

# 서지 및 전거 MARC 형식을 적용한 검색시스템 설계에 관한 연구

## A Study on the Retrieval System Design Applying the Bibliographic and Authority format

조수련(SOO-RYUN CHO)\*, 박재혁(JAE-HYEOK PARK)\*\*

### □ 목 차 □

- |                |                         |
|----------------|-------------------------|
| 1. 머리말         | 4.5 인덱스                 |
| 2. 시스템의 기본환경   | 4.6 검증                  |
| 3. MARC의 구성    | 4.7 문제점                 |
| 4. 포맷적용시 고려사항  | 5. 검색시스템의 설계            |
| 4.1 일반사항       | 5.1 일반사항                |
| 4.2 서지         | 5.2 전거레코드를 통한 서지레코드의 검색 |
| 4.3 전거         | 5.3 서지데이터의 출력           |
| 4.4 서지와 전거의 링크 | 6. 제안점 및 맺음말            |

### 초 목

MARC 포맷을 이용하여 DB를 구축하는 궁극적인 목적은 결국 구축된 데이터를 검색하는 데 있다. 따라서 본고에서는 실제로 검색시 발생하는 문제점과 좀더 검색효율을 향상시키기 위해 서지 및 전거데이터베이스 구축시 반드시 고려해야 할 문제점을 제시하고 이를 근거로 실제 검색시스템의 설계방안을 모색해 보았다.

### ABSTRACT

The Aim to construct bibliographic DB using MARC format is to retrieve the data. Therefore, this study shows the problems of retrieval in practical and points to be considered for the purpose to raise the efficiency of retrieval when constructing the bibliographic and authority data base. present how to design retrieval system based on the above the problems

\* 서울대학교 중앙도서관 전산실

\*\* 대림전문대학 문헌정보과 전임강사

■ 논문접수일 : 1995년 5월 1일

## 1. 머리말

이제 MARC(Machine Readable Catalog)는 도서관에서 뿐 아니라 자료를 정리하는 모든 곳에서 표준으로 인식되어 널리 사용되고 있다. MARC 포맷으로 데이터베이스를 구축하는 궁극적인 목표는 구축된 데이터를 검색하는데 있으므로 구축시 검색을 위해 여러 사항들을 고려하여야만 검색의 효율을 향상시키고 합리적인 검색시스템을 구현할 수 있다. MARC는 하나의 자료에 대한 정보를 체계적으로, 또한 일정한 기준에 따라 정리할 수 있도록 구성되어 있어 관리 및 구분이 쉽고 명확한 장점이 있는 반면에 모든 정보를 일정한 규칙에 맞도록 구분하여 작성하여야 하므로 정보를 무리하게 특정 구분에 맞추어야 하는 단점도 있다. 따라서 MARC 포맷을 사용하는 경우 어떻게 적용해야 검색시에 무리가 없을 것인가를 고려해 봐야 한다. 그러나 아직도 MARC 자체의 사용에만 급급하여 실제 구축된 자료가 검색에서 어떻게 활용될 것인가에 대해 충분히 고려되지 않은 채 구축에만 급급해온 것이 사실이다. 따라서 차후에 야기될 수 있는 검색의 문제점과 좀더 검색 효율을 향상시키기 위해 서지 및 전거 데이터베이스 구축시 반드시 고려되어야 할 사항들에 대해 논의하고자 한다. 이를 위해 본 고에서는 서울대학교 도서관 통합패키지 가운데 목록, 검색시스템을 근간으로 하여 기술하며 기본적인 환경은 다음과 같이 제한하기로 한다.

## 2. 시스템의 기본 환경

목록시스템에서 사용되는 포맷과 관련 규칙은 다음과 같이 제한한다.

### 2.1 사용포맷

- 서지포맷
  - 국내서
    - : KORMARC(KS C 5867)
  - 국외서(일서, 중국서 포함)
    - : USMARC
- 전거포맷
  - USMARC 전거 포맷(USMARC Format for Authority Data)

### 2.2 목록규칙

- USMARC : AACR2, AACR1, ISBD
- KORMARC : KCR3, 한국자동화목록기술 규칙, AACR2

### 2.3 표목의 표기원칙

- 로마자 : 그대로 표기
- 비로마자 : 국제 교류시 표준으로 인정되고 있는 표기방식에 따른다.
  - 일본어 : 수정Hepburn식 일본음 로마자
  - 중국어 : Wade-Giles식 중국음 로마자
  - 기타(노어 등) : LC Romanization Table에 근거하여 표기

### 3. MARC의 구성

MARC의 종류는 매우 다양하여 크게는 USMARC 계열과 UNIMARC 계열로 나뉘고 각각의 계열내에 국가별, 특정 기관별로 여러 가지의 MARC가 존재한다. 그 중 우리나라의 표준 MARC인 KORMARC는 USMARC 계열이며 UKMARC(영국), CANMARC(캐나다), OCLC-MARC 등도 USMARC계열에 속한다. 현재 우리나라에서는 USMARC와 KORMARC가 대종을 이루고 있으며 이들의 개요를 살펴보면 다음과 같다.

먼저, 서지 MARC의 적용범위는 <표 1>에 나타나 있으며 KORMARC인 경우 단행본과 연속간행물을 제외한 자료를 비도서자료로 구분하고 있으나 비도서자료에 대해서는 아직 시안만 발표되어 있다. 또한 각 자료유형은 모두 동일한 구성으로 되어 있으며, 전거레코드인 경우에도 표목의 구분에 따라 레코드 자체의 성격은 달라지나 구성은 동일하다.

한편 MARC의 구성은 다음과 같이 레코드

<표 1>

USMARC	KORMARC
Books(BK)	단행본
Archival and manuscript control(AM)	비도서자료
Computer file(CF)	
Maps(MP)	
Music(MU)	
Visual materials(VM)	
Serials(SE)	연속간행물

자체의 구성과 입력 데이터의 구성으로 구분할 수 있다.

#### 3.1 MARC 레코드의 구성요소

##### 1) 레코드 구조

: ANSI Z39.2(American National Standard for Bibliographic Information Interchange)에 따라 구현되어 있음.

##### 2) 내용표시자(Content Designation)

: 데이터 요소를 명확하게 식별하고 관리하기 위해 사용되는 각종 기호로 다음과 같은 종류가 있다.(USMARC Format for Bibliographic Data : Including Guide-lines for Content Designation 참조)

- 레코드종단기호(record terminator) : 레코드 끝을 나타내는 기호로 0x1D 사용
- 필드종단기호(field terminator) : 레코드 내에서 하나의 필드 끝을 나타내는 기호로 0x1E 사용
- 서브필드식별기호(subfield delimiter): 필드 내에서 하나의 서브필드 시작을 나타내는 기호로 0x1F 사용
- 필드요소식별기호(tag) : 가변장 필드의 의미를 나타내기 위한 기호로 세자리 숫자 사용
- 서브필드요소식별기호(subfield code): 필드내에서 각 서브필드의 의미를 나타내는 기호로 알파벳 소문자나 한자리의 숫자로 표현.
- 지시기호(indicator): 필드요소식별기호

에 추가 정보를 나타내기 위해 사용되며 두자리의 숫자 기호 사용

- 채움문자(fill character) : 실제 포맷 상에는 규정되어 있으나 입력할 수 없는 부분에 사용되며 0x7C(!)로 표시한다. 실제 적용 필드는 다음과 같다.

서지 tag 007, 008, 760-787 \$7

전거 tag 008, 4XX-5XX \$w

### 3) 데이터 내용

: 실제 데이터로 다양한 목록규칙 및 배열규칙이 적용된다.

## 3.2 MARC 데이터의 구성

MARC 데이터는 크게 다음과 같이 4가지로 구분한다.

### 1) LEADER

레코드 앞에 24자리 고정장필드로 레코드 길이 및 레코드의 기본적인 정보가 수록되어 있다. 서브필드나 지시기호는 사용되지 않으며 모든 요소는 일정한 자리를 가진다.

### 2) DIRECTORY

모든 tag에 대한 길이와 상대적인 시작 위치를 포함한다. 서브필드나 지시기호는 사용되지 않으며 모든 요소는 일정한 자리를 가지며 필드종단기호로 끝을 표시한다.

### 3) VARIABLE FIELDS

tag으로 구분되며 각각은 필드종단기호

를 갖는다. 모든 tag는 디렉토리부에 저장되고 각각의 tag에 대해 반복 및 필수 표시, 적용 자료유형의 제한이 가능하다. 가변길이 필드는 다음의 두가지로 구분할 수 있다.

#### 3.1) VARIABLE CONTROL FIELDS

구조적으로 가변길이 데이터필드와는 다르며 단일데이터나 일부 고정길이 데이터(tag 007, 008)를 수록한다. 서브필드나 지시기호는 사용되지 않으며 모든 요소는 일정한 자리를 가진다. 00X 필드가 여기 포함된다.

#### 3.2) VARIABLE DATA FIELDS

필드 내에 서브필드와 지시기호를 포함하는 가변길이 데이터필드.

## 4. 포맷적용시 고려사항

### 4.1 일반사항

MARC에는 시스템 구현시 도움이 되도록 고려된 부분이 많으므로 구현시 이를 충분히 활용해야 한다.

#### 1) MARC의 구분

MARC의 종류는 레코드 자체의 Leader /06의 코드로 구분이 가능하다.

