The Journal of Natural Science,
Pai Chai University, Korea
Volume 8: 109~114, 1995

메타 데이타베이스와 관리기의 설계 및 구현
- 통계 데이타베이스를 중심으로 -

안성옥
배재대학교 전자계산학과

The Design and Implementation of Meta database and Manager
- on the focus of a statistical database -

Sung-Ohk Ahn
Dept. of Computer Science, Pai Chai University

통계 데이타베이스의 효율적 이용을 위해 통계 분석을 위한 요약 정보를 미리 계산하여 저장함으로써 사용자에게 빠른 응답시간내에 통계 정보를 제공하려는 요약 데이타베이스와 이의 효율적인 관리와 사용의 편리를 위한 메타 데이타베이스가 생성되고 관리되어야만 한다. 요약 데이타베이스를 효율적으로 이용한 통계 분석 작업의 환경과 사용자의 편이성을 지원하기 위하여 계층 구조 형태인 데이타 사전/디렉토리의 형태로 독립적으로 운영되는 메타 데이타베이스와 관리기의 설계 및 구현 작업과 이를 이용한 운영 방법 등이 제시되었다.

For a effective management of statistical database, statistical summary information must be provided by accessing directly the precomputed summary data from summary database to store and manage multiple statistical measures and we store and manage meta database for supporting statistical analysis and providing users with statistical summary information. In order to support effectively the use of summary database, we do the design and implementation of meta database and manager having a hierarchical structure as a data dictionary/directory and operation method is presented.

Key words : meta database, summary database, statistical database

I. 서 론

메타 데이타란 데이타에 관한 정보를 총칭하는 것으로 다양한 방법으로 검색되고 운영되고 표현되는 데이타의 내용과 조직에 관한 체계적이고 설명적인 정보의 집합이다. 측정 데이타를 기준으로 볼때는 매개 데이타도 개별 측정 데이타의 성격을 규정하는 메타 데이타로 취급하기도 한다. 이러한 메타 데이타가 포함되고 있는 정보를 살펴보면 다음과 같다.[1,2,3,]

1) 명칭과 약어(Names와 Aliases) : 명칭은 메타 데이타 엔티티들을 위한 유일한 식별자(identifier)이며 약어나 동의어 등은 명칭을 대체할 수 있는 식별자이다.

2) 분류와 설명정보(Labeling과 Descriptive Information) : 명칭과 약어를 보충하는 분류, 주석, 색인용어, 코드화되어 있는 정보들을 포함한다.

3) Data Derivation과 Quality : 특수한 필드나 어트리뷰트에 관한 Derivation과 Quality 정보는 작성과 수정의 절차와 역사, 소오스 인용과 신뢰

성 평가등을 포함한다.

4) 보안사항(Security Specification) : 보안사항은 읽고, 삭제하고, 첨가하고 수정할 수 있는 프로그램이나 사용자의 종류와 범위등을 지적한다.

5) 논리구조(Logical Structure) : 구조정보는 데이타의 논리적 모델의 구조뿐만 아니라 다른 사용자들에 의해 다양한 뷰(views)들과 배열과 같은 복잡한 데이타 타입들을 묘사한다.

6) 접근경로와 연결명세(Access Path & Linkage Specification) : 메타 데이타는 다른 엔티티들이 어떻게 연결되어 있는지 등을 설명한다.

7) 처리절차(Processing Procedures) : 데이타의 변환, 오류시 표기, 구간화과정 등을 표기한다.

특별히 통계 데이타베이스를 취급하는 통계 메타 데이타는 아래와 같은 정보를 포함해야 한다. 여기서 통계 데이타베이스는 단순한 통계치일 뿐만 아니라 일반적인 통계 처리에서 필요한 통계 분석을 위해 주로 사용되는 데이타베이스를 말한다.

1) 값들의 비교를 위해 표준형태를 취하기 위한 가중식, 오류계산을 위한 오류식과 변수 생성식들이 있다.

2) 데이타의 측정단위, 경로선택 변수와 타임체크 변수 등이 있다.

