

한의학 용어 수집 및 관리 시스템 구축

차승준¹⁾ · 김상균^{2)†} · 송미영²⁾ · 최윤정¹⁾ · 엄동명^{4) 5)} · 이병욱³⁾ · 이규철^{1)*}

¹⁾충남대학교 컴퓨터공학과, ²⁾한국한의학연구원 정보연구센터,
³⁾부산대학교 한의학전문대학원 인문사회의학부,
⁴⁾원광대학교 한의과대학 원전학교실, ⁵⁾원광대학교 한국전통의학연구소

Development of Collecting and Managing System for Terminologies of Korean Oriental Medicine

Seung-Jun Cha,¹⁾ Kim Sang-Kyun,²⁾ Song Mi-young,²⁾
Yun-Jeong Choi,¹⁾ Dong Myung Eom,^{4) 5)}
Byung Wook Lee³⁾ & Kyu-Chul Lee^{1)*}

¹⁾Department of Computer Engineering, Chungnam National University

²⁾Information Research Center, Korea Institute of Oriental Medicine

³⁾Division of Humanities and Social Medicine, School of Korean Medicine, Pusan National University

⁴⁾Division of Oriental Medical Classics, College of Oriental Medicine, Wonkwang university

⁵⁾Research Center of Traditional Korean Medicine, Wonkwang University

Abstract

Background: The system for collecting and managing the terminologies of Korean oriental medicine is essential to facilitate researches of developing intelligent information architecture based on ontology.

Objectives: The purpose of this paper is to develop system for collecting and managing system for terminologies of Korean oriental medicine.

Methods: First we defined roles of users used in this system, and define the system workflow based on them. Then we developed web-based collecting and managing system in which multiuser can work simultaneously, and it is compose of relational database system.

· 접수: 2010년 4월 5일 · 수정접수: 2010년 4월 20일 · 채택: 2010년 4월 22일

* 교신저자: 이규철, 대전광역시 유성구 궁동, 충남대학교 컴퓨터공학과

전화: 042-821-6658, Fax: 042-822-4997, 전자우편: kclee@cnu.ac.kr

† 공동저자: 김상균, 305-811 대전광역시 유성구 엑스포로 483 한국한의학연구원 정보연구센터

전화: 042-861-1994, Fax: 042-861-5800, 전자우편: skkim@kiom.re.kr

Results : The terminologies of Korean Oriental Medicine can be efficiently collected, purified, inspected, and utilized by using proposed system.

Conclusions : The system, proposed in this paper, can be used the whole field of oriental medicines for basic technology and has great contributions to make intelligent information architecture of oriental medicine based on ontology concept.

Key words : Collecting system, Managing system, Korean oriental medicines, Terminology

I. 서론

한의학은 우리나라 고유의 의학으로 21세기 노령화 사회의 도래와 성인병 등의 질병구조에 대처하기 위한 많은 연구들이 진행되고 있다. 특히 한의학은 발전과 변화의 속도가 매우 빠른 학문 영역일 뿐만 아니라 인접 학문의 용어와도 관련성이 크기 때문에, 사용되는 용어 체계가 방대하고 용어관리가 상당히 복잡한 특징을 가진다. 이러한 복잡한 용어를 효율적으로 활용하고자 한국한의학연구원에서는 온톨로지를 기반으로 한의학 지능형 정보체계에 연구를 진행하고 있다.^{1) 2)} 이를 위해 다양한 출처에서 사용되고 있는 방대한 양의 용어들에 대한 정리가 선행되어야 한다.

한의학 용어들은 고서적들에서 다양한 형태로 존재하고 전수되어지고 있다. 기존의 한의학 용어 관리는 대부분 사전편찬이나 용어집 발간 형식을 통해 이루어졌다. 이를 위해 각각의 자원으로부터 용어가 개별적으로 수집되어야 하며 수집된 용어는 총괄자가 파일을 합치는 형태의 작업이 이루어진다. 이러한 방법은 용어 수집의 결과를 최종 총괄자의 능력에 의존하기 때문에 용어를 전체적으로 관리하거나 수집된 용어를 지속적으로 관리하기 어렵다는 한계가 있다.³⁾

이를 위해 웹기반의 한의학 용어관리 시스템¹⁾

을 구축하였다. 이는 개념 중심으로 용어를 수집하며, 수집자, 정제자, 검수자의 사용자를 정의하여 용어관리 시스템을 정의하였다. 또한 용어의 수집을 위해 사용자가 데이터베이스에 대한 전문적인 지식이 없어도 웹 브라우저를 이용해 서버에 접근하여 용어를 수집하고 관리할 수 있는 웹사이트를 제공한다. 하지만 사용자 역할 및 전체적인 워크플로우가 정의되어 있지 않아 정제자, 검수자들이 코드 등 특성을 부여하는 과정에서 혼란이 야기될 수 있으며, 또한 서로의 판단을 확인하고 논의하기 힘든 한계가 있다.

본 연구에서는 이러한 문제점을 해결하기 위해 기존 시스템을 확장하여 웹기반의 한의학 용어 수집 및 관리 시스템을 구축하였다. 우선 사용자의 역할에 따라 수집자, 정제자, 검수자, 관리자, 사용자를 구분하여 정의하여 각 사용자별로 필요한 요구사항을 처리하였으며, 정의한 역할에 따라 워크플로우를 정의함으로써 방대한 양의 용어들에 대해 정제자, 검수자의 판단의 차이를 최소화하여 효과적인 입력이 가능하도록 하였다.

개발된 시스템을 통해 다양한 출처에서 각각 사용되는 용어를 정리함으로써, 한의학 전 분야에 걸쳐 적용 개발되고 활용 가능한 기반 기술로 제공할 수 있다. 향후 전자 건강 기록(Electronic Health Records) 및 전자 의료 기록(Electronic Medical Record)과 같은 첨단 한의 의료 정보화를 위한 기반 기술로 사용할 수 있다.

또한 전문화된 한의학 지식을 효과적으로 저장

및 관리함으로써 온톨로지 기반의 한의학 지능형 정보체계에 큰 기여를 할 것이며, 추후 일반 사용자들도 이 시스템을 활용하여 한의학 관련 연구를 효과적으로 진행할 수 있다.

II. 본 론

1. 역할 및 시스템 워크플로우 정의

시스템을 구현하기에 앞서 시스템을 사용하는 사용자의 역할 및 이에 따른 워크플로우의 정의가 선행되어야 한다.

1) 역할(role) 정의

한의학 용어 수집 및 관리 시스템은 크게 용어의 수집, 정제, 검수단계로 구성되며, 또한 수집이 완료된 용어를 검색할 수 있는 시스템이 있다. 이에 따라 사용자의 역할을 수집자, 정제자, 검수자, 관리자, 사용자로 정의하였다.

정제자, 검수자, 관리자, 사용자로 정의하였다.