- 전거레코드 : z

- 소장레코드 : v,x,y

- 서지레코드 : z,v,x,y 이외의 code

#### 2) MARC의 확장성

〈표 2〉

	내 용	현재값	
LDR/10	지시기호 자리수	2	
LDR/11	식별기호 자리수	2	
LDR/20-23	entry map	필드길이	4
		필드시작위치	5
		실행위치	0
		미지정	0

시스템 설계시 확장성은 매우 중요하다. 예를 들어 현재 MARC에서 하나의 필드 길이가 최대 9999bytes로 정해져 있고 (LDR/20의 값이 4로 지정되는 경우) tag 520에 입력할 초록의 길이가 10kbytes가 넘는다면 MARC 포맷은 사용할 수 없게 될 것이다. 그러나 실제로 이러한 문제를 해결하기 위해 리더 부분에는 해당 레코드에서 사용할 필드의 길이를 사용자가 조종할 수 있도록 되어 있다. MARC는 이러한 부분을 충분히 고려하여 설계된 포맷으로 〈표 2〉에 나타나 있는 부분을 파라미터화하여 시스템을 구현하여야 한다.

3) local tag

9가 포함되는 tag은 모두 local tag으로 사용가능하다. 예를 들어 XX9, X9X, 9XX 등이며 이 경우에도 전체 포맷에서 사용되는 조기성을 고려하여야 한다.

4.2 서지

1) 검색항목의 결정

이용자들은 어떠한 검색항목이 있는가를 가장 먼저 알고자 할 것이며 일반적으로 저자, 서명, 주제, 출판사, 각종 코드어 등이 포함된다. 그러나 실제적으로 데이터의 내용을 잘 알고 있다면 검색항목을 결정하는 일이 그리 간단하지만은 않을 것이다. 따라서 서울대시스템에서의 검색항목은 뒤에서 언급하기로 하고 서지레코드를 실제로 적용하는 방법에 따라 달라지는 항목들만 우선 언급하기로 한다.

1.1) 저자/인명

: 서지 4xx, 6xx, 8xx의 인덱스 처리

이 경우는 a) 저자에 포함하여 처리하는 방법, b) 인명key로 재구성하는 방법, c) 주제,

〈표 2〉

tag	X00	X10	X11
1XX	기본표목-개인명	기본표목-단체명	기본표목-회의명
4XX	총서사항/부출표목-개인명*	총서사항/부출표목-단체명*	총서사항/부출표목-회의명*
6XX	주제명부출표목-개인명	주제명부출표목-단체명	주제명부출표목-회의명
7XX	부출표목-개인명	부출표목-단체명	부출표목-회의명
8XX	총서부출표목-개인명*	총서부출표목-단체명*	총서부출표목-회의명*

\* : KORMARC에는 없는 tag

총서명으로 접근하도록 처리하는 방법 등이 있으나 일반적으로 저자와 인명은 검색항목으로 명확하게 구분을 두어 사용되지 않는 것이 보통이다. 그러나 특정 tag를 대상으로 생각해 보면 명확한 구분이 가능하다. 저자 또는 인명과 관련된 tag는 'X00, X10, X11'이며 각각의 내용과 성격은 <표 3>과 같다.

위의 표에서 1XX와 7XX는 인명, 저자 어떤 항목으로 하여도 무방하다. 그러나 그외의 tag는 적용하기에 따라 달라진다. 먼저 6XX는 인명 항목에는 포함이 되나 저자항목에는 포함되지 않는다. 4XX와 8XX는 어떤 목록규칙을 사용할 것이냐에 따라 달라지는데 KORMARC의 경우 '한국자동화목록형식'을 적용하여 총서 관련으로는 440, 490, 830tag만을 사용하므로 서명항목으로 포함되거나 USMARC에서 AACR2를 적용하는 경우는 440, 800, 810, 811, 830tag를 사용하므로 800, 810, 811tag의 적용이 모호하다. 이 경우, tag적용은 인명 위주이며 경우에 따라 저자이거나 인명이므로 인명항목으로 구분하여야 한다.

예를 들어 총서명이 'Charles Dickens Works.'인 경우의 MARC와 목록규칙에 따른 적용은 다음과 같이 달라진다.

< USMARC에서 AACR2 사용 >

4901 \$aCharles Dickens Works.

800 \$aDickens, Charles \$tCharles Dickens Works.

< KORMARC 적용 >

440 \$aCharles Dickens Works.

위와 같이 동일한 총서명이 달리 표현되는 경우 첫번째 경우 800을 인명으로, 두번째 경우 440을 서명으로 적용하여야 할 것이다. 따라서 tag 400, 410, 411, 800, 810, 811을 사용할 경우 저자보다는 인명으로 검색항목을 잡아야 검색의 효율을 높일 수 있으며 KORMARC에서와 같이 총서사항으로 440, 830tag만을 사용하는 경우는 저자 항목을 선택하고 총서는 서명항목으로 검색하는 것이 바람직하다. 물론 저자명과 인명을 모두 사용할 수도 있으나 인명에는 저자명이 포함되므로 동일한 검색key가 중복으로 색인되는 문제점이 있다.

## 1.2) 서명/총서명

: 총서명의 변형 표기

서명과 총서명을 함께 서명으로 검색할 것인지 아니면 서명과 총서명을 구분하여 검색할 것인가에 대해 고려하여야 한다. KORMARC에서와 같이 총서명이 자체로서 전거통제를 받는 경우는 총서명을 굳이 서명과 구분할 필요가 없으나 USMARC에서 AACR2를 사용하는 경우 총서명 관련 일부 tag이 인명으로 검색된다면 실제 인명을 포함하고 있는 총서명을 검색하고자 할때 이용자에게 혼란이 올 수도 있을 것이다. 따라서 이런 경우 서명을 총서명과 구분하여 별도의 검색항목으로 처리함으로써 인명을 포함하는 총서명을 인명으로 또는 총서명으로 다양한 접근점을 제공할 수 있다.

이를 위해 4901tag에 원래의 서명을 입력하고, 830tag에 변형을 입력하면 되나 동양서의 경우 830은 전거통제를 받는 tag으로 로마나이즈된 표목이어야 하므로 실제의 일본어나

중국어 변형표기는 서지레코드 상에 존재하지 않아 4901과 830의 관계가 모호해질 뿐만 아니라 반복사용이 가능하여 변형표기와 변형표기의 로마나이즈를 모두 유지할 수 있다 하더라도 이 경우 전거통제 tag를 구분할 수 없다는 문제점이 있으므로 변형이 아닌 로마나이즈된 읽기를 830에 유지하고 전거에서 처리하는 것이 합리적이라 생각한다.

2) 전거통제 규칙

2.1) 서지레코드 상에서는 모든 서브필드를 사용하되 전거레코드 생성시 서브필드를 제한하여 동일표목에 대해 지나치게 여러개의 전거표목이 생성되는 것을 방지 하므로써 표목을 모아 통제받는다.

2.2) 기관명과 회의명에서는 次, 年, 地에 대한 사항을 통제받지 않는다.  
(계속적인 회의명은 회의명만, 비정기적인 회의명은 차,년,지까지 모두 전거통제

〈표 4〉

Tag	USMARC		KORMARC	
	undefined	uncontrolled	undefined	uncontrolled
100	u,4,6	e		e
110	u,4,6	c,d,e,n,p		c,d,e,n,p
111	u,4,6	c,d,n,p		c,d,n,p
130	6			
240	6			
243	6			
440	v,x;6		s,v,x	
600	u,2,3,4,6		2	
610	u,2,3,4,6		2	c,d,n,p
611	u,2,3,4,6		2	c,d,n,p
630	2,3;6		2	
650	c,d,e,2,3,6		2	
651	2,3;6		2	
700	u,x,3,4,5,6	e		e
710	u,x,3,4,5,6	c,d,e,n,p		c,d,e,n,p
711	u,x,3,4,5,6	c,d,n,p		c,d,n,p
730	x,3,5,6			
800	u,v,4,6	e		
810	u,v,4,6	c,d,e,n,p		
811	u,v,4,6	c,d,n,p		
830	6		v	

를 받게 되어있으나 구분하기 어려운 점이 있어 통제 받지 않는다.)

2.3) 역할지시어에 대한 사항은 통제받지 않는다.

2.4) 서지포맷상에는 정의되어 있으나 전거포맷에 정의되어 있지 않은 서브필드 (numeric subfield : 통제관련)는 모두 제외된다.

2.5) 전거통제를 받는 tag

전거통제를 받는다는 것은 해당tag의 내용과 동일한 내용을 갖는 tag이 전거레코드의 1XX(채택표목)으로 잡혀있다는 것을 의미하므로 작성시 전거표목을 확인하여야 하며, 없는 경우에는 반드시 신규로 동일한 전거레코드를 생성해 주어야 한다. 전거통제를 받는 tag과 전거통제시 제외되는 서브필드 리스트는 아래와 같다.

3) 표시상수와 표출어의 구분

표시상수는 tag이 갖는 의미를 누구나 쉽게 알도록 하거나 코드에 포함되어 있는 서브코드의 자릿수가 일정하여 구분표시를 입력할 필요가 없는 경우, 또는 목록카드 출력 등의 일정한 표기법이 적용되는 경우에 데이터 상에 입력하지 않고 출력시 자동적으로 생성되는 상수이며 표출어는 tag만으로는 데이터의 성격을 명확하게 구분할 수 없는 경우 MARC 데이터에 포함되는 단어이다. 특히 표출어의 경우는 KORMARC의 255 tag을 제외하고는 주기사항에만 사용되므로 검색과 직접적인 연관은 없

으나 표시상수를 데이터에 입력하는 경우 엉뚱한 검색어가 생성될 수 있으므로 반드시 구분하여 사용하여야 한다.

