

## OMIS의 키워드와 통제어 일치도에 관한 연구

A study of agreement rate between keyword and Thesaurus

李丙旭, 權寧奎

대구한의대학교 한의과대학

---

### Abstract

A study of agreement rate between keyword and Thesaurus

Byung-Wook Lee, Yung-Kue Kwon

College of Oriental medicine, Daegu Hanny University

#### Objective

We wish that oriental medicine theses can now be searched for by computer. So we have studied thesaurus and keywords in OMIS(Oriental Medicine Information System)

#### Methodologies

We have compared the keyword with the Thesaurus in the OMIS. So We have analyzed those into the agreement rate and the disagreement rate. Repeatedly, we have analyzed the agreement rate and the disagreement rate in chronological order.

#### Conclusions

1. As the most of authors don't know about thesaurus, So many keywords disagree with thesaurus.
2. Because thesaurus system is very difficult to general authors. So system administrators must build synonym data base, if many authors easily use thesaurus system.
3. We have to require institute editor to modify editorial policy. Presently most of institute require only one foreign language keywords, but it must be required foreign and Korean language keywords.
4. We must study not only oriental medicine thesaurus but also the other science thesaurus.

---

**Keywords** : keyword, thesaurus

---

### I. 序論

현재 국내에서 한의학 관련 문헌논문을 찾아

원문을 보는 방법은 여러 가지가 있다. 그중 대표적인 방법에는 대구한의대학교 제한동의학술원의 OMIS 서비스를 이용하는 방법과 한국한의학연구

원의 문헌검색시스템을 이용하는 방법 및 인터넷 검색을 통하여 원문제공 서비스 사업자의 서비스를 이용하는 방법이 있다.

그런데 이들 문헌검색 시스템들 중 인터넷 검색을 통하여 원문을 제공하는 방법은 원문의 검색방법으로 제목을 이용한 검색방법을 주로 이용하고 있다. 그리고 한국한의학연구원과 대구한의대학교의 인터넷 서비스의 경우에는 국문제목, 영문제목, 저자, 년도 및 각종 고급기능을 활용한 검색기능을 제공하고 있다. 그러나 이러한 방법들은 논문의 서지사항에 직접 기재된 내용이 아닌 경우에는 관련논문을 찾기 어렵다. 이러한 문제점을 해결하기 위하여 제시된 방법이 thesaurus를 이용한 검색이다. 하지만 이러한 검색방법이 지원 가능하도록 DB의 구조가 만들어 졌음에도 불구하고 실제로 활용되지 못하고 있는 실정이기 때문에 논문의 검색율이 그다지 높지 않은 실정이다. 따라서 본 연구에서는 문헌검색의 효율성을 제고시키기 위하여 대구한의대학교 제한동의학술원의 OMIS에 구축된 논문DB를 대상으로 시로시스(thesaurus: MeSH)의 활용도에 대하여 연구하였다.

## II. 本論

대구한의대학교 제한동의학술원의 OMIS는 1997년 한방치료기술개발연구지원사업단의 과제인 “인터넷상 한의약 학술정보시스템 구축 및 검색도구 개발”을 수행한 결과 만들어진 시스템을 기초로 지속적으로 데이터를 업데이트 시킨 문헌검색 시스템이다<sup>3)</sup>. 따라서 동일 과제 결과를 바탕으로 시스템을 구축한 한국한의학연구원의 문헌검색 시스템과 구조적으로 같다.

이 두 시스템은 기본검색 방법으로 제목, 저자,

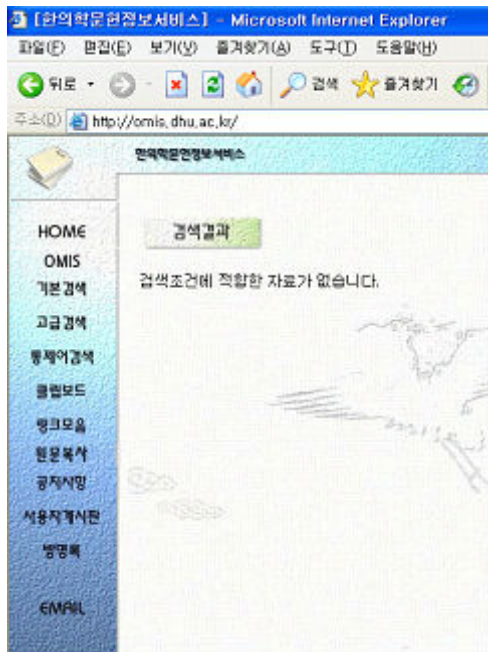
발행처, 발행년, 통제어 검색을 지원한다. 그리고 고급 검색을 통하여 세부적인 검색을 할 수 있도록 되어있다.

