

분담목록시스템을 위한 데이터 표준화(Ⅱ)

서울대학교 중앙도서관
조수련·조순영

□ 목 차 □

- | | |
|------------------|--------|
| 1. 서 언 | 4. 결 언 |
| 2. 서지DB의 관리 및 운영 | * 부 록 |
| 3. 전거DB의 관리 및 운영 | |

1. 서 언

국립대학도서관보 제12집에 분담목록시스템을 위한 데이터표준화 (I)안을 발표한 이래 4차례에 걸친 목록분과 회의가 개최되었다. 이 회의에서는 실제 DB구축상에서 발생하는 많은 문제점들이 토의되었고, 그 결정사항들은 현재 신규 및 소급 DB 구축에 적극적으로 반영되고 있는 것으로 안다. 4년여에 걸친 표준화 작업기간중 시스템의 개발 환경에 많은 변화가 있었고 또 회원 도서관들의 새로운 요구가 지속적으로 발생하여 표준화안의 부분적인 수정이 불가피하게 되었다. 이에 그 간의 결정사항과 보완사항을 재 정리하여 “분담목록시스템의 데이터 입력 편람”에 대신 한다.

표준화 작업은 앞으로도 계속적으로 진행하면서 시스템의 개발 상황에 따라 부분적으로 분담목록 시스템을 운영하고자한다. 따라서 '95년 5월 우선적으로 충남대학을 시범대학으로 지정하여 초기 센터 DB로 UPLOAD된 서울대학 DB를 DOWNLOAD 받을수 있도록 추진중이다. 그후 '95년 말까지 모든 회원 도서관에서 DOWNLOAD가 가능한 단계에 이르면 다시 회원대학의 자관 DB를 센터에 UPLOAD 시킬수 있도록 하여 본격적인 UNION DB구축을 꾀하려 한다. 뿐만 아니라 현재 국

공립대학으로 제한된 회원 도서관의 자격을 사립대학 도서관까지 대폭 확대하여 명실공히 한국의 대표적인 학술정보의 UNION DB를 구축할 것이다. 따라서 이후의 표준화는 각 도서관의 특정시스템에 구애됨이 없이 범 국가적인 차원의 분담 목록이 실현될 수 있도록 보다 일반적인 방향으로 나아갈 것이다.

2. 서지DB의 관리 및 운영

2. 1 언어별 포맷 적용

자료의 국내·외서의 기준은 발행지에 두지 아니하고 본문이 쓰여진 언어에 근거한다. 따라서 국내에서 발행된 영어 자료는 USMARC로, 외국에서 발행된 한글 자료는 KORMARC로 입력해야 한다. 동일한 자료라 할지라도 MARC를 다르게 적용할 경우 데이터 UPLOAD시 자동적인 중복 데이터 조사에 어려움이 있고, 다른 포맷을 사용하는 도서관에서는 DOWNLOAD시 포맷을 변환해야만 하기 때문에 원칙적으로 동일한 자료는 동일한 MARC로 입력하도록 규정한다. 각 도서관의 등록번호 부여 방법과는 관계없다.

2. 2 MARC FORMAT의 최신판 적용

USMARC과 KORMARC은 가장 최근에 출판된 판을 따른다. 센터에 이미 UPLOAD된 데이터는 센터에서 일괄적으로 UPDATE하지만 개정판 이후에 UPLOAD되는 데이터는 각 대학에서 UPDATE된 내용으로 올려야 한다. UPDATE판 적용 시기는 센터에서 수시로 고지한다. (단 초기에 일괄 UPLOAD되는 데이터는 센터에서 처리한다.)

〈참조〉 현재 적용중인 포맷의 판사항

-USMARC FORMAT FOR BIBLIOGRAPHIC DATA. 1994 ed.

USMARC SPECIFICATIONS FOR RECORD STRUCTURE, CHARACTER

- SETS, TAPES. 1990 ed.
USMARC CODE LIST FOR COUNTRIES. 1993
USMARC CODE LIST FOR GEOGRAPHIC AREAS. 1994
USMARC CODE LIST FOR RELATORS, SOURCES, DESCRIPTION
CONVENTIONS. 1993
USMARC CODE LIST FOR LANGUAGES. 1989
-USMARC FORMAT FOR AUTHORITY DATA. 1993 ed.
-USMARC FORMAT FOR HOLDING DATA. 1991 ed.
-한국문헌자동화목록형식: 단행본용. 1993
-한국문헌자동화목록형식: 연속간행물용. 1995

2. 3 목록규칙의 통일

동일 자료임에도 불구하고 각 도서관마다 다른 목록규칙을 적용함으로써 인명 및 총서 사항의 전거통제에 어려움이 있고, 데이터의 중복 조사 및 색인 생성의 과정이 복잡해지므로 점진적으로 센터에서 회원 도서관의 사용 현황을 고려하여 하나의 목록규칙으로 통일 하고자 한다. (단 이미 구축된 자료에 대해서는 센터에서 수정을 하고 본격적인 분담목록시스템의 가동 이후는 각 도서관에서 변경 적용하도록 한다.)