4) local tag의 적용

읽기항목을 위해 local tag을 사용하는 경우 전거레코드에 포함될 표목은 서지레코드에 입력하지 않아야 한다. 그렇지 않은 경우 전거를 통해 찾을 수 있는 표목에 대해 중복으로 인덱스를 생성하게 되기 때문이다. 예를 들어 다음의 경우에서 (X)로 표시된 부분은 서지레코드에서 제외를 하고 전거레코드의 4XX에 포함될 사항들이다.

예) KORMARC의 경우

700 10\$aK'ung, Hsiang-lin  
245 10\$a孔子聖蹟圖 / \$d孔祥林 ;  
700 10\$a공, 상림 (X)

예) 서울대학교의 경우

100 10\$aK'ung, Hsiang-lin  
245 10\$a孔子聖蹟圖 / \$d孔祥林  
700 10\$a공, 상림 (X)  
700 10\$a孔祥林 (X)  
700 10\$a공시양린 (X)

4.1) 9XX

변형된 표기나 기타 검색키가 될 수 있으면서 기존의 서지tag이나 전거레코드내에 입력할 수 없는 경우 사용한다. tag내의 서브필드 및 지시기호, 반복유무 등은 대응되는 일반 tag과 동일하게 적용한다.

※ 현재 전거레코드를 생성하지 않는 기관에서는 추후 전거레코드를 자동생성할 수 있도록 tag 9XX에 \$6 서브필드를 사



- 용하여 읽기 및 변형을 입력할 수 있다.
- 900/910/911 사용  
개인명, 단체명, 회의명을 위해 사용.
- 930 사용  
통일서명, 총서명을 위해 사용.
- 940 사용  
일반서명을 위해 사용.

< tag 900 사용 예 >

- 서지레코드  
100 \$aあいざわ, まちこ  
245 \$a藤田東湖 / \$c相澤まち子  
900 \$a상택마치자
- 전거레코드  
100 \$aあいざわ, まちこ  
400 \$a Aizawa, Machiko  
400 \$a 아이자와, 마치코
- ※ 전거레코드에 포함시킬 수 있는 비채택 표목은 기관마다 별도로 결정할 수 있다.

4.2) 090 MR

- \$6 : link field AN
- \$a : 청구기호 MN
- \$b : 도서기호 MN
- \$c : card 출력용 권연차 기호 AR
- ※ \$c에는 card 출력을 위해 창간호에 대한 volume 및 연도의 open entry 형식의 입력을 한다.

4.3) 093 MR

- \$6 : link field MN
- \$a : 등록번호 MR
- \$v : 권차사항 AR

- \$c : copy사항 AR
- \$x : 별칭기호 AN
- \$y : 분관표시 AN

※ 상세화면상에 093은 출력하지 않으므로 연결된 권차사항의 표시가 090이나 505에 기입하여 출력하도록 한다.

4.4) 기존포맷의 일부 변형 사용

-502 (USMARC)

- 1st indicator 0 : 석사학위논문
- 1 : 박사학위논문

- \$a : 학위명
- \$b : 학교명
- \$c : 학과명

※ 학위논문 검색을 위해 특정한 서브필드를 사용하여 적용.

-4901 (USMARC)

- 2nd indicator blank : 830에 변형된 총서명 입력
- 1 : 830에 읽기 입력 (romanize 등)

-507 (KORMARC)

원저자 및 원서명을 사용하기 위한 주기tag이나 원저자명인 경우 전거를 받아야 하며 이 경우 다른 저자명 관련tag과 서브필드의 구성이 다르면 전거통제를 받을 수 없으므로 100tag과 동일하게 서브필드를 확장하여 전거통제를 받고 단체명과 회의명은 별도로 7XX에 부출을 낸다.

-440, 830 (KORMARC)

※ \$a, \$v는 반복사용하지 않음.(tag을 반복 사용)

〈표 5〉

제한항목	관련MARC정보		비 고
	KORMARC	USMARC	
발행년도(단행본)	008/07-14	008/07-14	008/06의 코드에 따라 별도 처리 〈표 5.1〉
언어	008/35-37	008/35-37	'USMARC Code list for language' (3자리)
정부간행물	008/38-39	008/28	'정부기관명부호표'
대학간행물	008/26-27	X	'한국대학교명부호표'
회의간행물	008/29	008/29	0 : FALSE 1 : TRUE
기념논문집	008/30	008/30	0 : FALSE 1 : TRUE
전기	008/34	008/34	
자료유형	LDR 06/07	LDR 06/07	〈표 5.2〉
비치장소	X	X	대출시스템 데이터 사용
분관	X	X	대출시스템 데이터 사용
도착일	X	X	연속간행물시스템 데이터 사용
내용형식	008/24	008/24	KORMARC code 기준 (21개 항목)

5) 제한검색사항

MARC내에는 검색시 이용자가 좀더 특정한 결과를 얻고자 원하는 경우 사용할 수 있는 제한적인 정보가 많이 수록된다. 일반검색에서 사용될 표목은 보통 후방절단을 기본으로 하며

여러가지 검색이 될 수 있을 만한 표기를 함께 제공할 뿐 아니라 구두점 등의 특수기호를 생략하는 등 다각도로 검색에 편리하도록 가공되어 조금은 불명확한 데이터가 입력된다고 해도 검색에 큰 영향을 미치지 않는 반면 제한검색

〈표 5.1〉 년도제한 데이터 지정

code (008/06)	발행년도1 (008/07-10)	발행년도2 (008/11-14)	비 고
c	○	X	1개 제한항목 생성
d (US)	○	X	1개 제한항목 생성
m	○	○	범위지정가능
n	X	X	제한항목생성안함
q	X	X	제한항목생성안함
r	○	○	2개 제한항목 생성
s	○	X	1개 제한항목 생성
b (US)	X	X	제한항목생성안함

〈표 5.2〉 자료유형 구분(USMARC)

자료 유형	LDR/06	LDR/07
Book (BK)	a	a,c,m
Archival and Manuscript (AM)	b	
Computer File (CF)	m	
Maps (MP)	e,f	
Music (MU)	c,d,i,j	
Visual Material (VM)	g,k,o,r	
Serial (SE)	a	b,s

에서 사용되는 데이터는 입력된 그대로 사용되므로 검색시 결과에 크게 영향을 미치므로 제한사항으로 사용되는 데이터의 입력시에는 신중을 기해야 한다. 제한항목으로 사용되는 데이터는 〈표 5〉에 나타나 있다.

### 4.3 전거

#### 1) 서지와 전거의 관계

전거레코드에는 서지데이터와 관련이 있는 레코드와 서지데이터와는 무관한 레코드 두가지가 있다. 서지와 관계되는 경우에는 검색시

〈표 6〉

008/09	내 용	서지레코드와의 관련여부
a	채택표목	관련
b	참조레코드(보라참조 없음)	무관
c	참조레코드(보라참조 있음)	무관
d	세목레코드	무관
e	노드레이블레코드	무관
f	채택표목/세목레코드	관련
g	참조/세목레코드	무관

관련자료까지를 검색하여 검색의 효율을 높이는데 사용되며 서지와 무관한 경우에도 시소러스의 기능을 하거나 표목 선정에 있어 참고할 수 있는 중요한 데이터이다. 이러한 데이터의 구분은 아래의 〈표 6〉에서와 같이 008/09의 레코드 유형정보로 가능하다. 여기서 서지와 무관하다는 것은 전거 1XX 표목이 실제 서지 데이터에 나타나지 않음을 의미하며 실제 검색시에는 관련된 전거데이터 만이 사용된다.

〈표 7〉

표 목		각 표목별 표준 전거tag		
		확정표목	보라표목	도보라표목
인명 전거	개인명	100	400	500
	단체명	110	410	510
	회의명	111	411	511
	통일서명	130	430	530
	총서명	130	430	530
주제 전거	개인명	100	400	500
	단체명	110	410	510
	회의명	111	411	511
	통일서명	130	430	530
	일반주제 지명	150 151	450 451	550 551

〈표 8〉 (\*008/09 = a 인 경우에만 유효)

구 분	내 용	
	인 명	주 제
008/11	n	a,b,c,d,k,v,r,z
008/14	a	b
008/15	b	a

#### 2) 전거표목의 구분

이제부터는 서지데이터와 관련된 데이터가

운데에서도 확정표목이 있는 데이터만을 대상으로 하여 살펴본다. 전거표목은 크게 인명과 주제 두가지로 구분된다. 따라서 전거시스템도 크게 두가지로 구분되며 각 시스템 내의 사항과 관련tag은 <표 7>와 같다. 이러한 구분은 전거 MARC내의 데이터<표 8>를 가지고 자동적으로 하게 되므로 관련 정보를 정확히 입력해야한다.

단) 008/10의 경우 AACR2목록규칙에 따라 작성되었으면서(008/10 = c) 주제명 표목으로 사용이 가능한(008/15 = a) 레코드가 있을 수 있으므로 인명과 주제를 구분하는 key로 사용할 수 없음

### 3) 배열제외자수의 처리

전거데이터는 배열제외자수를 제외한 후에 작성되므로 실제로 제 2 지시기호에 배열제외자수를 기입하도록 되어 있으나 항상 0으로 채워지게 된다.