수집자는 한의학 서적으로부터 용어를 추출하고 수집하여 시스템에 입력하는 역할을 수행한다. 정제자는 수집자들이 수집한 용어에 대해서 수정 및 삭제의 역할과 용어를 관리하기 위한 코드를 부여하는 역할을 수행한다. 검수자는 정제자가 정제한 용어들을 대상으로 정제가 올바르게 수행되었는지 검수하여 용어 수집을 완료할 수 있다. 관리자는 시스템의 전반적인 사항을 제어해주는 역할로 사용자의 계정관리, 용어관리, 작업현황 관리의 역할을 수행한다. 사용자는 검수까지 끝나 완료된 용어를 대상으로 용어별, 분류별 검색을 수행한다.

2) 시스템 워크플로우 정의

시스템 내의 사용자들이 각자의 역할을 효과적으로 수행할 수 있도록 시스템의 진행 단계를 Figure 1.과 같이 7단계로 정의하였다.

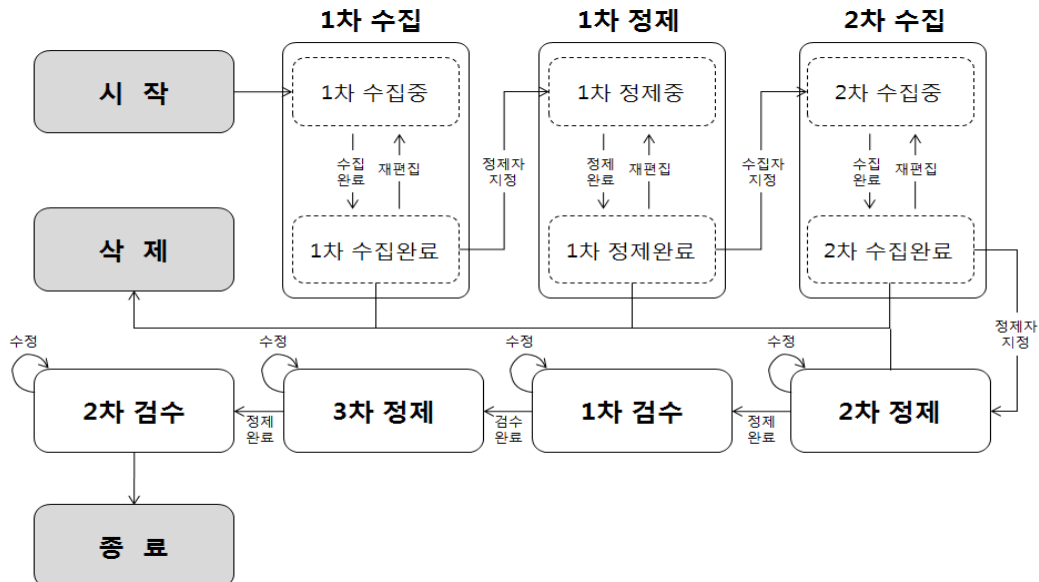


Figure 1. System Workflow

(1) 1차 수집

1차 수집에서 수집자는 한의학과 관련된 서적들에서 용어를 추출하여 시스템에 입력한다. 용어를 입력하기 위해서는 시스템에 입력된 용어를 대상으로 한글 또는 한자로 중복검색을 선행해야 한다.

1차 수집에서는 ‘한글’, ‘한자’, ‘분류1’, ‘출전’, ‘쪽수’, ‘용례’, ‘이미지’, ‘대표어’ 항목이 입력된다.

용어의 수집이 완료되지 못하고 수집자에 의해 임시 저장된 용어는 1차 수집중의 리스트에 저장되며, 수집이 완료된 용어는 1차 수집 완료 목록에 저장된다. 1차 수집 완료 목록에 저장된 내용은 관리자에 의해 정해진 1차 정제자에게 전달되며, 전달되지 않은 용어들은 수집자에 의해 재수정이 가능하다.

(2) 1차 정제

1차 수집이 완료되어 관리자에 의해 정해진 정제자에게 전달된 용어에 대해 1차 정제단계가 수행된다.

1차 정제에서는 1차 수집이 완료된 용어들을 바탕으로 목록 리스트에서 중복용어를 판단해 삭제하며, ‘한글’, ‘한자’, ‘분류1’, ‘용례’ 항목에 대해 잘못된 사항을 수정한다. 또한 CUI (Concept Unique Identifier) 대상 용어들을 선별하여 2차 수집으로 전달한다.

1차 수집과 마찬가지로 용어의 정제가 완료되지 못하고 정제자에 의해 임시 저장된 용어는 1차 정제중의 리스트에 저장되며, 정제가 완료된 용어는 1차 정제 완료 목록에 저장된다. 정제 완료 목록에 저장된 내용은 관리자에 의해 정해진 2차 수집자에게 전달되며, 전달되지 않은 용어들은 정제자에 의해 재수정이 가능하다.

(3) 2차 수집

1차 정제가 완료되어 관리자에 의해 정해진 수집자에게 전달된 용어에 대해 2차 수집단계가 수행된다.

2차 수집에서는 1차 정제에서 CUI 대상 용어로 판별되어 전달된 용어를 대상으로 ‘설명’ 항목을 입력한다. ‘설명’은 CUI 용어에만 입력되면 되는 사항이기 때문이며, 2차 정제 이후 단계에서는 CUI 용어들에 대한 정제와 검수가 수행된다.

용어의 수집이 완료되지 못하고 수집자에 의해 임시 저장된 용어는 2차 수집중의 리스트에 저장되며, 수집이 완료된 용어는 2차 수집 완료 목록에 저장된다. 수집 완료 목록에 저장된 내용은 관리자에 의해 정해진 2차 정제자에게 전달되며, 전달되지 않은 용어들은 수집자에 의해 재수정이 가능하다.

(4) 2차 정제

2차 수집이 완료되어 관리자에 의해 정해진 정제자에게 전달된 용어에 대해 2차 정제단계가 수행된다.

2차 정제에서는 2차 수집에서 설명정보가 입력된 CUI 대상 용어에 분류2를 입력하고, 용어를 구분하기 위한 코드를 부여한다. 코드는 CUI, LUI(Lexical Unique identifier), SUI (String Unique Identifier) 3종류로 구성되어 있다. 또한 각 코드는 CUI가 LUI를 LUI가 SUI를 포함하는 계층적인 구조를 가진다.

(5) 1차 검수

정제자에 대해 관리자가 지정한 검수자에게 2차 정제 완료된 용어가 전달되면 1차 검수가 수행된다.

1차 검수에서 검수자는 ‘한글’, ‘분류1’, ‘설명’, ‘분류2’, ‘코드보기’ 항목에 대한 정제사항을 검수하고 추가 정제가 필요한 경우 ‘검수코멘트’를 입력한다. 추가 정제 사항이 없거나 작은 수정만 필요할 경우에는 경우 바로 최종검수완료 시킬 수 있다. 1차 검수 완료된 용어는 관리자에 의해 설정된 정제자에게 전달되어 3차 정제를 수행한다.