<참조> 1. 국내 대학도서관 사용 목록규칙 현황

양서: AACR2 (42%), KCR (41%)

국내서: KCR3 (60%), KCR2 (22%)

2. OCLC 사용 목록규칙

AACR2

2. 4 레코드의 길이

하나의 서지레코드는 12,000 byte, Tag별 길이는 4,000 byte(등록번호 약150개 까지 가능)로 제한한다. 따라서 그 길이를 초과하는 데이터는 레코드를 2개로 분리 하여 입력한다.

2. 5 레코드 및 필드의 종단 기호

레코드 종단기호는 ASCII 1D(16), 필드 종단기호는 ASCII 1E(16)를 사용함으로 같은 코드값을 가지는 서지 데이터가 있는 경우 다른 기호로 대체 입력 해야한다.

2. 6 LEADER 및 고정장 필드에서 시스템이 자동생성하는 부분

구 분	화면출력	내 용	생성값	비 고
LDR	00-04	레코드길이		시스템 자동생성
	05	레코드 상태	n:신규 c:수정 d:삭제	작업구분에 따라 시스템 자동생성
	06	유형	a	시스템 default
	07	서지수준	a:단행본 s:연속 간행물	자료구분에 따라 시스템 자동생성
	10	지시기호자리수	2	시스템 setting
	11	식별기호자리수	2	시스템 setting
	12-16	기본번지		시스템 자동생성
	17	입력수준	blank	시스템 default
	18	목록기술형식	a	시스템 default
	19	연관레코드조건	blank	시스템 default
	20-23	엔트리맵	4500	시스템 자동생성
001		전거레코드 유일번호		시스템 자동생성
005		최종처리일시		시스템 자동생성
008	00-05	입력일자		시스템 자동생성

USMARC 단행본

구	분	화면출력	Default 생성값
008	06	STATUS	s
	07-14	DATES	19
	15-17	CTRY	
	18-21	ILLUS	a
	22	AUD	
	23	FORM	
	24-27	CONT	
	28	GOVT	
	29	CONF	0
	30	FEST	0
	31	INDX	1
	33	FICT	0
	34	BIO	
	35-37	LANG	eng
38	MODF		
39	SORC	d	

USMARC 연속간행물

구	분	화면출력	Default 생성값
008	06	STATUS	c
	07-14	DATES	19 9999
	15-17	CTRY	
	18	FREQ	m
	19	REGL	r
	20	ISDS	
	21	STYP	p
	22	MEDM	
	23	FORM	
	24	MATL	
	25-27	CONT	
	28	GOVT	
	29	CONF	0
	33	ALPHA	a
	34	S/L ENT	0
	35-37	LANG	eng
	38	MODF	
	39	SORC	d

KORMARC 단행본

구	분	화면출력	Default 생성값
008	06	연도유형	s
	07-10	발행년도	19
	15-17	발행국	ulk
	18-21	삼도	
	22	지식수준	
	23	자료형태	
	24-25	내용형식	
	26-27	대학명	
	28	수정여부	
	29	회의간행	0
	30	기념논문	0
	31	색인	0
	32	목록전거	a
	33	문학형식	
	34	전기	
	35-37	언어	kor
38-39	정부기관		

KORMARC 연속간행물

구	분	화면출력	Default 생성값
008	06	연도유형	c
	07-10	발행년도	19 9999
	15-17	발행국	ulk
	18	발행빈도	a
	19	정규성	r
	20	ISDS	8
	21	연간유형	p
	22	월발간형식	
	23	발간형식	
	24-25	내용형식	
	26-27	대학명	
	28	수정여부	
	29	회의간행	o
	32	목록전거	a
	33	표제언어	k
	34	목록기술	o
	35-37	언어	kor
	38-39	정부기관	

2. 7 전거관련

(1) 번역서의 표목

하나의 표목은 하나의 전거레코드를 원칙으로하므로 번역서의 경우에도 그저작의 원서와 동일한 표기를 해준다. 즉 이문열의 작품은 어느 언어로 번역되어도 그 표목은 “이 문열”이고 WILLIAM SHAKESPEARE의 작품은 어느 언어로 번역되어도 “SHAKESPEARE, WILLIAM”이 표목이 된다.

ex) 서지레코드

1001 \$aShakespeare, William

240 0\$aHamlet.

24510\$a햄릿 / \$d윌리엄 셰익스피어 著 ; \$e김경수 譯

(2) 서지TAG 6XX의 표기

주제명은 인명과 별도로 전거 통제를 받지만 주제명중 6XX에 해당되는 인명은 인명전거의 1XX와 동일하게 표기한다.

