

# 학술논문의 참고문헌 자동매핑 방법에 관한 연구

## Study on Automatic Mapping Method for Reference of Scholarly Papers

한정민\* · 장현철\*\* · 김진현\*\*\* · 예상준\*\*\*\* · 김상균\*\*\*\*\* · 김철\*\*\*\*\* · 송미영\*\*\*\*\*

**Jeong-Min Han · Hyun-Chul Jang · Jin-Hyun Kim · Sang-Jun Yea · Sang-Kyun Kim · Chul Kim · Mi-Young Song**

### 차 례

1. 서론	5. 제안시스템 결과 분석
2. 연구방법	6. 결론
3. 시스템 설계 시 고려사항	· 참고문헌
4. 참고문헌 자동매핑시스템 설계 및 구현	

\* 한국한의학연구원 전통의학연구본부 정보연구센터 연구원  
(Researcher, Information Research Center, TKM Information Research Division KIAM, goal@kiom.re.kr)

\*\* 한국한의학연구원 전통의학연구본부 정보연구센터 선임연구원  
(Senior Researcher, Information Research Center, TKM Information Research Division KIAM, hcjang@kiom.re.kr)

\*\*\* 한국한의학연구원 전통의학연구본부 정보연구센터 연구원  
(Researcher, Information Research Center, TKM Information Research Division KIAM, kjh970203@kiom.re.kr)

\*\*\*\* 한국한의학연구원 전통의학연구본부 정보연구센터 연구원  
(Researcher, Information Research Center, TKM Information Research Division KIAM, tomita@kiom.re.kr)

\*\*\*\*\* 한국한의학연구원 전통의학연구본부 정보연구센터 선임연구원  
(Senior Researcher, Information Research Center, TKM Information Research Division KIAM, skkim@kiom.re.kr)

\*\*\*\*\* 한국한의학연구원 전통의학연구본부 정보연구센터 선임연구원  
(Senior Researcher, Information Research Center, TKM Information Research Division KIAM, chulnice@kiom.re.kr)

\*\*\*\*\* 한국한의학연구원 전통의학연구본부 정보연구센터 책임연구원(교신저자)  
(Corresponding Author, Principal Researcher, Information Research Center, TKM Information Research Division KIAM, smyoung@kiom.re.kr)

• 논문접수일자 : 2010년 2월 10일  
• 최종심사일자 : 2010년 4월 7일  
• 게재확정일자 : 2010년 4월 12일

## 초 록

학문의 발전과 주제의 다양화로 인하여 각계의 연구자들은 자신에게 필요한 정보를 정확하게 찾을 필요성이 커지고 있다. 그리하여 본 논문에서는 효율적인 참고문헌 추출 방법으로 중복된 참고문헌을 비교 분석하여 자동으로 매핑해주는 시스템을 구축하고, 한의학 사전을 통한 한자의 오타를 교정할 수 있는 방법을 연구하였다. 이러한 방법을 적용함으로써 참고문헌의 중복입력과 한자오류를 개선할 수 있었다.

## 키 워 드

학술논문, 정보검색시스템, 참고문헌, 인용지수

## ABSTRACT

With the advanced learning and the diversity of topics, researchers on each area keenly feel the need of precise and a quick discovery of required information at any time. This study presents a way of constructing the automatic mapping system that can compare and analyze duplicated data and that describes the result by building an effective reference extraction method and another way of correcting the wrong form of used Chinese characters with Traditional Korean Medicine dictionary. With this innovation, data duplication on references and Chinese characters errors can be fixed. Under the situation that a number of references of newly published papers that can continuously be extracted,

## KEYWORDS

Treatise, Information Retrieval System, Reference, Citation Index

## 1. 서 론

과학기술이 세분화하고 인터넷 서비스의 발전으로 인하여 새롭게 생산되는 정보의 양은 기하급수적으로 증가하고 있다. 하지만, 새롭게 생산된 정보는 기존의 정보를 기반으로 가공되는 부분이 많아 저작물에 대한 선·후관계를 명확히 할 필요가 있다. 새롭게 생산되는 정보는 개인의 Blog나 Web-Cafe를 통해 복

제되고 있어 데이터의 중복 문제가 발생되고 있다. 한의학 분야에서 신규로 생산되는 정보들도 체계적인 수집이나 관리가 미흡한 상태로 기존의 데이터와 중복되는 문제에 대한 해결책이 필요하다.

이를 해결하기 위한 방법으로 생산된 정보에 대한 중복여부 검증과 연구의 선·후관계를 파악할 수 있는 시스템의 필요성이 커지고 있다. 특히, 한의학 분야의 학문적 특성, 연

구, 저작활동 및 인용문헌에 대한 제반특성을 알아 볼 수 있는 계량적 분석의 제공이 더욱 필요하게 되었으며, 이를 위하여 문헌과 저자들에 대한 국내저자의 생산성 조사, 인용문헌 분포, 주제 분석을 수량화할 수 있는 시스템 구축이 필요하다.

하지만, 대다수의 한의학 논문이 하나의 한글에 대해 다른 한자를 사용하고 있고, 논문입력에 대한 표준이 마련되어 있지 않아, 한의학 논문 인용지수 검색시스템을 구축하는 데 많은 비용과 시간이 투자될 수밖에 없다. 그리하여 본 논문에서는 효율적인 참고문헌 입력 방법으로 중복된 참고문헌을 비교분석하여 자동

으로 매핑하거나 추천해 주는 미들웨어를 구현하였다.

## 2. 연구방법

### 2.1 연구대상

본 연구는 한국한의학연구원에서 서비스하고 있는 전통의학정보포털(OASIS)의 한의학 논문을 기준으로 참고문헌 입력에 대한 작업 소요 시간과 오류율을 비교하였다. 연구를 위한 한의학 논문의 선정은 학술지의 특성으로

〈표 1〉 연구대상 학회지별 논문건수

학회지명	기존방법에 적용될 논문건수	신규방법에 적용될 논문건수
경락경혈학회지	30건	30건
대전대연구소논문집	18건	18건
대한본초학회	41건	41건
대한스포츠한의학학회지	24건	23건
대한암한의학학회지	4건	3건
대한약침학회지	29건	29건
대한예방한의학학회지	23건	22건
대한중풍학회지	5건	5건
대한침구학회지	49건	48건
대한한방내과학회지	48건	48건
대한한방부인과학회지	40건	40건
대한한방소아과학회지	18건	18건
대한한의정보학회지	7건	7건
대한한의진단학회지	10건	11건
대한한의약방제학회지	17건	17건
대한한의학원전학회지	37건	37건
대한한의학회지	42건	42건
사상체질의학회지	24건	24건
소문학회지	16건	15건
척추신경추나학회지	10건	10건
한방비만학회지	8건	8건
총계	500건	496건

인해 다양한 학술지를 선정하였다. 학회지 발행연도는 2008년에 발행된 논문으로 하였고, 한의학의 다양한 분야를 싣고 있는 대한한의학회 학술지와 한자가 비교적 많은 '대한한의학회지'를 포함한 21개의 학회지를 대상으로 하였다.