## 4) 전거레코드 작성시 고려 사항

### 4.1) 1XX 채택표목을 가지지 않는 전거레코드의 허용 여부

일반적으로 국외서인 경우 참조용 도구가 많이 있으므로 채택표목을 정하는 일은 국내서보다 수월하지만 국내서에 있는 표목에 대해 채택표목을 정하기란 쉬운일이 아니다. 그렇다고 해서 서지레코드의 전거통제받는 tag를 입력하지 않을 수는 없으므로 확정표목으로 할 수 없는 경우 008/09의 레코드 유형을 'b'로 하여 전거레코드를 구성하는 것도 한가지

방법이 될 수 있다.

### 4.2) 변형이 없는 전거레코드의 생성

변형이 없는 경우는 실제 전거의 기능인 표목의 grouping이나 검색의 효율 향상에 영향을 미치지 않아 특별한 의미를 가지지는 않으나 추후 변형이 발견된 다음에 서지레코드를 대상으로 하여 확정표목을 다시 설정하기 번거로우므로 생성하는 것이 바람직하다.

### 4.3) 전거 1xx 표목이 변경되는 경우 처리 방법

이 경우 이론상 처리할 수 있는 방법은 첫째, 1xx 표목변경시 기존 서지 및 인덱스 모두 변경하는 것이고, 둘째, 새로운 레코드생성후 5xx로 연결하여 기존 서지 및 인덱스 변경할 필요없이 5xx \$W에서 처리하는 것이고, 셋째, 1XX 변경 후 변경 전 표목을 4xx에 삽입하여 기존 서지 및 인덱스를 변경하고 4xx \$W에서 처리하는 방법이 있다.

### 4.4) 4XX와 5XX의 적용 범위 설정

#### (1) 4XX의 적용 범위

- 로마자의 타 언어 읽기의 경우 변형이 서지사항에 발생할 때 마다 추가  
e.g. Shakespeare→셰익스피어  
→셰익쓰피어  
→セイスピラ  
→サイスピマ
- 원어표기 추가  
e.g. Dostoevskii→ДОСТОЕВСКИЙ

- 타 언어표기는 서지사항에 나타날 때  
마다 추가  
e.g. Tanaka, Hiroshi → たなか ひろし  
→ 田中 廣

(2) 5XX의 적용범위

- 단체명, 회의명 등이 바뀐 경우 사용  
e.g. 문교부 → 교육부 / 문화부  
전거레코드1 전거레코드2 전거레코드3  
110 \$a문교부 110 \$a교육부 110 \$a문화부  
510 \$a교육부 510 \$a문교부 510 \$a문교부  
510 \$a문화부
- 2개의 표목이 공존하는 경우나 관련  
주제의 표현  
특별히 008/09 데이터가 a(확정표목레  
코드)가 아닌 주제관련 레코드 형식을  
지원하지 않는 경우에는 5XX를 사용하  
여 관련주제를 표현할 수 있을 것이다.

도의 tag도 있으나 <표 9> 대상이 되는 레코드  
는 연간물이나 총서 등 다권본으로 정리된 경  
우에 한하며 저자명에 대한 전거레코드나 주제  
명에 대한 전거레코드에는 적용할 수 없다.

<표 9>

TAG	반복여부	적 용 대 상 (전거레코드구분/레코드유형(008/09))
014	반복	총서 / a
020	반복	총서 / a
022	반복	총서 / a
060	반복	총서 / 모두
070	반복	총서 / 모두
082	반복	총서 / 모두
090	반복	총서 / a
(local)	반복	총서 / a

4.4 서지와 전거의 링크

전거레코드와 서지레코드를 연결하는 방법  
에는 서지의 각 전거표목마다 관련된 전거레코  
드의 통제번호를 갖는 직접적인 방법과 동일  
데이터를 데이터베이스에 기록하여 연결하는  
간접적인 방법이 있을 수 있다. 그러나 직접 연  
결방법인 경우 서지데이터가 길어질 뿐 만 아  
니라 표목을 수정할 때 마다 전거통제번호까지  
함께 수정해 주어야 하는 번거로움이 있다. 따  
라서 대부분의 시스템에서는 전거와 서지에 있  
는 동일한 표목을 DB상에서 인식하여 연결할  
수 있는 간접적인 방법이 많이 사용되고 있다.  
또한 표목 외에도 MARC상에 동일 데이터를  
유지함으로써 연결레코드를 식별할 수 있는 용

4.5 인덱스

인덱스의 생성은 검색에 있어 가장 중요한  
도구를 제공하는 부분으로 목록데이터가 아무  
리 완전하게 구축되어 있다고 하더라도 인덱스  
생성정책이 잘못되어 있다면 검색되지 않는 무  
용지물이 되어 버릴 것이다. 인덱스는 일반적  
으로 이용자가 많이 사용하는 항목을 key로 만  
들어 검색이 수월한 형태로 재조직한 것을 말  
하는데 도서관 장서 검색시스템에서 많이 사용  
되고 있는 항목(prefix)별 유형과 실제 색인어  
구성 단위는 다음과 같이 나누어 볼 수 있다.

1) 인덱스 항목별 유형

1.1) 저자(또는 인명) : 앞서 서지레코드에

서 언급

1.2) 서명

1.3) 출판자

1.4) 주제

1.5) 각종 분류기호 : 기관별로 사용하고 있는 분류기호외의 모든 분류기호를 포함한다.

e.g. 청구기호, DDCN, NALN, NLMN 등

1.6) 각종 통제기호 : 분류기호나 통제기호 이외에 특정 기관등에서 부여하는 각종 식별기호를 의미한다.

e.g. CODEN, STRN, Report Number, ISSN, ISBN

2) 색인어 구성단위

다음의 구절과 어절은 저자명, 서명, 출판자, 주제명에 한하여 적용되며 각종 기호들은 별도의 단위로 구분하여 처리하는 것이 편리하다.

2.1) 구절(Phrase) : 특정한 tag이나 서브필드에서 추출된 2개 이상의 어절을 포함하는 문자열.

e.g. Papillomaviruses and Human Disease

2.2) 어절(Keyword) : 추출된 구절에서 공백문자나 특정 분리문자로 분리된 각개의 문자열로서 위의 구절에서 추출되는 어절은 다음과 같다.

e.g. Papillomaviruses, Human, Disease

3) 구성단위별 색인 처리

3.1) 구절의 색인처리

구절을 색인처리하기 위하여는 먼저 적용할 처리규칙을 정하고 구절의 성격에 따라 달리하여야 할 부분을 명확히 구분하여야 한다.

3.1.1) 구절의 구분

구절은 크게 전거통제를 받는 구절, 즉, 전거통제를 통하여 검색할 구절과 그렇지 않은 구절로 나눌 수 있는데 이 경우 동일한 tag와 서브필드에서 구절을 추출한다고 해도 전거를 받기 위한 구절과 그렇지 않은 구절은 내용이 달라지게 된다.

- 전거통제를 받는 구절

전거통제는 기본적으로 MARC상에 기입한 그대로 받아야 한다. 즉, 목록규칙에 따라 일정하게 정해져 있는 순서를 따르지 않는 경우 타 기관이나 전거유틸리티에서 down-load받은 전거 데이터를 그대로 사용하여 검색할 수 없기 때문이다. 따라서 전거통제용으로 구절을 생성하는 경우 위에 정의되어 있는 몇몇 서브필드(표 4)를 제외한 입력된 모든 서브필드가 색인 대상이 된다.

- 전거통제를 받지 않는 구절

전거통제를 받는 tag의 일부 서브필드와 전거통제 받지 않는 모든 tag에서 추출된 구절이 포함된다. 이 경우 서브필드는 입력된 순서와는 무관하게 추출 가능하며 서브필드의 반복에 따라 융통성있는 색인 추출도 가능하다.

다음의 예에 동일한 tag에서 전거통제를 받는 구절과 그렇지 않은 구절이 달리 추출되는 예를 보면 다음과 같다.

<예>

440 0\$Janua Linguarum.\$pSeries maior,\$x0075-3114 :\$v100

→ 전거통제를 받는 구절

Janua Linguarum. Series maior, (\$v, \$x는 전거통제 받지 않는 서브필드임)

→ 전거통제를 받지 않는 구절

@Janua Linguarum. Series maior,  
@Series maior,  
@Janua Linguarum. v100

### 3.1.2) 추출된 구절의 색인화

상기의 3.1.1) 기준에 따라 추출된 구절은 여러가지의 정규화 규칙에 따라 조정된 후 색인으로 등록된다. 각각에 적용되는 정규화 규칙에는 다음과 같은 종류가 있다.

- ㉔ 소문자를 대문자로 바꾼다.
- ㉕ 모든 구두점과 일부 특수기호를 삭제한다.  
: 2바이트 코드(<) 0x7f)나 alphanumeric 외의 1바이트 코드 삭제.
- ㉖ 영문과 영문, 영문과 타언어 사이의 공백문자는 하나만 인정하고 그 외의 경우 모두 무시한다.  
: 즉, 2바이트 코드사이의 공백문자는 모두 삭제한다. 따라서 한글과 한글 사이의 공백은 삭제되는데 이로써 서로 다르게 표기된 띄어쓰기로 인한 검색 결과에서의 탈락을 방지할 수 있다.
- ㉗ 앞, 뒤의 () 내용을 삭제한다.

