

SISAC 바코드를 이용한 최적의 서지DB제작시스템 설계 및 구현

노경란, 권오진, 예용희, 신영호
한국과학기술정보연구원 정보시스템부

Design and Implementation of Optimal Bibliographic DB by SISAC Barcode

Noh Kyung-Ran, Kwon Oh-Jin, Yae Yong-Hee, Shin Young-Ho
KISTI, Information System Division

요 약

과거 서지DB 제작시스템은 자료의 수집부터 DB제작에 이르기까지 자료입수지연 및 서지데이터입력오류, 중복레코드생성, 권호정보의 부정확한 기입, 철자오류와 관련된 많은 문제들을 내포하고 있었다. 본 연구는 이와같은 문제들을 최소화하기 위해 SISAC기반 서지DB 제작시스템을 설계·구현하였다. 이 시스템은 기존 가용자원을 최대한 활용하기 위해 SISAC 바코드를 스캐닝하여 자료의 입수처리를 신속하게 하고, 서지데이터를 입력할 때 기존에 구축된 데이터를 공유함으로써 데이터 중복입력으로 인한 시간소모적인 작업을 줄인다. 그리고 개방형 환경에서 양질의 능력을 지닌 DB제작 전문가를 선정하는데 투명성을 제공하고, 이들 DB제작전문가를 활용함으로써 데이터의 품질향상에 기여한다.

1. 서론

기하급수적으로 증가하는 1차 정보에 효율적으로 접근하기 위해, 양질의 정보를 선별적으로 분석 가공하여 2차 서지 데이터베이스를 구축·활용한다. 과거의 서지DB 제작시스템은 데이터베이스간 연계를 고려하지 않은 채 폐쇄적이고 독립적인 DB제작입력시스템을 사용하였으며, 서지 데이터베이스의 품질보다는 구축량에 치중하였다. 이로 인해 데이터베이스에 수록된 데이터의 부정확성, 중복레코드 및 오류데이터의 발생, 데이터의 일관성 결여, DB구축까지 많은 시간이 소요되는 등 여러 가지 문제가 발생하였다. 또한 DB 제작에 활용가능한 인력이 제한되고 주제분야별 세부 DB제작 전문가를 활용하는데 한계가 있었다.

이러한 DB제작과정에서 발생하는 문제들을 해결하기 위해 SISAC 바코드를 이용한 최적의 서지데이터베이스 제작시스템을 설계 구현하였다. 이 제작시스템은 과거 폐쇄적이고 독립적으로 구축·운영되던 학술지목록DB, 학술지목차DB, 학술지체크인DB를 SISAC Barcode를 사용하여 연계활용함으

로써 서지DB제작에 필요한 서지데이터요소를 추출한다. 또한 개방환경에서 주제전문가들을 공개적으로 모집하고 선정함으로써 세부 주제분야별 전문가를 활용한다.

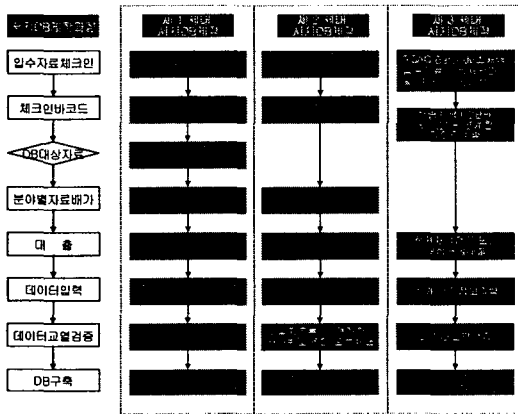
이 개방형 제작시스템은 여러 가용 DB를 연계활용함으로써 데이터의 정확성, 완전성, 최신성, 완전성을 향상시킨다. 웹환경에서 DB를 제작함으로써, 과거 폐쇄적 DB제작시스템이 지니고 있던 시간적, 공간적 물리적 한계를 극복하고 공개된 환경에서 인적자원과 시스템자원이라는 가용자원을 최대한 활용한다. DB제작에 참여하는 작성자, 평가자, 관리자는 물리적 위치에 제한받지 않고 서지DB 제작시스템에 접근할 수 있다.

2. 서지데이터베이스 제작시스템

전통적인 서지데이터베이스 제작시스템은 서지DB 제작대상이 되는 정보자료가 도착하면 이 학술 자료를 검색한 후 발행년, 권/호정보를 입력하여 체크인이라는 입수처리를 수행한다. 체크인 이후 데이터베이스 제작대상자료인지 판단하여 자료의 배가작

업이 이루어지고, 제작대상자료를 주제분야별로 배치시킨다. 데이터 작성자는 자료입수를 확인한 후 대상자료를 대출한다. 데이터 작성자는 대상자료의 수록기사를 분석하여 서지레코드를 작성하며, 데이터 교열자는 이 작성파일을 교열교정하고 작성된 데이터를 평가한다. 데이터는 데이터베이스로 로딩되고 이용자는 정보검색시스템을 이용하여 구축된 데이터베이스를 검색한다.