또한, 각 대상학회지에서 입력한 논문 996건을 하나의 학회지에 실린 논문을 50:50으로 나누어 기존의 입력방식과 시스템을 이용한 입력방식으로 나누어 실험을 실시하였다.

전통의학정보포털(OASIS)은 전통의학 관련 학술논문에 대한 메타데이터 및 원문 서비스와 논문의 인용 및 피인용 지수(Impact Factor)를 제공하고 있으며, 프로젝트 검색 및 분석, 한의약 통계를 제공하는 우리나라 최대의 전통의학포털사이트로 한의학 관련 연구에 큰 의미가 있다고 판단하였다.

## 2.2 선행연구

박은자(1999)는 “온라인 자료의 인용 및 참고문헌 수록 양식과 국내 대학 및 학술잡지에서 사용하고 있는 인용 및 참고문헌 수록 양식 조사연구”에서 연구논문이나 리포트를 쓸 때 참고할 수 있도록 미국심리학회(APA)양식과 미국언어학회(MLA)양식에서 제시된 최근 온라인 정보원의 인용 및 참고문헌 수록방법을 조사하여 소개하였다. 그리고 각 학회에서 작성한 참고문헌 양식을 비교하여 일관성 있게 지키지 않은 경우가 많았다고 결론 내리고 이를 극

복하기 위한 방안을 제시하지 못하고 있다.

남영준(2005)은 국가차원의 인용색인 데이터베이스 구축에 필요한 참고문헌 서지기술 표준을 제안하였다. 문헌정보학 분야의 학술지 규정을 분석하여 참고문헌에 포함된 서지 요소는 크게 저자사항, 서명사항, 발행사항의 하부요소로 구성되어 있다고 하였으며, 기본 포맷과 단행본, 연속간행물, 전자자료 등으로 구분하여 표준안을 제시하여 표준화 하자고 주장하였으나, 시스템적인 추천이나 교정 등 대안을 제시하지 않고 있다.

강병규(2003)는 기존의 LATEX 시스템에 대한 고정 환경과 제약사항을 설명하고 XML의 구조적 특성을 이용한 참고문헌 자동 삽입 시스템을 제안하였다. W3C의 표준인 XML을 이용하여 정보 검색, 삽입, 수정, 삭제가 가능하며 HTML로 변환할 경우 웹 게시가 가능하다 하여 효율적인 참고문헌 입력표준과 관련하여 XML활용방안을 고려해 볼 필요가 있었다. 하지만, 참고문헌을 추출하고 입력하여 매핑 테이블생성보다는 기존의 참고문헌에 대한 외부환경에 서비스를 하기 위한 포맷정의만 제시하여 구축에 필요한 내용 제시는 미약하였다.

신행자(2005)는 XML을 이용하여 웹 기반 원격 교육 시스템에서 강의 내용에 참조될 참고문헌 지원 시스템을 설계하고 구현하였다. 각 대학이나 연구소에 분산되어 있는 여러 데이터 소스로부터 필요한 정보만을 실시간으로 입력하여 수집, 통합, 통계처리 할 수 있도록

확장해야 한다고 하였다. 본 논문에서 제시하고자 하는 참고문헌 입력에 대한 정확성과 용어에 대한 추천 기능 등은 포함되지 않아 차이점을 보이고 있다.

### 3. 시스템 설계 시 고려사항

#### 3.1 한의학 참고문헌의 유형분석

참고문헌에 대한 인용지수 검색 시스템을 구축하기 위해서는 논문의 선·후관계를 파악할 수 있도록 논문 간의 정확한 매핑이 필요하다. 하지만 한의학 논문을 발행하는 각 학회의 참고문헌 표기 기준이 다를 뿐만 아니라, 표기 기준이 정해져 있지 않아 저자들마다 다른 방식으로 기재하고 있는 것이 많다. 뿐만 아니라, 같은 학회지라 하더라도 학회지명을 한자만으로 기록하는 경우와 한글만으로 기록하는 경우, 학회지명의 완전지명이 아닌 약칭만 적는 경우가 있어, 참고문헌 입력 작업자들은 저자가 기록한 참고문헌이 맞는지 일일이 파악해야 하는 문제점이 있다. 이러한 방법으로 참고문헌을 입력하였다 하더라도 기존에 서비스하고 있는 시스템과의 매핑도 별도로 이뤄져야 하므로 하나의 참고문헌을 매핑하기 위해서는 2~3배의 작업 노력이 필요하게 된다.

이러한 특성에 따라 각 논문지의 발행 지침과 표기방식을 파악하여 메타데이터를 입력해야 한다. 본 연구를 위해 분석한 한의학 논문

의 몇 가지 특성은 다음과 같다.

첫째, 한의학의 특성상 논문에 표기되는 대부분의 용어가 한자어로 표기되어 있다. 이를 처리하기 위해서는 시스템에서 각 한자 코드를 지원해야 하지만, 표준코드인 유니코드의 범위에서 벗어나는 한자들이 많아 처리 방법에 대한 정리가 필요하다.

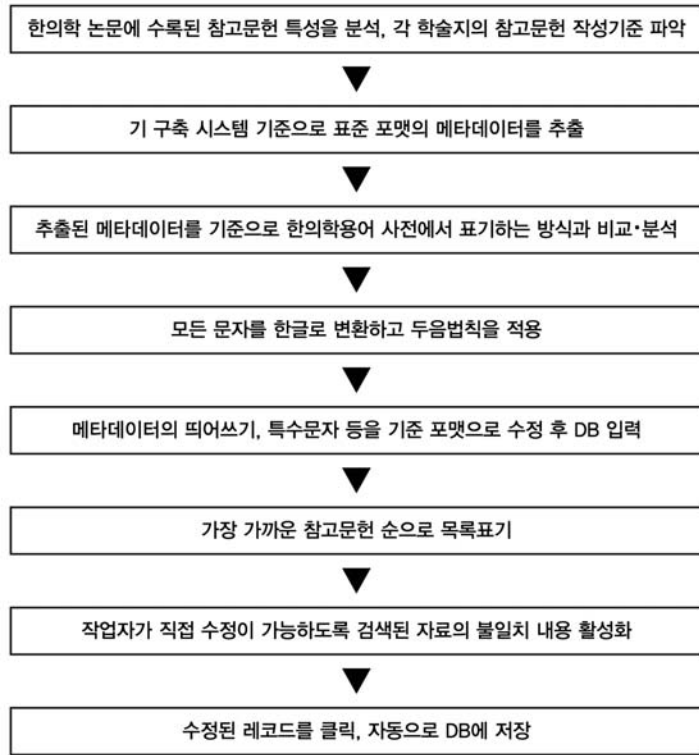
둘째, 참고문헌 표기 기준이 학회마다 다르다. 또한, 한자와 한글 및 영어로 혼합 표기하는 경우에는 같은 논문이라 하더라도 문자열 비교 검색으로는 같은 논문으로 인식하지 못한다.