020tag(ISBN) \$a의 ()내 정보 삭제등의 경우 사용.

- ㉘ 구절 앞, 뒤 부분의 공백은 그대로 인정한다.  
010tag(LC Control Number) \$a의 prefix로 사용되는 공백을 인정하기 위한 규칙.
- ㉙ 2바이트 코드 중 예외로 지정된 코드를 삭제한다.
- ㉚ 모든 공백을 삭제한다.
- ㉛ 한문은 한글로 변환한다.  
: KORMARC 245tag \$a 서명 상의 한문을 한글로 변환하여 색인을 생성한다.
- ㉜ 특수 2바이트 기호를 한글표기로 변환한다.  
: 변환하고자 하는 코드는 기관별로 정하며 예로 그리스 문자인 'ϑ'를 '세타'로 변환하여 색인 생성.
- ㉝ 특수 2바이트 기호를 영문으로 변환한다.  
: 변환하고자 하는 코드는 기관별로 정하며 예로 그리스 문자인 'δ'를 'delta'로 변환하여 색인 생성.
- ㉞ 하이픈(-)을 삭제한다.  
: 서명내에 'MS-DOS'와 같은 구절이 있는 경우 'MSDOS'와 같이 변환하여 검색시의 하이픈 사용과 무관하게 검색이 가능하게 한다.
- ㉟ 배열제외자수 및 관칭에 따라 구절 앞부분의 일정한 길이를 삭제한다(<표 10>참조).  
: 배열제외자수는 서명관련 tag에만 제한이 되며 \$a에 한하여 적용한다. 따라서 440tag에서 \$p(하위총서명)을 별도의

구절로 추출하고자 한다면 440은 배열 제외자수 적용tag이지만 추출하고자 하는 서브필드 \$a가 아니므로 적용할 수 없다.

㉞ 총서명 \$v의 불림사항을 전개하여 개별화한다.

: 수서 복본조사 등에서 사용할 수 있도록 각권별 색인을 한다. 각권 표시기호 (vol., v., no., 권 등)는 무관하게 실제 번호만을 대상으로 하여 색인한다.

<예> 440 4\$aThe critical idiom :\$v24-25  
 → critical idiom 24  
 → critical idiom 25

이상의 정규화 규칙과 순서를 적절히 조합하여 사용하면 모든 tag에 대한 구절정규화가 가능하다. 다음에서 몇가지 예를 들어 정규화 규칙에 대해 명확히 한다.

<예 1>

USMARC "70010\$aJung, C. G.\$q(Carl Gustav).\$d1875-1961,\$eformer owner."에 대해 전거통제용으로 색인하고자 하는 경우 정규화 작업 순서

- 1) 색인하고자 하는 서브필드 추출(전거통제용이므로 모든 서브필드가 대상이 되며 기입된 순서대로 추출)
- 2) 배열제외자수여부 판정(배열제외자수 적용tag 아님 <표 10>)
- 3) 전거통제시 제외되는 서브필드 삭제(\$e 삭제 <표 4>)  
 → Jung, C. G. (Carl Gustav), 1875-1961.

4) 적용할 정규화 규칙 및 순서

㉠→JUNG, C. G. (CARL GUSTAV), 1875-1961,

㉡→JUNG C G CARL GUSTAV 1875 1961

㉢→JUNG C G CARL GUSTAV

<표 10>

Tag	지 시 기 호		
	USMARC	KORMARC	
	(non-filing)	(관제())	
130	1		•
211	2		•
214	2		•
222	2		•
240	2		•
242	2		•
243	2		•
245	2	1	0 : ( )제외 (기본표목)
			1 : ( )제외
			2 : ( )포함, 제외 각각 생성
440	2	1	0 : ( )제외
			1 : ( )포함
507	.	1	0 : 인덱스 생성 안함
			1 : ( )제외
			2 : ( )포함, 제외 각각 생성
630	1		•
730	1		•
740	1		•
830	2		•



1875 1961←색인구절

〈예 2〉 KORMARC “24520\$a(建築)構造計劃/\$d金炫山 著”에 대해 전거통제를 받지 않는 구절을 생성하고자 하는 경우 정규화 작업 순서

- (1) 색인하고자 하는 서브필드 추출(\$a의 내용을 서명으로 추출)
- (2) 배열제외자수여부 판정(배열제외자수 적용tag이며 \$a 포함)
  - (2.1) ()안의 내용 포함하여 색인  
→(建築)構造計劃/
  - (2.1.1) 적용할 정규화 규칙 및 순서
    - ㉠→建築 構造計劃
    - ㉡→建築 構造計劃←색인구절
    - ㉢→건축 구조계획←색인구절
  - (2.2) ()안의 내용 제외하고 색인  
→構造計劃/
  - (2.2.1) 적용할 정규화 규칙 및 순서
    - ㉠→構造計劃
    - ㉡→構造計劃←색인구절
    - ㉢→구조계획←색인구절
- (3) ㉠규칙을 적용하는 경우 한글 변환하지 않은 구절도 색인구절로 등록하여 특정 한자를 정하여 검색하고자 하는 경우에도 검색이 가능하도록 한다.

### 3.1.2) 어절의 색인처리

어절은 기본적으로 구절색인 생성시 공백문자나 특수기호를 기준으로 분리하여 생성한다. 여기서 특수기호란 기관이나 색인의 종류에 따라 달리 적용이 가능하다. 일반적으로 모든 구두점은 분리기준으로 사용되지만 구절색인 생성시 원하는 구두점과 기호를 공백으로 변경하

는 처리가 선행되므로 굳이 특수기호를 별도로 지정하지 않아도 될 것이다. 다만 하이픈의 경우는 서명, 저자명의 경우 달리 적용되어야 한다. 즉, 서명에서는 분리기호로 사용할 수 있으나 저자명인 경우에는 분리기호로 사용할 경우 의미없는 어절색인이 다수 생성되기 때문이다.

#### 3.1.2.1) 서명 어절색인시 하이픈 처리

“MS-DOS 입문”이라는 서명이 있는 경우 이용자가 검색할 만한 구절색인과 어절색인은 모두 생각해 보면 다음과 같을 것이다.

#### 〈구절색인〉

- MS-DOS 입문
- MSDOS 입문
- MS DOS 입문

#### 〈어절색인〉

- MS, DOS, MSDOS, 입문

첫번째 구절색인은 실제 검색시 검색시스템 상에서 색인작성시와 동일하게 하이픈을 공백문자화하거나 삭제함으로써 검색이 가능하므로 별도의 구절색인으로 생성하지 않는다. 두번째와 세번째 구절색인은 하이픈을 공백문자화한 것과 삭제한 것의 두가지로 생성한다. 이렇게 구절 색인이 생성되면 어절색인은 생성된 2개의 구절색인에서 공백을 기준으로하여 분리해내면 된다.

#### 3.1.2.2) 인명 어절색인시 하이픈 처리

중국인명을 로마나이즈하는 경우 음절을 명확히 구분하기 위해 공백이나 하이픈이 사용된다. 그러나 공백을 사용하는 경우 실제 어절생

성 기준과 혼동될수 있으므로 하이픈으로 연결하여 표시한다. 따라서 인명에 사용되는 하이픈은 삭제하여 음절마다 어절색인이 되지 않도록하여야 한다.

〈예〉 "Chu, Hsiao-tsang"

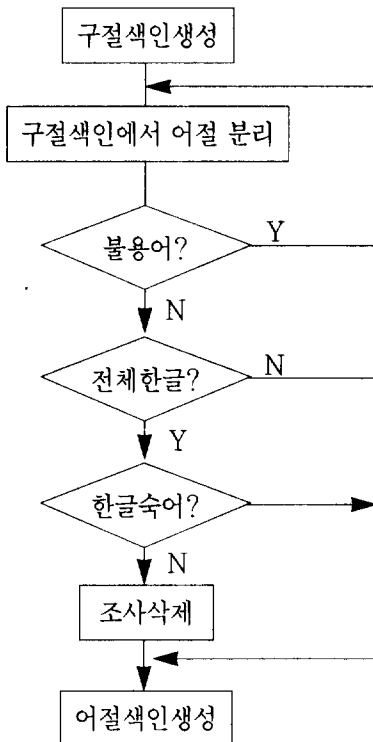
구절색인→ Chu Hsiaotsang

어절색인→Chu

Hsiaotsang

### 3.1.2.2) 적용화일

다음 화일들은 목록시스템에서 색인을 구축하거나 검색어를 조정하기 위하여 사용되는 화일들로 서지데이터 상의 정보를 다음의 화일을 근거로 하여 수정하여 색인화하고 검색시에도 이용자에 의해 입력된 검색어에 대해 동일한 처리가 되어야만 검색이 가능하다.



### (1) 불용어 화일

불용어란 어절색인으로 부적합한 어절을 의미하며 이는 어절색인 생성시 삭제된다.

### (2) 한글조사화일

어절검색시 어절색인의 끝부분에 있는 조사를 삭제하기 위해 필요하다. 조사 삭제시에는 길이가 긴 조사부터 비교하여 삭제하여야 한다.

