 병기, 증상, 병명 등이 혼재되어 있으며, 병인과 증상, 병기와 증상, 병명과 수반증상, 증상과 양상, 여러 증상들의 결합 등 병렬, 인과, 종속과 같은 다양한 구조를 가지고 있으므로, 주치데이터의 정제, 분해는 각 개념을 하위 세부개념으로 분해함으로써 온톨로지내에서 병증과 약재, 처방 경혈 지식 간의 연계성을 높일 뿐 아니라, 다양한 시스템에서 온톨로지로의 접근을 용이하게 하고, 향후 한의학지식을 단순한 문자열 기반 연계의 한계를 넘어 의미기반으로의 지식 간 연계에 활용할 수 있게 하므로, 한의학 전문가가 다양한 구조로 이루어진 주치용어를 최소의미단위로 분해하는 연구가 반드시 필요하다.

한편, 주치 내 최소의미단위 용어들은 같은 상황의 용어라도 증상의 개념화 및 데이터를 어떻게 활용할 것인지의 방향에 따라 전문가 간의 용어 분해 및 분석수준이 달라질 수 있으며, 많은 한의학서적이 한자로 표기되어 있기 때문에 복잡하고 함축적인 한의학지식 특성상 문자열이 같더라도 문맥에 따라 혹은 개별 데이터의 내용에 따라서도 여러 의미를 가질 수 있다. 즉, 최소의미단위에 관한 객관적인 기준설정이 어렵고, 전문가에 따라 같은 주치를 다르게 분해할 수 있기 때문에, 이러한 작업에는 되도록 여러 명의 전문가가 참여하고 다른 전문가의 의견을 실시간으로 참고해서 주치분해에 적용하는 등 객관성 확보의 필요성이 있다. 또한 수천 개의 용어 데이터 정제, 분해 시 개개인의 인지력을 보조하여 좀 더 일관적이고 효과적으로 연구를 수행할 수 있는 지원도구가 필요하다고 판단되어진다.

기존의 관련 연구를 살펴보면, 데이터를 정제, 분해하고 데이터 간 연계율을 높이려는 연구는 꾸준히 있어왔다[3-5]. 기 연구들은 지식 간 연계율을 높이기 위해 데이터를 정제, 분해한다는 점은 의의가 있으나 정제, 분해한 결과에 있어 객관적인 기준을 제시하지는 못하는 한계가 있다. 또한 치법용어 및 효능용어에 대한 데이터 정제 분해 및 대표술어 변환에 관한 오[5]등의 연구가 있으나, 증상용어와 관련된 해당 연구는 미비하다. 그리고 치법용어 및 효능용어는 주치용어에 비해 상대적으로 구조와 의미가 명확하나, 주치의 경우 그 복합구조로 인하여 다수의 전문가의 합의에 기반한 최소의미단위 분해의 객관성 확보가 더욱 필요하다.

본 연구에서는 함축적이고 복잡한 한의학지식의 특성상, 이를 해석 및 정제, 가공 시 다양해 질 수 있는 용어분해결과에 대하여, 여러 전문가의 합의를 기반으로 귀납적으로 객관성을 확보하고, 관련 이슈사항 발생을 최소화하기위하여 용어분해 지원도구를 개발하였으며, 현재 한국한의학연구원에서 구축한 약재,처방 온톨로지 내의 약재 처방의 주치 4756건에 대하여 지원도구를 활용하여 주치용어 분해를 수행함으로써 의미단위결과에 대하여 객관성을 확보하고자 하였다.

연구대상 및 방법

주치데이터의 정제, 분해결과에 대한 보편성과 객관성을 확보하고, 용어 분해, 정제를 효율적으로 지원하기 위하여, 복수의 연구자가 동시에 작업할 수 있는 환경을 기반으로 하여 용어 정제, 분해하는 워크플로우를 설계, 웹기반의 주치용어 분해 지원도구를 개발하였다.