셋째, 참고문헌 표기법은 단행본의 경우 저자명(발간년도), 서명, 출판사, 소재도시, 출판사명으로 표기하고, 논문의 경우 저자명(발간년도), 논문명, 시작쪽수-끝쪽수로 정하고 있지만, 대부분의 학회에서는 발간되는 논문은 이를 무시하는 경우가 많다.

넷째, 한자에 대한 입력의 오류들이 종종 발생한다. 저자들이 참고문헌을 기록할 때 표준명칭이 아닌 잘못된 한자를 사용하지 않는 경우이다. 예를 들어 “인삼”의 경우, “人蔘”으로 입력해야 하나 “人蔘”으로 기록하는 등 오타들이 많이 발견되었다.

#### 3.2 참고문헌 자동매핑시스템의 제안

작업자들의 작업량과 시간을 줄이기 위한 개선방안으로 참고문헌 자동매핑시스템의 기능을 <그림1>과 같이 제안한다.



〈그림 1〉 참고문헌 자동매핑시스템 기능

## 4. 참고문헌 자동매핑시스템 설계 및 구현

### 4.1 데이터 표준화 처리

논문에 인용된 참고문헌을 추출하기 위해서는 참고문헌의 내용과 형식이 정확히 일치해야 한다. 그리하여, 추천하기 위해 각 참고문헌의 각 항목을 표준화된 포맷으로 정의하였다.

#### 4.1.1 파일코드

각 논문에 포함되어 있는 참고문헌을 추출

하는 경우 추출된 참고문헌이 어떤 논문에 포함되었는지 구별이 필요하다. 그리하여 한의학논문에 대한 파일코드는 앞에서부터 학회코드, 발행연도, 권, 호, 논문 수록 순서로 구성하였다(〈그림 2〉 참조).



〈그림 2〉 파일코드의 구성 예시

#### 4.1.2 순번

참고문헌 추출에서 순번은 작업자의 검색과 저장에 중요한 항목이 아니다. 하지만, 논문에 첨부된 항목의 개수가 다르기 때문에 차후 인용지수 매핑이나 위치정보를 위해 필요하므로 같은 논문에는 추출된 참고문헌 정보가 저장될 때마다 숫자 1씩 증가된다. 새로운 논문에서 입력되는 참고문헌은 숫자가 새로 시작된다.

#### 4.1.3 저자명

참고문헌의 저자명은 성과 이름은 모두 붙여서 저장하였으며 한자로 되어 있는 경우에는 한자와 한글이 같이 입력되도록 하였다. 저자가 다수일 경우 모두 입력하지만, 주 저자 한 명만 입력하고 “외 0명”으로 되어 있을 때는 저자명만 입력하고 저자 외 항목으로 분리·저장하였다. 영문저자일 경우 “첫 번째 저자 두 번째 저자” 순서로 작성하되 첫 번째 저자와 두 번째 저자 사이에 한 칸 공란을 두며 두 번째 저자는 붙여서 작성하고, 첫 글자는 대문자, 나머지는 소문자로 입력되도록 하였다.

저자명을 도치하여 기록한 경우에는 알 수

있는 저자명은 성 이름약칭(Shin YW)과 같은 형식으로 수정 처리하였다. 하지만, 알 수 없는 저자명은 그대로 입력하였다. 저자와 저자 이름 사이에 “가운데 점(.)”이 있을 경우에는 “침표(.)”로 바뀌어서 입력하였다(〈표 2〉 참조).

#### 4.1.4 저자 외

저자가 여러 명일 경우 저자명 뒤에 “등”, “외 0명”으로 쓰여진 것은 저자이름이 일부만 나와 있는 경우 〈표 3〉과 같이 표기한다. 예를 들어, “김덕님, 신경숙 등”으로 작성되어 있거나, “Halici Z et al”로 표기되어 있을 경우 〈표 3〉의 아래와 같은 형식으로 수정하였다.

#### 4.1.5 제목

같은 제목을 한글, 영문, 한자로 다양하게 표기하고 있는 국내 문헌의 경우, 한자를 한글로 전환하고 영문명은 그대로 사용하였다. 국내에서 발행된 영문저널의 경우 영문명을 그대로 사용하였다. “大韓韓醫學會”로 기록한 경우, “대한한의학회”로 정의하였다.

또한 특수문자는 그대로 입력하였으나 위첨

〈표 2〉 저자명의 표준 수정 예시

-약칭 형식: Shin YW, Bae EA, Cai XF, Lee JJ, Kim DB
-저자: 홍길동 · 성춘향 → 홍길동, 성춘향

〈표 3〉 저자 외 항목의 수정 예시

저자명	저자 외
김덕님, 신경숙	등
Halici Z	et al

자나 아래첨자 등 태그로 구성된 글자는 태그를 삭제하고 목에 쉼표(.)와 콜론(:), 세미콜론(;),은 그대로 입력하되 마침표(.)는 삭제하였다.

또한, “《”, “[”, “f” 등 특수 문자를 사용하는 경우 인용된 논문에 저자에 의해 변형될 가능성이 높아 모두 제거하였다.

영문제목에서 대/소문자의 경우 시스템에서 소문자를 기본으로 변경처리 하였다.

#### 4.1.6 웹사이트 주소

참고문헌에 수록된 인용일은 입력하지 않고 웹 주소는 그대로 입력되도록 하였다.

#### 4.1.7 발행처

수록학회지일 경우, 학회지명을 그대로 입력하지만 한자는 한글로 변경하여 입력하였다. 또한 소속 학회나 대학교의 이름을 간략히 적었을 경우 정식명칭으로 수정하여 입력하였다.