3) 조건처리와 데이타의 적절한 범위를 위한 부재 데이타(missing data)가 명세되어야 한다.

4) 수치속성에 대한 기술 통계량인 최대값, 최소값, 합, 평균, 분산, 표준편차등의 식과 선택변수가 명세되어야 한다.

이 논문에서는 통계 데이타베이스를 효율적으로 사용하기 위해 사용자의 편이성을 지원하고 요약 정보를 효율적으로 운영하기 위하여 메타 데이타베이스와 관리기의 설계 및 구현작업과 이를 이용한 운영 방법등이 제시 되었다.

II. 메타 데이타베이스의 구조 및 특성

1. 메타 데이타베이스의 구조

메타 데이타베이스는 시스템의 실제 운영을 위한 모든 정보 특히 요약 데이타베이스를 효율적으로 운영하기 위한 정보들을 저장하며, 데이타를 정의하고 서술하는 데이타 사전과 데이타베이스 각 성분의 위치와 구조적 정보를 제공하는 데이타 디렉토리가 결합된 데이타 사전/디렉토리 (Data Dictionary/Directory : DD/D)의 형태로 구성된다.[2] 이러한 데이타 사전/디렉토리에서 사용하는 메타정보는 기본 메타 데이타와 정보 메타 데이타로 나누어진다.

기본 메타 데이타는 화일과 매개 데이타의 이름, 코드화된 매개 데이타의 값, 레코드의 갯수와 각 필드의 크기 등이 저장되며 정보 메타 데이타는 그래프형태의 구조를 명시하고 각 기본 메타 데이타의 구간화, 확장된 필드속성, 통계적으로 유도된 데이타와 수치 분류화정보 등이 저장된다. 이를 위해 기본 메타데이타는 전역 메타 데이타베이스로 저장하고 하부구조로 정보 메타 데이타를 저장하여, 실시간 응답이 가능하고 통계 데이타의 분석에 적합한 Fig. 1 과 같은 계층구조 형태인 데이타 사전/디렉토리 시스템으로 구성하였다. 상위단계의 메타 데이타베이스는 원시 또는 요약 데이타베이스에서 하위단계의 연결에 필요한 정보를 저장하며 사용자 질의용어와 원시와 요약 데이타베이스의 전체적인 관리를 수행한다. 하위단계의 메타 데이타베이스는 요약 데이타 테이블 요구시 행과 열의 인덱스에 대한 정보와 기존 통계량의 이름들을 저장하여 관리한다.

이 논문에서는 사용자가 알기 쉽게 정보를 제공받을 수 있게 하기 위하여 기본 메타 데이타와 정보 메타 데이타의 기술 통계량 이름을 한글로 표현하여 코드화된값과 대응하도록 구현하였다.

또한 요약 데이타베이스를 쉽게 운영하기 위한 요약값에 관한 정보와 코드도 한글로 구현하였다. 그리하여 요약 테이블(Automatic Summary Table : AST)구현시 필요한 정보인 테이블 이름, 행 어트리뷰트 집합에 속하는 범주 속성의 이름, 열 어트리뷰트 집합에 속하는 범주속성의 이름, 각행과 열에 속하는 범주속성 데이타들의 레벨수준, 유도된 기술 통계량의 이름들이 저장되어 있어 요약 데이타 테이블 요구시 명시된 스키마에 따라서 올바른 오퍼레이션이 쉽게 수행되도록 보조하고, 시각적인 디스플레이 형태가 되도록한다.

이러한 메타 데이타의 사용을 기능별로 살펴보면 다음과 같다.

첫째, 데이타 정의 : 메타 데이타의 중요한 사용은 데이타 정의로서 메타 데이타 어트리뷰트들은 내부 데이타 화일 뿐만 아니라 외부 데이타를 묘사하는 데에도 사용될 수 있다.

둘째, 도큐멘테이션(Documentation) : 데이타를 설명하는 온라인 또는 인쇄형태의 도큐멘테이션을 위해 사용된다.