개발언어는 JAVA, JSP, 트러스트폼 (TrustForm)을 사용하였고 DB는 오라클을 사용하였다. 서버 사이드는 JAVA, JSP를 이용하여 데이터를 처리하였으며, 클라이언트 사이드는 트러스트폼을 이용하였다. X인터넷 솔루션인 트러스트폼 (TrustForm)[6]은 클라이언트/서버(C/S)환경의 풍부한 기능을 웹상에서 구현할 수 있을 뿐 아니라 최소한의 데이터 전송으로 네트워크 운영경비를 낮춰 업무생산성을 향상시킬 수 있다. 또한 웹이 내세우는 장점인 배포와 관리의 용이성을 제공함으로써 보다 효과적으로 웹을 개발하고 관리할 수 있게 해준다.

용어의 정제, 분해 지원도구의 유용성 및 활용성 검토를 위해, 한국한의학연구원에서 구축한 약재, 처방 온톨로지 내의 약재의 주치용어 3627건과 처방 주치용어 4256건에 대하여 동일문자열 일치에 대한 중복제거 등의 정제를 거친 총 4756건을 대상으로, 개발된 지원도구를 활용하여 복수의 한의학전문가가 주치데이터를 의미단위로 정제, 분해 및 상호 검증을 통해 작업을 하였다.

결 과

1. 용어 정제, 분해 지원도구 시스템 개발

시스템의 프로세스 개요는 Fig. 1과 같으며, Fig. 2는 로그인 화면과 메인화면의 일부를 나타낸 것이다. 주요 기능 및 특징은 다음과 같다.

1) 웹 기반 주치용어분해 지원 시스템

인터넷이 접속되는 환경에서는 어디서나 사용할 수 있다. 기존 방식처럼 개인별 작업 중인 엑셀파일을 따로 관리하지 않아도 된다.

2) 여러 사용자 동시 사용 가능

여러 사용자가 동시에 주치분해작업을 진행할 수 있어 다른 사용자가 작업한 내용을 실시간으로 확인하고 참고 할 수 있다. 또한 엑셀로 각자 작업했을 때 통합하는 과정을 하지 않아도 되어, 시간과 오류를 줄일 수 있다.

3) 사용자 간의 주치분해 결과, 상태, 비고, 수정일 비교 기능

여러 사용자가 작업한 주치분해 결과, 상태, 비고, 수정일을 비교할 수 있는 화면을 제공함으로써 각 내용간의 공통점, 수정할 정보 등을 쉽게 찾을 수 있다.

4) 사용자 권한에 따른 수정대상 설정 기능

사용자는 권한이 주어진 분해정보만을 수정할 수 있다. 단, 관리자 권한을 부여 받은 사용자는 관리자 화면에 접속하여, 다른 사용자가 작업한 분해정보를 수정할 수 있다.

5) 하나의 주치정보에 대해 3개의 분해 결과 입력 기능

하나의 주치정보에 대해 최대 3개까지 분해정보를 관리할 수 있으므로 여러 사용자가 하나의 주치정보에 대해 각자 분해결과를 저장할 수 있다.

6) 단축키 제공

• F2, '/', 1 등의 단축키를 제공하여 주치분해작업을 빠르고 쉽게 할 수 있다.

• 분해결과 마지막에 1을 붙여주면 상태가 보류로 바뀐다.

• 단축키 F2를 누르면 주치분해 대상 정보가 결과에 복사된다.

• 분해결과가 '/'로 끝나면, 병증 여부를 N으로 자동 인식한다.

• 주치정보 조회 시 검색어를 ','로 분리해서 조회버튼을 선택하면 최대 3개의 탭으로 나누어진 결과를 보여준다.

7) 상태정보(완료, 결과다름, 보류, 초기)별 주치 조회 기능

• 완료 : 동일분석자의 작업이 완료되었거나, 분석자간 주치분해가 일치된 경우의 최종상태완료에 대하여 각 완료결과들을 조회할 수 있다.

• 결과다름 : 하나의 주치분해대상에 대해 작업자 간 분해 결과가 하나라도 다른 경우 '결과다름'이 되며, 이러한 리스트들을 조회할 수 있다.

• 보류 : 하나의 주치분해대상에 대해 한 사용자라도 보류라고 입력하면 조회된다.

• 초기 : 작업되지 않은 주치대상이 조회된다.

8) 최종 의미단위결과 추출 및 병증여부, 개수 정보 조회 기능

최종단위결과, 병증여부, 개수를 보여주는 화면과 주치 하나당 최종결과, 병증여부, 개수, 상태, 분해결과 화면을 제공한다.