(6) 3차 정제

1차 검수를 마쳐진 용어에 대해 관리자가 지정한 정제자가 3차 정제 단계를 수행한다.

3차 정제에서는 1차 검수에서 입력된 검수코멘트를 바탕으로 용어에 대한 입력정보를 수정한다. 용어 정제자는 수행한 정제사항에 대해 '3차 정제사항'을 입력하고 정제 완료한다. 3차 정제 완료된 용어는 1차 검수를 수행한 용어 검수자에게 전달되어 2차 검수를 수행한다.

(7) 2차 검수

2차 검수에서는 1차 검수에서 입력된 '검수코멘트'가 3차 정제에서 올바르게 반영되었는지, '3차 정제사항'을 확인하고 수집된 정보에 대해서 최종 검수를 수행한다. 정제가 올바르게 반영되었다면 2차 검수에서 최종검수완료를 하고 용어 수집을 완료한다. 추가적인 정제단계가 필요할 경우 검수자는 개인적으로 정제자에게 연락하여 논의를 수행하여야 한다. 이는 시스템에서 더 이상 정제-검수의 단계를 진행하지 않기 때문에 정제자와 검수자의 편차가 생기지 않기 위해 반드시 수행해야 한다.

2. 데이터베이스 설계

데이터베이스는 앞에서 정의한 요구사항에 맞도록 Figure 2.와 같이 20개의 테이블로 구성하였다. 각 테이블에 대한 설명은 다음과 같다.

1) KIOM_TERM

한의학 용어의 기본적인 정보를 가지고 있는 테이블이다. KIOM_CODE 테이블을 제외한 모든 테이블은 KIOM_TERM에 대한 외래키를 가지고 있다.

- ID: 용어의 식별자
- KOR: 용어의 한글 표기
- CLASS1: 용어의 분류1 정보

- CLASS2: 용어의 분류2 정보
- REFERENCE: 용어의 출전 정보
- PAGE: 출전에 명시된 쪽수 정보
- EXAMPLE: 용어의 용례 정보
- IMAGE: 용어와 관련된 그림 정보
- REPRE: 용어를 대표할 수 있는 대표어 정보
- USERID: 현재 용어를 처리하고 있는 사용자 정보
- STATE: 용어의 처리 상태 정보

2) KIOM_TERM

한의학 용어의 한자 정보를 가진다. 이는 한글과 한자에 서로 다른 SUI를 입력할 수 있도록 구분한 것이다.

- ID: 한자 용어의 식별자
- TERMD: 한자 용어와 연관된 한글 용어의 ID
- S_KOR: 해당 한의학 용어의 한자 표기 정보를 가진다.

3) KIOM_COMMENT

한의학 용어에 대한 설명정보와 1차 검수에서의 검수사항, 2차 정제에서의 정제사항 정보를 저장한다.

- ID: KIOM_TERM 테이블의 ID를 참조
- DESCRIPTION: 용어의 설명정보
- PURIFIER: 1차 검수자가 입력한 검수코멘트의 정보
- INSPECTOR: 3차 정제자가 1차 검수 내용을 보고 정리한 정제의견에 대한 정보

4) KIOM_MORIGINAL

분류1이 약재에 해당하는 용어는 추가적인 중 기원식물 정보를 저장한다.