(3) 학위논문주기의 학위 수여기관 및 학과

서지 TAG502의 \$b, \$c는 전거통제를 안받는 TAG이지만 수여기관 및 학과명이 제한 검색 대상이므로 검색시 누락이 되지 않도록 전거에서의 1XX와 동일한 표기를 한다.

ex) Seoul National University. Graduate School (x)

서울대학교 대학원 (o)

2. 8 서명기입

KORMARC에서 합집의 두번째 이하 서명은 Tag 245 \$a의 반복 사용으로 처리되나 여러개의 \$a가 모두 하나의 지시기호를 동일하게 적용할 수 없기때문에 부출의 필요가 있는 서명에 대해서는 다시 Tag 740에 기입해준다.

ex) 서지레코드

24511\$a(신편)가정보건강학 ; \$a(한국인을위한)건강식단 / \$d유종명.
740 0\$a한국인을위한 건강식단

2. 9 USMARC Leader 18번째 자리 Descriptive Cataloging Form에 한국목록규칙과 KORMARC 기술규칙을 추가

k : 한국목록규칙 m : KORMARC 기술규칙

2. 10 USMARC에 KDC tag 삽입

056 \$a 한국십진분류기호 (R)
\$2 판표시

2. 11 삭제된 레코드의 제어 번호를 tag 019에 기입하여 상호 참조가 가능하도록한다.

019 bb : Processed duplicate (R)

1b : Identified but unprocessed duplicate

\$a : Center control no. of deleted record or record scheduled for merge

ex) 019 \$a4781464

2. 12 일서, 중국서 관련

1) 한자의 약자, 간자, 고어 등은 정자(正字)로 대신 입력하고 tag 890에 한글로 기입해준다. 없는 한자는 한글로 대체한다.

890(반복不可) \$h(반복可)

ex) 890 \$h245 한

2) 전거통제를 받는 Tag은 읽기가 전거레코드의 4XX에서 처리되기때문에 별도로 9XX를 사용할 필요가 없으나 전거 통제 대상이 아닌 서명 등의 읽기는 Tag 940에 Romanize 한다.

3) 일서, 중국서의 서명이 띄어쓰기가 되어 있지 않아 단어별 서명키워드의 생성이 곤란하므로 Tag 940에서 띄어쓰기를 하고 불용어에 해당하는 조사 등을 생략해 준다.

4) 일어에서의 장음과 중국어에서의 음독 부호는 ‘~’과 ‘#’으로 대체 입력한다.

5) USMARC Tag 4901은 다르게 부출되는 총서명에 사용된다. 그러나 일서, 중국서에서는 Tag 440과 830이 전거 통제를 받는 문제를 고려하여 4901을 확대 적용한다.

4901 : 다르게 부출되는 총서명

49011 : 단순한 읽기

ex) ① 4901 \$a日本岩波文庫全書 ; \$v第三卷

830 0 \$aIwanami shinsho ; \$vv.3

※ 830 0에서 岩波新書가 생략된것

② 49011 \$a日本岩波文庫全書 ; \$v第三卷

830 0 \$aNihon Iwanami bunko zensho ; \$vv.3

2. 13 다권본의 소장사항

다권본은 별도의 소장 레코드를 갖지 않으므로 각 도서관의 소장사항에 대해서는 Tag599에 Library has나 본관 소장등의 표출어를 사용하여 소장사항을 기입한다. Tag599에 입력된 소장사항은 Tag850에 소장기관별로 보여주어 상호대차시 활용할 수 있도록한다. Tag850에 입력되는 소장기관명 코드는 USMARC에서도 한국도서관부호표를 사용한다. (단 일괄 UPLOAD된 데이터에 대해서는 센터에서 처리한다.)

ex) 599 \$aLibrary has: \$vv.1, 3

599 \$a본관소장: \$v1-3권

2. 14 총서사항(I안에서의 변경)

4XX, 8XX tag의 \$v를 검색키로 사용하기위해서 개별적으로 색인을 생성했으나 \$v의 표기와 범위등에 예외적인 상황이 많이 발생하여 종전과 같이 MARC FORMAT을 그대로 적용한다.

ex)4901 \$aAdvances in chemical physics ; \$v13, 18, 19
830 0 \$aAdvances in chemical physics ; \$v13, etc.

3. 전거DB의 관리 및 운영

3. 1 전거DB의 구분

인명과 주제명으로 구분하여 관리되며 인명에는 단체명, 회의명, 총서명이 포함된다. 총서명 전거레코드는 레코드 생성시 default값의 적용 등을 위해서 시스템 내부적으로는 구분처리 되나 실제 별도의 DB를 갖지는 않는다.

3. 2 사용되는 레코드 유형

008/09에서 구분되는 레코드 유형 중 f(Established heading and subdivision record)와 e(Node label record)를 제외한 모든 유형을 사용한다.

3. 3 전거표목의 관리

- 1) 동일전거시스템(Name, Subject)내에서 두개 이상의 동일 채택표목 허용 불가
- 2) 하나의 전거레코드 내에는 동일 표목 허용 불가
- 3) 동일전거시스템 내에서 동일한 표목이 서로 다른 전거레코드의 1XX와 4XX에 존재할 수 없음.

- 4) 서지레코드 내의 전거 관련 표목은 반드시 대응 전거레코드를 가져야 함
- 5) 서지레코드에 포함되지 않은 전거레코드 허용
- 6) 확정된 채택표목을 잡을 수 없는 레코드(008/09가 'a'이면서 1XX 표목을 잡을 수 없는 경우)는 참조레코드로 처리 (008/09 b,c)

3. 4 전거통제대상 tag 및 subfield

서지format상에는 정의되어 있으나 전거포맷에 정의되어 있지 않은 서브필드(numeric subfield : 통계관련)는 모두 제외된다.