같은 학술지지만 사용자에게 의해 축소하여 기록하거나 영문명을 단·복수로 표기한 경우 전체 학술지명을 기준으로 하고, 영문명은 단수로 표기하였다. “대한동의생리병리학회”로 기록한 경우 “대한동의생리학회/대한동의병리학회”로 수정처리 하였다.

또한, 학회지명에 대해 약어 사용과 저자명을 도치하여 기록한 경우에는 저자가 “침구학회”라 기록하였더라도 “대한침구학회”로 학회의 전체 이름으로 수정하였다. 다음 <표 4>는 참고문헌 발행처에 대한 수정 예시이다.

#### 4.1.8 발행연도

발행연도의 경우 ‘YYYY-MM-DD’로 통일하여 작성하였다. 하지만, 발행 월이나 일자가 명기되지 않은 경우에는 연도만 기록하였고, 발행연도나 일자가 기간으로 정해져 있는 경우, 가장 빠른 일자를 기록하였다. 또한, 발행

<표 4> 발행처 항목의 수정 예시

대한당뇨병학회 21차 추계학술대회	→ 대한당뇨병학회
동국한의학연구소논문집	→ 동국대학교
대한동의생리병리학회	→ 대한동의생리학회·대한동의병리학회
동의생리병리학회지	→ 동의생리병리학회지(그대로 써준다)
대한동의생리학회	→ 대한동의생리학회
대한동의병리학회	→ 대한동의병리학회
경희대	→ 경희대학교
서울대학교출판부(국)	→ 서울대학교
식품의약품안전청고시 제2008-24호	→ 식품의약품안전청
J가 앞에 붙을 수도 있고 중간에 붙을 수 있으며, 반드시 대문자로 입력한다.	
경희대학교대학원 석사학위논문	→ 경희대학교
원광대 박사학위논문	→ 원광대학교
원광대학교 대학원	→ 원광대학교



연도 뒤에 “2000 spr” 과 같이 나올 경우 “spr”과 같은 문자는 삭제하였다. 다음 <표 5>는 발행연도에 대한 수정 예시이다.

<표 5> 발행연도 항목의 수정 예시

표준포맷 : 2005-12-30 Oct-08 → 2008-10 2005 2008 Jun 15~30 → 2008-06-15 1996-1999 → 1996
--

4.1.9 권호

권호에 대한 기본포맷은 ‘Vol 00-0’로 통일하였다. 권호 내용 중 권수나 호수만 있는 경우 권수나 호수만 표기하였고, 페이지, 권호

수: 페이지 형식으로 되어 있을 경우 순번을 따로 해서 각각 입력하도록 처리하였다. 권호가 통권으로 나와 있는 경우 Vol을 삭제하고, 그대로 입력한다. 또한 숫자의 형식이 아닌 문자가 들어가 있는 경우는 원문 그대로 표기하였다. <표 6>은 권호에 대한 수정 예시이다.

4.1.10 페이지

페이지의 기본 포맷은 ‘00-00’로 정의하였다. 하지만 페이지가 연결되지 않고 여러 개일 경우 “십표(.)”로 구분하여 기록하도록 하였고 페이지 구분에서 뒤의 자리가 생략된 경우 아래 <표 7>과 같이 수정하였다. 다음 <표 7>은 페이지에 대한 수정 예시이다.

<표 6> 권호 항목의 수정 예시

① 15권 2호 → Vol 15-2												
② 247(1) → Vol 247-1												
③ 3(4-6) → Vol 3-4, Vol 3-5, Vol 3-6												
④ 제 7권 → Vol 7												
⑤ 10호 → Vol -10												
⑥ (10) → Vol -10												
⑦ 홍길동, 의학입문, 서울출판사, 1:11, 2:25												
<table border="1"> <thead> <tr> <th>저자명</th> <th>제목</th> <th>발행처</th> <th>페이지</th> </tr> </thead> <tbody> <tr> <td>홍길동</td> <td>의학입문 1</td> <td>서울출판사</td> <td>11</td> </tr> <tr> <td>홍길동</td> <td>의학입문 2</td> <td>서울출판사</td> <td>25</td> </tr> </tbody> </table>	저자명	제목	발행처	페이지	홍길동	의학입문 1	서울출판사	11	홍길동	의학입문 2	서울출판사	25
저자명	제목	발행처	페이지									
홍길동	의학입문 1	서울출판사	11									
홍길동	의학입문 2	서울출판사	25									
⑧ 通卷(5) → 通卷-5												
⑨ 13(suppl 3) → Vol 13(suppl 3)												

<표 7> 페이지 항목의 수정 예시

표준포맷 : 277-285 227-285, 356, 366 1557-62 → 1557-1562 H459-67 → H459-467
--

#### 4.1.11 구분코드

저자들은 논문을 쓸 때 다양한 곳에서 참고 문헌을 인용한다. 대부분의 학회지나 일반서적은 그 구분이 명확하지만, 한의학 관련 학술지 중에서는 정확한 특성을 띄지 않는 것들이 많아 구분코드로 미리 명시하였다. <표 8>은 구분코드에 대한 예외 사항과 코드를 정의한 표이다.

위에서 설명한 방식으로 정확한 매핑을 위해 표준을 정하고, 입력되는 데이터에 대해 자동 수정되도록 시스템을 구성하였다(<표 9> 참조).

작업자들이 데이터의 공백을 찾아 제거하기는 매우 힘들기 때문에 데이터가 입력되는 동시에 공백을 제거하도록 하였다. 또한 영문데이터는 그대로 입력하였지만, 작업자가 키워드로 검색을 할 경우 입력된 단어와 추천된 단어는 모두 소문자로 변환하여 비교하도록 설계하였다.

한자로 입력된 경우에는 새로운 한글 필드를 만들고 한자를 한글로 변환 입력되도록 하였다. 또한, 모든 글자 사이의 공백과 특수문자 및 html 코드를 제거하고 “-”로 표시된 글자는 공백으로 수정하였다.

