

문헌정보식별기술—선부호기술에 대한 리해

윤 향 희

위대한 령도자 김정일동지께서는 다음과 같이 교시하시였다.

《오늘 세계적으로 과학기술은 매우 빠른 속도로 발전하고있으며 사회발전과 인간생활에서 과학기술의 역할은 날을 따라 더욱더 커가고있습니다.》(《김정일선집》 증보판 제20권 352페이지)

세계적으로 새로운 과학연구성과들이 대량적으로, 련속적으로 이룩되고 그것을 반영한 문헌의 량과 종류도 급격히 늘어나고있는 조건에서 문헌정보관리사업에서는 많은 노력과 시간, 자금 등이 소모되게 된다.

끊임없이 축적되는 방대한 량의 문헌정보들을 수요자들이 신속정확하게 리용할수 있도록 하자면 새로운 과학기술을 제때에 받아들여 문헌정보관리사업을 부단히 현대화해나가야 한다.

오늘날 많은 문헌정보기관들에서는 선부호기술을 문헌정보식별에 적극 리용하고있다.

선부호기술은 수많은 요소들로 이루어진 어떤 모임속에서 개개의 요소를 식별하는 기계수감기술의 일종이다. 즉 너비와 반사률이 서로 다른 검은선과 흰선(혹은 검은막대기와 흰막대기라고도 한다.)모양의 기계수감부호렬로 정보내용을 표기하여 수감기구체계에 의하여 정보가 식별되게 하는 기술이다.

일반적으로 선부호는 량쪽끝에 시작과 마감을 표시하는 시작바이트와 끝내기바이트가 있고 시작바이트뒤에 자료바이트, 비교검사바이트가 순서대로 배열된다.

보통 사람의 눈으로 식별할수 있는 자료표식을 선부호의 아래면에 기입해준다.



검은색막대기와 흰색막대기의 배열을 보면 보통 비대칭2진배열이며 그것은 선부호읽기장치가 쌍방향으로 읽도록 해준다.

선부호는 구조형식과 부호작성방식에 따라 코드39, 교차25, EAN/UPC, 코드93, 코드128, 코드49 등으로 구분할수 있다.

선부호에는 자모식으로 된것도 있고 수자식으로 된것도 있으며 자모식과 수자식이 결합된것도 있다. 그러므로 식별하려는 대상의 특성과 규모에 따라서 선부호코드종류를 선택하여 리용한다.

국제적인 상품생산 및 류통분야에서는 EAN/UPC선부호를 많이 리용하며 국부적인 자

동화관리체계에서는 코드39가 리용되는데 그 차이를 보면 다음과 같다.

구분	자료류형	부호 총개수	부호길이 방식	부호의 련속 가능성	비교검사 기능	2진수 자리수
코드39	자모수자식	43	비고정	불련속	있다	9
EAN/UPC	수자식	10	고정	련속	있다	7

표를 통하여 알수 있는것처럼 EAN/UPC선부호는 자료류형이 수자식이고 부호의 총개수는 10개이며 부호길이가 고정되어있고 비교검사기능을 가지고있다. EAN/UPC선부호에서 2진수코드의 자리수는 7이다.

EAN/UPC선부호는 크게 기본선부호와 보충선부호로 이루어지며 대체로 기본선부호만을 리용하는 경우가 보다 보편적이다. 기본선부호의 구조는 시작바이트, 자료바이트, 비교검사바이트, 끝내기바이트로 되어있다.

EAN/UPC선부호에서는 필요한 경우 즉 가격이나 권차(호수)번호 등 보충적인 자료를 나타내는 보충선부호를 사용할수 있다. 보충선부호에는 두글자보충선부호, 다섯글자보충선부호가 있는데 도서에는 다섯글자보충선부호를, 잡지에는 두글자보충선부호를 사용한다.

문헌정보관리기관들에서는 EAN/UPC선부호를 리용하여 ISBN선부호와 ISSN선부호를 표시한다.

ISBN선부호는 ISBN(International Standard Book Number. 국제표준도서번호)을 기계가 인식하여 처리할수 있도록 작성된 기계수감부호이다.

ISBN은 국제적으로 서로 다른 종의 도서를 구별하는 유일식별기호이다.

ISBN은 10개의 수자로 되어있는데 첫번째 수자부터 아홉번째 수자까지는 해당 도서의 출판지, 출판사, 도서종번호코드이고 열번째 수자는 비교검사수이다.

