

데이터웨어하우스를 이용한 도서관 관리자 의사 결정 시스템 개발에 관한 연구

최인선

우송정보대학 전산정보계열
ischoi@zerobit.woosonginfo.ac.kr

Library Decision Support System
Using DataWarehouse

In-Seon Choi

Dept. of Computer and Information Science, Woosong Information College

요 약

도서관 이용자의 요구에 적합한 도서관 자료를 수집하고 처리함과 함께 소장되어 있는 자료가 이용자의 요구에 어느 정도 만족되는 지를 도서관의 경영 관리자는 알고자 한다. 이는 도서관의 재정지출에 효율화 및 경비 절감에 크게 기여할 수 있기 때문이다. 그러나 도서관 데이터는 방대하고 다양함으로 비교 및 분석에 어려움이 따른다. 그러므로 본 시스템은 시계열성이 강조되며 데이터의 조회만이 일어나는 데이터웨어하우스 기법을 이용하여 관리자의 의사결정에 필요한 정보를 제공하는 시스템을 개발하였다.

1. 서론

최근 도서관등 정보 제공 기관에서는 그들이 소장한 자료에 대한 이용자의 만족도와 그와 관련한 예산 지출이 효과적으로 되고 있는 지에 대한 요구가 정보제공기관의 경영자의 입장에서 중요하게 대두 되어졌다. 그러나 정보제공기관의 데이터는 무척 방대하고 다양함으로 1) 정보제공기관의 자료가 과연 사용자의 필요를 만족시키고 있는가? 2) 수집된 자료가 일정기간동안 대출되어진 비율 3) 이용자로부터 전혀 이용되지 않은 자료를 수집하고 처리하는데 소요된 경비등과 같은 정보를 얻기가 쉽지 않다. 그러나 이러한 정보는 도서관의 관리자 입장에서 무척 필요하며 또한 이것은 정보제공기관의 재정을 효과적으로 운용하는데 크게 기여하게 된다. 본 연구는 이러한 요구를 위하여 읽기 전용의 데이터들로 이루어져 의사 결정에 도움을 주기 위한 조회만이 가능하고 갱신이나 삭제가 일어나지 않는다는 것을 기본 전제로 하는 데이터 웨어하우스 기법을 이용하여 설계 구현 되었다. 본 연구는 특별히 미국 플로리다 주립대학교내의 5 개 도서관에 관한 데이터를 사용하였고, 본 시스템은 플로리다 대학의 도서관 자원제공부서(Resource Services Department) 을 위해 개발되었다

2. 관련 연구

2.1 사용자 요구 분석

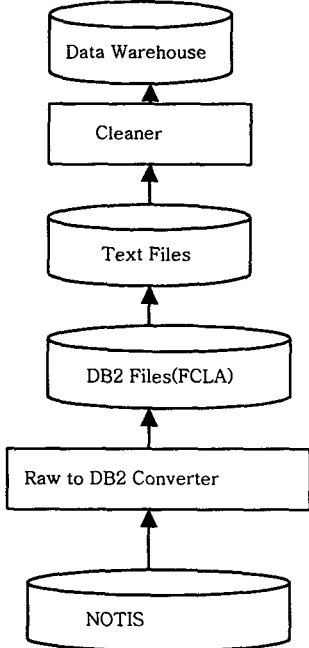
도서관 자원 제공 부서의 관리자는 효율적으로 도서관 재정을 운용하기 위하여 다음과 같은 기능을 시스템에게 요구한다.

- 1) approval title의 대출 과 비대출 비율
- 2) firm order title의 대출과 비대출 비율
- 3) 비대출 title중 분실된 title의 비율
- 4) 비대출 title 을 수집하고 처리하는데 소요된 재정
- 5) 비대출 title의 original, shared 그리고 copy cataloging 의 수
- 6) 대출된 title 의 주제 분야와 비대출 title 의 주제 분야
- 7) 일정 기간내에 주제별로 일정 횟수 만큼 대출된 title은 무엇인가?