<표 8> 구분코드 항목의 수정 예시

종류	코드	비고
학회지	1	동국논집, 경희의학, 해회의학, **대학교 논문집, 정신신체의학
학술대회	2	심포지엄
일반서적	3	고려의학
학위논문	4	대학원
웹사이트	5	
연구보고서	6	
저널	7	동서의학, 최신의학, 발행처 없이 권호수가 나온 경우
기타	8	

<표 9> 참고문헌 추출 표준 항목 예

구분	내용
파일코드	AD2008230102.pdf
순번	1
저자명	홍길동, 성춘향, 이몽룡
저자 외	외 3명
제목	인삼의 사포닌 성분에 대한 .....
웹사이트	http://www.fdagov/cder/guidance/index.htm
발행처	한국영양식량학회지
발행연도	1994
권호	Vol23-2
페이지	181-186
구분코드	1

## 4.2 참고문헌 자동매핑시스템 설계

### 4.2.1 참고문헌 자동매핑 흐름도

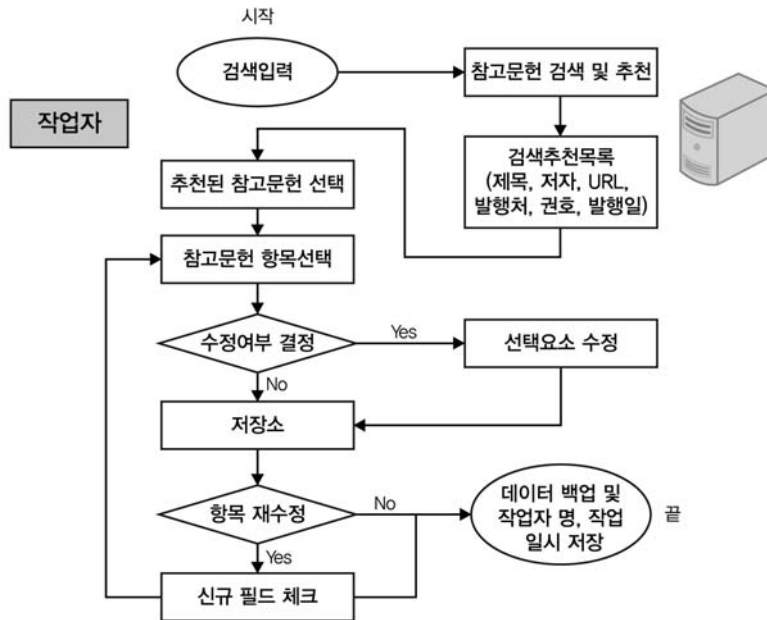
참고문헌 입력 및 추천시스템은 <그림 3>의 참고문헌 입력 흐름도와 같이 진행된다.

참고문헌 인용검색 시스템을 구축하기 위해서는 참고문헌 입력 작업이 선행되어야 한다. 기존에는 참고문헌 입력에 대해 작업자가 일정한 포맷에 맞게 참고문헌을 입력하였지만, 참고문헌 자동매핑시스템은 입력하고자하는 참고문헌에 대한 제목의 몇 글자만 넣으면 시스템에서 매치되는 데이터가 있는지 검색하고 일치할 확률이 높은 순으로 정렬해 준다. 만약, 검색결과가 일치한다면 작업자는 일치하는 내용만 클릭하면 자동으로 시스템에 저장

되도록 하였다.

하지만, 정확히 일치할 확률은 적기 때문에 시스템은 작업자가 입력한 키워드를 기준으로 참고문헌을 추천해 준다. 전통의학정보포털(OASIS)에서 서비스하고 있는 약 2만여 건의 한의학논문을 기준으로 1차 추천을 하고, 작업자가 입력한 참고문헌을 기준으로 2차 추천을 하도록 하였다. 참고문헌은 한의학 논문 외에 일반서적, 웹사이트, 해외 논문 등 다양하기 때문에 1차와 2차를 구분하여 추천하도록 설계하였다.

또한, 논문작성자에 따라서 같은 참고문헌이라 하더라도 한자, 한글 및 모든 학회지명을 입력했거나 약칭만 입력한 것이 있어 작업자에게 추천된 데이터를 기준으로 가공하거나



<그림 3> 참고문헌 입력 흐름도

새로 입력할 수 있도록 하였다. 수정이 필요한 항목을 수정 후 저장 버튼을 클릭하면 새로운 참고문헌이 저장되고 시스템은 작업자와 논문 간의 관계를 설정해 준다.

#### 4.2.2 ER-Diagram

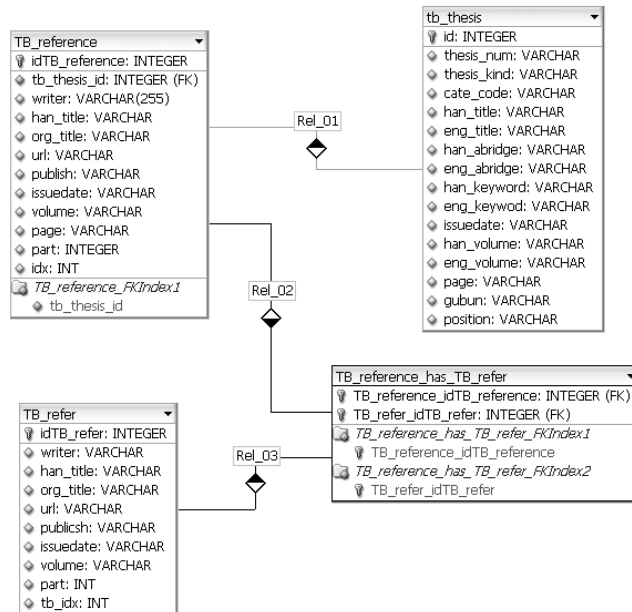
참고문헌 입력 비교시스템의 DB는 3개의 “Tables”로 구성하였다.

TB\_THESIS 테이블은 전통의학정보포털 (OASIS)에서 서비스하고 있는 약 2만여 건의 한의학 논문에 대한 index번호, 저자, 발행연도, 제목, 권호, 학술지명 등으로 구성하였고 참고문헌 입력 및 추천의 1차 기준이 되는 테이블이다.

TB\_REFERENCE 테이블은 TB\_THESIS

에 있는 한의학 논문에 첨부되어 있는 참고문헌들을 모아 놓은 테이블이다. 참고문헌 입력 작업자가 새로 입력하거나 입력한 데이터의 품질검증이 끝난 데이터는 최종적으로 이 테이블에 저장되게 된다.

TB\_REFERER 테이블은 참고문헌 입력 작업자가 키워드를 입력하여 검색된 데이터 및 신규로 입력한 참고문헌의 임시테이블이다. 원본데이터는 온전하다 하더라도 작업자의 실수에 의한 오류를 줄이기 위한 테이블이다. 또한, 참고문헌에 등록되어 있는 한의학 용어에 대한 매핑을 위해 용어사전 테이블을 가지고 있으나 데이터 입력에 사용되기 보다는 검색어의 형식을 걸러주는 역할을 하므로 <그림 4>의 ER-Diagram에서는 표현하지 않았다.