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	

3. 5 LEADER 및 고정장 필드에서 시스템이 자동생성하는 부분

다음은 시스템이 자동으로 생성 및 수정해 주는 부분이며 시스템 setting값은 수정이 불가능하고 시스템default 값은 시스템이 가장 많이 사용되는 값으로 생성하고 경우에 따라 수정할 수 있도록 한 값이다.

구 분		화면출력	내 용	생성값	비 고
LDR	00-04		레코드길이		시스템 자동생성
	05	RECSTS	레코드 상태	n:신규 c:수정 d:삭제	작업구분에 따라 시스템 자동생성
	06		유형	z	시스템 setting
	07-09		미정의		
	10		지시기호자리수	2	시스템 setting
	11		식별기호자리수	2	시스템 setting
	12-16		기본번지		시스템 자동생성
	17	ENCOD	입력수준	n	시스템 default
	18-19		미정의		
	20-23		엔트리맵		시스템 자동생성
001		전거레코드 유일번호		시스템 자동생성	
005		최종처리일시		시스템 자동생성	
008	00-05		입력일자		시스템 자동생성

008 부분은 전거레코드가 어느 데이터베이스에 포함되는지 또한 어떤 레코드 유형인가(008/09)에 따라 다음과 같이 달리 관리된다.

(1) 인명

서지레코드의 1XX, 7XX에서 생성된 레코드.

구	분	화면출력	내 용	생성값 (008/09)			비 고
				a	b	c	
008	06	GEOSUBD	지리구분	n	n	n	
	07		로마자번호자표	e	e	e	
	09	KIND	레코드종류	a	b	c	
	10	DESCAT	기술목록형식	c	c	c	
	11	SUBJ	주제명표목표	n	n	n	
	12	SERIES	총서형태	n	n	n	
	13	SERNUM	총서번호유무	n	n	n	
	14	NAMEUSE	주표목사용	a	b	b	
	15	SUBJUSE	주제부출표목사용	a	b	b	
	16	SERUSE	총서부출표목사용	b	b	b	
	17		주제세목구분	n	n	n	
	28		정부기관				(blank)
	29	REFEVA	참조평가	a	n	n	
	31	UPDATE	레코드 갱신	a	a	a	default
	32	DIFNAM	목록전거	a	n	a	
	33	ESTLEV	표목수준	a	n	n	
	38	MODF	수정레코드				(blank)
	39	SORC	목록정보원	d	d	d	

(2) 총서명

서지레코드의 4XX, 8XX에서 생성된 레코드

구 분		화면출력	내 용	생성값 (008/09)			비 고
				a	b	c	
008	06	GEOSUBD	지리구분	n	n	n	
	07		로마자번호자표	e	e	e	
	09	KIND	레코드종류	a	b	c	
	10	DESCAT	기술목록형식	c	c	c	
	11	SUBJ	주제명표목표	n	n	n	
	12	SERIES	총서형태	a	a	a	default
	13	SERNUM	총서번호유무	a	a	a	default
	14	NAMEUSE	주표목사용	b	b	b	
	15	SUBJUSE	주제부출표목사용	b	b	b	
	16	SERUSE	총서부출표목사용	a	b	b	
	17		주제세목구분	n	n	n	
	28		정부기관				(blank)
	29	REFEVA	참조평가	a	n	n	
	31	UPDATE	레코드 갱신	a	a	a	default
	32	DIFNAM	목록전거	n	n	n	
	33	ESTLEV	포목수준	a	n	n	a:default
38	MODF	수정레코드				(blank)	
39	SORC	목록정보원	d	d	d		

(3) 주제명

서지레코드의 6XX에서 생성된 레코드

구	분	화면출력	내 용	생성값 (008/09)					비 고
				a	b	c	d	g	
008	06	GEOSUBD	지리구분	n	n	n	n	n	
	07		로마자번호자료	n	n	n	n	n	
	09	KIND	레코드종류	a	b	c	d	g	
	10	DESCAT	기술목록형식	n	n	n	n	n	
	11	SUBJ	주제명표목표	a	a	n	a	a	
	12	SERIES	총서형태	n	n	n	n	n	
	13	SERNUM	총서번호유무	n	n	n	n	n	
	14	NAMEUSE	주표목사용	b	b	b	b	b	
	15	SUBJUSE	주제부출표목사용	a	b	b	b	b	
	16	SERUSE	총서부출표목사용	b	b	b	b	b	
	17		주제세목구분	n	n	n	a	a	
	28		정부기관						(blank)
	29	REFEVA	참조평가	a	n	n	n	n	
	31	UPDATE	레코드 갱신	a	a	a	a	a	default
	32	DIFNAM	목록전거	n	n	n	n	n	
	33	ESTLEV	표목수준	a	n	n	n	n	
38	MODF	수정레코드						(blank)	
39	SORC	목록정보원	d	d	d	d	d		

4. 결 언

도서관에서 사용하는 목록규칙은 여러가지가 있지만 그 목록규칙이 다양한 데이터의 내용을 모두 수용하지는 못한다. 특히 국내 60%이상의 도서관이 사용하고 있는 한국목록규칙에서는 더욱 그렇다. 그래서 우리는 더욱 목록의 표준화를 꾀해야한다. 각대학 마다의 고유한 규칙이 있음에도 불구하고 효율적인 분담목록의 운영과

양질의 센터 DB를 구축하기위한 우리의 수고는 분담목록이 존재하는한 앞으로도 계속 되어야 할것이다. 어찌면 이제부터 더 본격적인 표준화 작업이 진행되어야 할 것이다. 최고의 최대의 데이터베이스를위하여.

