

미국국립의학도서관분류법(NLMC) 일본어판 데이터베이스 소개

한국의학도서관협회 편집위원회 (역)

I. 들어가며

일본의학도서관협회(이하 JLMA)는 2019년 6월 10일에 미국국립의학도서관분류법(NLMC) 일본어판 데이터베이스(이하 NLMC-DB)의 웹 사이트(<http://jmla1927.org>)를 구축하여 서비스를 시작했다(그림 1).

본 글에서는 NLMC-DB의 작성을 담당한 워킹그룹(이하 본 WG)의 서비스 개요와 사용 방법 팁, 그 구조를 소개한다. 또한 서비스 개시 후의 이용 실태를 보고하고 향후의 과제와 전망에 대해서 기술하였다.

NLMC-DB는 2019년 9월부터 서비스를 시작하였고 JMLA회원만 이용 가능하다. JMLA 회원전용 웹페이지에 이용을 위한 아이디와 비밀번호가 게시되어 있다.

II. NLMC-DB의 개요와 검색방법

NLMC-DB 개요와 검색방법에 대해서 소개하겠다.

1. NLMC-DB란

NLMC-DB란 미국국립의학도서관에서 만든 의학, 보건, 간호학 분야의 분류법(National Library of Medicine Classification: NLMC)을 일본어로 번역하여 웹에서 이용하는 서비스로, NLMC의 2016년판을 번역하여 2017년 6월에 책자 (이하 책자판)로 출판한 것을 데이터베이스화 했다[1].

NLMC-DB는 책자판의 분류법과 색인을 각각 검색하여 이용할 수 있다.



그림 1. NLMC-DB TOP 페이지.

Received October 15, 2019, Revised November 20, 2019, Accepted December 16, 2019

Corresponding author: Jung Mi Park

Medical Information Library, Kangbuk Samsung Hospital, 29 Saemunan-ro, Seoul 03181, Korea
 Tel: 82-2-2001-2793, Fax: 82-2-2001-2799, E-mail: jm0827.park@samsung.com

본 논문은 저자 Satoshi Hashimoto가 "Introduction of Japanese Version of National Library of Medicine Classification Database"라는 제목으로 2019. Journal of the Japan Medical Library Association. 66(3):212-216.에 게재한 내용임.

Copyright © 2019 The Korean Medical Library Association. All rights reserved

또한 책자판의 번역과 NLMC-DB의 작성은 둘 다 수탁사업으로 진행한 것으로 본 WG는 번역작업을 진행한 워킹그룹(이하 번역 WG) 멤버의 일부로 구성되어 있다. 번역 WG의 활동과 작업내용에 대해서는 번역 WG그룹장이었던 코지마가 보고하고 있으며, 관심있는 분은 코지마의 자료를 참조하기 바란다[2,3].

2. 분류 검색

분류검색에서는 QZ, WA 등의 분류기호와 분류번호를 조합해서 검색한다. 해당 검색결과가 있는 경우 분류표안의 해당 부분과 대응하는 표제어도 표시되어 책자판의 분류표처럼 자유롭게 전후를 확인할 수 있다. 해당 검색결과가 없는 경우에는 검색에 사용한 분류기호 분류표의 최상부가 표시된다.

표시된 분류표는 QZ, WA 등의 분류기호가 있다. 분류기호 간의 이동은 되지 않는다. 다른 분류기호에 속한 경우에는 재검색을 할 필요가 있다.

3. 색인검색

1) 색인검색의 개요

색인 검색에서는 임의의 키워드로 검색을 하면 그 키워드를 포함한 색인어가 50음순으로 표시되어 합산되며, 그 색인어가 대응하는 분류와 연관 검색어도 표시된다. 또한 검색결과는 분류표의 해당 부분으로 이동하거나 연관 검색어에서 색인을 재검색 할 수도 있다.

색인어는 일영병기되어 있어 일본어와 영어로도 검색할 수 있다. 일본어의 명칭은 주로 [의학용어 시소러스 제8판] [4] 및 [일본의학회 의학용어 사전 WEB판] [5]를 기준으로 수록했다. 또한 영어로 검색한 경우에도 결과의 표시는 번역어의 50음순으로 표시된다.