<그림 4> ER-Diagram

4.2.3 시스템 구현 화면

〈그림 5〉는 참고문헌 입력 및 추천시스템에서 「인삼」으로 검색된 참고문헌의 추천결과 화면이다. 이 시스템은 작업자의 업무부하를 줄이기 위한 시스템이기에 검색입력은 제목으로만 한정하였다. 작업자에게는 검색어에 대한 리스트 정보만 보여주지만, 〈그림 5〉의 정보들은 시스템에 의해 추천된 순서로 정렬된

결과 화면이다.

논문작성자의 오타나 내용의 불명확으로 인한 입력은 시스템 스스로 판단하기 어렵다. 그리하여 한의학 표준용어를 기준으로 오타에 대한 용어를 추천해 주고 작업자는 용어의 오류여부를 반자동으로 수정하도록 하였다. 작업자가 추천된 목록 내용 중에서 가장 가깝다고 판단되는 참고문헌을 선택하면 〈그림 6〉과

The System for Reference Extraction									
TITLE: 인삼									
클릭하시면 클립보에 복사됩니다. 그리고 그대로 엑셀에 붙여 넣으세요.									
저자, 저지, 제목, URL, 발행처, 발행일, 불용, 페이지, 구분, 파일용인 호, 업로드									
김민수	김민수	인삼의 효능	대한한약학회지	2003-06-20	제20권 제3호	131-140	1	9285	
김민수	김민수	인삼의 효능	대한한약학회지	2005				4	
김민수	김민수	인삼의 효능	대한한약학회지	1996-12-15	제99권 제1호	355-360	1	8072	
김민수	김민수	인삼의 효능	대한한약학회지	2008-06-30	제27권 제2호	103-110	1		
김민수	김민수	인삼의 효능	대한한약학회지	2007-03-30	제22권 제1호	41-48	1	16150	
김민수	김민수	인삼의 효능	대한한약학회지	2002-06-29	제23권 제2호	78-87	1	3137	
김민수	김민수	인삼의 효능	대한한약학회지	1999-04-30	제5권 제1호	39929	1	8434	
김민수	김민수	인삼의 효능	대한한약학회지	1996	제16권 2호	126-135	1	15749	
김민수	김민수	인삼의 효능	대한한약학회지	2003-09-30	제18권 제3호	41-50	1	4143	
김민수	김민수	인삼의 효능	대한한약학회지	1999	제19권 01호	409-427	1	14912	
김민수	김민수	인삼의 효능	대한한약학회지	2007-03-30	제22권 제1호	35-40	1	18149	
김민수	김민수	인삼의 효능	대한한약학회지	2006-12-31	제21권 제4호	123-131	1	16133	
김민수	김민수	인삼의 효능	대한한약학회지	2006-12-31	제21권 제4호	133-141	1	16134	
김민수	김민수	인삼의 효능	대한한약학회지	1994-02-28	제4권 제1호	8553			
김민수	김민수	인삼의 효능	대한한약학회지	1997-02-28	제7권 제1호	30-35	1	6695	
김민수	김민수	인삼의 효능	대한한약학회지	1996-08-31	제10권 제2호	147-152	1	3510	
김민수	김민수	인삼의 효능	대한한약학회지	1990-10-10	제9권 제2호	33-44	1	2349	
김민수	김민수	인삼의 효능	대한한약학회지	1998				4	
김민수	김민수	인삼의 효능	대한한약학회지	1990-12-15	제12권 제1호	219-233	1	0172	
김민수	김민수	인삼의 효능	대한한약학회지	1988-12-15	제11권 제1호	15-35	1	8120	
김민수	김민수	인삼의 효능	대한한약학회지	1994-08-25	제11권 제1호	509-517	1	8588	
김민수	김민수	인삼의 효능	대한한약학회지	1993-07-15	제10권 제1호	353-370	1	8552	
김민수	김민수	인삼의 효능	대한한약학회지	1975	제12권 제1호	19-20	1	11449	
김민수	김민수	인삼의 효능	대한한약학회지	1999-12-31	제21권 제4호	314-320	1	9601	

〈그림 5〉 참고문헌 검색결과 화면

참고문헌 DB입력

DB저장    DB삭제

파일명	BA2008170113.pdf	순번	29
저자명	허준	저자외	
제목	對譯東醫寶鑑		
웹사이트			
발행처	법민문화사		
발행년도	1999	권호	
		페이지	360, 1539, 1584
		구분	3

〈그림 6〉 참고문헌 수정 및 입력화면

같은 화면이 나타나고 작업자는 수정이 필요한 항목만을 수정하여 저장하도록 하였다.

뿐만 아니라, 한글로 검색하여도 한자와 같이 검색되며 작업자는 검색목록을 클릭하기만 하면 참고문헌을 입력하거나 편집모드로 변경하여 사용할 수 있다.

## 5. 제안시스템 결과 분석

### 5.1 실험대상 및 테스트 컬렉션

이 장에서는 참고문헌 입력 및 추천시스템을 이용하여 작업한 결과에 대한 정확성과 업무처리 속도를 비교하여 평가하였다.

실험 대상 논문은 2008년에 발행한 한의학 논문으로 기존방법에는 500건의 논문, 신규방법에는 496건의 논문을 선정하였다. 논문 1편에 약 29.4건의 참고문헌을 가지고 있지만, 참고문헌의 개수 차이로 인해 신규방법에 적용한 논문편수는 기존방법에 적용된 논문보다

4편이 적다. 하지만, 참고문헌을 기준으로 기존방법과 신규방법 모두 1만4,700건으로 동일하게 적용하였다. 또한, 각 학회지의 고유한 특성이 있어 같은 학회지의 논문을 50:50으로 나누어 기존방법과 새로운 방법에 적용하였다. 경락경혈학회의 경우, 2008년도 발행된 논문 60편 중 30편은 기존방법으로, 나머지 30편은 새로운 방법을 적용하여 실험하였다.

또한, 실험은 참고문헌의 건수를 기준으로 참고문헌의 길이, 난이도, 영문, 한문 포함 여부 등은 고려하지 않고 각 실험환경에 참고문헌 1만4,700건씩으로 나누어 실시하였다.

또한, 작업자의 작업의 능숙 여부와 외부환경도 고려치 않고 각각의 실험환경에 2명씩 나누어 실시하였다. 참고문헌 입력에 대한 실험환경은 다음 <표 10>과 같다.