〈 부 록 〉

서지 레코드 검증 테이블

(General condition 008/09=a)

BIBLIO. TAGS	VALIDATE AGAINST AUTHORITY TAGS	CONDITION
100	100, 400	008/14=a
110	110, 111, 151, 410, 411, 451	008/14=a
111	111, 151, 411, 451	008/14=a
130	130, 430	008/14=c
100/110/111 + 240 (US)	100, 110, 111, 400, 410, 411	008/14=a
100/110/111 + 243 (US)	100, 110, 111, 400, 410, 411	008/14=a
700/710/711 + 240 (KOR)	100, 110, 111, 400, 410, 411	008/14=a
440	130, 440	008/16=a
600	100, 400	008/15=a
610	110, 111, 151, 410, 411, 451	008/15=a
611	111, 151, 411, 451	008/15=a
630	130, 430	008/15=a
650	150, 450	008/15=a
651	110, 151, 410, 451	008/15=a
700	100, 400	008/14=a
710	110, 111, 151, 410, 411, 451	008/14=a
711	111, 151, 411, 451	008/14=a
730	130, 430	008/14=c
830	130, 430	008/16=a

전거 레코드 검증 테이블

General condition(AID. 008 /15=AID1.008 /15)

AUTHORITY TAGS (AID)	VALIDATE AGAINST AUTHORITY TAGS (AID1)	
		Condition AID=AID1
100	100	400
110	110	410
111	111	411
130	130	430
150	150	450
151	151	451
400		100
410		110
411		111
430		130
450		150
451		151
500		100
510		110
511		111
530		130
550		150
551		151

인덱스 규칙

< MAT : material type >

- 1 : BOOKS (BK)
- 2 : ARCHIVAL & MANUSCRIPTS CONTROL (AM)
- 3 : COMPUTER FILE (CF)
- 4 : MAPS (MP)
- 5 : MUSIC (MU)
- 6 : VISUAL MATERIALS (VM)
- 7 : SERIALS (SE)
- * : ALL

< TAG >

- 대상 tag 및 indicator 지정
- 구성 : (L +) tag[3] (+ 1st_indicator[1])
 - L : local tag 또는 일부 변경 사용 tag
- 1st indicator가 지정되어 있는 경우 해당하는 경우에 한하여 index 추출

< ATTRIBUTE >

- tag attribute (2 자리)
 - M : Mandatory
 - A : Applicable
 - R : Repeatable
 - N : Non-Repeatable
- index attribute (4 자리)
 - o 100 미만 tag (code 부분)
 - tag 내용 표시
 - o 100 이상 tag
 - 1st : author
 - 2nd : title
 - 3rd : subject
 - 4th : publisher

< SUBFIELD >

- 대상 subfield 지정

- all : 입력된 모든 subfield
- / : 반복되는 경우 개별적으로 생성
- [] : tag은 있으나 해당subfield가 없을 수 있음
- 9 : qualifier
- 대상 subfield는 모두 or 조건으로 처리

< AUTHORITY >

- 전거통제 여부
- C : Code (전거통제와 무관)
- Y : 전거통제
- N : 전거통제 받지 않음

< CODE >

- normalize 처리시 예외로 규정된 code

< NORMALIZATION >

- normalization rule 지정

< Phrase 선정 규칙 >

- Tag을 합하여 phrase를 생성하는 경우
: 해당 tag이 반복되는 경우 첫번째 tag을 선정하여 대상으로 한다.
- Authority가 N인 경우 table에 지정된 순서대로, Y인 경우 입력된 순서대로 phrase를 생성한다.

Algorithm 1.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 모든 구두점과 일부 특수기호를 삭제한다.
3. 소문자는 대문자로 바꾼다.
4. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시 한다.