- ID: KOIM_TERM 테이블의 ID를 참조

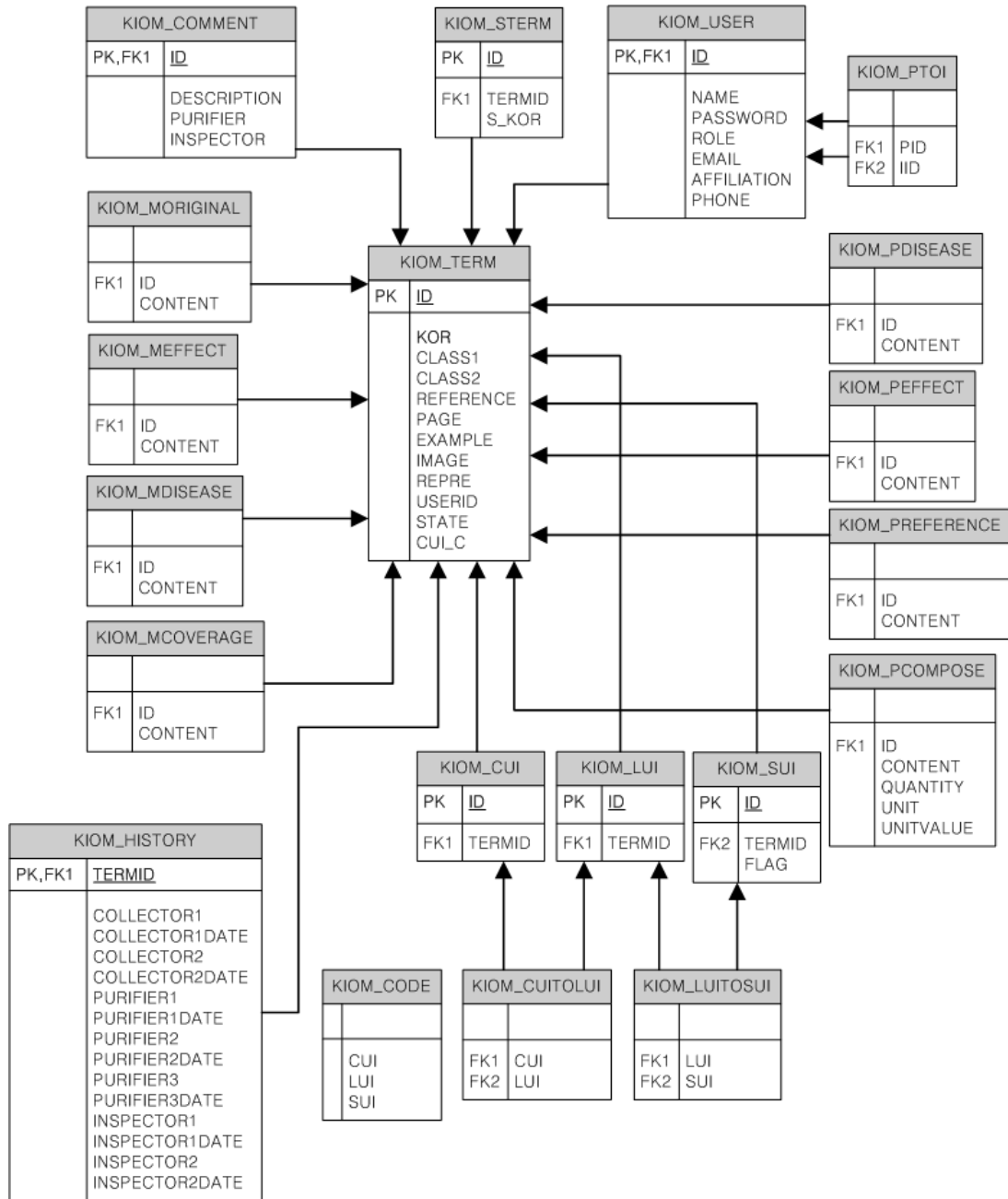


Figure 2. Design of the Database

- CONTENT : 약재의 기원식물 정보
- 5) KIOM_MEFFECT
- 분류1이 약재에 해당하는 용어는 추가적인 중 효능 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 약재의 효능 정보
- 6) KIOM_MDISEASE
- 분류1이 약재에 해당하는 용어는 추가적인 중 관련병증 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 약재의 관련병증 정보
- 7) KIOM_MCOVERAGE
- 분류1이 약재에 해당하는 용어는 추가적인 중 약용부위 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 약재의 약용부위 정보
- 8) KIOM_PDISEASE
- 분류1이 처방에 해당하는 용어는 추가적인 중 관련병증 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 처방의 관련병증들 정보
- 9) KIOM_PEFFECT
- 분류1이 처방에 해당하는 용어는 추가적인 중 효능 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 처방의 효능 정보
- 10) KIOM_PREFERENCE
- 분류1이 처방에 해당하는 용어는 추가적인 중 출전 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 처방의 출전 정보
- 11) KIOM_PCOMPOSE
- 분류1이 처방에 해당하는 용어는 추가적인 중 구성약재 정보를 저장한다.
- ID : KIOM_TERM 테이블의 ID를 참조
 - CONTENT : 처방의 구성약재 정보
 - QUANTITY : 구성약재의 용량정보
 - UNIT : 구성약재의 용량 단위 정보
 - UNITVALUE : 해당 단위의 기준 정보
- 12) KIOM_CODE
- 등록된 CUI, LUI, SUI 코드에 대한 카운트 정보를 저장한다. 이는 각 코드를 발급받을 때 사용된다.
- CUI : 등록된 CUI의 카운트 정보
 - LUI : 등록된 CUI의 카운트 정보
 - SUI : 등록된 CUI의 카운트 정보
- 13) KIOM_CUI
- 부여된 CUI 코드 및 관련 용어의 ID 정보를 저장한다.
- ID : CUI 코드
 - TERMIID : KIOM_TERM 테이블의 ID
- 14) KIOM_LUI
- 부여된 LUI 코드 및 관련 용어의 ID 정보를 저장한다.
- ID : LUI 코드

- TERMID: KIOM_TERM 테이블의 ID

15) KIOM_SUI

부여된 SUI 코드 및 관련 용어의 ID 정보를 저장한다.

- ID: SUI 코드
- TERMID: KIOM_TERM이나 KIOM_STERM 테이블의 ID
- FLAG: 한글, 한자 구분 정보

16) KIOM_CUITOLUI

CUI와 LUI의 관계정보를 저장한다.

- CUI: KIOM_CUI 테이블의 ID를 참조
- LUI: KIOM_LUI 테이블의 ID를 참조

17) KIOM_LUITOSUI

LUI와 SUI의 관계정보를 저장한다.

- LUI: KIOM_LUI 테이블의 ID를 참조
- SUI: KIOM_SUI 테이블의 ID를 참조

18) KIOM_HISTORY

용어가 각 단계별로 진행시에 담당한 사용자와 사용일에 대한 정보를 저장한다.

- TERMID: KIOM_TERM 테이블의 ID
- COLLECTOR1: 1차 수집자의 ID
- COLLECTOR1DATE: 1차 수집의 완료일
- COLLECTOR2: 2차 수집자의 ID
- COLLECTOR2DATE: 2차 수집의 완료일
- PURIFIER1: 1차 정제자의 ID
- PURIFIER1DATE: 1차 정제의 완료일
- PURIFIER2: 2차 정제자의 ID값
- PURIFIER2DATE: 2차 정제의 완료일
- PURIFIER3: 3차 정제자의 ID
- PURIFIER3DATE: 3차 정제의 완료일

- INSPECTOR1: 1차 검수자의 ID
- INSPECTOR1DATE: 1차 검수의 완료일
- INSPECTOR2: 2차 검수자의 ID값
- INSPECTOR2DATE: 2차 검수의 완료일

19) KIOM_USER

한의학 용어 관리 시스템에서의 사용자 정보를 저장한다.

- ID: 사용자의 ID
- NAME: 사용자의 이름 정보
- PASSWORD: 사용자의 비밀번호 정보
- ROLE: 사용자의 담당역할 정보
- EMAIL: 사용자의 이메일에 대한 정보
- AFFILIATION: 사용자가 소속된 기관에 대한 정보
- PHONE: 사용자의 전화번호 정보

20) KIOM_PTOI

정제자와 검수자의 용어 전달을 위해 관계에 대한 정보를 저장한다.

- PID: 정제자와 검수자의 관계에서 정제자의 ID
- IID: 정제자와 검수자의 관계에서 검수자의 ID

3. 시스템 구현

설계한 데이터베이스를 사용하여 다수의 사람들의 용어를 입력할 수 있는 웹사이트 시스템을 구현하였다.

1) 구현 환경

본 시스템의 구현 환경은 다음과 같다

- CPU: Intel Xeon 3.0GHz
- 메모리: 4GB

○ 1차 수집_수집중

용어 기본정보

한글	사군자탕	분류1	약재	[선택]
한자	四君子湯			
출전	방제학	쪽수	48	
용례	사군자탕 용례			
이미지	[찾아보기...]	대표어	사물탕	

약재 추가정보

기원식물	[추가]	[X]	약용부위	[추가]	[X]
효능	[추가]	[X]	주치병증	[추가]	[X]

처방 추가정보

출전					
구성약재	[추가]	용량	무게	▼ 釐 ▼ [X]	
효능	[추가]	[X]	주치병증	[추가]	[X]

[임시저장] [수집완료]

Figure 3. Screen for Collecting System

- 운영체제 : Fedora 8
- 데이터베이스 : Oracle 9i
- 서버환경 : Apache Server 2.0, Apache Tomcat 5.0
- 구현언어 : Java 6.0

2) 사용자 인터페이스 구현

(1) 수집자

1차 수집을 위해 수집자는 중복된 용어가 있는지 중복검사를 수행하게 된다. 중복이 없을 경우 Figure 3.과 같은 수집화면을 통해 새로운 용어를 수집하게 된다. 한글, 한자, 분류1, 출전, 쪽수, 용례, 이미지, 대표어 항목을 입력할 수 있는 필드가 있다. 그 외의 수집자에서의 기능은 다음과 같다.

① 수집 단계의 구분 - 수집 중, 수집완료

용어 수집자가 용어를 수집하는 단계는 1차 수집과 2차 수집으로 구성되어 있다. 2단계로

정의되어 있는 수집단계는 수집중과 수집완료 단계로 구분하였다.