특히 문헌정보의 키워드 외에 통제어를 설정하게 되면 논문저자들의 개인적 판단에 따라 설정하게 되는 키워드 선정에 따른 검색을 저하를 해결할 수 있으며 관련된 자료들을 좀 더 효율적으로 검색이 가능하다. 즉 제목에 언급되지 않은 관련 정보를 찾을 수 있다. 그러나 이 두 시스템 모두 통제어 검색은 정상적으로 실행되지 않고 있다. 대구한의대학교의 OMIS의 경우에는 통제어 관련 필드에 논문 저자들이 기록한 키워드를 입력한 상태로 서비스가 진행되고 있고 한의학 연구원의 문헌검색 시스템의 경우에는 키워드도 입력되어 있지 않다. 따라서 기존의 통제어 시스템에 사용되는 용어와 키워드의 불일치로 인하여 정상적인 통제어 검색이 정상적으로 이루어지지 않고 있다.

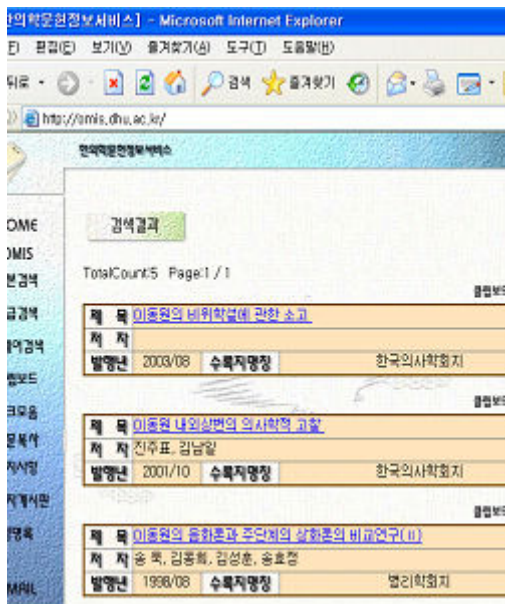
### 1. 제목검색과 통제어 검색 비교

2005년 5월 10일 현재 대구한의대학교의 OMIS에는 11437건의 한의약 관련 논문정보가 저장되어 있다. 이를 활용하여 李杲(이고)에 대한 통제어, 제목검색을 활용한 검색결과를 비교하면 다음과 같다.

제목검색을 이용하여 李杲(이고)라는 단어로 논문을 검색하면 논문 제목에 ‘이고’라는 글자가 들어간 경우만 출력되기 때문에 ‘관련 조건에 적합한 자료가 없습니다.’라는 결과가 출력된다.

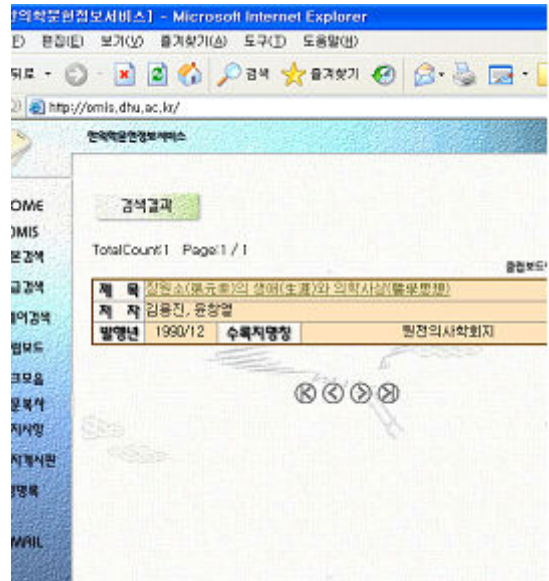


‘이동원’ 이라는 단어로 제목검색을 실시하면 “이동원의 비위학설에 관한 소고”, “이동원 내외상변의 의사학적 고찰”, “이동원의 음화론과 주단계의 상화론의 비교연구” 등의 결과를 얻는다.