세째, 데이타 선택 : 메타 데이타는 데이타 선택의 다양한 타입을 지원하며 데이타의 저장위치를 계산할 수 있게 한다. 그러나, 그것은 보다 더 복잡한 메타 데이타와 그 메타 데이타를 사용하기 위해 복잡한 데이타 관리 소프트웨어를 요구한다. 즉, 메타 데이타 색인들을 생성하고 유지하기 위한 자동 색인 메카니즘같은 편리성이 지원되어야 한다.

네째, 데이타 운영(Manipulation) : 데이타 운영루틴들은 메타 데이타의 어트리뷰트들, 카테고리 세트들을 확인하는 것 뿐만 아니라 알고리즘의 중요한 부분으로 사용한다.

다섯째, 데이타 디스플레이(Display) : 입력과 출력시 데이타를 알기쉽고 편리하게 디스플레이 할 수 있는 정보를 제공한다.

2. 메타 데이타베이스의 특성

이 논문에서 구현된 메타 데이타베이스의 특성은 다음과 같다.

1) 메타 데이타의 데이타 타입 : 메타 데이타는 주로 도서 목록과 같은 문서화(textual) 형태로서 효과적인 사용형태는 표준적이고 고정된 길이의 데이타 타입들 뿐만 아니라 가변 길이 텍스트같은 확장되고 복잡한 데이타 타입들도 포함한다.

2) 메타 데이타 구조 : 융통성 있는 확장 가능 계층 데이타 구조(open-ended hierarchical data structure)이다. 각 노드들은 메타 데이타 엔티티들의 다른 타입들의 정보의 계층모임들(clusters)을 표시하고 이 노드들은 명백한 연결(linkage)이나 연합(association)을 나타내는 선(line)에 의하여 구조를 형성한다.

3) 메타 데이타에서의 논리연결(logical linkage) : 메타 데이타는 보다 더 확장된 정보나 다른 영역의 키나 포인터로 제공될 다른 메타 데이타의 어트리뷰트들의 이름을 포함하여 연결시키며, 또한 데이타베이스로부터 데이타 원소 수준으로 어떤 총체적 메타 데이타 어트리뷰트들의 자동 계승(automatic inheritance)의 기능을 가진다.

4) 메타 데이타의 액세스형태 : 대부분의 데이타베이스에서 사용자는 데이타의 선택을 위해 메타 데이타의 일부분만이 자주 사용된다.

5) 메타 데이타 갱신 : 메타 데이타 값의 갱신은 새로운 데이타베이스가 생성될 시 그에 따른 메타 데이타 값들이 첨가되며 또한 존재하는 데이타베이스의 구조, 화일 또는 약어첨가 등의 변화에 따라 변경된다. 그리하여 메타 데이타 갱신은 빠르고 동시성이기 보다 새로운 메타 데이타와 타입들이 쉬운 삽입을 위한 편리성이 제공되었다.

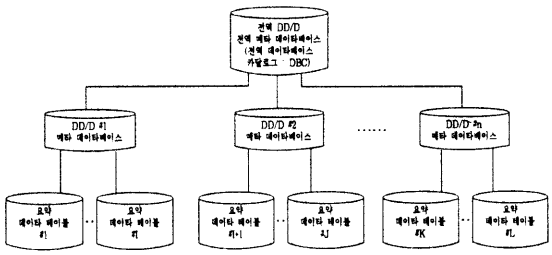

Fig. 1 Architecture of Data Dictionary/Directory having a hierarchical structure

III. 메타 데이타베이스 관리기와 시스템의 운영

1. 경로 결정 알고리즘

메타 데이타베이스의 활용을 통한 메뉴방식에 의한 사용자 질의에 해당하는 코드가 입력되면 시스템의 경로결정 모듈에서 경로결정을 위한 선택값들을 생성한다. 이러한 과정을 요약하면 다음과 같다.