실례로 ISBN이 0-262-15054-9일 때

0 : 도서를 출판한 나라 혹은 출판장소

262 : 도서의 출판사

15054 : 도서의 종번호

9 : 비교검사수

여기에서 비교검사수는 입력된 값의 정확성을 담보하기 위한 수로서 컴퓨터에서 자동적으로 계산되어 할당된다. 만일 일치하면 입력된 식별기호는 오류가 없다는것이며 일치하지 않으면 오류로 판단하고 식별기호를 다시 확인하여야 한다.

비교검사수를 계산하는 방법은 대단히 많은데 현재 세계적으로 널리 쓰이는것만 해도 100여가지나 된다.

ISBN의 비교검사수를 계산하는 대표적인 방법의 하나를 보기로 하자.

① 식별기호를 이루는 매 수자에 오른쪽으로부터 왼쪽으로 가면서 위치번호를 매긴다.

우에서 실례로 든것처럼 ISBN이 0-262-15054-9일 때

기호	0	2	6	2	1	5	0	5	4
번호	10	9	8	7	6	5	4	3	2

- ② 서로 대응한 값들을 곱하여 얻은 적들을 모두 더한다.

$$(0 \times 10) + (2 \times 9) + (6 \times 8) + (2 \times 7) + (1 \times 6) + (5 \times 5) + (0 \times 4) + (5 \times 3) + (4 \times 2) = 134$$

- ③ 얻어진 총합을 모의수값 11로 나누어 나머지를 구한다.

$$134 / 11 = 12 \quad \text{나머지 } 2$$

- ④ 11에서 나머지로 되는 수를 뺀다.

$$11 - 2 = 9$$

이러한 ISBN식별기호를 선부호화하여 도서에 인쇄한다.

도서에 인쇄되는 ISBN선부호는 모두 13개의 수자로 되어있는데 첫번째 수자부터 세번째 수자까지는 도서를 표시하는 전용코드이며 네번째 수자부터 열두번째 수자까지는 해당 도서의 출판지, 출판사, 도서종번호코드이다. 열세번째 수자는 비교검사수이다.

도서에 인쇄되는 ISBN선부호의 비교검사수를 계산하는 방법은 다음과 같다.

- ① 선부호의 열두개 수자에 각각 수 1과 3을 엇바꾸어 할당한다.

번호	13	12	11	10	9	8	7	6	5	4	3	2
기호	1	3	1	3	1	3	1	3	1	3	1	3

- ② 해당한 수자에 할당된 수값을 각각 곱하여 얻어진 12개의 적들을 모두 합한다.

$$S1 + S2 + S3 + S4 + S5 + S6 + S7 + S8 + S9 + S10 + S11 + S12 = S(n)$$

- ③ 10에서 위에서 얻은 수 $S(n)$ 의 하나자리수를 뺀다. 이때 얻어진 차가 비교검사수이다.

$$10 - S(n) \text{의 하나자리수} = \text{비교검사수}$$

실례로 ISBN번호가 978994610043 2인 도서인 경우 978은 도서라는것을 표시하는 코드번호, 9946은 출판지역코드번호, 1은 출판기관코드번호, 0043은 도서종코드번호, 2는 비교검사수이다.



이때 비교검사수는 다음과 같이 구할수 있다.

$$\begin{aligned} &① (9 \times 1) + (7 \times 3) + (8 \times 1) + (9 \times 3) + (9 \times 1) + (4 \times 3) + (6 \times 1) + (1 \times 3) + \\ &(0 \times 1) + (0 \times 3) + (4 \times 1) + (3 \times 3) = 108 \end{aligned}$$

② $10 - 8 = 2$

ISBN선부호는 국제적으로 서로 다른 도서의 종을 식별하기 위한 기호인것으로 하여 한 종당 많은 부수를 관리하고있는 문헌정보관리기관들에서는 리용에 불편한 측면도 있다.

이때 문헌정보관리기관들에서는 자체의 전용식별기호를 설계제작하여 리용한다.

자기 기관이 소유하고있는 여러가지 류형의 도서들을 종합적으로 관리하면서도 도서의 종뿐아니라 개별적인 부수까지도 구체적으로 관리할수 있도록 식별기호를 설계제작하는것은 보다 복잡한 사업이다.

여기서 리용하여야 할 도서식별기호는 비반복적이고 해당 종의 매 부수에 이르기까지 다 식별할수 있는 유일식별자로 되여야 하며 공통적인 특징을 가지는 도서들과의 련관성을 보장할수 있도록 설계되여야 한다.

일반적으로 도서식별기호는 13개의 자모나 수자 혹은 그것들의 결합으로 설계한다.

우선 인쇄본인가 비인쇄본인가, 어느 어종인가, 공개본인가 비공개본인가, 도서인가 련속간행물인가 등 여러가지 정보자원류형항목들로서 종을 식별하는 부호들을 구성하고 그 다음 한종안에서 개별적인 부수를 식별하기 위한 기호들을 구성하는 방법으로 마지막까지 계단식으로 쪼개내려가면서 구조화하는 방법이 있다.