서지DB 제작시스템은 데이터베이스의 개방성뿐만 아니라 인력이라는 자원의 동적 연계·활용성 측면에서 폐쇄형 DB 제작시스템, 개방형 DB 제작시스템으로 구분된다.

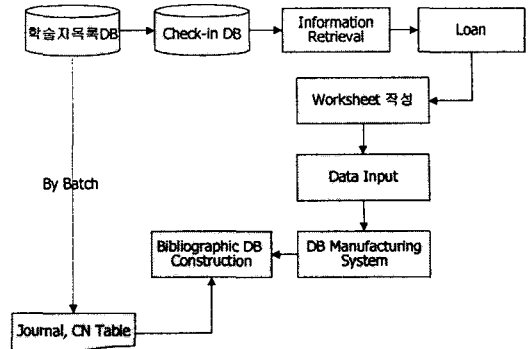


〈그림 1〉 서지데이터베이스 제작시스템

2.1 폐쇄형 서지DB 제작시스템

① 제1세대 서지DB 제작시스템

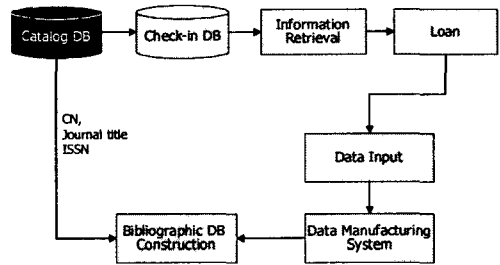
초기 데이터베이스 제작시스템은 다른 서지정보 DB와의 연계성을 전혀 고려하지 않고 독립적인 DB 제작시스템으로 운영되었다. DB 작업자는 서가를 무작위로 브라우징하여 DB 대상저널을 찾아 대출하였다. 그리고 서지정보를 데이터 작성자가 직접 수작업으로 입력하였다. 또한 데이터의 가공에 필요한 분류표, 용어사전, 소장정보를 표현하는 서가번호테이블을 별도로 운영관리하였다. 이와같은 DB 제작시스템으로 생성된 서지정보는 데이터 입력단계에서부터 철자오류, 데이터오류, 레코드 중복 생성 등과 같은 문제들이 발생하였으며, 데이터베이스 구축까지 많은 시간이 소요되었다.



〈그림 2〉 제1세대 서지DB 제작시스템

② 제2세대 서지DB 제작시스템

제 2세대 서지DB 제작시스템은 기존의 구축된 타DB와의 연계성을 모색하기 시작하였다. 이 시스템은 학술지목록DB와의 연계성을 고려하여 생성된 서지정보에 대해 자료의 소장정보를 나타내는 서가번호 및 학술지의 ISSN의 입력오류를 탐지하고 오류를 교열교정하였다. 그러나 이 시스템은 학술지와 관련해서 발생하는 일부 문제를 해결했을 뿐, 제1세대 서지DB 제작시스템이 가지고 있던 여러 문제들을 여전히 지니고 있었다.



〈그림 3〉 제2세대 서지DB 제작시스템

2.2 개방형 서지DB 제작시스템

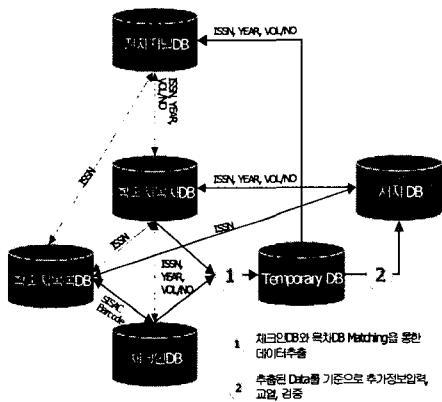
① 제3세대 서지DB 제작시스템

제3세대 서지DB 제작시스템은 기존의 가용자원을 최대한 활용한 제작시스템으로, 다른 DB와의 연계성을 고려하여 데이터의 정확성, 완전성, 최신성을 개선시킨 시스템이다. 학술지목록시스템과의 연계활용뿐만 아니라 SISAC Barcode(Z39.56)를 스캐닝함으로써 정보자료의 자동체크인과 동시에,

DB대상자료의 각 권호에 대해 서가번호, ISSN, 권호정보, 발행년정보, DB대상자료 유무를 수록한 바코드를 자동생성한다. 이 시스템은 자료의 입수처리에 대한 오버헤드를 줄이고, Z39.56이라는 표준에 입각한 권호/발행년정보를 활용함으로써 DB간 연계성을 향상시키고 서지데이터의 품질을 개선시킨다. 또한 SISAC Barcode를 이용한 체크인시 생성된 ISSN, 권호정보, 발행년정보를 활용하여 학술지목차 DB로부터 데이터를 추출함으로써 서지데이터의 입력에 소요되는 부하 및 입력오류를 최소화하였다.