용어 수집자가 용어 정보를 입력하던 중 임시저장 기능을 사용하여 입력중인 정보를 저장할 수 있고, 추후에 다시 수정할 수 있다. 임시 저장된 용어들은 수집 중 목록에 나열되어 수집 완료된 용어와는 별도로 관리된다.

수집이 완료된 용어들은 수집중인 용어들과는 별도로 수집완료 목록에 관리된다. 이를 통해 관리자에 의해 정제자가 지정되어 정제단계로 전달되기 이전까지 추가 수정할 수 있다. 시스템 관리자는 용어에 대한 정제자를 지정할 때 수집중과 수집완료를 구분하여 용어를 효율적으로 관리 할 수 있다.

② 분류1에 따른 추가 정보입력

분류1에 입력되는 값에 따라서 약재인 경우 기원식물, 약용부위, 효능, 주치병증 항목을, 처방인 경우 출전, 구성약재, 효능, 주치병증 항목을 추가적으로 입력할 수 있다.

분류1이 처방인 경우 입력되는 4가지 항목 (출전, 구성약재, 효능, 주치병증) 중 구성약재 항목은 약재와 용량을 함께 입력하는 항목이다. 구성약재의 용량을 입력할 때 용량을 표시하는 숫자와 단위를 분리하여 입력 할 수 있도록 구분하였다. 숫자와 단위가 별도로 관리되어 숫자만 따로 분석하는 경우 데이터의 활용을 용이하게 하였다.

③ 수집 목록 확인

수집자는 시스템에서 자신이 수집한 용어에 대한 목록을 확인할 수 있다. 출력된 내용은 화면으로 확인이 가능하며 또한 엑셀(excel)

파일로 출력하여 확인할 수 있다.

(2) 정제자

정제자의 단계는 1차 수집된 용어들에 대해 수정, 중복된 용어 삭제 및 CUI 대상 용어를 선정하는 1차 정제와 설명정보가 추가된 CUI 대상용어에 코드를 부여하는 2차 정제, 검수자의 코멘트를 반영하는 3차 정제로 구성된다. 이에 따른 정제자의 기능은 다음과 같다.

- ① 1차 정제의 용어 수정 및 중복용어 삭제
1차 수집에서 여러 명의 용어 수집자가 다양한 서적으로부터 용어를 추출하면 오타 및 중복

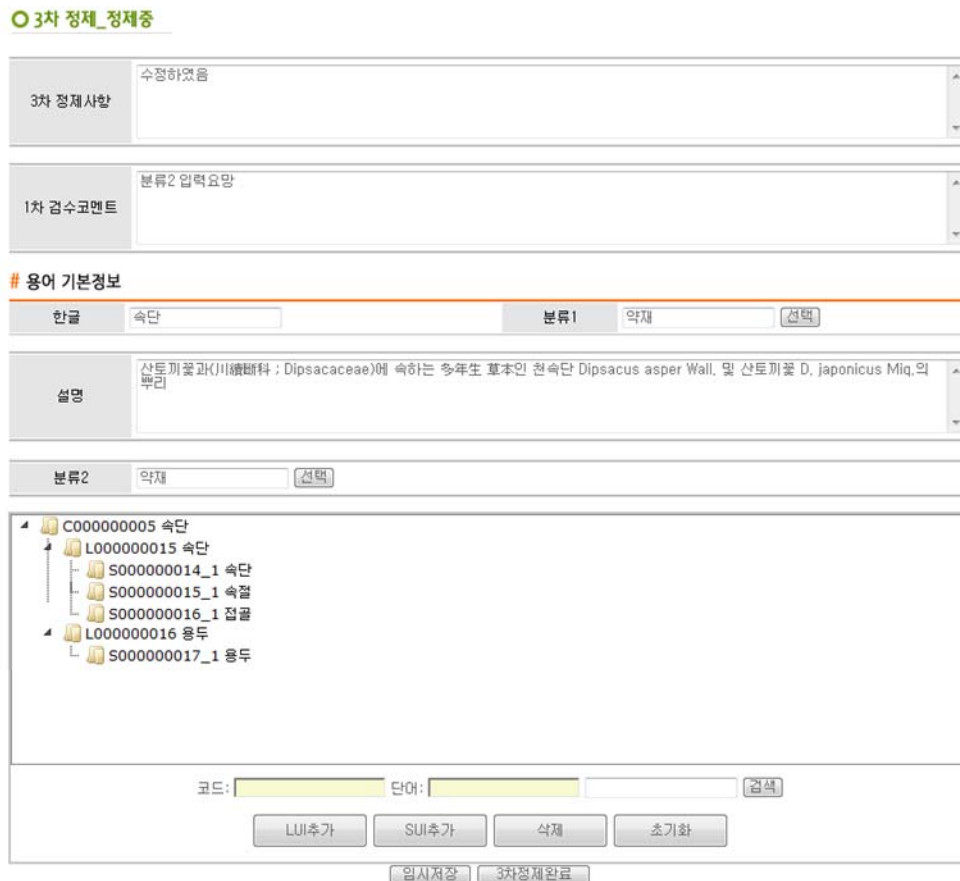


Figure 4. Screen for Purifier and Inspector System

용어가 입력될 수 있다. 이러한 용어들에 대한 수정 및 삭제를 1차 정제에서 수행하게 된다.

② CUI 대상 용어 선택

용어 정제자는 1차 정제 완료된 용어들 중 설명 정보 입력을 위해 CUI 대상 용어를 선별해서 2차 수집 단계로 전달한다. CUI 대상 용어란 정제자가 판단하기에 CUI 코드를 부여할 수 있는 용어를 의미한다. CUI 대상 용어를 선택하는 것은 설명정보를 부여하고 코딩에 사용될 용어를 선택함으로써 이후 단계에서 다루어야 하는 용어의 수를 줄일 수 있다.

③ 코드 부여

2차 정제에서 정제자는 용어의 관리를 용이하게 하고자 각 용어에 코드를 부여하였다. 코드는 CUI, LUI, SUI, 3종류로 구성되고 CUI-LUI-SUI의 계층 구조로 정의된다. 각각을 정의하는 방법은 Table 1~3.과 같다.

CUI(Concept Unique Identifier) 코드는 용어에 대한 개념을 식별해주는 최상위 코드로, 개념 식별자, 일련번호, 구분, 분류자로 구성되어 있다. 개념 식별자로는 'C'를 사용하고 일련번호는 숫자 9자리로 000,000,001 ~ 999,999,999 까지 사용한다(Table 1).

LUI(Lexical Unique Identifier) 코드는 동일한 용어를 묶어주는 코드로, 어휘 식별자, 일련번호로 구성되어 있다. 어휘 식별자로는 'L'을 사용하고 일련번호는 숫자 9자리로 000,000,001 ~ 999,999,999까지 사용한다(Table 2).