Algorithm 2.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 제 1 indicator가 0이면 ()내용 제외하고 1이면 ()내용 포함하여 index 생성한다.
3. 모든 구두점과 일부 특수기호를 삭제한다.
4. 소문자는 대문자로 바꾼다.
5. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 3.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. indicator에 관계없이 처음 ()내용을 제외하고 index한다.
3. 모든 구두점과 일부 특수기호를 삭제한다.
4. 소문자는 대문자로 바꾼다.
5. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 4.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 제 1 indicator의 non-filing character 만큼을 phrase 앞에서 제외한다.
3. 모든 구두점과 일부 특수기호를 삭제한다.
4. 소문자는 대문자로 바꾼다.
5. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 5.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 제 2 indicator의 non-filing character 만큼을 phrase 앞에서 제외한다.
3. 모든 구두점과 일부 특수기호를 삭제한다.
4. 소문자는 대문자로 바꾼다.
5. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 6.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 제 1 indicator가 1인 경우 처음 () 내용을 함께 INDEX하고 0인 경우 () 내용을 제외한다.
3. 440의 \$v 처리 : 모든 XXX는 무시한다.
 - "\$v [XXX.] n,XXX. n"
첫번째 n만 인정하여 하나의 phrase를 만든다.
 - "\$v [XXX.] n,n,..."
n의 갯수만큼 phrase를 만들고 뒤에 각각의 n을 추가한다.
 - "\$v [XXX.] n1-n2"
n2-n1+1 만큼의 phrase를 만들고 뒤에 n1에서 n2까지 각각의 n을 추가한다.
4. 모든 구두점과 일부 특수기호를 삭제한다.
5. 소문자는 대문자로 바꾼다.
6. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 7.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 제 2 indicator의 non-filing character 만큼 phrase 앞에서 제외한다.
(제 2 indicator가 space인 경우 0으로 간주한다.)
3. 440, 8XX의 \$v 처리 : 모든 XXX는 무시한다.
 - "\$v [XXX.] n,XXX. n"
첫번째 n만 인정하여 하나의 phrase를 만든다.
 - "\$v [XXX.] n,n,..."
n의 갯수만큼 phrase를 만들고 뒤에 각각의 n을 추가한다.
 - "\$v [XXX.] n1-n2"
n2-n1+1 만큼의 phrase를 만들고 뒤에 n1에서 n2까지 각각의 n을 추가한다.
4. 모든 구두점과 일부 특수기호를 삭제한다.
5. 소문자는 대문자로 바꾼다.
6. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 8.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. 앞, 뒤의 () 내용을 삭제한다.
3. 정의된 특수기호 외의 구두점 및 일부 특수기호를 삭제한다.
4. 영문과 영문, 영문과 타언어 사이의 space는 하나만 인정하고 모두 무시한다.

Algorithm 9.

1. 지정된 subfield에 따라 phrase를 추출한다.
2. indicator에 관계없이 처음 () 내용을 제외하고 index한다.
3. 830의 \$v 처리 : 모든 XXX는 무시한다.
 - "\$v [XXX.] n,XXX. n"
첫번째 n만 인정하여 하나의 phrase를 만든다.
 - "\$v [XXX.] n,n,..."
n의 갯수만큼 phrase를 만들고 뒤에 각각의 n을 추가한다.
 - "\$v [XXX.] n1-n2"
n2-n1+1 만큼의 phrase를 만들고 뒤에 n1에서 n2까지 각각의 n을 추가한다.
4. 모든 구두점과 일부 특수기호를 삭제한다.
5. 소문자는 대문자로 바꾼다.

1. INDEX RULE FOR USMARC

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	001	MN		C		8
*	010	AN LCCN	[a]	C		8
123456	020	AR ISBN	[a]	C		8
37	022	AR ISSN	[a]	C	-	8
147	027	AR RPTN	a	C	/-	8
7	030	AR CODN	a	C		8
*	050	AR LC	ab	C	.	8
*	060	AR NLMN	a/	C	.	8
*	070	AR NALN	a/	C	.	8
13567	082	AR DDCN	a/	C	.	8
147	088	AR RPTN	a	C	/-	8
*	L 090	MR DDCN	a	C	.	8
*	L 092	AR DDCN	a/	C	.	8
*	L 093	MR SNUI	a/	C	-	8
*	100	AN YNNN	abcdfgq	Y		1
*	110	AN YNNN	abgk	Y		1
*	110	AN NYNN	p/	N		1
*	111	AN YNNN	aeg	Y		1
*	111	AN NYNN	p/	N		1
*	130	AN NYNN	anp	Y		4
*	130	AN NYNN	p/	N		1
7	210	AN NYNN	a	N		1
3	211	AR NYNN	a	N		5
7	212	AR NYNN	a	N		1
13	214	AR NYNN	a	N		5
37	222	AR NYNN	a	N		5

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	240	AN NYNN	a	N		5
*	240	AN NYNN	g	N		1
*	240	AN NYNN	p/	N		1
*	2421	AR NYNN	a	N		5
1256	243	AN NYNN	a	N		5
1256	243	AN NYNN	p/	N		1
*	245	MN NYNN	a	N		5
37	246	AR NYNN	a	N		1
7	247	AR NYNN	a	N		1
*	260	AN NNNY	b/	N		1
134567	440	AR NYNN	anp	Y		5
134567	440	AR ISSN	x	N		7
134567	440	AR NYNN	p/	N		1
123456	L 502	AR DS2B	b(학위수여기관)	C		1
123456	L 502	AN DS2B	c(학과 및 전공)	C		1
*	600	AR NNYN	abcdfgqxyz	Y		1
*	600	AR NNYN	t	N		1
*	610	AR NNYN	abgkxyz	Y		1
*	610	AR NNYN	p/	N		1
*	610	AR NNYN	t	N		1
*	611	AR NNYN	aegxyz	Y		1
*	611	AR NNYN	t	N		1
*	611	AR NNYN	p/	N		1
*	630	AR NNYN	anpxyz	Y		4
*	630	AR NNYN	p/	N		1
*	650	AR NNYN	axyz	Y		1
*	651	AR NNYN	axyz	Y		1
*	700	AR YNNN	abcdfgq	Y		1
*	700	AR NYNN	t	N		1