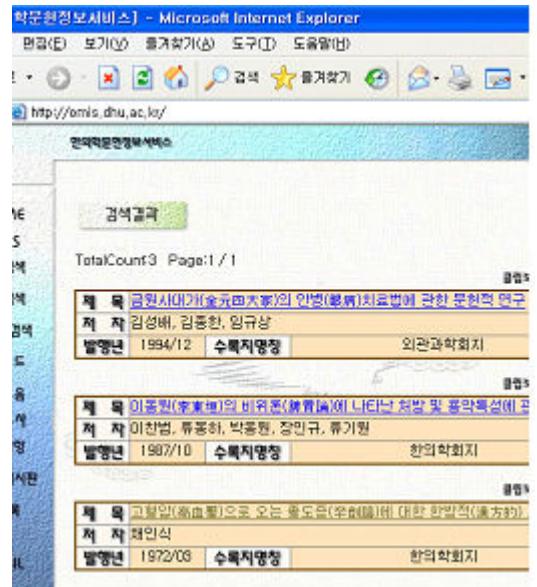


통제어를 활용하여 검색하면, 비록 논문 제목에 ‘이고’ 라는 명칭이 없더라도 “장원소의 생애

와 의학사상” 이라는 논문정보가 출력된다.



‘이동원’ 으로 검색하면 “금원사대가의 안병 치료법에 관한 문헌적 연구”, “이동원의 비위론에 나타난 처방 및 용약특성에 관한 고찰”, “고혈압으로 오는 졸도증에 대한 한방적 고찰” 등의 결과를 얻을 수 있다.



이처럼 제목검색과 통제어 항목을 통한 검색결과가 서로 다르게 나올 수 있다. 그러나 이론상 통제어 검색이 더욱 좋은 정보를 제공해야

함에도 불구하고 만족스러운 결과를 얻을 수 없는 것으로 나타났다..

통제어 검색에서 이러한 불만족스러운 결과를 얻게 된 이유는 다음과 같다.

1) 李杲(이고)나 李東垣(이동원)은 통제어 목록에 없다. 따라서 ‘이고’ 나 ‘이동원’에 대한 정보를 하나의 용어로 표현하는 기준이 없으므로 통제어 항목에 저자나 문헌정보 입력자가 임의로 정보를 입력하였다.

2) 통제어 목록 검색에서 기존의 통제어 목록을 보여주며 검색하는 방식을 지원하고 있지 않다.

3) 원래 저작에 통제어에 대한 기재가 이루어지지 않고 있다. 다만 키워드라는 항목이 있으며, 이 또한 저자 임의로 선정하는 경우가 많다.

## 2. OMIS의 통제어 목록과 한의학 관련 논문들의 키워드 비교

한의학연구원의 문헌정보 시스템에는 한의학 관련논문의 키워드나 통제어를 입력하지 않은 상태로 시스템이 운영되는 관계로 한의학 관련 논문들의 키워드와 기존의 한의학 관련 통제어의 비교는 OMIS를 대상으로 실시하였다. 또한 키워드와 기존의 통제어의 비교는 분석의 편의상 "like \*& 키워드 &\*"의 문법을 사용하지 않고 전체 Cell이 일치하는 것을 기준으로 분석하였다.

2005년 5월 10일 현재 OMIS에 입력된 논문정보는 11,437건이 입력되어 있었고, 그 중 통제어 필드에 자료가 입력된 것은 5,015건이었다. 물론 이것은 통제어 필드에 입력되어 있는 내용이지만 실제로는 논문 저자가 작성한 키워드 목록을 입력한 것으로 볼 수 있다. 그리고 사용된 키워드 숫자는 25,448개가 사용되었으며, 총13,706종의 단어가 사용되었다. 또한 13,706종의 단어 가운데

기존의 통제어 목록에 기록된 11,810개의 단어와 일치도를 살펴보면, 12,423개의 단어가 일치하지 않아 90.6%가 통제어와 다르다는 결론이 나왔다. 이에 대한 통계는 다음과 같다.