SUI(String Unique Identifier) 코드는 모든 용어에 부여되는 코드로, 문자열 식별자, 일련번호, 구분, 문자열 언어 형식으로 구성되어 있다. 1-segment는 문자열 식별자는 'S'와 9자리 숫자의 조합으로 총 10자리의 코드 값을 가지며, 일련번호로 증가시킨다. 작업자의 내부 코드를 분리하는 기호로 underbar(_)를 사용한다. 2-segment의 1자리는 문자열의 언어 형식을 구분하는 것으로, 한글은 '0', 번자는 '1', 간자는 '2', 속자는 '3', 기타는 '4'의 코드를 부여한다(Table 3).

④ 코드 및 분류 계층구조 트리형식 표현

코드 및 분류가 가지고 있는 계층구조를 사용자가 쉽게 인식하고 관리를 용이하게 하도록 트리구조로 표현하였다. 2차 정제 단계에서 용어에 대한 코드를 부여하기 위해 코드 검색할 때 CUI, LUI, SUI 각 코드의 계층구조를 트리형식으로 표현하여 각 코드의 하위코드 포함여부를 손쉽게 알 수 있도록 하였다. LUI

Table 1. Rule of Generating CUI Code

	자리수	구별자 명	구별기호	비고
1 segment	0	개념식별자	C	
	1~9	Numbering	000,000,001 ~ 999,999,999 (일련번호)	

Table 2. Rule of Generating LUI Code

	자리수	구별자 명	구별기호	비고
1 segment	0	개념식별자	L	
	1~9	Numbering	000,000,001 ~ 999,999,999 (일련번호)	

Table 3. Rule of Generating SUI Code

	자리수	구별자 명	구별기호	비고	
1-segment	0	문자열식별자	S		
	1~9	Numbering	000,000,001 ~ 999,999,999 (일련번호)		
구분		구분기호	_(underbar)		
2-segment	1	문자열언어 형식	0	한글	
			1	번자	
			2	간자	
			3	속자	
			4	기타	

코드의 추가는 LUI를 포함하는 CUI를 선택하여 추가할 수 있도록 하였고, CUI 코드 아래의 LUI 코드 이동은 드래그&드롭으로 이동할 수 있게 하였다. 또한 각 코드의 코드명 수정 및 삭제도 코드명 옆에 기능 버튼을 두어 트리 구조 안에서 바로 수정할 수 있도록 하였다.

(3) 검수자

용어 검수자는 2차 정제 완료된 용어에 대해 검수를 진행하는 1차 검수와 3차 정제 완료된 용어에 대해 검수를 진행하는 3차 검수로 구성된다.

검수는 코드, 한글, 분류1, 분류2, 설명 항목을 대상으로 이루어지며, 용어인쇄 버튼을 이용하여 검수자가 검수한 모든 용어를 인쇄할 수 있다. 검수자에서의 기능은 다음과 같다.

① 검수 코멘트 입력

1차 검수자는 2차 정제가 완료된 용어를 검수하는 과정에서 검수 코멘트를 작성할 수 있다. 이는 용어에 대한 많은 수정이 필요할 경우, 검수자가 직접 수정하는 것이 아닌 정제자에게 확인 후 수정을 요청하는 것이다. 입력된 검수코멘트를 바탕으로 정제자는 3차 정제를 수행하게 된다.

② 최종 검수 완료

1차 검수자는 2차 정제가 완료된 용어를 검수할 경우 적은 수정만이 필요한 경우 본인이 수정하여 최종 검수를 완료하여 용어에 대한 수집을 마칠 수 있다. 또한 2차 검수자는 1차 검수 코멘트와 3차 정제 코멘트를 바탕으로 다시 검수를 수행하고 수정하여 최종 검수를 완료할 수 있다. 최종 검수가 완료된 용어는 더 이상 시스템의 워크플로우 안에서 수정되지 않는다.

(4) 관리자

관리자는 시스템의 전반적인 부분에 대해 관리한다. 구현된 기능은 다음과 같다.

① 사용자 관리

한의학 용어 수집 및 관리 시스템을 사용할 수 있는 사용자를 관리한다. 용어 수집자, 용어 정제자, 용어 검수자의 역할별로 사용자를 분리하여 새로운 사용자를 추가하거나 기존의 사용자의 정보를 수정 및 삭제 할 수 있다(Figure 5).

② 수집용어 작업자 지정

용어 수집자에 의해 시스템에 입력된 용어들은 관리자에 의해 지정된 용어 정제자에게



Figure 5. Managing of Users

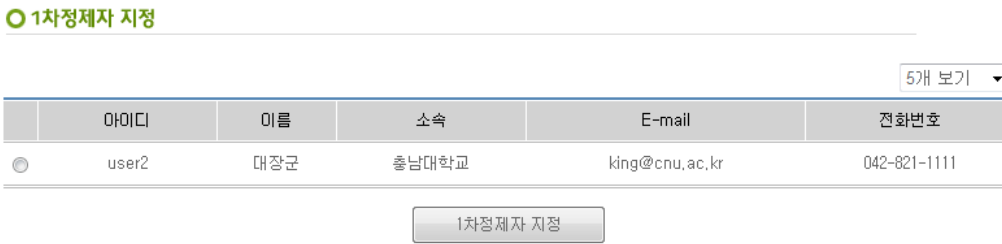


Figure 6. Appointment Terminology to User



Figure 7. Managing of Terminology

전달된다. 시스템 관리자는 용어관리를 통해 각 작업자들의 현재 진행 중인 용어나 진행 완료된 용어의 분량을 파악하여 용어 정제자에게 배분하여 전달된다. 작업자 지정은 1차 수집에서 1차 정제로 용어가 전달될 때 용어 정제자가 지정되고, 1차 정제에서 2차 수집으로 용어가 전달될 때 용어 수집자가 지정되고, 2차 수집에서 2차 정제로 용어가 전달될 때 용어 정제자가 다시 지정된다. 용어 정제자와 용어 검수자는 용어가 아닌 사람을 기반으로

지정하도록 하였다(Figure 6).

③ 용어 관리

시스템을 사용하는 사용자의 역할자별, 개인별 작업현황을 알 수 있도록 용어를 관리하는 기능을 두었다. 수집자별, 정제자별, 검수자별, 전체 단계별로 작업현황을 확인할 수 있다(Figure 7).

수집자별 작업현황에서는 1차 수집 중, 1차 수집완료, 2차 수집 중, 2차 수집완료로 4단계

별 작업건수와 각 단계의 속하는 용어를 목록 형태로 확인할 수 있다. 1차 수집 중, 1차 수집 완료 목록에는 한글, 한자, 분류1, 출전, 쪽수, 대표어 항목을 포함하며, 2차 수집 중 목록에는 한글, 한자, 분류1, 용례의 항목을 포함한다. 2차 수집완료의 목록에는 한글, 한자, 분류1, 설명의 항목을 포함한다. 작업현황에서 용어 수집자를 선택하면 선택한 수집자가 각 4단계에서 진행하고 있는 용어를 확인할 수 있다.