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	710	AN YNNN	abgk	Y		1
*	710	AR NYNN	p/	N		1
*	710	AR NYNN	t	N		1
*	711	AR YNNN	aeg	Y		1
*	711	AR NYNN	t	N		1
*	711	AR NYNN	p/	N		1
*	730	AR NYNN	anp	Y		4
*	730	AR NYNN	p/	N		1
123456	740	AR NYNN	a	N		4
37	760	AR NYNN	s	N		1
37	760	AR NYNN	t	N		1
37	760	AR ISSN	x	N		1
37	762	AR NYNN	s	N		1
37	762	AR NYNN	t	N		1
37	762	AR ISSN	x	N		1
137	765	AR NYNN	s	N		1
137	765	AR NYNN	t	N		1
137	765	AR ISSN	x	N		1
137	767	AR NYNN	s	N		1
137	767	AR NYNN	t	N		1
137	767	AR ISSN	x	N		1
137	770	AR NYNN	s	N		1
137	770	AR NYNN	t	N		1
137	770	AR ISSN	x	N		1
137	772	AR NYNN	s	N		1
137	772	AR NYNN	t	N		1
137	772	AR ISSN	x	N		1
*	773	AR NYNN	s	N		1
*	773	AR NYNN	t	N		1

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	773	AR ISSN	x	N		1
137	775	AR NYNN	s	N		1
137	775	AR NYNN	t	N		1
137	775	AR ISSN	x	N		1
*	776	AR NYNN	s	N		1
*	776	AR NYNN	t	N		1
*	776	AR ISSN	x	N		1
37	777	AR NYNN	s	N		1
37	777	AR NYNN	t	N		1
37	777	AR ISSN	x	N		1
137	780	AR NYNN	s	N		1
137	780	AR NYNN	t	N		1
137	780	AR ISSN	x	N		1
137	785	AR NYNN	s	N		1
137	785	AR NYNN	t	N		1
137	785	AR ISSN	x	N		1
137	787	AR NYNN	s	N		1
137	787	AR NYNN	t	N		1
137	787	AR ISSN	x	N		1
134567	800	AR YNNN	abcfgqnpt	Y		1
134567	810	AR YNNN	abgknpt	Y		1
134567	811	AR YNNN	aegnpt	Y		1
134567	830	AR NYNN	anp	Y		5
134567	830	AR NYNN	p/	N		1
*	L 900	AR YNNN	abcfgq	N		1
*	L 900	AR NYNN	t	N		1
*	L 910	AR YNNN	abgk	N		1
*	L 910	AR NYNN	t	N		1
*	L 910	AR YNNN	p/	N		1

2. INDEX RULE FOR KORMARC

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	001	MN SNUN		C		8
*	020	AR ISBN	[a]	C		8
*	022	AR ISSN	[a]	C	-	8
*	027	AR RPTN	a	C	/-	8
7	030	AR CODN	a	C		8
*	050	AR LC	ab	C	.	8
*	082	AR DDCN	a/	C	.	8
*	088	AR RPTN	a	C	/-	8
*	L 090	MR DDCN	a	C	.	8
*	L 092	AR DDCN	a/	C	.	8
*	L 093	MR SNUI	a/	C	-	8
*	100	AN YNNN	abcdfgq	Y		1
*	100	AN YNNN	fa	N		1
*	110	AN YNNN	abgk	Y		1
*	110	AN NYNN	p/	N		1
*	111	AN YNNN	aeg	Y		1
*	111	AN NYNN	p/	N		1
*	130	AN NYNN	anp	Y		3
*	130	AN NYNN	p/	N		1
s	210	AN NYNN	a	N		1
s	222	AR NYNN	a	N		3
*	240	AN NYNN	a	N		3
*	240	AN NYNN	g	N		1
*	240	AN NYNN	p/	N		1
*	2450	MN NYNN	a/	N		3
*	2451	MN NYNN	a/	N		3

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	2452	MN NYNN	a/	N		1
*	2452	MN NYNN	a/	N		3
s	246	AR NYNN	a	N		1
s	247	AR NYNN	a	N		1
*	260	AN NNNY	b/	N		1
*	4400	AR NYNN	anp	Y		3
*	4401	AR NYNN	anp	N		2
*	4401	AR NYNN	anp	Y		3
*	440	AR ISSN	s	N		6
*	440	AR NYNN	p/	N		1
*	5001	AR NYNN	befghijkmr	N		3
*	5002	AR NYNN	befghijkmr	N		1
*	5002	AR NYNN	befghijkmr	N		3
*	5001	AR YNNN	ls	N		3
1	502	AR NNNN	b(학위수여기관)	N		1
1	502	AR NNNN	c(학기 및 전공)	N		1
*	L 5071	AR YNNN	abcdfgq	Y		1
*	L 5071	AR NYNN	t	N		3
*	L 5072	AR NYNN	t	N		3
*	L 5072	AR NYNN	t	N		1
*	600	AR NNYN	abcdfgqxyz	Y		1
*	600	AR NNYN	fa	N		1
*	610	AR NNYN	abgkxyz	Y		1
*	610	AR NNYN	p/	N		1
*	611	AR NNYN	aegxyz	Y		1
*	611	AR NNYN	p/	N		1
*	630	AR NNYN	anpxyz	Y		3
*	630	AR NNYN	p/	N		1
*	650	AR NNYN	axyz	Y		1