년대	논문수	키워드입력논문	입력비율
1960년대	274	269	98%
1970년대	409	189	46%
1980년대	1183	706	60%
-1994까지	1704	1083	64%
1999까지	3677	716	19%
2000-	4190	2052	49%
계	11437	5015	44%

표1. 년대별 키워드 입력 논문 수

년대	키워드 수	키워드 종류	키워드 사용빈도
1960년대	1671	1097	1.52
1970년대	790	612	1.29
1980년대	4180	2532	1.65
-1994까지	6576	3622	1.82
1999까지	3325	2511	1.32
2000-	8906	5965	1.49
계	25448	13706 (총계아님)	1.86

표2. 년대별 키워드 수와 종류

년대	키워드수	통제어와 일치하는 키워드	일치율
1960년대	1097	276	28%
1970년대	612	169	28%
1980년대	2532	596	24%
-1994까지	3622	808	22%
1999까지	2511	251	10%
2000-	5965	127	2%

표3. 년대별 키워드와 통제어의 일치율

### 1) 1960년대 논문

1960년대 논문은 모두 274편이 기록되어 있었

고 여기에서 키워드 항목이 기재된 논문은 269편이고 총 1671개의 키워드를 입력하였으며, 1097종의 단어가 사용되었다. 1097종의 단어 가운데에는 기존의 통제어와 일치하는 것이 276종이 있었으며, 일치하지 않는 것이 821종의 단어가 있었다. 일치하지 않는 단어에는 통제어에 없는 영문표기가 67종으로 전체 불일치 용어의 8%를 차지했으며, ‘(’, ‘-’, ‘y’ 등과 같은 특수 문자를 포함한 용어가 535종으로 전체 불일치 용어의 65%를 차지했고, 일반적인 한의학 관련 용어를 사용한 것이 219종 27%를 차지했다.

#### 2) 1970년대 논문

1970년대 논문은 모두 409편이 기록되어 있었고 여기에서 키워드 항목이 기재된 논문은 189편이고 총 790개의 키워드를 입력하였으며, 612종의 단어가 사용되었다. 612종의 단어 가운데에는 기존의 통제어와 일치하는 것이 169종이 있었으며, 일치하지 않는 것이 443종의 단어가 있었다. 일치하지 않는 단어에는 통제어에 없는 영문표기가 30종으로 전체 불일치 용어의 7%를 차지했으며, ‘(’, ‘-’, ‘y’ 등과 같은 특수 문자를 포함한 용어가 149종으로 전체 불일치 용어의 34%를 차지했고, 일반적인 한의학 관련 용어를 사용한 것이 264종 60%를 차지했다.

#### 3) 1980년대 논문

1980년대 논문은 모두 1183편이 기록되어 있었고 여기에서 키워드 항목이 기재된 논문은 706편이고 총 4180개의 키워드를 입력하였으며, 2532종의 단어가 사용되었다. 2532종의 단어 가운데에는 기존의 통제어와 일치하는 것이 596종이 있었으며, 일치하지 않는 것이 1936종의 단어가 있었다. 일치하지 않는 단어에는 통제어에 없는 영문

표기가 240종으로 전체 불일치 용어의 12%를 차지했으며, ‘(’, ‘-’, ‘y’ 등과 같은 특수 문자를 포함한 용어가 386종으로 전체 불일치 용어의 20%를 차지했고, 일반적인 한의학 관련 용어를 사용한 것이 1310종 68%를 차지했다.

#### 4) 1990-1994년 논문

1990-1994년까지의 한의학 관련 논문은 모두 1704편이 기록되어 있었고 여기에서 키워드 항목이 기재된 논문은 1083편이고 총 6576개의 키워드를 입력하였으며, 3622종의 단어가 사용되었다. 3622종의 단어 가운데에는 기존의 통제어와 일치하는 것이 808종이 있었으며, 일치하지 않는 것이 2814종의 단어가 있었다. 일치하지 않는 단어에는 통제어에 없는 영문표기가 375종으로 전체 불일치 용어의 13%를 차지했으며, ‘(’, ‘-’, ‘y’ 등과 같은 특수 문자를 포함한 용어가 635종으로 전체 불일치 용어의 23%를 차지했고, 일반적인 한의학 관련 용어를 사용한 것이 1804종 64%를 차지했다.