9) 주치분해 결과 정보의 엑셀 내보내기 기능

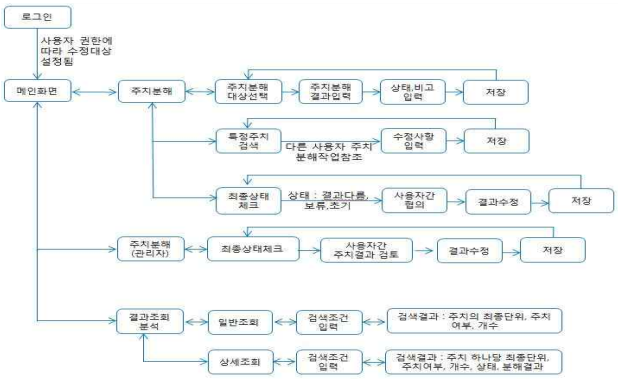

Fig. 1. Process of Disassembling indication terms of medicinal treatment.

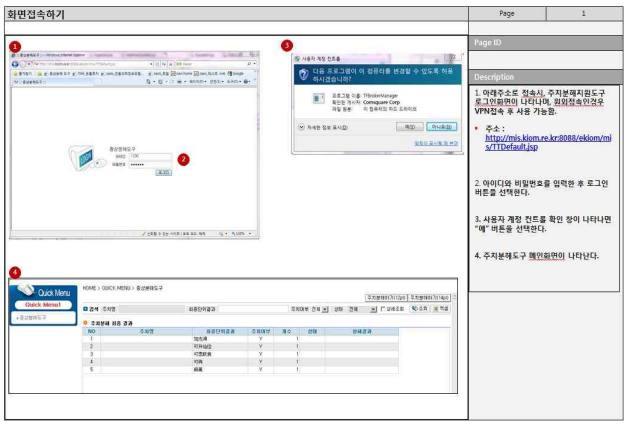

Fig. 2. Login and main screen.

2. 분해지원도구를 활용한 약재처방 주치용어 정제, 분해

용어 정제 분해 지원 도구의 메인화면에서 주치분해 기능을 선택하면 주치분해 화면으로 들어갈 수 있다. 각 분석자는 주치분해 시 각 해당 비고란에 이슈 및 분석 사항을 입력할 수 있으며, 분해결과를 포함한 모든 내용에 대하여 다른 분석자들이 이를 참고할 수 있다. 단, 입력과 수정권한은 본인의 작업내용만 가능하고, 다른 사람의 입력내용에 대하여 수정 권한은 없으며, 관리자 모드에서는 모든 분석자의 데이터에 대하여 수정, 삭제 권한이 있다. 온톨로지 내의 본초, 처방의 주치에 대하여 한의학 전문가3명이 최소의미단위 처리의 공통된 기본 지침을 바탕으로, 개발된 분해지원도구를 활용하여 주치용어를 정제, 분해하였다.

1) 동일분석자의 주치용어 정제, 분해

Fig. 3은 메인화면에서 주치분해하기 기능을 선택했을 때 보여지는 화면으로, 주치용어인 咽乾口渴에 대한 실제 정제, 분해 화면이다. 분해화면은 크게 주치리스트와 검색 화면의 2단 평행 구성으로 되어 있다. 주치리스트 화면에서는 咽乾口渴에 대하여 咽乾, 口渴을 각각 의미단위로 분석한 결과를 분해결과란에 입력할 수 있으며, 분해결과항목 뒤의 각 숫자들은 해당작업자의 번호를 의미한다. 의미단위로 분해한 이후 하단의 검색화면에서 의미단위결과인 咽乾과 口渴을 각각 포함하고 있는 모든 주치 리스트를 검색할 수 있어서 일관적인 기준을 적용하여 한번에 관련 주치들을 효율적으로 분해할 수 있도록 지원한다.

메인화면과 마찬가지로 주치분해 화면 상단에 위치한 검색 기능을 통해 주치, 분해결과의 최종상태, 분석내용 등을 검색, 조회할 수 있으며, 해당 조회결과를 엑셀파일형태로 저장할 수 있다.