2.2 데이터 수집

본 시스템의 데이터는 실제 플로리다 주립 대학교의 도서관 자료를 이용하였다. 이용된 데이터는 실제 미국의 150여개 대학에서 사용하고 있는 NOTIS 도서관 관리 시스템에 raw format으로 존재하였다. NOTIS 시스템은 LUIS를 사용하여 cataloging, 대출 과 공공 접근(public access)의 기능을 제공하고 있는 시스템이다. NOTIS의 raw format 데이터는 플로리다주의 여러 주립 대학교의 도서관에 자동화 서비스를 제공하는 기관인 FCLA(Florida Center of Library automation)에 DB2 파일로 변환되어져 있다. 그러므로 본 시스템에서는 FCLA의 DB2 파일을 본 시스템이 구축되는 컴퓨터로 텍스트 파일의 형태로 파일 전송을 받았다. 그러나 전송된 데이터는 파일 변환 과정에서 원하지 않는 포맷으로 변환된 부분이 있었다.

그래서 정확한 데이터를 얻기 위하여 전송된 데이터는 정화(cleaning)과정을 거쳐 데이터 웨어하우스의 현재 상세 데이터로 입력될 준비를 갖추게 된다.



<그림 1.> 데이터웨어하우스를 위한 데이터 수집 flow

2.3 도서관 관리자 의사 결정 시스템

2.3.1 설계 및 개발 개요

데이터 수집 모듈에서 데이터웨어하우스로 입력된 데이터는 현재 상세 데이터(current detail data)가 된다. 현재 상세 데이터는 사용자가 요구하는 메뉴에 따라서 요약데이터가 되어진다. 이 단계 디자인에서 고려되어질 사항은 1) 어느 시간대 상에 데이터가 요약 데이터가 될 것인가? 2) 현재 상세 데이터중 요약데이터에 포함되어질 항목은 무엇인지가 요약데이터 생성 모듈에서 처리되어진다. 또한 과거 상세 데이터 모듈에서는 가끔씩 필요에 의해 액세스 될 수는 있으나 사용자의 관심 밖의 대상이 되는 데이터는 시간 개념에 따라 현재 상세 데이터를 과거 상세 데이터로 바꿔게 하는 모듈이다. 사용자가 원하는 레포트는 요약 데이터 모듈에서 생성된 요약데이터에 근거하여 Report Generator에서 출력된다.

2.3.2 데이터 구조

현재 상세 데이터는 5개의 관계 테이블로 구성되며 각각은 다음과 같다.

Ufpay_Wh : 도서관에서 구입한 자료에 관한 예산 지출 관련 테이블

{Invoice_num, Order_num, Ord_div, Bib_key, Create_date, Lst_update, Paid_amt, Usd_amt, Fund_key, Exp_class, Fiscal_year, Copies, Type_pay}

Ufcopy_Wh : 대학내 5개 도서관에 소장되어 있는 동일한 자료에 관한 테이블

{Bib_num, Process_unit, Cpyct1_num, Item_count, Main_location, Sub_location, opy_num, Catalog_state, Call_num_type, Copy_state, Call_number}

Uford_Wh : 도서관 자료 주문에 관한 테이블
{Ord_num, Bib_key, Ord_div, Order_date, Ord_vended, Copyct1_num, Acr_stat, Fund_key}

Ufbib_Wh : 도서관 자료의 서지정보에 관한 테이블
{bib_key, Title, Author, Publisher, Pub_date}

Ufcirc_Wh : 도서관 자료의 대출 기록 테이블
{bib_key, Charges, Browses, Notes, Cta_code}

현재상세데이터 구조에 근거하여 요약데이터는 사용자의 요구에 따라 사용자 View에서 새로운 요약데이터 테이블이 생성되기도 하고 없어지기도 한다. 다음은 요약데이터 테이블과 필드일 수 있는 예를 보인다.

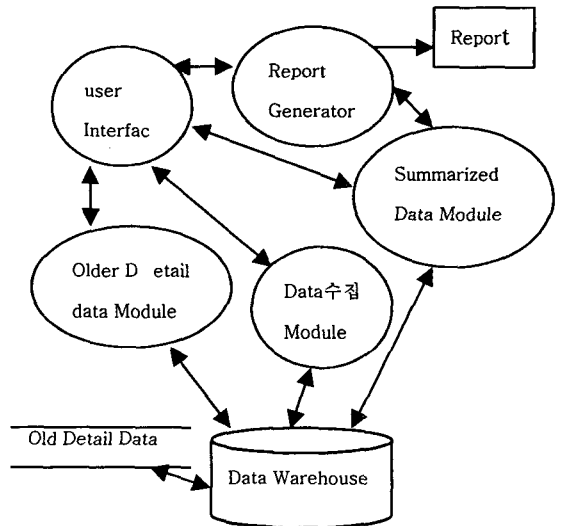
Ufpay : Ufpay_Wh에서 요약된 테이블
{bib_key, Create_date, Paid_amt, Fund_subject}