#### 4) 1995-1999년 논문

1995-1999년까지의 한의학 관련 논문은 모두 3677편이 기록되어 있었고 여기에서 키워드 항목이 기재된 논문은 716편이고 총 3325개의 키워드를 입력하였으며, 2511종의 단어가 사용되었다. 2511종의 단어 가운데에는 기존의 통제어와 일치하는 것이 251종이 있었으며, 일치하지 않는 것이 2260종의 단어가 있었다. 일치하지 않는 단어에는 통제어에 없는 영문표기가 1562종으로 전체 불일치 용어의 69%를 차지했으며, ‘(’, ‘-’, ‘y’ 등과 같은 특수 문자를 포함한 용어가 91종으로 전체 불일치 용어의 4%를 차지했고, 일반적인 한의학 관련 용어를 사용한 것이 607종

27%를 차지했다.

#### 4) 2000년 이후 논문

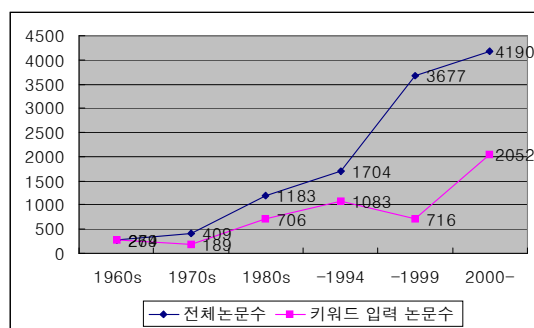
2000년 이후 지금까지의 한의학 관련 논문은 모두 4190편이 기록되어 있었고 여기에서 키워드 항목이 기재된 논문은 2052편이고 총 8906개의 키워드를 입력하였으며, 5965 종의 단어가 사용되었다. 5965종의 단어 가운데에는 기존의 통제어와 일치하는 것이 127종이 있었으며, 일치하지 않는 것이 5838종의 단어가 있었다. 일치하지 않는 단어에는 통제어에 없는 영문표기가 5465종으로 전체 불일치 용어의 94%를 차지했으며, ‘(’, ‘-’, ‘\`’ 등과 같은 특수 문자를 포함한 용어가 428종으로 전체 불일치 용어의 7%를 차지했고, 일반적인 한의학 관련 용어를 사용한 것은 입력 데이터 형식이 복잡하여 분석하지 못했다.

### III. 考察

지금까지 한의학 관련 논문에 기록된 키워드 입력비율, 사용빈도, 통제어와 일치율, 통제어와 불일치 키워드의 비율을 살펴보고 각 년대별 키워드의 특징을 살펴보면 다음과 같다.

#### 1. 키워드 입력 비율의 변화

	1960s	1970s	1980s	-1994	-1999	2000-
전체논문수	274	409	1183	1704	3677	4190
키워드 입력논문수	269	189	706	1083	716	2052
비율	98%	46%	60%	64%	19%	49%



키워드 입력 비율의 변화를 살펴보면 1960년대가 가장 높은 98%를 차지하고 있는데, 이는 논문 저자가 직접 입력한 것이 아니고 논문에 대한 서지사항을 입력하는 시스템 관리자에 의하여 의도적으로 입력된 것이므로 다른 년대에 비하여 상대적으로 높은 입력율이 나타났다. 그리고 이러한 입력율은 대체적으로 높아지다가 1995년 이후 갑작스러운 입력율 저하가 나타난다. 이는 1995년 이후 다양한 학회지의 등장으로 많은 수의 논문들이 저술되었으나, 이당시의 논문들에 대해서는 시스템 관리자에 의한 입력이 이루어지지 않았으며, 저자가 선정한 키워드를 입력하는 방식을 취하였기 때문에 일부 키워드 항목에 대한 편집 규정이 마련되지 않은 학회지의 논문은 키워드 입력이 이루어지지 않았다. 때문에 입력율의 갑작스러운 저하가 나타났다.

#### 2. 키워드의 사용빈도와 통제어 일치율의 변화

	1960s	1970s	1980s	-1994	-1999	2000-	전체
키워드	1671	790	4180	6576	3325	8906	25448
종류	1097	612	2532	3622	2511	5965	1376
빈도	1.52	1.29	1.65	1.82	1.32	1.49	1.86
통제어 일치율	25%	28%	24%	22%	10%	2%	9%

키워드의 사용빈도를 살펴보면 약 45년간의 통계상으로 13706종의 키워드가 총 25448번 기록된 것이므로 1개의 키워드가 평균 1.86번 사용되었음을 알 수 있다. 그리고 년대별 사용빈도는 더욱 떨어져서 최저 1.29회(70년대)에서 최고 1.82회(90년대 전반)가 사용되었다. 또한 통제어 일치율은 70년대를 정점으로 해마다 감소하여 2000년대 전반기에는 2%까지 떨어졌다.