첫째, FLAG변수를 사용하여 원시 데이타베이스와 요약 데이타베이스중 어느 곳에서 검색할 것인지를 결정한다. 일반적으로 통계 데이타베이스는 90%이상이 분석을 위한 요약값을 원하기 때문에 일반적으로 요약 데이타베이스로 부터 질의에 대한 응답을 얻는 것이 대부분이지만, 모든 경우의 질의에 응답하도록 하기위해 원시 데이타베이스로 부터도 검색할 수 있도록 하였다. 단 원시 데이타베이스가 국가적 정책 또는 개인의 사생활과 관련하여 보안이 요구되는 경우 보호 메카니즘을 사용하여 권한을 가진 사용자만이 액세스 할 수 있도록 시스템이 보안되어야 한다. 또한 원시 데이타베이스의 양이 너무 방대한 센서스 데이타 같은 경우는 요약 데이타베이스만 시스템에서 관리하고 원시 데이타베이스는 보조기억장치등에 저장만 하여도 된다.

둘째, 같은 FLAG 변수를 사용하여 요약값을

요구하는 사용자 질의가 시스템에 생성됨과 동시에 제공된 구조에 의해 자동 생성된 요약 데이타베이스로 부터 구해지는가 아니면 다른 구조의 요약세트를 원하는 가의 여부를 판단하여 경로를 결정한다.

세째, CHECK 변수를 사용하여 요구하는 요약값의 테이블 형태를 결정해 주어 요약 테이블 출력모듈로 그 값을 전달한다.

네째, ST_VAL변수를 사용하여 시스템이 제공하는 기술 통계량의 값 중 질의에서 요구하는 값을 대응시켜 해당모듈로 그 값을 전달한다.

다섯째, LEE 변수를 사용하여 사용자가 요구하는 다른 구조의 요약세트가 이미 다른 사용자에 의해 요구되어 요약 데이타베이스에 갱신되었는 가를 판단하여 경로를 결정한다. 이와같은 경로결정 알고리즘을 의사코드(pseudo-code)로 표현하면 Fig. 2 와 같다.

```
Algorithm for path - decision

PRODUCT := 1;
do INDEX = 1,LAST
/* LAST는 입력 CODE의 갯수 */
 PROUCT :=PRODUCT * CODE(INDEX);
if PRODUCT = 0 then
 {
    FLAG := 1; /* 자동 요약값 */
    I := LAST;
    do while (I > 0 AND CODE(I) = 0)
      I := I - 1;
    MULTY := 1;
    do J = 1, I
      MULTY := MULTY * CODE(J);
    if MULTY = 0 then
    {
      FLAG := -1; /* 다른 구조 요약값 */
      LEE :=0;
      if 요약 데이타베이스에 존재하지 않으면
      then
        LEE :=1
    }
  }
/* 요약 테이블 형 결정 */
CHEK := 0; /* 기본 요약 테이블 */
do IC = 1, LAST
if CODE(IC) > MACODE(IC)  then
  {
    CHECK := 2; /* 다중 요약 테이블 */
    KCODE(IC) := 1;
    LCODE(IC) := MACODE(IC)
  }
 else
   KCODE(IC) = LCODE(IC) := CODE(IC)
end /* DO */
    else FLAG := 0; /* 원시 데이타 베이스로
부터의 검색 */
if FLAG ^= 0 then
 if CODE(LAST + 1) = 0 then
   ST_VAL := 기술 통계량 전부;
 else
  {
    if CODE(LAST + 1) = 1 then
       ST_VAL := AUT_DATA.KEYVL;
                 /* 키값*/
    if CODE(LAST + 1) = 2 then
       ST_VAL := AUT_DATA.CONT;
                 /* 빈도수*/
    if CODE(LAST + 1) = 3 then
       ST_VAL := AUT_DATA.SUM;
     if CODE(LAST + 1) = 4 then
       ST_VAL := AUT_DATA.MEAN;
     if CODE(LAST + 1) = 5 then
       ST_VAL := AUT_DATA.MAX;
     if CODE(LAST + 1) = 6 then
       ST_VAL := AUT_DATA.MIN;
     if CODE(LAST + 1) = 7 then
       ST_VAL := AUT_DATA.SDV
  }
```

Fig. 2 Path-decision algorithm

2. 관리기의 운영

경로 결정 알고리즘에 의하여 결정된 선택값들에 의해 메타 데이타베이스와 원시 데이타베이스와 요약 데이타베이스에서 필요한 값을 편리하게 찾아가는 시스템의 경로를 3가지 경우로 나누어 살펴보면 아래와 같다.