색인 작업 중의 분류도 검색 대상이다. 분류 검색을 하면 대응하는 색인어를 찾을 수 있다. 이런 경우에는 분류기호와 분류 번호의 사이에 공백(스페이스바)을 구분한다.

2) 색인 검색의 유의사항

검색 키워드를 공백(스페이스바)으로 단락을 지어 입력하면 AND 검색이 된다. OR이나 AND 등의 불리언 연산자 검색의 경우에 (*, !, " ") 등의 검색 기호는 사용할 수 없다. 또 소리 검색도 되지 않는다. 검색대상이 되는

색인어의 계층 구조는 색인 표제어(=상위개념 예: 종양) 레벨이다. 색인 표제어의 세목(=하위개념 예: 종양→약물용법)은 검색의 대상이 아니다.

또한 시소러스와 같은 동의어 등의 다양한 어휘에 대응하는 것은 아니다. 덧붙여 번역이라는 성질 상 반드시 이용자가 생각하는 번역어가 색인어로 사용되었다고 할 수 없다.

3) 색인검색의 팁

논문 검색 데이터베이스에서 검색결과 원하는 결과를 얻지 못하는 경우에 사용하는 색인검색 사용 팁을 소개하겠다.

예를 들어 [재해 간호]라는 키워드의 분류를 검색을 하면 그 단어 자체로는 표제어가 아닌 [간호] 또는 [재해]가 세목으로 나타날 가능성이 있다. 재검색을 하면 재해의 세목에 재해→간호 WY154.2가 검색된다. 이와 같이 어디까지나 색인이라는 것을 잊지 말고 검색어를 수정하면서 재검색을 하면 원하는 검색결과를 얻을 수가 있다.

색인검색은 초기상태에서는 부분 일치로 검색을 하도록 되어 있지만 검색 키워드의 입력 창의 왼쪽 옆에서 전방 일치나 후방 일치 검색방법을 변경할 수 있다. 예를 들면 [당뇨병]을 언어에 포함하는 단어(전당뇨병 상태 등)를 피해서 당뇨병에서 시작하는 단어(제1형 당뇨병 등)만을 검색하고 싶을 때는 전방 일치로 바꾸는 것이 좋다. 또 [증후군]으로 끝나는 언어를 찾고 싶을 때 등은 후방일치로 변경하는 것이 좋다.

검색 키워드의 입력 창에 키워드를 넣으면 서제스트 기능에 의해 전방일치로 색인어의 사전 검색이 진행되어 최대 10 단어후보가 표시된다. 예를 들면 [재생의료]라는 키워드의 분류 단어를 알고 싶다고 하자. 입력 창에 [재생]까지 입력하면 후보로 [재생의학]이 나오지만 [재생의료]는 나오지 않는다. 재생의료는 색인에 등록되어 있지 않다. Regenerative Medicine의 번역어는 의중지 시소러스에서 통제어로 [재생의학]이 사용되고 있어 NLMC-DB에서도 사용하고 있기 때문이다. 이와 같이 자신이 생각하는 번역어가 등록되어 있지 않는 경우에도 사전 검색의 후보를 보는 것으로 색인중의 번역어를 예측하여 적절한 검색을 할 수 있다.

4. NLMC-DB의 화면 구성

NLMC-DB는 사용 기기에 따라 최적의 화면 구성으로 이용할 수 있다. 이에 따라 노트북만이 아니라 태블릿이나 스마트폰으로도 다양한 검색이 가능하다. 노트북 이용자가 사용하고 있는 브라우저 표시폭을 좁히는 것으로 스마트폰 상의 화면구성을 재현할 수 있으니 어떻게 보이는지 궁금하신 분들은 테스트해보기 바란다.

III. NLMC-DB의 데이터와 구조

이번 장에서는 데이터베이스 실제 작성의 측면에서 NLMC-DB에 대해서 설명하겠다.

1. NLMC-DB 데이터

데이터베이스화에는 책자판 작성시의 데이터를 사용하였다.

엑셀 작업을 한 데이터는 데이터베이스화를 할 때 필요한 기본적인 구성요소를 갖추고 있었다. 하지만 책자판 작성의 작업을 시작했을 때부터 데이터베이스화를 염두에 둔 것이 아니라 어디까지나 인쇄물을 제작하기 위한 데이터였기 때문에 그대로는 사용하지 못하는 부분도 많이 있어 수정이 필요했다. 또한 책자판은 NLMC의 2016년판을 번역한 것이었다. 그로 인해 현상의 NLMC-BD

에는 2017년 이후의 개정은 반영되어 있지 않다.