Fig. 3. Disassembling "咽乾口渴".

2) 다수분석자의 주치용어 정제, 분해

Fig. 4는 다수분석자의 주치용어 정제, 분해 시 분해결과가 일치하는 경우의 예이다. 咽乾口渴의 예를 보면 분해결과가 咽乾, 口渴로 분석자1과 분석자3의 결과가 일치하므로 주치분해의 최종상태가 완료로 표시되어진다. 비교적 명확한 구조의 주치데이터이므로 전문가들 간 이견이 없어 최종분해결과가 완료라고 표시되어 있다.

Fig. 4. Example of agreements among experts(咽乾口渴).

Fig. 5. Example of different results among experts(風熱頭痛).

Fig. 5는 다수분석자의 주치용어 정제, 분해 시 분해결과가 일치하지 않은 예이다.

風熱頭痛의 경우 病因이 風熱이며 동시에 風熱證으로 辨證이 완료된 상태로, 임상상 진단의 관점에서 보면 風熱頭痛은 최소의미단위일 수 있다. 한편 데이터 연계 관점에서는 風熱과 頭痛으로 각각 분해하면 관련병증, 약재, 처방과의 확장 연계율이 높아지고, 향후 병인병기와 병증 데이터로 각각 의미 분류가 가능해지므로, 風熱頭痛의 경우 전문가 간 관점 차이로 인해 다르게 분해 될 수 있는데, 작업자1이 風熱을 포함한 용어를 일괄 정제, 분해하는 과정에서 風熱頭痛을 하나의 최소의미단위로 처리한 이 후에, 작업자3이 Fig. 5와 같이 頭痛을 포함한 용어를 일괄 정제, 분해하는 경우, 風熱頭痛이 頭痛을 포함한 용어이므로 Fig. 5에서 하단의 頭痛 탭에서처럼 風熱頭痛이 분해대상에 포함되어 보여진다. 작업자3은 용어 분해 시 작업자1이 風熱頭痛을 최소의미단위로 이미 분해한 결과를 참조할 수 있다. 이때 작업자3이 風熱頭痛에 대해 앞서 설명한 이유로 風熱과 頭痛로 각각 분해한다면, 작업자1과 작업자3의 분해 값이 일치하지 않으므로, 최소의미단위의 최종분해상태는 “결과다름”이 된다. 이러한 “결과다름” 리스트들은 Fig. 6과 같이 필터링하여 한꺼번에 리스트업 한 후 다양한 분야의 전문가 간 논의를 거쳐 목적에 맞는 주치분해의 객관적 합의점을 찾을 수 있게 된다. 논의를 거쳐 조정된 결과는 관리자화면에서 수정, 적용하게 되고 이후 최종결과값은 완료로 바뀌게 된다. 즉, 주치분해 도구를 활용하여 주치분해를 하는 경우, 복수의 작업자들은 다른 전문가의 의견을 참고함과 동시에 상호 검증방법으로 활용하면서, 분해 결과에 대한 객관적 합의를 이끌어 낼 수 있다. 또한, 각 상태항목에서 이슈사항처리가 필요하여 분해처리 보류한 주치들 역시 일괄적으로 보류리스트를 검색하여 전문가 논의 및 자문 등을 통하여 적절한 결론을 얻은 후 관리자 모드에서 이를 일괄 적용할 수 있다.

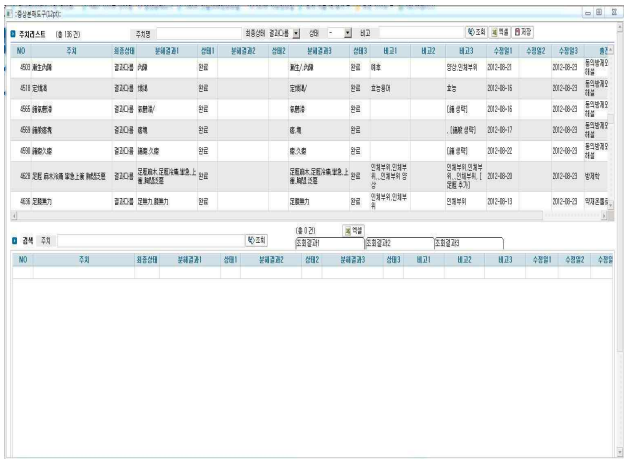

Fig. 6. Example for list of different results.