### 5.2 성능측정 및 실험결과

새로운 방법의 시스템 검증방법은 참고문헌 입력에 대한 작업시간 단축과 정확성 여부를

<표 10> 참고문헌 건수에 따른 작업량 비교

구분	기존방법 적용 환경	제안 시스템 적용 환경
학술지 건수	21건	21건
논문 건수	500건	496건
참고문헌 건수	14,700건	14,700건
발행연도	2008년	2008년
작업자	2명	2명

기준으로 하였다. 작업자들에게 똑같은 작업 시간을 할당하고 참고문헌 건수 1만4,700건을 모두 완료할 때까지를 측정기준으로 삼았다. 작업자들은 하루 8시간씩 업무시간에만 실험을 하였다.

### 5.2.1 평가 방법

참고문헌 자동매핑시스템을 사용한 결과와 사용하지 않고 입력한 결과에 대한 평가방법은 참고문헌 입력속도와 입력된 데이터의 정확성을 검사하였다.

입력속도는 참고문헌 기준으로 3,000건씩 입력하는 시점에서 걸린 시간을 기록하였고 최종 데이터 입력 시점의 평균 시간 차이를 비교하였다.

또한, 매핑 데이터의 정확성 검사는 입력속도 비교와 같은 방식으로 3,000건이 입력되는 시점에서 입력된 데이터에 대해 실험에 참가하지 않은 다른 작업자에게 검수작업을 의뢰하였다. 오류에 대한 검수 지침서를 만들고 각 항목에 입력되어야 할 표준을 작성하였다.

각 항목을 참고문헌이 입력되는 순서에 따른 순번증가와 파일명, 그리고 논문에 등재되어 있는 참고문헌의 위치정보, 저자명, 제목, 발행처, 웹 인용일 경우 웹 주소, 발행연도, 권호, 페이지, 참고문헌의 종류에 따른 구분으로 구별하여 검수하였다.

각 항목들 중에 1개라도 오타나 입력형식이 잘못되었을 경우, 오류로 간주하고 계산하여 오류율을 계산하였다. 예를 들어 참고문헌에

11개의 항목 중 1개가 오류이든지 3개가 오류이든지 오류건수는 1개로 계산하였다.

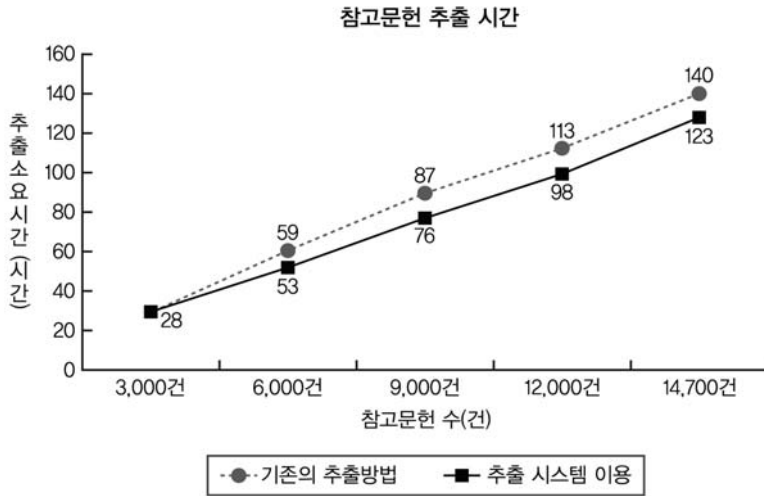
또한, 논문작성자가 참고문헌을 입력할 때 범한 오류는 용어의 경우 간략히 수정되도록 되어 있다. 이에 대하여 오류율에 대한 계산은 수정된 내용이 옳은 것으로 판단하였다. 즉, 원문과 추출된 참고문헌의 내용이 다르더라도 한의학 용어에 대한 오타수정은 옳은 것으로 처리하였다.

예를 들어 인삼(人蔘)을 인삼(人參)으로 입력한 경우, 인삼(人蔘)이 정확한 표기이므로 오류가 아닌 것을 평가하였다.

또한, 입력속도가 빠르더라도 오류가 많다면, 결코 효율적인 시스템이라 할 수 없으므로 두 가지 경우를 동시에 비교하였다.

### 5.2.2 참고문헌 입력속도

〈그림 7〉에서 보는 바와 같이, 작업자의 참고문헌 입력 작업속도는 1만4,700건을 완료하는 시점에서 17시간을 단축한 결과를 보였다. 약 3,000건을 입력할 때까지는 작업자가 직접 입력하고 매핑하는 기존 방식과 제안한 시스템을 이용한 자동매핑 입력방식에서 큰 차이를 보이지 않았다. 그렇지만, 시간이 흐름에 따라 작업시간은 차이를 보였다. 처음에 차이를 보이지 않았던 것은 기존의 입력방법이 작업자들에게 더 익숙하였던 것으로 분석된다. 또한, 작업자가 직접 입력하고 기준데이터와 직접 매핑해야 하는 복잡한 작업과정을 거치는 절차의 문제도 있는 것으로 판단된다.

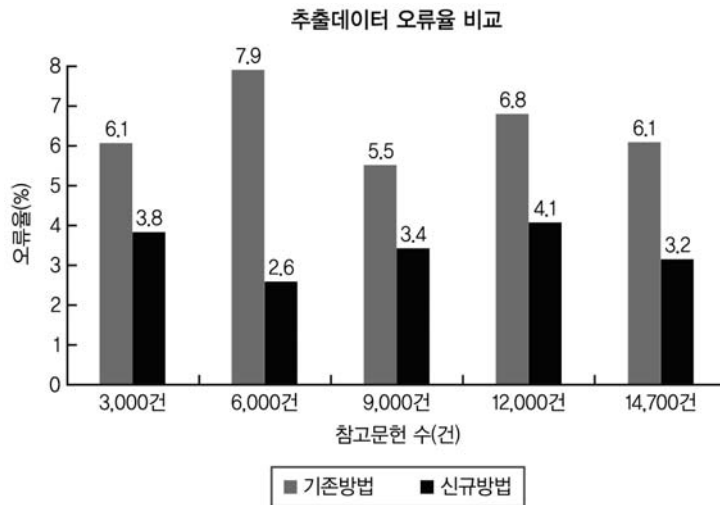


〈그림 7〉 참고문헌 건수에 따른 작업량 비교

### 5.2.3 입력 및 추출된 데이터의 정확도

입력 및 매핑된 데이터에 대한 정확도 검사 결과, 작업자가 3,000건의 참고문헌 입력 작업에서 6.1%와 3.8%의 오류율을 보였고, 3,000건부터 6,000건까지의 참고문헌 입력

작업에서는 7.9%와 2.6%의 오류율을 보였다. 〈그림 8〉에서 보는 바와 같이, 입력된 데이터의 정확도는 제안된 시스템을 사용하지 않고 기존의 방식으로 직접 입력한 결과는 평균 6.48%의 오류율을 보였다. 반면 시스템을



〈그림 8〉 참고문헌 건수별 정확도 비교



사용하여 작업한 결과는 평균 3.42%로 약 3.6%의 차이를 보였다. 이러한 결과는 모든 구간에서 시스템을 사용했을 경우, 사용하지 않았을 경우보다 적은 오류율을 보여 작업속도에 이어 유용하게 사용할 수 있을 것으로 파악되었다.