용어 정제자별 작업현황에서는 1차 정제 중, 1차 정제완료, 2차 정제 중, 3차 정제 중, 4단계의 작업건수와 각 단계의 용어를 목록 형태로 확인할 수 있다. 1차 정제 중 목록에는 코드, 한글, 한자, 분류1, 출전, 쪽수, 용례 항목을 포함한다. 2차 정제 중 목록에는 한글, 한자, 분류1, 분류2, 설명의 항목을 포함하며, 3차 정제 중에는 한글, 한자, 분류1, 분류2, 설명, 검수의견의 항목을 포함한다. 용어 정제자를 선택하면 선택한 정제자가 각 단계에서 진행하고 있는 용어목록을 확인할 수 있다.

용어 검수자별 작업현황은 1차 검수 중, 2차 검수 중 2단계로 작업건수와 각 단계의 용어

목록을 확인할 수 있다. 1차 검수 목록은 CUI 코드, 한글, 한자, 분류1, 분류2, 출전, 쪽수를 포함하며, 2차 검수 목록은 CUI코드, 한글, 한자, 분류1, 분류2, 설명, 정제사항의 항목을 포함한다. 용어 검수자를 선택하면 선택한 검수자가 각 단계에서 진행하고 있는 용어 목록을 확인할 수 있다.

용어 단계별 작업현황은 시스템 전체의 작업단계인 7단계에 대한 작업 건수와 각 단계에서 진행 중인 단어 목록을 확인할 수 있다. 각 단계별 목록은 수집자, 정제자, 검수자별 작업현황에 정의된 항목들을 보여준다.

④ 용어 출력

관리자는 시스템 전체적으로 수집/정제/검수되고 있는 용어에 대한 용어를 확인할 수 있다. 모든 용어에 대한 한글, 분류1, 설명, 상태에 대한 정보를 확인할 수 있으며, 각 용어에 대한 코드부여 현황을 확인할 수 있다(Figure 8).

(5) 사용자

사용자는 수집 완료된 용어를 대상으로 용

○ 용어 출력

한글	분류1	설명	상태	
속단	약재	산토끼꽃과(川續斷科 : Dipsacaceae)에 속하는 多年生 草本인 현속단 Dipsacus asper Wall. 및 산토끼꽃 D. japonicus Miq.의 뿌리	2차 정제중	<input type="button" value="보기"/>
조소	약재		1차 정제완료	<input type="button" value="보기"/>
속절	약재		2차 정제중	<input type="button" value="보기"/>
접골	약재		1차 수집완료	<input type="button" value="보기"/>
용두	약재		3차 정제중	<input type="button" value="보기"/>
남초	약재		1차 정제완료	<input type="button" value="보기"/>
강활	약재	미나리과(? 形科 : Umbelliferae)에 속한 多年生 草本인 강활 Ostericum koreanum Maximowicz 의 뿌리	2차 정제중	<input type="button" value="보기"/>
강형	약재		2차 수집중	<input type="button" value="보기"/>
호강사자	약재		2차 수집완료	<input type="button" value="보기"/>
호황사자	약재		3차 정제중	<input type="button" value="보기"/>
강활	약재		1차 검수중	<input type="button" value="보기"/>
사물탕	처방	補血調經하는 처방. 月經不調, 膈脹絞痛, 崩中漏下, 胎動不安 등에 쓴다	1차 검수중	<input type="button" value="보기"/>
사군자탕	처방	益氣健脾하는 처방. 氣虛로 인한 心腹脹滿, 不思飲食, 腸鳴泄瀉 등에 쓴다	3차 정제중	<input type="button" value="보기"/>

Figure 8. Printing of Terminology

어를 검색할 수 있다. 이에 대해 구현된 기능은 다음과 같다.

① 용어별 검색

용어별 검색은 검색이 완료된 용어를 대상으로 사용자가 키워드를 입력하여 적합한 용어를 찾는 것이다(Figure 9).

용어별 검색의 결과는 키워드와 조건이 일치하는 용어의 한글/한자에 대한 검색을 통해 구성된다. 목록에서 용어를 선택하면 용어가 포

함되어 있는 모든 코드의 계층구조(CUI, LUI, SUI)를 트리형식으로 보여준다. 트리 구조 내에서 CUI를 선택하면 설명 정보를 확인할 수 있고, SUI를 선택하면 용례 정보를 확인할 수 있다.

② 분류별 검색

분류1을 이용하여 수집된 용어에 대해서 검색을 할 수 있게 하였다(Figure 10). 트리구조로 표현된 분류의 항목을 선택하면 해당 분류



Figure 9. Searching Terminology using Keyword



Figure 10. Searching Terminology using Classification

에 속한 모든 용어 목록을 확인할 수 있다.

III. 고찰

국내외로 용어를 용이하게 관리하기 위한 용어 관리 시스템들이 많이 연구되고 있다. 관련된 시스템으로는 크게 의료 용어 관리시스템, 일반 용어 관리 시스템으로 나눌 수 있다.

의료(양방) 용어 관리 시스템은 의료 용어에 대한 개념 및 관계를 정의하고 용어를 관리할 수 있도록 하는 솔루션 제공하는 것으로 Medicomp의 MEDCIN,⁴⁾ 3M의 Healthcare Information System,⁵⁾ Apelon의 Distributed Terminology System,⁶⁾ 서울대학교의 KOSTOM 브라우저/에디터⁷⁾가 있다. 이들은 양방 의료 용어가 개념을 기반으로 하고 있으며 계층 구조를 가지고 있기 때문에 용어 관리 시스템이라기보다는 온톨로지 관리 시스템의 성격을 가진다.

일반/전문 용어 관리 시스템은 한국어 용어 또는 전문 분야의 용어들을 통합 관리하기 위한 다양한 기능들을 지원하는 것으로, 대표적으로 울산대학교의 U-Win와 KISTI의 국립국어원 전문용어 지원 시스템이 있다. 이들은 한국어 어휘들에 대한 의미적, 개념적 연결구조로 어휘망을 형성한 것으로 한명의 입력자가 용어를 입력하도록 개발되어 방대한 용어에 대한 작업에 적용하기 어려우며, 또한 전문 분야에 맞게 구축되어 있기 때문에 한의학 도메인에서 바로 적용할 수 없다.

한의학 용어관리 시스템¹⁾은 분산되어 관리되기 어려운 한의학 용어를 효과적으로 수집하도록 하는 인터넷 기반의 시스템으로 다음과 같은 특징을 가지고 있다. 우선 개념을 중심으로 용어체계를 구축하여 개념 단위로 ID를 부여하였으며, 개념 정의에 있어서 설명과 용례에 대한 정보를 저장하고 관리할 수 있고, 유연한 분류체계 관리기능을 통하여 용어 체

계를 구성한다. 하지만 수집자, 정제자, 검수자로 사용자의 역할이 명확히 구분되어 있지 않으며 수집 및 관리를 위한 워크플로우가 정의되어 있지 않다.