또한 일정한 정도까지 구조화하고 그다음부터는 기계적인 부호들을 구성하는 설계방법도 있다.

실례로 어느 한 기관에서 설계한 정보자원식별기호를 보면 맨처음 두개의 수자는 어종을, 그다음 세개의 수자는 출판년도를, 그다음 두개의 수자는 주제를, 그다음 여섯개의 수자는 등록순서번호를 부호화하는 방법으로 설계하였다.

실례로 어떤 도서에 KO99837123456이라는 식별기호를 달아준다고 하자.

여기서 KO는 어종을, 998은 출판년도를, 37은 주제, 123456은 등록순서번호를 나타낸다. 즉 이 도서는 조선어로 1998년에 출판된 컴퓨터에 관한 정보자원이라는것을 나타낸다.

ISSN선부호는 ISSN(International Standard Serial Number. 국제표준련속간행물번호)을 기계가 인식하여 처리할수 있도록 작성된 기계수감부호이다.

ISSN은 국제적으로 서로 다른 종의 잡지를 구별하는 유일식별기호이다.

ISSN은 8개의 수자로 이루어지는데 첫번째 수자부터 일곱번째 수자까지는 출판물의 순서번호이고 마지막수자는 비교검사수이다. 읽기에 편리하게 하기 위하여 앞의 4개 수자와 뒤의 4개 수자사이에 《-》를 리용하여 련결한다.(례 : ISSN 1727-9399)

비교검사수가 10이라면 대문자 X로 대신한다.

ISSN선부호는 기본부분과 보조부분을 련결하는 방식을 취한다.

련속간행물에 인쇄되는 ISSN선부호에서 기본부분은 모두 13개의 수자로 이루어지는데 첫번째 수자부터 세번째 수자는 해당 련속간행물을 표시하는것이며 네번째 수자부터 열번째 수자는 해당 련속간행물의 품종번호, 열한번째와 열두번째 수자는 적용되는 선부호의 특성변화값이다. 그리고 열세번째 수자는 비교검사수이다.

ISSN선부호에서 비교검사수를 계산하는 방법은 ISBN선부호에서와 같다.

련속간행물에 인쇄되는 ISSN선부호에서 보조부분을 반영하는것은 두개의 수자이다.

실례로 《천리마》 잡지 주체105(2016)년 3호의 ISSN선부호는 9771727939003 >03이다.

여기서 977은 연속간행물코드번호, 1727939는 이 연속출판물의 품종코드번호, 00은 적용되는 선부호의 특성변화값코드번호, 3은 비교검사수, 03은 연속출판물의 호수를 나타낸다.



문헌정보관리사업에 이러한 선부호기술을 리용하면 문헌정보관리의 효과성을 더욱 높일 수 있다.

우선 문헌정보수집에서 신속성과 정확성을 보장할 수 있다.

일반적으로 문헌수집은 구입단위들에 제출한 예약주문명세서와 문헌현물을 대조확인하는 공정이 여러 단계에서 진행된다. 즉 수집자는 구입지에서 인계인수할 때 그리고 자기 기관에 문헌들을 들여온 다음 등록자에게 인계인수할 때 대조공정을 거쳐 그 정확성 여부를 확인해야 한다. 이때 선부호를 리용하여 문헌현물과 예약주문서에 올라있는 목록기록들을 자동대조하면 수동적으로 대조하는 것보다 시간을 절약하면서도 정확성을 보장할 수 있다.

또한 문헌정보조직에서 ISBN선부호를 리용하여 문헌의 목록작성에 필요한 일련의 마당값들을 자동입력함으로써 목록작성의 신속성과 정확성을 보장할 수 있다.

서지기입규칙에는 문헌의 목록을 작성하는데서 문헌류형, ISBN, 출판사, 출판지 등의 항목들이 없어서는 안 될 필수항목으로 규정되어 있다.

선부호기술을 리용하여 필요한 자료들을 자동적으로 입력하면 사람이 해당한 문헌을 놓고 필요한 자료값을 찾아 컴퓨터건반을 리용하여 수동적으로 입력하는 것보다 결정적으로 빠르며 실수률도 적어진다.

우리는 선부호기술을 비롯한 여러가지 과학기술에 대한 폭넓은 이해를 가지고 문헌정보관리사업에 제때에 적극 받아들이므로써 나라의 과학기술을 발전시키는데 적극 이바지해나가야 할 것이다.

실마리어 선부호기술, 문헌정보관리