#### 5.2.4 실험결과

본 연구의 실험결과, 직접 입력하는 것보다 시스템을 이용하는 것이 참고문헌 입력속도와 오류율에서 모두 좋은 결과를 보였다. 참고문헌 기준으로 3,000건 입력 시점에서는 1시간 정도 입력시간이 절약되었을 뿐만 아니라, 작업이 완료되는 시점에서는 17시간을 단축하였다. 오류율에서도 6,000건을 입력하는 시점에서는 최대 5.3%까지 줄인 결과를 얻었다(〈표 11〉 참조). 그리하여 본 연구에서는 참고문헌 자동매핑시스템을 이용하는 것이 이용하지 않는 것보다 모든 부분에서 효율적이라는 결과를 얻었다.

## 6. 결론

최신 기술들이 빈번히 출현하고 발전함에 따라 생산된 정보에 대한 중복여부 검증과 연구의 선·후관계를 파악할 수 있는 시스템의 필요성이 커지고 있다. 특히, 한의학 분야의 학문적 특성, 연구, 저작활동 및 인용문헌에 대한 제반특성을 알아 볼 수 있는 계량적 분석의 제공이 더욱 필요하게 되었다. 이를 위하여 문헌과 저지들에 대한 국내저자의 생산성 조사, 인용문헌 분포, 주제 분석을 수량화할 수 있는 시스템 구축이 필요하다.

그리하여 참고문헌 인용정보 검색시스템을 구축하기 위해서는 각 논문에 수록되어 있는 참고문헌을 입력하는 작업이 선행되어야 한다. 이러한 작업은 참고문헌 입력과 검수를 해야 하므로 작업자의 많은 노동력을 필요로 하게 된다. 이에 본 논문에서는 기존의 엑셀이나 문서편집기에 직접 입력하는 방식을 탈피하여 시스템을 이용, 보다 효율적인 방법을 제안하였다.

〈표 11〉 구간별 입력속도와 오류율 편차

대분류	입력방법	3,000건	6,000건	9,000건	12,000건	14,700건
입력속도 (시간)	직접입력	29	59	87	113	140
	시스템이용	28	53	76	98	123
	편차	-1	-6	-11	-15	-17
오류율(%)	직접입력	6.1	7.9	5.5	6.8	6.1
	시스템이용	3.8	2.6	3.4	4.1	3.2
	편차	-2.3	-5.3	-2.1	-2.7	-2.9

이를 위해 각 한의학 관련 학회에서 발행되는 학술지를 참고로 참고문헌 패턴을 분석하여 한의학 참고문헌 표준포맷을 정의하였으며 정의된 표준 포맷에 맞게 각 학회의 데이터를 가공하여 기준 데이터베이스를 만들었다.

이를 적용함으로써 기존의 작업방법에 비해 참고문헌 입력에 대한 작업 진행 시간과 오류를 줄이는 결과를 얻었다. 하지만, 본 시스템은 시스템 스스로 판단하고 매핑해 주기보다는 기존에 작성했던 참고문헌을 찾아주고 각 필드의 항목 중 완벽한 데이터는 자동으로 처리해 주지만, 완벽하게 매치되지 않는 데이터는 우선순위가 높은 것을 추천해 줄 뿐만 아니라, 작업자가 수정해야 할 부분을 표시해 줌으로써 작업자는 보다 편리하게 작업을 진행할 수 있도록 하였다.

참고문헌 인용검색뿐만 아니라, 한의학 논문 맵(Paper Map)을 구현하기 위해서는 한의학 학술논문에 대한 수집 및 정제, 데이터 구축이 선행되어야 한다. 뿐만 아니라, 국내 전통지식과 관련된 논문자료를 체계적으로 확보하여 데이터베이스를 구축함으로써 전통지식과 관련된 유사 지식 검색 및 참고자료로 활용할 수 있을 것으로 예상된다.

## 참고문헌

김병규, 강무영, 박재원, 강지훈. 2003. XML을 이용한 논문 참고문헌 자동 삽입 처리기의 설계 및 구현. 『한국정보과학회 학술발표

논문집』, 30(2III): 283-285.

김태수, 남영광, 최석두. 1999. 국내 학술지 인용색인을 위한 데이터요소의 기술형태 분석: 정보관리학회지를 중심으로. 『정보관리학회지』, 16(2): 183-199.

남영준, 조현양, 배순자. 2005. 참고문헌의 서지 기술 표준에 관한 연구. 『한국문헌정보학회지』, 39(4): 261-279.

박은자. 1999. 온라인 자료의 인용 및 참고문헌 수록 양식과 국내 대학 및 학술잡지에서 사용하고 있는 인용 및 참고문헌 수록 양식 조사연구. 『정보관리학회지』, 16(2): 81-104.

신규민, 한요섭, 김래현, 차정원. 2009. 기계 학습을 이용한 인용문헌 추출. 『한국정보과학회 학술발표논문집』, 36(1C): 331-335.

신행자, 박경환. 2002. XML 데이터 처리 기반의 참고문헌 지원 시스템의 설계 및 구현. 『한국정보과학회 학술발표논문집』, 27(1B): 726-728.

양승오. 1990. 최근 3년간 대한방사선의학회지에 실린 논문의 주제별 분석 및 참고문헌의 인용지수 등에 대한 연구. 『대한방사선의학회지』, 26(4): 804-810.

한국한의학연구원 오아시스. [인용 2010.02.10].  
〈<http://oasis.kiom.re.kr>〉.

Dezhong Yao, Li Wang, Robert Oostenveld. 2005. "A comparative study of different references for EEG spectral mapping." *Physiological measurement*, 26(3): 173-184.

Rolf Gruetter and Ivan Tkáč, 2000, "Field Mapping Without Reference Scan Using Asymmetric Echo-Planar Tech-

niques," *Magnetic Resonance in Medicine*, 43(2): 319-323.