이번의 서비스 개시에 수정작업을 마치지 못한 부분도 있다. 따라서 NLMC의 2017년 이후의 개정에 대해 대응을 해야 하는지 말아야 하는지에 대한 문제도 있다. 이에 대해서는 V. 과제와 전망에서 이야기하겠다.

2. NLMC-DB의 구조

작성을 담당한 WG는 한시적이므로 장기적인 지원이 어렵다. 검색의 예러나 관리의 수고를 없애기 위해 되도록 심플한 구성으로 데이터베이스를 구축하였다.

NLMC-DB는 MySQL을 이용한 관계형 데이터베이스(relational Database)를 구축하였다. 분류표를 격납하는 class 테이블과 테이블의 검색 속도를 빠르게 하는 index 테이블 두 가지로 나뉜다.

표 1에 간략화하여 발췌한 index테이블의 구조를 표시하였다. 책자판 작성시에는 회색부분에는 데이터가 없었다.

책자판의 색인의 순서는 가타카나(단어의 읽는 소리를 표기한)와 번호로 생성된 소트용 소리열(sort)로 제어되어 있었다. 책자판은 그로 충분했지만 데이터베이스로써는 그대로의 구조로는 검색 시의 표제어와 세목을 어떻게 묶을지가 문제였다. 대처법으로 여러가지를 생각했지만 최종적으로는 그림에 있는 대로 세목에도 표제어의 정보를 가지도록 데이터를 수정하였다. 이에 따라 색인 표제

표 1. Index 테이블 구조

dbno ID	Index_j 색인 표제어	Subhead_j 색인 부표목	Refer_number 참조 번호	Refer_index_j 참조 색인어	See_j 음/를(도) 보라	Result 번역결과	Sort 정렬 소리열
27957	당뇨병		WH 810-850			당뇨병(diabetes mellitus) WK810-850	당뇨병 가타카나 발음 10631
27958	당뇨병	임신중	WQ 248			임신중 WQ248	당뇨병 가타카나 발음 10631@10632
27959	당뇨병	간호	WY 155			간호 WY155	당뇨병 가타카나 발음 10631@10633
27960	당뇨병	치료	WK 815-825			치료 WK 815-825	당뇨병 가타카나 발음 10631@10634
27961	당뇨병			요독증	음/를(도) 보자	요독증을/를(도) 보라	당뇨병 가타카나 발음 10631@10635
27962	당뇨병			식사-당뇨병	음/를(도) 보자	식사-당뇨병을/를(도) 보라	당뇨병 가타카나 발음 10631@10636
27963	당뇨병			전 당뇨병 상태	음/를(도) 보자	전 당뇨병 상태를/를(도) 보라	당뇨병 가타카나 발음 10631@10637

어 열(index-j)을 검색하고 HIT한 결과 열(result)을 ID (dbno) 순으로 표시하는 것만으로 데이터베이스가 성립하게 되었다.

이 검색처리는 다음과 같은 매우 간단한 검색용SQL로 기술할 수 있다. 이로 인해 검색으로 생기는 반복되는 오류가 적고 검색 속도도 빨라졌다.

```
select result from index where index_j like '% 키워드%' order by dbno
```

검색용의 SQL의 기본적 기술 구조는 이하와 같다.

```
Select (A) from (B) where (C)(D);
```

(A)에서 호출 대상 컬럼과 그 호출대상을 지정하고 (B)에서 검색 대상의 테이블을 지정한다. (C)에서 검색의 조건을 기술하고 (D)에서 검색결과의 표시 측정 등을 지정한다. (A)(B)에는 복잡한 식을 기술할 수도 있으나 그 문장의 경우에는 심플하게 컬럼이나 테이블을 지정하고 있다. 또한 그 문장의 경우 (C)는 index_j like '키워드%' 다.

index_j가 검색 대상 컬럼, like는 부정요소를 포함한 검색방법의 선언, 싱글 포메이션 안이 검색 키워드가 되어 %부분에 임의의 문자열을 허용한다. 그 문장의 경우 (D)는 ID (dbno) 순으로 검색결과를 표시한다.