### (3) 한글속어화일

어절색인 끝부분의 조사를 제거하는 경우 조사 원어절에 포함되어 있는 부분이 조사로 삭제되는 것을 방지하기 위하여 사용

e.g. "가정의(家庭醫)"에서 "의"를 조사로 삭제하는 경우

## 4.5 검증

### 1) 포맷에 대한 검증

#### 1.1) 길이 check

1.1.1) LDR length check : 24 bytes

1.1.2) directory check

- directory length가 12의 배수인가 확인

- 'directory length / 12 == no. of tag' 확인

(tag delimiter 갯수로 확인 가능)

1.1.3) 008 tag length check : 40 bytes

#### 1.2) tag 단위

1.2.1) mandatory tag check

〈표 11〉

TAG	Subfield	길이	MARC	비 고
010(LCCN)	a	12	KORMARC	
012(국립중앙도서관 통제번호)	a	14		
020(ISBN)	a	10		
022(ISSN)	a	9		
030(CODEN)	a	6		
090(청구기호)	a	3		.이전까지
092(분류번호)	a	3		.이전까지

1.2.2) non-repeatable tag check

alphabet 소문자나 numeric 외의

1.2.3) directory tag갯수만큼 tag  
delimiter가 있는가 check

code가 있는가 확인

1.2.4) tag delimiter가 없는 것

2) 데이터에 대한 검증

1.3) subfield 단위

2.1) 고정 길이 데이터 길이 〈표 11〉 참조

1.3.1) subfield delimiter 다음에 소문자  
알파벳이거나 숫자가 아닌것

001, 005, 008/00-06은 시스템이 생성  
하므로 확인하지 않음.

1.3.2) mandatory subfield check

2.2) 지시기호 확인

1.3.3) non-repeatable subfield check

– 지시기호가 없는 tag(00X tag)에 지  
시기호가 포함되어 있는 경우

1.4) record length check

– 지시기호가 있어야 할 tag(010이상의  
tag)에 지시기호가 없는 경우

: LRD/00-04와 실제 record length  
check

3) 구절색인 생성시의 확인

1.5) delimiter 및 code check

3.1) 동일전거시스템(Name, Subject)내에

1.5.1) delimiter check

–record delimiter : record 끝 0x1D  
확인

서의 동일 확정표목 허용 안됨

–tag delimiter : 0x1E (tag 갯수만큼  
있는지 확인)

3.2) 하나의 전거레코드 내에는 동일표목  
허용 불가

1.5.2) subfield code check

3.3) 서지레코드 내의 전거 관련표목은 반

: subfield delimiter (0x1F) 다음에

드시 대응 전거레코드를 가져야 함

3.4) 서지레코드에 포함되지 않은 전거레코

드 허용

3.5) 선거레코드에 포함되어 있는 '~도보  
라표목'은 반드시 다른 선거레코드에 확  
정표목으로 등록되어 있어야 한다.

#### 4.6 문제점

우리나라의 경우 이용자들이 검색어로 사용  
하는 언어가 다양하여 하나의 표목으로 여러  
표기를 통제하기란 무척 어려움이 있다. 또한  
MARC 포맷도 KS규격인 KORMARC와  
USMARC를 병행하여 사용해야 하므로 어려  
움은 더욱 크다고 하겠다. 현재 포맷을  
KORMARC나 USMARC 한가지로 할 수 없  
는 이유는 다음과 같다.

- 1) 서지 및 선거 유틸리티를 통해 다운로드  
받아 사용가능한 포맷이 USMARC이므  
로 이를 전적으로 배제하고 신규로 목록  
을 작성한다는 것은 도서관 전체의 효율  
적 측면에서도 바람직하지는 못하다.
- 2) 아직까지 KORMARC 서지 포맷에 대한  
선거 포맷이 없으므로 선거통제를 받고  
자 하는 경우 USMARC 포맷을 사용하  
여야 한다.
- 3) KORMARC 서지포맷상에서 AACR2가  
완벽하게 지원되지 않는다.
- 4) 해외 도서관과의 분담목록이나 데이터  
교류를 위해 국외서인 경우는 USMARC  
를 사용.

따라서 현재 국공립대학간의 목록표준화 기  
구(국공립대학교 도서관장협의회 산하 목록분  
과위원회)에서는 앞서 밝힌 바와 같이 국내서  
는 KORMARC, 국외서(중국서, 일서 포함)는

USMARC를 사용하기로 합의한 바 있다. 그  
러나 이 경우 다음과 같은 문제들이 생긴다.

- (1) 선거시스템을 KORMARC와 USMARC  
서지포맷 모두에 대해 통합함으로써 발생  
하는 문제
  - (1.1) USMARC 선거포맷에 없는 서브필  
드를 KORMARC 서지포맷에 맞추어 추  
가하여 사용  
e.g. KORMARC에는 100tag에 \$f歷朝  
가 있으나 USMARC에는 없음.
  - (1.2) USMARC와 KORMARC 서지포맷  
에서 다른 용도로 사용되는 동일 서브필  
드의 적용  
e.g. KORMARC의 100tag \$g한국 및 중  
국의 世系이나, USMARC의 100tag  
\$g는 Miscellaneous information임.
- (2) 동양서를 USMARC포맷으로 정리하는  
경우 nonfiling 처리의 문제

USMARC상에서의 배열제외자수는 단순히  
관사나 정관사 등이 구절색인의 앞부분에 사용  
되는 것을 막기위한 용도로 현재는 최대 9자리  
로 제한되어 있다. 그러나 동양서인 경우 국내  
서와 마찬가지로 관제, 관칭이 사용되어야 한  
다. 따라서 9자리 넘는 관제, 관칭의 처리문제  
가 대두된다.

#### (3) 중국서, 일서의 총서명 처리

중국서, 일서는 국제적으로 통용되는 표기  
를 사용함으로써 기관 간의 데이터 교류를 원  
만하게 하기위해 로마나이즈된 표기를 표목으  
로 잡는다. 따라서 총서명의 경우 총서명표목  
으로 사용 가능한 총서명도 tag490을 사용하  
여 tag8XX에 로마나이즈된 표기로 부출기입  
을 하고 tag830의 표목으로 선거통제를 받게

된다. 이러한 적용은 엄밀한 의미에서 tag830에 나가는 로마나이즈된 표목은 읽기의 성격을 갖고 있어 변형표기를 기입하는 tag830을 사용하는 것은 올바른 MARC의 적용방법이라고 할수 없으나 전거통제를 받아야 하므로 임의 적용해야 한다. 물론 다음과 같이 대체하여 사용가능한 tag과 방법이 있으나 그 각각에도 나름대로의 문제점이 있기 때문이다.

### (3.1) 대체문자를 기입하는 tag880의 사용 (USMARC)

실제로 USMARC상에서는 영어 이외의 타 국어를 사용한 도서를 적용할 때 모든 표목을 영문 로마나이즈 표기로 기입하고 각각의 원어 표기를 tag880에 기입하는 예를 제시하고 있으며 이와 반대로 원어표기로 표목을 잡고 그에 대한 로마나이즈 표기를 tag880에 기입하고 tag880의 표목을 색인으로 생성하는 방법이 있을 수 있다. 다만, 이 경우 총서명과 같이 전거통제를 받는 사항은 로마나이즈한다는 원칙에 벗어나므로 동일하게 전거통제를 받을 수 없다.

### (3.2) local tag으로 지정된 읽기 tag의 사용 (tag930)

이 경우도 첫번째 방법과 마찬가지로 전거 통제를 받을 수 없다.

### (3.3) vernacular tag의 사용

OCLC에서 영어 이외의 언어에 대해 사용하는 vernacular tag을 사용하여 모든 표목에 대해 동일tag을 반복하여 원어와 영어 로마나이즈 표기를 병행하는 방법이다. 전거통제시 반복되는 첫번째 tag에 대해 전거통제를 받는다는 식의 정의만 되면 문제가 없겠으나 동양서 총서명에 한하여 발생하는 문제를 해결하기

위해 전거레코드에 기입될 모든 표목을 중복되게 서지레코드에 포함시키는 것은 불합리하다.

## 5. 검색시스템의 설계

4장까지의 내용에서와 같이 검색을 할 수 있는 환경이 모두 갖추어졌다고 보면 다음은 구축된 데이터베이스와 몇몇 관련 화일등을 이용하여 검색할 수 있도록 검색시스템을 구성하는 과정이 남는다. 본고에서는 일반적인 검색시스템에서 제공되는 모든 명령(EXPAND, SORT, LIST 등)을 대상으로 하지 않고 구절 및 어절 검색에 대해서만 언급한다.

### 5.1 일반사항

앞서 밝힌 바와 같이 검색시스템에서도 검색어가 입력되면 목록시스템에서 색인을 생성하기 위해 거친 과정을 그대로 거쳐 검색어를 수정하여야만 올바른 검색이 가능하다.

#### 1) 검색식의 구성

일반적으로 검색을 위한 검색식은 검색항목을 구분하는 검색prefix와 검색어 또는 검색구로 구성된다. 예를들어 저자명구절을 "AU"라는 검색prefix로 정의한다면 "AU = Dieter, George E."와 같이 검색식을 구성하게 된다.

#### 2) 검색의 종류

2.1) 단일검색 : 하나의 검색 prefix를 사용하는 검색

2.2) 조합검색 : 단일검색의 결과를

Boolean논리에 의해 조합하는 검색

〈예〉 "AK = Friedrich \* TI = The analysis of the law"

저자명어절과 서명구절의 AND 조합

2.3) 다중집합의 조합검색 : 결과는 조합검색과 동일하나 검색식 구성시 이미 결과를 가지는 검색집합에 부여된 유일한 이름이나 번호로 조합하는 검색

〈예〉 S1 : AK = Friedrich

S2 : TI = The analysis of the law

S1 + S2 : S1결과집합과 S2결과집합의 OR 조합

2.4) 제한검색 : 2.1)-2.3)에 제시된 검색에 일정한 기준에 의해 제한을 가하고자 하는 경우 사용. 즉, 검색결과에 대해 특정한 조건의 부분집합을 구하고자 할때 사용.