Ufcopy : Ufcopy_Wh에서 요약된 테이블
{Main_location, Sub_location, Copy_num}

Uford : Uford_Wh에서 요약된 테이블
{Ord_num, Bib_key, Ord_vended, Fund_code}

Ufbib : Ufbib_Wh에서 요약된 테이블
{Bib_key, Title}

Ufcirc : Ufcirc_Wh에서 요약된 테이블
{Bib_key, Charges, Lost, Ctg_code}



<그림 2.> 시스템의 기능적 자료 흐름도

2.3.3 구현

사용자의 시간 설정과 필드 선택에 의해 요약된 데이터는 사용자의 질의에 여러 결과물을 제공한다.

사용자의 요구는 데이터웨어하우스의 현재 상세 데이터로부터 생성된 요약 데이터 테이블로부터 질의된다.

이런 질의를 나타내는 SQL 문장의 예는 다음과 같다.

```
1) 대출되어진 approval title
EXEC SQL SELECT COUNT(Bib_key)
      INTO :count_charge
      FROM UFCIRC
      WHERE charges IS NOT NULL and Bib_key IN
            (SELECT Bib_key
             FROM UFORD
             WHERE fund_code = " 07" );
Apv_circulation = count_charge/count_total * 100
```

```
2) 비대출 자료들중 분실된 title의 비율
EXEC SQL SELECT COUNT(Bib_key)
      INTO :count_lost
      FROM UFCIRC
      WHERE charge IS NULL
      AND lost IS NOT NULL;
EXEC SQL SELECT COUNT(Bib_key)
      INTO :count_total
      FROM UFCIRC
      WHERE charge IS NULL;
Lm_non_circulation = count_lost/count_total * 100
```

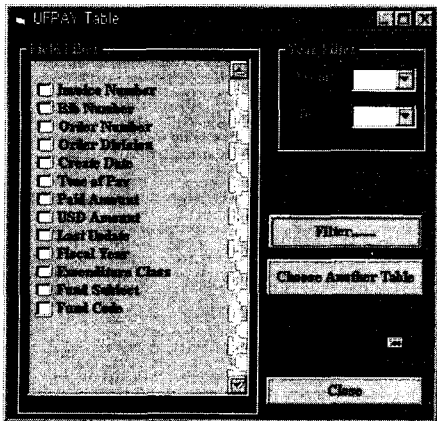
```
3) 대출된 title의 주제 분야
SELECT DISTINCT fund_subject
FROM UFPAY
WHERE Bib_key IN (SELECT Bib_key
                  FROM UFCIRC
                  WHERE charges IS NOT NULL);
ORDER BY fund_subject;
```

3. 결론

오퍼레이셔널 시스템의 데이터는 액세스하는 순간에 정확해야만 의미가 있는 것에 비해 데이터웨어하우스 기법을 이용하여 구축한 본 시스템은 일정 기간동안의 데이터를 분석하여 그에 관한 결과를 제공하는 것으로서 도서관 관리자의 의사결정에 필요한 정보에 대한 요구에 적합하였다. 본 시스템은 Pentium Pc 에서 Visual Basic 5.0을 이용하여 구현 하였다.

4. 참고 문헌

[1] America Online, Inc., "Data Modeling for Data Warehouses, White Paper, 1996
 [2] Elmasri, R., and Navathe, S. B., Fundamentals of Database Systems(2nd), Benjamin/Cummings Publishing, 1994.
 [3] Hammer, J., et.al, "The Stanford Data Warehousing Project, " IEEE Data Engineering Bulletin, June, 1995.
 [4] Hansen, G., Hansen, J., Database Mangement and Design, Prentice Hall, 1992
 [5] S.Kelly, Data Warehousing, John Wiley & Sons, 1996.
 [6] 데이터 웨어하우스와 의사결정지원시스템, 컴퓨터 월드, 1997



<그림 3.> 현재상세 데이터로부터 요약데이터를 생성