IV. NLMC-DB의 이용실태

2019년 6월 10일 NLMC-DB 서비스를 제공한 후 한 달이 지난 7월 9일까지의 이용통계 집계 결과 다음과 같이 나타났다.

30일간 탐 페이지의 방문 수는 612회이다. 평균을 내면 하루에 약 20회의 방문되었지만 이용의 96%는 월요일부터 금요일의 평일에 집중되어 있어 평일 방문 수의 평균은 27회였다.

검색은 2,372회로 그 중 분류검색이 1,101회 색인 검색이 1,271회였다.

복수의 IP주소에서 100회 넘는 검색을 하여 이미 업무에 NLMC-DB를 사용하는 기관이 있다는 것을 알 수 있었다. 한편 색인의 검색결과에서 조금 슬픈 실태도 알게 되었다. 색인 검색 내 약 3할정도의 359회 검색 건수 0으로 나타났다. II. 3. 2)으로 설명된 색인 검색 특유의 영향인 것으로 보인다.

색인검색의 검색 대상은 전자판의 색인어로 시소러스와 같은 어휘의 풍부함과 일반적인 Web검색과 같은 자유도는 없다. 전자판의 번역 작업을 할 때에는 4만어가 넘는 검색어의 통역(번역어의 통일, 조정)으로 작업 수행시 정신이 아찔해졌지만 데이터베이스의 어수로써는 부족하다는 것을 짐작할 수 있었다.

V. 과제와 전망

NLMC-DB의 수정과 갱신, 향후 전망에 대해서 서술하겠다.

1. NLMC-DB의 수정

NLMC-DB에는 다음과 같은 수정의 여지가 있다.

1) 오자, 탈자, 실수의 수정

책자판에서는 데이터 작성 후 인쇄하여 교정작업을 했다. 이번에 사용한 데이터에는 이 교정을 반영하지 않았기 때문에 오자, 탈자가 남아있는 부분이 있다. 또한 데이터베이스로써 구축했을 때 표제어 레벨이 맞지 않는 실수가 있어 향후 수정을 생각하고 있다.

2) 색인어 수정

책자판 작성시에 원본의 색인이었던 언어만이 아니라 번역 WG 독자적으로 분류표에 있는 언어와 동의어를 보충한 것이 있다. 이것들은 책자판에 필요한 최소 데이터(result, sort 예 등)일뿐, 데이터베이스로 필요한 데이터로는 부족하기도 하고 틀린 부분도 있다. 검색 시에 예러나 노이즈로 표시되어 이후 수정을 검토하고 싶다.

3) 읽기의 정비

II. 3. 2) 색인 검색의 유의사항에 기술된 것처럼 읽기만으로는 검색이 되지 않는다. 책자판 작성시의 작업 데이터의 읽기 정보와 정렬순의 소트를 위해 적절하게 편집되어 있고 읽기는 완전하지 않고 숫자와 기호가 섞여 있다. 이로 인해 읽기 정보는 데이터베이스에 존재하고 있으나 실제로 검색에 사용할 수 없다. 읽기 검색을 사용하기 위해서는 이후 정보를 갱신할 필요가 있으나 작업량이 많아 현재는 갱신할 예정은 없다.

4) 인터페이스의 수정

인터페이스에 대해서 대규모의 수정은 어려우나 소규

모르면 가능하다고 생각된다. 예를 들면 NLMC에 익숙하지 않은 이용자는 QZ, WA라는 분류기호에 대응하는 분류를 알 수 없어 단독으로 주제 분류 검색을 사용하기 어렵다. 이 점은 분류 검색의 분류 기호 창에 분류 명을 병기하는 등의 방법으로 개선의 여지가 있다고 생각된다.

2. NLMC-DB의 갱신

III. 1. NLMC-DB의 데이터와 구조에서 언급한 것처럼 NLMC-BD는 NLMC의 2016년판을 바탕으로 이 후의 개정은 반영되지 않았다. 매년의 개정판도 번역되면 그만한 것이 없지만 개정의 내용에 따라 데이터베이스 구조가 바뀌는 부분도 생기므로 그에 맞는 체제를 갖출 필요가 있다. 이 때문에 아쉽지만 개정판의 갱신 여부는 현재로는 예정이 없다.