고 찰

한의학지식의 경우 관련 문헌들에서 대부분 한자로 기록되어 있고, 의학용어이므로 사용되는 용어 체계가 방대하며, 오랜 세월 경험적지식이 추가 되는 형태로 지식이 계승되어 왔기 때문에 기술된 용어가 상당히 복잡한 특징을 지닌다. 이러한 지식을 온톨로지와 같은 IT기술을 기반으로 데이터베이스화하고 실용화 시스템에서 활용하기 위해서는 전문가들이 직관적으로 이해하던 지식을 분해하고 구조화하는 연구가 필요하다.

본 연구에서 개발, 활용한 주치용어 정제, 분해 도구의 주요 목적은 함축적인 한의학지식에 대하여 여러 전문가가 용어를 정제, 분해함에 있어서 그 결과에 대해 전문가의 합의를 이끌어 내어, 되도록 객관적인 최소의미단위의 용어 분해 결과를 얻기 위한 것이다.

한의학연구원의 온톨로지 연구에서는 이 시스템을 이용하여 본초, 처방의 주치데이터를 의미 단위로 분해하였다. 분해한 데이터를 온톨로지를 시각화하여 보여주는 gruff 프로그램(4.1.5 버전)을 통하여 확인해보면, 咽乾口渴의 경우 문헌 원문에 충실한 온톨로지를 구축하였을 때 口渴로 접근할 수 있는 처방과 병증은 각각 1건과 27건 인데 반해(Fig. 7), 연구결과를 적용하면 口渴에 대하여 약재 25건, 처방 11건, 병증 48건이 口渴로부터 접근될 수 있으므로(Fig. 8)[7], 단순 연계율로만 계산했을 때 咽乾口渴의 완료된 분해결과 적용 후 온톨로지 내 지식 연계율은 27건에서 84건으로 300%이상 향상되게 된다. 이와 같이 작업되어진 결과

데이터들은 의미단위로 지식을 확장연계하기 위하여 주치용어를 증상용어와 비증상용어로 1차 분해 결과이며, 각 증상별 특성을 고려해야하는 증상개념화를 위한 2,3차 분해 연구 및 주치 간 관계 표현연구 등은 본 논문내용에서는 포함되어 있지는 않는다.

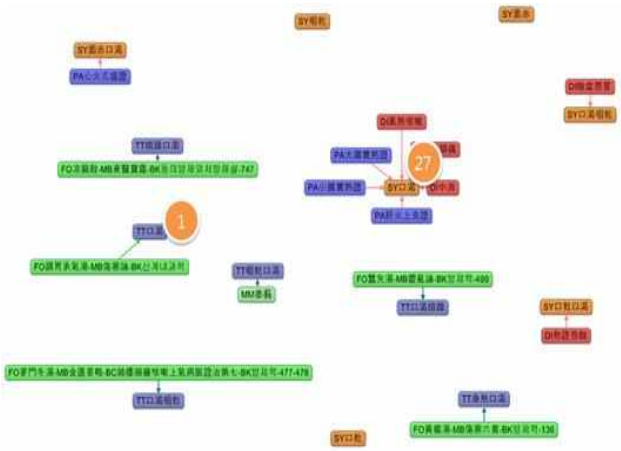

Fig. 7. Data connection before disassembling terms.

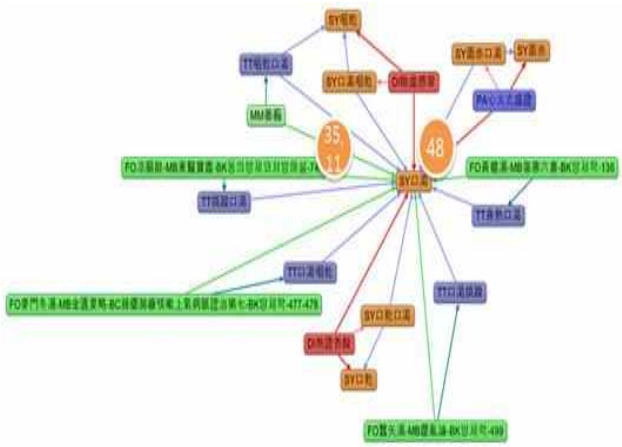

Fig. 8. Data connection after disassembling terms.