〈예〉 "SU = Career Exploration /PY 1980-1990"

주제구절에 의한 검색결과 중 출판년도가 1980년에서 1990년 사이인 서지정보 검색.

### 3) 검색기법

검색기법 중 가장 일반적으로 사용되는 것은 절단검색으로 다음과 같은 종류가 있다.

3.1) 후방절단 : 우측절단이라고도 하며 앞에서부터 입력한 부분까지만 같으면 뒷부분은 문자나 길이에 상관없이 모두 검색이 가능하며 구절 및 어절 모두에 적용된다. 코드어 검색을 제외한 모든 검색에서 지정하지 않아도 우측절단은 기본으

로 제공된다. 코드어인 경우는 분류 기호(Call Number, DDCN 등)에는 기본으로 제공되어 유,강,목 등의 단위로 검색이 가능하나 ISBN, ISSN, CODEN, Control Number 등 유일하면 앞, 뒤의 번호가 특별한 관련을 갖지 않는 경우는 후방절단기호가 입력된 경우에 한하여 제공된다.

3.2) 전방절단 : 좌측절단이라고도 하며 기호가 사용된 부분은 문자나 길이에 무관하게 뒷부분이 같은 어절을 검색한다. 구절에도 적용은 가능하나 동일 구절에 대해 지나치게 많은 부분 구절을 색인으로 작성하거나 복잡한 알고리즘을 적용해야 하므로 많이 사용되지는 않는다.

3.3) 중앙절단 : 양측은 동일하고 중간만 다른 구절 및 어절을 검색할 경우 사용되며 임의의 한문자, 임의의 n개 문자를 대체하여 검색한다.

3.4) 양측절단 : 구절 및 어절의 입력된 중앙부분만 동일하면 양측은 임의의 한문자나 임의의 n개 문자로 대체하여 검색한다.

〈예〉 # : 임의의 한문자

? : 임의의 n개 문자

## 5.2 전거레코드를 통한 서지레코드의 검색

저자 및 총서명, 주제명은 전거레코드 내의 보라표목과의 비교를 거쳐 확정표목으로 대체되어 검색된다. 각각의 검색어에 따라 검색 대상이 되는 전거레코드도 구분이 되는데 저자 및 총서명인 경우는 인명전거레코드로, 주제명

은 주제명전거레코드만이 전거통제를 받는 대상이 된다. 이렇게 검색시 서지와 전거레코드가 직접적이면서도 자동적으로 연결되어 전거통제를 받아 검색해야 하므로 데이터베이스 설계시 이러한 사항들이 모두 고려되어야 한다.

## 1) 데이터베이스 구성

전거통제로 검색되는 항목은 구절이므로 구절색인에 대한 데이터베이스의 구성에 대해서만 살펴본다.

### 1.1) 구절색인 데이터베이스의 구성

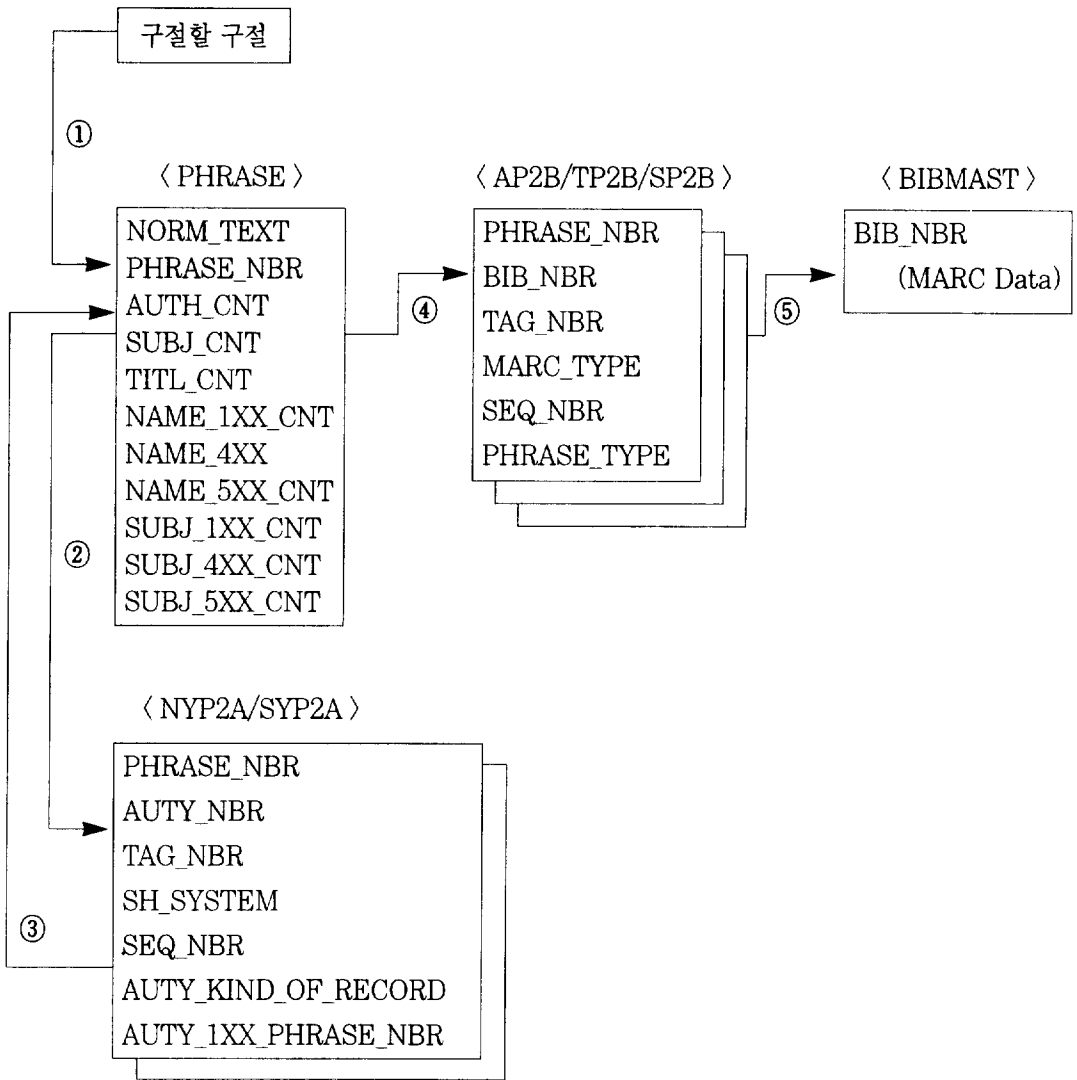
- 구절번호 : 구절색인의 primary key로 유일한 번호
- 서지 저자명 구절 갯수 : 서지레코드 저자명 관련tag에서 생성된 본 구절의 갯수
- 서지 주제명 구절 갯수 : 서지레코드 주제명 관련tag에서 생성된 본 구절의 갯수
- 서지 서명 구절 갯수 : 서지레코드 서명 관련tag에서 생성된 본 구절의 갯수
- 서지 출판자명 구절 갯수 : 서지레코드 출판자명 관련tag에서 생성된 본 구절의 갯수
- 전거 인명확정표목 구절 갯수 : 인명 전거레코드의 1XXtag에서 생성된 본 구절의 갯수
- 전거 인명보라참조표목 구절 갯수 : 인명 전거레코드의 4XXtag에서 생성된 본 구절의 갯수
- 전거 인명도보라참조표목 구절 갯수 : 인명 전거레코드의 5XXtag에서 생성된 본 구절의 갯수
- 전거 주제명확정표목 구절 갯수 : 주제명

전거레코드의 1XXtag에서 생성된 본 구절의 갯수

- 전거 주제명보라참조표목 구절 갯수 : 주제명 전거레코드의 4XXtag에서 생성된 본 구절의 갯수
- 전거 주제명도보라참조표목 구절 갯수 : 주제명 전거레코드의 5XXtag에서 생성된 본 구절의 갯수 상기의 각각 필드의 용도는 다음과 같다.
- 서지나 전거레코드에서의 구절 갯수 : 검색시 검색건수를 먼저 출력한 후 결과에 대한 간략화면 및 상세화면을 출력하므로 실제 검색결과를 구성하기 전의 검색식과 검색건수를 출력할 수 있다. : 새로운 구절생성시 검증이 가능하다. 예를 들어 저자명 전거를 받는 tag에서 생성된 구절레코드의 전거인명 보라참조표목 구절 갯수가 1 이상인 경우 등록하고자 하는 구절은 확정표목이 아니므로 에러처리가 가능하다.

### 1.2) 구절색인과 서지데이터베이스의 링크 관련 링크화일은 저자명, 서명, 주제명의 세가지가 있으며 각각의 구성은 다음과 같다.

- 구절번호 : 구절색인에서 생성된 구절번호
- 서지레코드번호 : 구절이 생성된 서지레코드 번호
- tag번호 : 구절이 생성된 tag번호
- MARC종류 : 구절이 생성된 서지레코드의 MARC 유형 (KORMARC, USMARC)



- tag일련번호 : 서지레코드 내에 동일 tag 이 여러개인 경우 일련번호
- 구절종류 : 저자, 서명, 주제명 구절 구분 검색시 구절색인에서 검색된 구절번호로 해당 링크화일을 통해 서지레코드에 접근이 가능하다.