본 연구는 한의학 용어관리 시스템을 바탕으로 한의학 용어 수집 및 관리에 요구사항을 모두 만족시킬 수 있도록 시스템을 구성하였다. 우선 사용자의 역할을 수집자, 정제자, 검수자, 관리자, 사용자로 구분하여 정의하였다. 이를 통해 용어에 대한 단순히 수집뿐만 아니라 수집된 내용을 검색하여 제공할 수 있으며 또한 관리자를 통해 전체적인 수집 프로세스의 관리가 가능하다.

또한 정의된 사용자 역할을 바탕으로 워크플로우를 정의하였다. 워크플로우가 정의되지 않으면 정제자, 검수자들이 코드 등의 특성을 부여하거나 검수하는 과정에서 많은 혼란이 야기될 수 있으며, 또한 정제, 검수의 과정에서 편차가 심하게 나타나거나 오히려 현실에서 사용되는 개념으로부터 멀어지는 문제가 발생할 수 있다. 본 연구에서 정의한 워크플로우를 통해 정제자, 검수자들이 서로의 판단을 확인하고 논의하여 수집된 용어의 완성도를 높일 수 있다.

하지만 2차 검수 이후는 최종 검수로 더 이상 정제의 단계를 갖지 않는다. 현재 정의된 워크플로우에서는 2차 검수 후 추가적인 정제를 통한 수정이 필요하다면 정제자와 검수자의 직접적인 연락을 통한 수정을 권고하고 있지만, 추후 시스템의 운용을 통해 용어의 수집을 테스트하여 추가적인 정제 단계가 필요한 경우 적절한 단계의 추가가 필요하다.

IV. 결론

용어 관리 시스템은 웹 기반으로 여러 사용자가 용어를 수집하고 관리하기 위한 시스템

으로써, 한의학 용어를 수집, 정제하고 관리하는데 활용하기 위한 것이다.

한의학 용어 관리 시스템의 사용자는 수집자, 정제자, 검수자, 관리자, 사용자로 나누어지며 각 사용자에게 따른 워크플로우는 다음과 같다. 우선 수집자가 용어를 1차 수집을 통해 한글, 한자, 분류, 출전, 쪽수, 용례, 이미지, 대표어 등 용어 정보를 수집하면 1차 정제를 통해 중복 여부 등의 검사를 수행하며 2차 수집 여부를 결정한다. 2차 수집이 필요한 용어 즉, 대표어가 될 수 있는 용어는 2차 수집 과정을 통해 용어의 설명을 입력하게 되며, 2차 정제에서는 수집된 용어에 대해서 정제가 이루어지고 용어 코드가 할당된다. 2차 정제는 2차 수집된 용어를 중심으로 대표어를 선정하고 대표어에 대한 이명 및 다른 용어들 간의 관계를 맺어준다. 2차 정제를 거친 용어는 검수자가 1차 검수를 통해 용어 검수를 수행하게 되며 수정이 필요하면 3차 정제 및 2차 검수를 통해 최종적으로 용어 수집, 정제, 검수가 완료된다. 또한 관리자는 사용자 관리, 분류체계 관리, 사용자 업무 배정, 용어 관리 및 출력을 수행할 수 있으며, 사용자는 용어에 대한 용어별 검색과 분류별 검색을 수행할 수 있다.

이와 같은 용어 관리 시스템 프로세스는 자바와 JSP를 기반으로 웹 어플리케이션을 구축하였으며 용어 관리 시스템의 모든 용어는 Oracle 데이터베이스를 통해 관리함으로써 여러 사용자가 동시에 접속해서 작업할 수 있도록 하였다.

또한 용어 코드를 할당하는데 있어 용어들 간의 상호 관계를 검색하고 연결하기 용이하도록 하기 위해서 트리 기반의 라이브러리를 이용하여 시스템을 구현함으로써 용어 정제 및 검수자가 트리 형태의 인터페이스를 통해 쉽게 용어 정제 및 검수가 가능하도록 하였다.

본 연구를 통해 구축한 시스템은 온톨로지 기반의 지능형 정보시스템 개발을 위한 선행

작업으로, 실제 용어에 대한 정보를 수집/관리할 경우 수집자, 정제자, 검수자 간의 서로의 판단을 확인하고 논의할 수 있게 워크플로우를 정의했다는 데 의의가 있다.

시스템을 구축함으로써 생기는 기대효과는 다음과 같다. 개발된 시스템을 바탕으로 전통 지식 데이터베이스의 가치를 극대화할 수 있다. 이는 한의학 분야의 전통지식 데이터베이스를 구축함에 있어서 시스템을 통해 수집된 용어를 사용함으로써 자원의 현대화는 물론 고급화된 자원의 효율적 가치 활용도를 높일 수 있다. 또한 분산된 자원의 효율적인 활용 방안으로도 사용될 수 있다. 분야별, 출처별로 세분화되어 서로 분산되어 존재하는 용어 자원들을 시스템을 통해 효과적으로 통합함으로써 실제적인 자원의 활용이 가능한 최대한의 환경을 구현한다. 마지막으로 전통지식의 통합된 서비스 설계를 가능하게 한다. 향후 한의학의 전반적인 분야의 자원들을 단계별로 체계화하여 관리함으로써 사용자 중심의 통합되고 현대화된 전통 의학 정보의 서비스가 가능하도록 할 수 있다.

연구비 지원

본 연구는 한국한의학연구원의 2009년 ‘온톨로지 기반 한의학 지능형 정보체계 연구’ 과제의 연구비지원에 의해 수행되었습니다

참고문헌

1. 고유미, 엄동명. 의미 네비게이션을 지원하는 온톨로지 기반 한의학 논문 검색 시스템 설계 연구. *한국한의학연구원논문집*. 2005:11(2):1-12
2. 이현실. 온톨로지 기반 한의학 처방 지식 관리시스템 설계에 관한 연구. *중앙대학교*

- 대학원. 2003: 183
3. 이병욱, 엄동명. 한의학 용어 관리 시스템 구축 연구. *대한예방한의학회지*. 2009: 13(2): 115-128
 4. P.S. Goltra, Chantilly. *MEDICIN A New Nomenclature for Clinical Medicine*. Springer. 1997
 5. 3M Company Health Information System. Capturing and Integrating Patient Health-care Information to Improve Accessibility. Research and data for Status report. 2001
 6. Apelon Distributed Terminology System; Available from: <http://www.apelon.com/products/dts.htm>
 7. 보건의료정보표준화위원회. 보건의료표준용어 (KOSTOM, KOrean Standard Terminology Of Medicine). 2009