한편, 인적자원측면에서는 웹이라는 개방형 환경에서 주제전문가들을 모집하고 일정기준에 의해 DB작성자를 선정하여 계약을 체결하고, DB제작지침에 관한 교육을 실시한다. 데이터 관리자는 작성된 데이터의 교열검증단계를 거친 후 데이터를 로딩하며, 구축된 DB로부터 각종 작업통계를 생성한다.

이 서지DB 제작시스템은 기존에 공개된 데이터베이스를 최대한 연계활용하여 데이터의 중복입력, 데이터 입력오류, DB제작까지 소요되는 시간지연(time lag), 중복레코드의 생성방지 등 기존 서지DB제작시스템이 지니고 있는 문제들을 최소화하였다.



<그림 4> 데이터베이스의 동적·다차원 연계활용

3. SISAC바코드를 이용한 서지데이터베이스 제작

SISAC 바코드를 이용한 서지데이터베이스 제작방식은 기존에 구축된 개방형 DB들을 다차원적으로 연계활용한다. 또한 공개된 환경에서 선정된 주제전문가들에 의한 데이터가공 및 교열검증작업으로 데이터베이스의 품질을 향상시킨다. DB구축이 완료되면 데이터와 가공전문가에 대한 정량/정성평가를 내리고, 이 평가결과를 가공전문가에게 피드백한다.

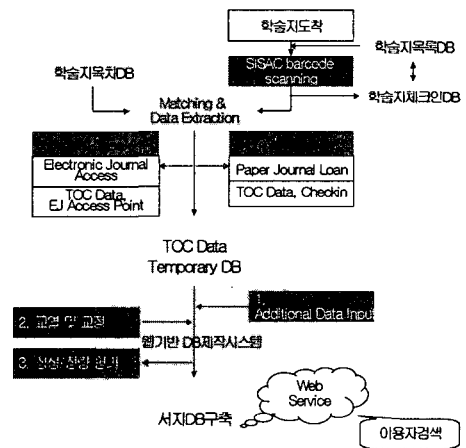
3.1 정보자료입수처리

학술지목차DB와 연계하여 SISAC 바코드 스캐닝을 통한 입수자료의 자동체크인으로 신속한 자료 입수처리한다. 표준화된 발행년, 권호정보패턴을 사용함으로써 서지정보의 정확성을 높이고, 서지DB대상 자료의 정확하고 신속한 판단한다.

SISAC 바코드는 SICI 코드의 Serial Item Identifier를 기계가독형 바코드로 표현한 것이다. SICI 중 Serial Item Identifier의 데이터요소는 ① ISSN, ② 연대기호 (Chronology), ③ 권호명 (Enumeration), ④ 부록/색인, ⑤ Standard Version Number, ⑥ Check Character이다.



<그림 5> SISAC Barcode 예



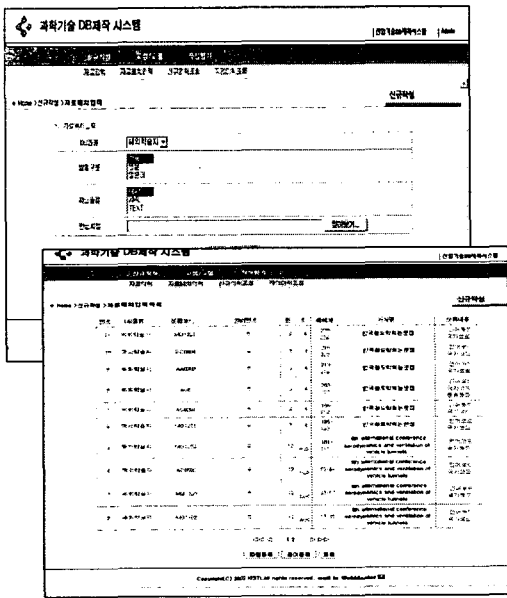
<그림 6> SISAC 바코드를 이용한 서지DB제작시스템

3.2 서지데이터추출

대상자료의 자동체크인이 이루어지면 학술지목차 DB로부터 입수자료의 "ISSN, Vol., No"를 사용하여 기본 서지정보인 기사명, 저자명, 수록저널명, 권호정보, 발행년, 페이지정보를 추출한다. 자료 입수정보는 데이터 작성자에게 이메일로 공지되며, 추출된 데이터는 Temporary DB에 저장된다. 데이터 작성자가 웹환경에서 대상자료의 원문에 접근할 수 있으면 전자저널 access point 정보가 이메일에 첨부된다. 전자저널에 접근할 수 없는 데이터 작성자는 대상자료를 대출한다.