이러한 통제어 일치율의 저하는 하나의 개념에 대하여 저자들이 다양한 키워드를 사용함으로써 키워드를 활용한 검색의 효율을 급속하게 떨어뜨린다고 볼 수 있다. 즉 다양한 키워드를 사용함으로써 평균 키워드 사용빈도가 2개에도 미치지 못하기 때문에 키워드를 활용한 검색을 사용할 경우 자신의 논문 외에 다른 논문이 2편 이상 검색될 확률이 극히 적다고 볼 수 있다. 물론 키워드 검색에 있어서도 키워드 전체일치가 아닌 부분일치 조건을 사용하면 좀 더 많은 결론을 얻을 수 있겠지만 검색의 정확도가 떨어진다고 볼 수 있다. 이러한 문제점을 해결하기 위해서는 통제어 체계를 적극 활용하는 방안을 강구하여 검색효율을 높여야 한다. 또한 통제어 체계의 활용도를 높이기 위해서는 통제어 체계와 유사어에 대한 정리가 이루어져야 할 것이다.

3. 키워드의 통제어 불일치율에 대한 분석

	1960s	1970s	1980s	-1994	-1999	2000-	전체
불일치율	75%	72%	76%	78%	90%	98%	91%
영문	8%	7%	12%	13%	69%	94%	57%
특수문자	65%	34%	20%	23%	4%	7%	16%
일반용어	27%	60%	68%	64%	27%	-	-

1960년대와 1990년대 후반기의 통제어 불일치율을 분석해보면, 60년대에는 논문의 키워드를 설

정할 때 실험논문에서 사용한 ‘α’, ‘β’, ‘γ’ 와 같은 특수문자를 포함한 용어를 사용하였기 때문에 현재 구축된 한의학 이론체계 중심의 통제어와 일치하지 않는 경우가 대부분이었다. 그러나 90년대 후반에 이르러서는 60년대와 같은 경우 보다는 학회지의 편집 규정상 외래어를 사용한 초록 형식을 도입한 경우가 증가함에 따라서 한의학 용어를 영문으로 표기한 경우가 급속하게 증가하였다. 또한 실험논문의 급속한 증가도 키워드와 기존의 통제어 체계와 맞지 않는 결과를 양산하였다. 때문에 키워드를 이용한 검색효율이 매우 많이 떨어졌다.

4. 통제어 활용 방안

통제어와 키워드의 불일치로 인한 문헌검색 효율성의 저하 문제의 해결을 위해서 가장 이상적인 방법은 모든 문헌의 저자가 통제어 시스템에 대한 지식을 바탕으로 논문에 대한 키워드를 선정하는 것이다. 그러나 통제어 시스템 자체에 대한 지식을 습득하여 활용하는 것 역시 손쉬운 일이 아니기 때문에 모든 논문 저자가 이에 대한 지식을 쌓을 필요는 없다.

하지만 기존의 시스템에서 소개하고 있는 통제어가 트리구조(tree structure)로 되어 있으므로 논문 저자들이 자신의 논문주제에 적합한 통제어를 주제어로 선정한다면 검색효율을 최대한 높일 수 있을 것이다.

이차적으로 해결할 수 있는 방법은 각 학회의 논문 편집 책임자와 편집실무자들이 DB를 구축할 시점에 해당 논문의 주제에 맞는 통제어를 선정하는 방식이다. 하지만 모든 학회가 직접 DB를 구축하지 않고 있기 때문에 DB구축에 관여하는 기관의 시스템관리자의 지원도 필요하다. 따라서 이러한 문제를 해결하기 위해서는 시스템 관리자

와 논문 편집자의 협력이 필요하다.

우선 시스템 관리자 입장에서 고려할 문제로는 기존의 통제어 시스템에 유사용어에 대한 데이터 베이스의 확충문제와 한의약 용어의 영문 표기 방법에 대한 표준안 마련문제가 있다.

논문 편집자 입장에서 고려할 문제로는 초록의 작성에 있어서 키워드는 외국어와 한국어어를 동시에 작성하도록 편집규정을 바꾸는 문제이다.