3. NLMC-DB의 향후 전망

NLMC-DB의 향후 전망에 대해서 아래 2가지를 소개한다.

1) 일반 공개

현재 NLMC-DB는 JLMA 회원 한정으로 공개하고 있다. 이는 책자판과 전자판의 판매를 위한 배려이지만 일정기간이 지나 불필요해지는 시기에 이용제한을 풀고 일반인도 이용할 수 있도록 만들고 싶다.

2) API의 실장(実装)

현재의 데이터베이스는 스탠드 아론 시스템으로 확장성이 없다. 향후 구체적인 계획은 없으나 API 기반 연계 솔루션을 도입하면 외부 기능 서비스로 확대되어 편리성이 향상되지 않을까 생각한다.

예를 들면 Open Search에 대응하여 외부에서의 분류 검색에 대해서 그 분류에 대응하는 분류표의 표제 정보를 반영할 수 있게 되는 서비스 등이 있다. 엑셀의 web-service 함수 등에서 참조할 수 있게 되면 프로그래밍의 기능에 대한 지식이 없더라도 엑셀 상의 분류를 준비하면 바로 그 분류에 대응하는 표제어를 알 수 있다. 목록의 담당자를 시작으로 편리하다고 느끼는 이용자가 있을 것으로 생각된다.

다만 이러한 개발에는 큰 노력이 필요하다. 이에 맞는 요구가 있다고 해도 각 기능이 어떤 형태로 구현되는지

심사숙고가 필요하며 이후 NLMC-DB의 이용정보의 축적, 분석을 한 후 검토를 해야 한다고 생각한다.

VI. 마지막으로

시간이 촉박하여 NLMC-DB의 설명이 부족한 점이 있다. 데이터베이스로써는 미숙하고 수정해야하는 부분이 많고 아직 발전의 여지가 많이 남아 있지만 본 WG로써는 우선 그 역할을 해내야 한다고 생각한다. 다만 올해도 본 WG는 활동을 계속하여 V. 과제와 전망에 기술한 것처럼 데이터와 인터페이스를 수정 할 예정이다. NLMC-DB에 이상한 점, 의견 등이 있으시면 본 WG 앞으로 연락 주시길 바란다.

NLMC-DB의 작성은 JMLA에서는 지금까지 유례가 없는 일이었으며 많은 시행착오를 겪으며 시행되었다. 당초의 예상과 비교하면 생각하지 못한 문제를 많이 겪었다. 그런 만큼 본 서비스가 이후 지금의 형태를 유지하거나 형태를 바꾸더라도 오래도록 많은 사람이 이용하기를 바란다.

마지막으로 이 자리를 빌려 NLMC-DB의 구축 작업에 협력해 주신 분들께 감사드립니다.

NLMC-DB의 실제의 구축작업에는 마츠자카 시게노리님, 키시 유키님을 비롯하여 주식회사 인포레스타 여러 분께서 수고해 주셨다. 실제로 데이터 서비스를 제공하는 서버 계약에 대해서는 JMLA 사무국의 우사미님, 사무국장 타니자와님께도 도움을 받았다. 또 관련 서비스 전반에 대해서는 JLMA 기획, 홍보 위원회 위원장 마유즈님께 조언을 받았다. 실제 NLMC-DB의 구축과 서비스 제공은 많은 분의 도움으로 가능하였다.

REFERENCES

1. The Japan Medical Library Association. U.S. National Library of Medicine (NLMC) 2016 edition Japanese version. Tokyo: Japan Med Lib Assoc; 2017. Japanese.
2. Kozumi Tomomi. U.S. National Library of Medicine Classification 2016 edition Japanese version, Journal of the Japan Medical Library Association 2017;64(2):65. Japanese.
3. Kozumi Tomomi. U.S. National Library of Medicine Classification 2016 edition Japanese version: Original text-book and how to use the Japanese version, Hospital Librarian

- 2018;43(1):33-38. Japanese.
4. Committee of Central Japan Medical Journal. Japan-English Medical Terminology. [internet]. 2019. [cited by 2019 Aug.5]. Available from: www.jamas.or.jp/database/thesaurus.html/.
 5. Japanese Academy of Medical Sciences. Japanese Medical Terminology Web version. [internet]. 2019. [cited by 2019 Aug.5]. Available from: jams.med.or.jp/dic/mdic.html/.