3.3 서지데이터가공

서지데이터를 가공하는 데이터 작성자들은 웹상에서 공개선정된 주제전문가들로 이용자 인증과정을 거쳐 Temporary DB에 접근한다. 이들 주제전문가들은 학술지목록DB에서 생성된 기본 서지정보에 원문정보를 분석한다. 주제분류테이블을 보고 주제분류코드를 부여하며, 색인어 및 주요어를 추출한다. 마지막으로 원문을 보지않고 내용을 명확히 파악할 수 있도록 원문을 대체할 수 있는 초록을 작성한다. 이와 같은 심층 가공정보는 DB가공지침서에 따라서 작성한다.



〈그림 7〉 DB구축을 위한 데이터입력 및 확인화면

3.4 서지데이터검증

서지정보의 가공입력이 완료된 데이터는 교열검증 단계를 거친다. 기본 서지데이터의 추출로 인해 데이터입력오류는 과거 시스템에 비해 현격히 감소하였으며, 데이터 가공시 발생한 논리적, 물리적 오류를 찾아 정정한다. 데이터 작성자와 마찬가지로 데이터의 교열검증인력은 웹에서 공개선정된 주제전문가들로 구성되며, 이들은 이전 단계의 가공입력자보다 한층 더 전문지식을 지니고 있는 고급주제전문가들이다. 교열교정자는 가공한 데이터의 내용과 형태 측면에서 중복레코드탐지, 데이터값 입력오류를 학술지목록DB와 연계하여 체크한다. 이 단계를 거친 후 데이터는 정확성, 적합성, 수록데이터의 완전성

등이 향상된다.

3.5 DB 구축

DB 작성자가 생성한 데이터를 교열교정하고, 검증을 통해 데이터의 우수한 부분과 미흡한 부분을 파악하고, 미흡한 부분에 대한 보완을 통해 품질을 향상시킨다. 그리고 DB작성자에 대해서는 데이터 품질평가에 따라 등급을 부여하고 우수 DB작성자에 대해서는 인센티브를 제공하고, 일정기준에 미달하는 DB작성자에 대해서는 재교육을 실시하거나 DB작성자 계약을 해지한다. 이와 같은 품질평가후 피드백을 통해 DB작성자는 DB작성시 노력을 더 기울이게 되고, 제작되는 DB의 품질은 향상된다.

이 작업통계를 기반으로 데이터를 계량적으로 측정하여 데이터 작성자나 교열교정 전문가들에 대한 평가를 정량적 평가와 함께 정성적 평가를 내린다. 이와같이 내려진 평가결과를 DB제작에 참여했던 주제전문가에게 피드백함으로써 양질의 적합한 주제전문가를 선별할 수 있다.

4. 결론

본 논문에서 설계구현한 SISAC 바코드를 이용한 서지DB 제작시스템은 SISAC Barcode를 기반으로 공개된 가용자원을 최대한 활용하여 최적의 서지DB를 제작하는 시스템이다. 구축되어 있는 DB를 연계활용하여 필요한 서지데이터를 추출하고, 데이터 입력오류를 체크한다. 그리고 웹이라는 개방환경에서 데이터 가공전문가를 선정활용한다. 이로써 DB 제작과정에서 발생할 수 있는 오류를 최소화함으로써 데이터의 정확성, 완전성, 최신성, 완전성을 향상시킨다.

향후 DB제작에 참여하는 각분야의 주제전문가를 효율적으로 관리할 수 있는 평가방법개발 및 SICI 기반 DOI를 이용하여 외부DB의 연계활용을 연구하여 DB제작시스템에 반영하는 것이 요구된다.

참고문헌

- [1] 권오진, 노경란. 2000. "SISAC 바코드를 이용한 자동체크인 시스템의 설계 및 구현" 한국정보처리학회 추계학술발표 논문집 7(2) : 453-456.
- [2] 한혜영. 1999. "국내 학술지 체계논문을 위한 SICI기반 DOI 체계," 이화여대 문헌정보학과 창립 40주년 기념논문집, 157-186.
- [3] 정의섭, 권오진. "컨텐츠제작시스템 및 그 방법." 한국특허 공개번호 : 2000-058637.