이와 더불어 한의학 영역 이외의 개념에 대한 통제어 적용방안에 대한 연구도 진행되어야 할 것이다. 물론 한의학 영역이외의 개념이나 용어는 직접 개발하는 것보다 MedLine에서 채택하고 있는 MeSH(Medical Subject Headings)을 인용하는 방식이 효율적이다.

이미 세계보건기구 서태평양지구(WHO-WPRO) 전통의학부서에서는 지난 5월 북경에서 정보표준화를 위한 1차 회의를 개최하여 중국을 비롯한 한국, 일본의 MeSH현황을 파악하고 현재 세 나라의 MeSH 자료를 수집하고 비교하여 중복되는 용어를 제외한 뒤 공통안을 만들기 위한 작업이 진행되고 있다<sup>4)</sup>.

일본의 경우 모든 의학용어는 MeSH를 그대로 이용하고 단지 보험에 적용이 되고 있는 처방명만 추가하여 활용하고 있으며, 중국은 이미 TCM-MeSH 2판의 출판이 마무리 단계에 있으며 중의약학 논문DB는 중국중의연구원 신식연구소에서 유지하고 있다. 중국은 서양의학과 관련된 DB는 신식연구소와 별개로 유지되고 있기 때문에 TCM-MeSH에는 중의약학과 관련된 통제어 외에도 서양의학과 관련된 통제어까지 확장하고 있는 실정이다.

우리나라의 경우 중국의 TCM-MeSH(제1판) 통제어를 기본으로 하고 우리나라 실정에 필요한 용어를 보충한 통제어시안 보고서가 1999년 제출

된 이후 추가적인 연구나 보완연구가 제대로 진행되지 않았을 뿐만 아니라 통제어를 직접 사용하면서 문제점을 해결하는 노력 또한 진행되지 않고 있는 실정이다.

국내 통제어와 관련된 위와 같은 문제가 해결되고 한중일 간의 통합안이 마련된다면, 모든 논문 저자에게 통제어에 대한 지식을 요구할 필요가 없으며, 그들의 표현에 대한 자유를 많이 제약하지 않으면서도 효과적이고 적중률 높은 문헌검색 결과를 얻을 수 있을 것이다.

또한 앞으로 한중일 삼국간의 통제어 통합안이 마련되면 한중일을 중심으로 한 전통의학논문에 대한 통합검색시스템 구축방안도 조만간 논의될 가능성이 높아질 것으로 보여진다. 더 이상 한의학논문이 한의학전공자나 국내용이 아니라 세계 의학자들에게 전통의학논문으로 소개될 시점이 되어가게 될 것으로 예상된다. 통제어는 수단이나 방법이지 목적이 아니므로 좋은 논문, 전통의학을 연구하는 혹은 전통의학을 이용하고자 하는 세계 의학자들의 관심을 끌 수 있는 논문이 발간되고 체계적으로 관리할 수 있는 구조가 이루어져야 할 것이다.

#### IV. 結論

1. OMIS 시스템의 키워드와 통제어 시스템의 불일치는 논문 저자 혹은 문헌정보 관리자의 통제어 시스템에 대한 이해가 부족하기 때문이다.

2. 시스템 관리자 입장에서 통제어 시스템과 각 문헌들의 키워드를 활용한 문헌검색의 효율성을 높이기 위해서는 기존의 통제어 시스템에 유사용어에 대한 데이터베이스를 구축하여 유사용어 사용시 통제어를 활용하여 결과를 산출 할 수 있는



시스템을 구축해야 한다.

3. 논문 편집자 입장에서 통제어 시스템과 각 문헌들의 키워드를 활용한 문헌검색의 효율성을 높이기 위해서는 초록 작성시 외국어 키워드와 한글 키워드를 동시에 표기하도록 편집규정을 수정하는 것이 바람직하다.

4. 비한의학 영역의 용어에 대한 표기 방법에 대한 연구도 필요하다.

### 참고문헌

1. <http://omis.dhu.ac.kr>
2. <http://jisik.kiom.re.kr/>
3. 경산대학교 한의과대학; 인터넷상 한의약 학술 정보시스템 구축 및 검색도구 개발 2차년도 최종 보고서, 한국한의학연구원 서울 1999.
4. 1st Informal Consultation On Information Standardization Of Tranditional Medicine, 17-19 May 2005, Beijing, China