- 1.3) 구절색인과 전거데이터베이스의 링크 관련링크화일은 인명, 주제명의 두가지이며 각각의 구성은 다음과 같다.
- 구절번호
  - 전거레코드번호
  - tag번호
  - 주제명표목표



1) 간략화면의 구성

				데이터 규칙	길 이
일련번호				검색결과와 일련번호	3
서 명	USMARC			440/800/810/811/830 \$v + 245 \$a	45
	KORMARC			440/830 \$v + 245 \$a	
저 자	USMARC			100/110/111/700/710/711	10
	KORMARC			507 \$a/700/710/711	
년 도	연속간행물 /컴퓨터화일	008/06	c	"CONT"	4
			d	"CEAS"	
			u	"UNKN"	
	이외의 자료				008/07-10
청구기호				090	12
자료유형				LDR/06-07	2
대출가능여부				대출시스템 데이터	1

- \* /는 나열된 tag 가운데 하나를 선택
- \* +는 두개의 구절을 연결
- \* 년도의 경우 260 \$c를 보여주면 일서의 경우 '昭和'만 출력되는 경우가 있으며 연속간행물 같이 계속 되는 출판물인 경우 출판년이 의미가 없으므로 상태를 표시한다

1) 상세화면의 구성

LABEL		단행본	연속간행물
별치기호		093 \$x	
청구기호		090 *	
대출가능여부		대출시스템 데이터	
저 자	USMARC	100/110/111/700/710/711 *	
	KORMARC	507 \$a/ 700/710/711 *	
서 명		245 *	
권호사항			362 \$a
원서명	USMARC	240 \$a	
	KORMARC	507 \$t	
판사항		250 \$a	
출판사항		260 *	
형태사항		300 *	
총서사항		440/490 *	
간 기			310 \$a
ISBN	020 \$a		
ISSN	022 \$a		

CODEN		030 \$a
일반주기	500 *	
학위논문주기	502 *	
내용주기	505 \$a	
초록	520 \$a	
변경전표제	7800 \$a\$t / 580 \$a	
변경후표제	7850 \$a\$t / 580 \$a	
주제명	600-650 *	
소장사항		연속간행물 소장데이터

※ tag022는 USMARC상에서 단행본용으로는 지정되어 있지 않으나 ISSN을 갖는 단행본이 있으므로 연속간행물과 함께 적용.  
 ※ 변경 전, 후표제사항은 tag780,785의 제1지시기호가 0이아니면 580 \$a 데이터 출력.

-tag일련번호  
 -전거레코드종류  
 -해당1XX의 구절번호  
 전거레코드 검색시 구절색인에서 검색된 구절번호로 해당 링크화일을 통해 전거레코드에 접근한다.

2) 전거레코드를 통한 서지레코드의 검색  
 1)에서 언급된 3개의 데이터베이스를 이용하여 전거통제 후 서지레코드를 검색하게 된

다. 접근 경로와 순서는 다음과 같다.

- 입력된 검색식이 "AU = 홍길동"인 경우  
 ① : 정규화된 "홍길동"으로 PHRASE table검색하여 구절번호 얻음.  
 PHRASE table에서 hit된 레코드의 NAME\_1XX\_CNT가 1 이상이면  
 ② : 검색된 구절번호로 \*YP2A table 검색하여 해당구절의 확정표목의 구절번호 얻음.

제2지시기호	변경전(780)	변경후(785)
0	Continues	Continued by
1	Continues in part	Continued in part by
2	Supercedes	Superceded by
3	upercedes in part	Superceded in part by
4	Formed by the union of	Absorbed by
5	Absorbed	Absorbed in part by
6	Absorbed in part	Split into A and B
7	Separated from	Merged with A to form B
8		Changed back to

- ③ : ②번에서 얻은 구절번호로 PHRASE table 재검색 후 구절번호 얻음.
- ④ : ①②③번에서 얻은 구절번호로 \*P2B (저자검색인 경우 AP2B) 검색하여 해당 서지번호 얻음.
- ⑤ : 서지번호로 서지master table에서 해당 레코드에 접근 : 서지레코드 검색시 authority data의 name/subject 구분하여 검색

은 표시상수를 적용한다.

〈예〉 24504\$aThe Southeastern college art review and newsletter.  
78501\$aSoutheastern College Art Conference.\$tSECAC newsletter\$w (OCOLC)4973820  
→ Continued in part by: Southeastern College Art Conference. SECAC newsletter.

### 5.3 서지데이터의 출력

검색된 서지사항을 보여주는 방법은 여러가지가 있다. 크게는 간략서지와 상세서지로 구분이 되며 상세서지는 Label형식, tag형식, card형식 등이 일반적이다. 간략서지는 하나의 서지를 한 라인에 출력해야 하므로 주요정보를 잘 편집하여 출력하여야 하며 일반이용자를 위한 상세 화면양식으로는 Label형식이 가장 일반적이다.

상세화면 구성 항목은 기관에 따라 달라질 수 있으며 각 매체별로도 달라질 수 있다. 상세 화면 구성시 중요한 것은 표시상수의 적용이다.

〈예〉 주제명의 각 tag내의 세목을 나타내는 \$x, \$y, \$z 데이터는 -(dash)로 표시하여야 한다.

650 0\$aNuclear energy\$xHistory  
→ Nuclear energy — History

#### 3) 연속간행물 관련주기의 적용

3.1) 제1지시기호가 '0' 인 경우 780, 785tag의 데이터를 사용하고 다음과 같

3.2) 제1지시기호가 '0' 이 아닌 경우는 580tag의 주기사항을 출력하며 이경우는 위의 표시상수를 적용하지 않고 tag내에 직접 표출어를 입력한다.

〈예〉 580 \$aContinued in 1982by:U.S. exports. Schedule E commodity groupings by world area and country.  
78510\$tU.S. exports. Schedule E commodity groupings by world area and country\$w(DLC) 84641135  
→ tag580내용 출력

## 6. 제안점 및 맺음말

책을 정리한다는 것은 이용자들의 검색을 전제로 하는 작업이므로 정리시 당연히 어떻게 하면 이용자들이 좀더 쉽고 정확한 검색을 할 수 있을까에 대해 충분히 고려하여 Format을 적용하여야 할 것이다. 이에 대해 지금까지 MARC Format을 달리 적용함에 따라 검색에 어떠한 영향을 주는가에 대해 알아 보고 각 문제점들에 대한 해결방법도 모색해 보았다. 끝

으로 비록 근본적인 해결책은 아니나 다만 그와 관련하여 다음과 같은 제안을 하며 글을 맺으려 한다.

1) MARC간의 호환성이 증시되어야 한다. 물론 한가지 포맷으로 통일하여 사용하는 방법도 있겠으나 국제간의 활발한 정보 교류를 위해서는, 제공되는 유틸리티가 가장 많은 USMARC를 사용하지 않을 수 없으며, 반면에 국가표준 MARC로서 국내서의 특성을 반영하고 있는 KORMARC도 또한 사용하지 않을 수 없기 때문이다.

2) 종래의 도서관에서는 통제어휘인 표목이 검색의 주종을 이루고 있으므로 표목에 대한 통제가 필수적이었고 그에 따라 발생되는 문제점들은 위에서 언급된 바 있다. 그러나 앞으로는 자연어를 그대로 검색어로 사용함으로써 전거 등의 통제시 발생하는 문제 해결과 함께 좀더 융통성있는 검색기법을 제공해야 할 것이다.

3) MARC적용의 한계점을 극복하도록 해야 한다. MARC로 사용이 가능한 데이터의 최대 길이는 100Kbyte이다. 따라서 초록, 색인 등은 어느 정도 수용이 되겠지만 전문DB를 구축하기 위한 Format으로는 적합하지 않다. 또한 방법이 없는 것은 아니나 연속간행물 기사별 검색을 위한 정보를 제공하기에도 적합하지 않다. 전문DB와 연속간행물 기사 단위의 서비스를 위해서는 기존MARC format에 다른 형태의 전문 및 기사의 내용을 위한 별도의 format을 연결(link)하여 사용하거나 MARC Format을 확장해야 하며 더 나아가 전혀 새로운 형태의 Format 개발에도 주력해야 한다.

4) 하이퍼텍스트 등 새로운 형태의 정보서비스 기법과 다양한 뉴미디어를 도입하여 대이용자 서비스를 향상시켜야 할 것이다.

## 참고문헌

- 국립중앙도서관(1993). 한국문헌자동화목록 형식 : 단행본용. 서울 : 국립중앙도서관
- 서울대학교 중앙도서관(1992). 도서관전산화 S/W를 위한 시스템설계 및 기능에 관한 연구 : 목록, 검색 편. 서울 : 서울대학교 중앙도서관.
- 서울대학교 중앙도서관, 한국컴퓨터(주)(1992). 서울대학교 학술정보시스템 사용자 Manual. 서울 : 서울대학교 중앙도서관.
- 서울대학교 중앙도서관(1994). 전산화시스템의 개요. 서울 : 서울대학교 중앙도서관
- Library of Congress(1987). USMARC formats for bibliographic data Including Guidelines for Contents Designation. Washington
- Library of Congress(1987). USMARC formats for Authority data Including Guidelines for Contents Designation. Washington.