전문가의 합의에 기반한 분해결과 데이터는 높은 신뢰도를 제공하며, 이와 같이 병증과 약재, 처방, 침구와의 연계율을 높여 사용자에게 보다 확장된 데이터의 연계를 보여 줄 수 있고, 향후 실용화 시스템의 서비스방향에 따라 활용도도 높아지게 한다. 이러한 도구의 활용은 한의지식 처리에 있어서 한의전문지식의 특성 및 문헌 내 한자표기의 함축적 표현으로 인한 주관적 해석을 가급적 배제하고, 한의지식 해석의 객관적 기준을 제시하는 기초연구의 일환으로 의의가 있다고 할 수 있다. 또한 효능, 치법용어 등 한의학 용어데이터를 분해, 분석하는 연구를 효율적으로 지원할 수 있으며, 현재 수작업으로 진행한 검수, 검증작업에 대한 기능을 추가함으로써 주치분석 검증도구로 확장할 수 있다. 그리고 이러한 연구를 통한 용어데이터들은, 향후 한의학 문헌 자료들을 DB로 구축함에 있어서 1차 자동으로 용어를 분해, 분석하는 등의 기능을 통해 전문가의 작업을 도움으로써 데이터마이닝 등 인포매틱스 기술을 이용한 한의학 정보화 연구에도 기여할 수 있다고 하겠다.

결 론

이 연구에서는 주치데이터를 정제하고, 최소의미단위로 분해하는 연구를 효율적으로 지원하기 위하여 웹기반의 용어 분해 지원 시스템을 개발하였다. 이 시스템을 활용하여 주치분해를 하는 경우, 복수의 작업자들은 다른 전문가의 의견을 참고함과 동시에 상호 검증방법으로 활용하면서, 분해 결과에 대한 객관적 합의를 이끌어 낼 수 있다. 이러한 도구는 함축적이고 복잡한 한의학지식의 특성상, 이를 해석 및 정제, 가공 시 다양해 질 수 있는 용어분해결과에 대하여 전문가의 합의를 기반으로 귀납적으로 객관성을 확보하고, 관련 이슈사항 발생을 최소화 할 수 있게 지원한다. 우리는 개발된 시스템을 활용하여 현재 한국한의학연구원에서 구축한 한의학 온톨로지 내의 본초, 처방의 주치데이터 4756건에 대하여 정제, 분석하였으며, 분해된 주치 데이터는 약재 처방과 병증의 연계율을 높여 지식간의 연계가 확장 될 수 있는 기반을 마련하였다.

감사의 글

본 연구는 한국한의학연구원 “온톨로지 기반 한의학 지능형 정보체계 연구”(과제번호:K13090) 과제의 지원을 받아 수행되었습니다.

참고문헌

1. 尹用甲. 圖解增補 東醫方劑와 解說. 의성당. p 168, 620, 2007.
2. Hyunchul Jang, Jinhyun Kim, Sang-Kyun Kim, Chul Kim, Soon-Hee Bae, Anna Kim, Dong-Myung Eum, Mi-Young Song. Ontology for Medicinal Materials Based on Traditional Korean Medicine. Bioinformatics. pp 2359-2360, 2010.
3. 오용택, 이병욱, 김은하. 본초 효능 용어에 관한 연구. 대한한의학원전학회지 23(5), 2010.
4. 이병욱. 한의 진단 처방 지원시스템을 위한 용어 연계에 관한 연구. ㈜PNU동제메디칼. 2012.
5. 오용택, 전병욱, 장현철, 김창석, 김진현, 김상균, 송미영. 대표 술어로의 변환을 통한 치법 용어의 본초, 처방의 효능 용어 연계에 관한 연구. 대한동의생리병리학회지 24(5):99-105, 2011.
6. TrustFormSystem. http://www.comsquare.kr/index.asp
7. 장현철. 온톨로지 기반 한의학지능형 정보체계 연구(4차년도). 한국한의학연구원. p 108, 2012.