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	650	AR NNYN	axyz	Y		1
*	651	AR NNYN	axyz	Y		1
*	700	AR YNNN	abcdfgq	Y		1 4)
*	700	AR YNNN	fa	N		1
*	700	AR NYNN	t	N		1
*	710	AR YNNN	abgk	Y		1 5)
*	710	AR NYNN	t	N		1 5)
*	710	AR NYNN	p/	N		1 5)
*	711	AR YNNN	aeg	Y		1 6)
*	711	AR NYNN	p/	N		1
*	711	AR NYNN	t	N		1
*	730	AR NYNN	anp	Y		3
*	730	AR NYNN	p/	N		1
1	740	AR NYNN	a	N		3
7	760	AR NYNN	s	N		1
7	760	AR NYNN	t	N		1
7	760	AR ISSN	x	N		1
7	762	AR NYNN	s	N		1
7	762	AR NYNN	t	N		1
7	762	AR ISSN	x	N		1
*	765	AR NYNN	s	N		1
*	765	AR NYNN	t	N		1
*	765	AR ISSN	x	N		1
*	767	AR NYNN	s	N		1
*	767	AR NYNN	t	N		1
*	767	AR ISSN	x	N		1
*	770	AR NYNN	s	N		1
*	770	AR NYNN	t	N		1
*	770	AR ISSN	x	N		1

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	772	AR NYNN	s	N		1
*	772	AR NYNN	t	N		1
*	772	AR ISSN	x	N		1
*	773	AR NYNN	s	N		1
*	773	AR NYNN	t	N		1
*	773	AR ISSN	x	N		1
*	775	AR NYNN	s	N		1
*	775	AR NYNN	t	N		1
*	775	AR ISSN	x	N		1
*	776	AR NYNN	s	N		1
*	776	AR NYNN	t	N		1
*	776	AR ISSN	x	N		1
7	777	AR NYNN	s	N		1
7	777	AR NYNN	t	N		1
7	777	AR ISSN	x	N		1
*	780	AR NYNN	s	N		1
*	780	AR NYNN	t	N		1
*	780	AR ISSN	x	N		1
*	785	AR NYNN	s	N		1
*	785	AR NYNN	t	N		1
*	785	AR ISSN	x	N		1
*	787	AR NYNN	s	N		1
*	787	AR NYNN	t	N		1
*	787	AR ISSN	x	N		1
*	830	AR NYNN	anp	Y		3
*	830	AR NYNN	p/	N		1
*	900	AR YNNN	abcdfgq	N		1
*	900	AR YNNN	fa	N		1
*	900	AR NYNN	t	N		1

MAT	TAG	ATTRIBUTE	SUBFIELD	AUTHORITY	CODE	NORMALIZE
*	910	AR YNNN	abgk	N		1
*	910	AR NYNN	t	N		1
*	911	AR YNNN	aeg	N		1
*	911	AR NYNN	t	N		1
*	911	AR NYNN	p/	N		1
*	930	AR NYNN	anp	N		3
*	930	AR NYNN	p/	N		1
*	940	AR NYNN	a	N		3
*	949	AR NNYN	anp	N		1
*	960	AR NNNY	b/	N		1

※ 093 yx + 090 abc/ + 093 vc MR CALN N 8

3. INDEX RULE FOR AUTHORITY

MAT	TAG	ATTRIBUTE	SUBFIELD	CODE	NORMALIZE
	010	AN LCCN	[a]		8
	020	AR ISBN	[a]		8
	022	AR ISSN	[a]		8
	050	AR LC	ab		8
	060	AR NLMN	a/		8
	070	AR NALN	a/		8
	082	AR DDCN	a/		8
	083	AR DDCN	a		8
	090	AR DDCN	a		8
	092	AR DDCN	a/		8
	100	AN YNNN	abcfgqxyz		1
	110	AN YNNN	abgkxyz		1
	111	AN YNNN	aegxyz		1
	130	AN NYNN	anpxyz		5
	150	AN NNYN	axyz		1
	151	AN NNYN	axyz		1
	400	AR YNNN	abcfgqxyz		1
	410	AR YNNN	abgkxyz		1
	411	AR YNNN	aegxyz		1
	430	AR NYNN	anpxyz		5
	450	AR NNYN	axyz		1
	451	AR NNYN	axyz		1
	500	AR YNNN	abcfgqxyz		1
	510	AR YNNN	abgktfhlmsrxyz		1
	511	AR YNNN	aegxyz		1
	530	AR NYNN	anpxyz		5
	550	AR NNYN	axyz		1
	551	AR NNYN	axyz		1