1) CASE 1의 경로

플래그 값이 1인 경우로 시스템에 의해 제공된 구조에 의해 자동 생성된 요약 데이타베이스로

부터 검색한다. 이때, ST_VAL 변수의 값에 의해 기술통계량의 종류가 결정되어 이미 시스템에서 계산되어 저장된 값이 요약 데이타베이스로부터 검색된다.

즉, 이와같이 통계량들을 통합적으로 저장함으로써 이러한 통계량들을 요구하는 질의에 대해서 매번 원시 데이타베이스를 접근하여 계산하지 않고 직접 요약 데이타만을 접근하므로 빠른 응답시간을 보장한다. 또한 CHECK 변수의 값에 의해 기본 요약 테이블 또는 다중 요약 테이블 중 요약값의 출력 형태를 결정하여 요약 테이블 출력모듈로 부터 시각적인 형태의 2차원 테이블로 출력하도록 한다.

이와같은 경로운영 과정은 Fig.3과 같다.

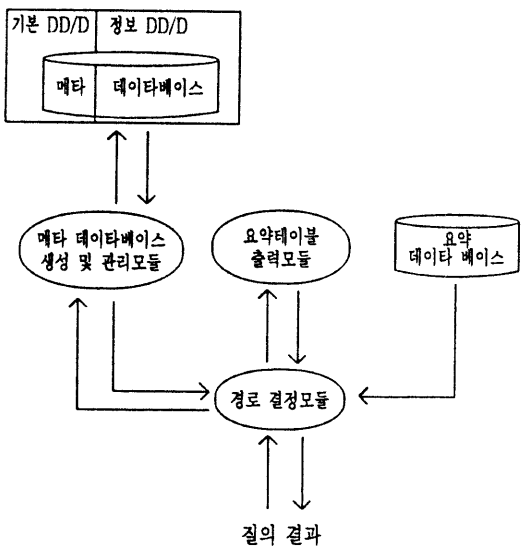

Fig. 3 Path of CASE 1

2) CASE 2의 경로

플래그 값이 0 인 경우로 원시 데이타베이스에서 직접 검색한다.

Fig. 4와 같은 경로로 운영된다.

3) CASE 3의 경로

플래그값이 -1인 경우로 사용자가 자동 생성된 요약 데이타베이스와는 다른 구조의 요약세트를 요구할때 원시 데이타베이스의 요약 데이타 세트 갱신모듈로 부터 새로운 요약세트를 생성하여 제공하며 또한 요약 데이타베이스를 갱신하다.

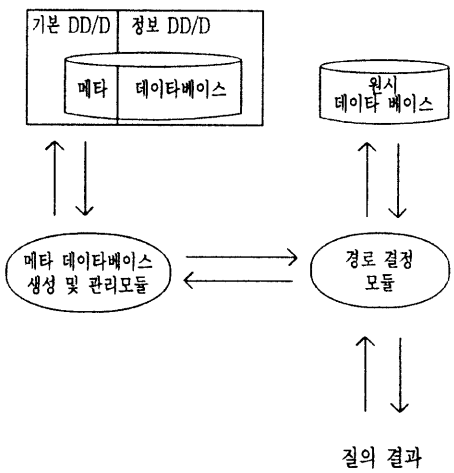

Fig. 4 Path of CASE 2

단, LEE 변수의 값이 0이므로, 이미 다른 사용자에 의해 질의가 요구되어 계산되어 저장된 요약 세트의 값은 요약 데이타베이스에서 검색만 한다. (Fig.5의 --> 표시) 또한 CASE 1의 경로와 마찬가지로 ST_VAL 변수의 값에 의해 기술통계량의 종류가 결정되고, CHECK변수의 값에 의해 기본 요약 테이블 또는 다중 요약 테이블중 요약값의 출력형태를 결정한다.

이 경로의 운영은 Fig. 5와 같다.

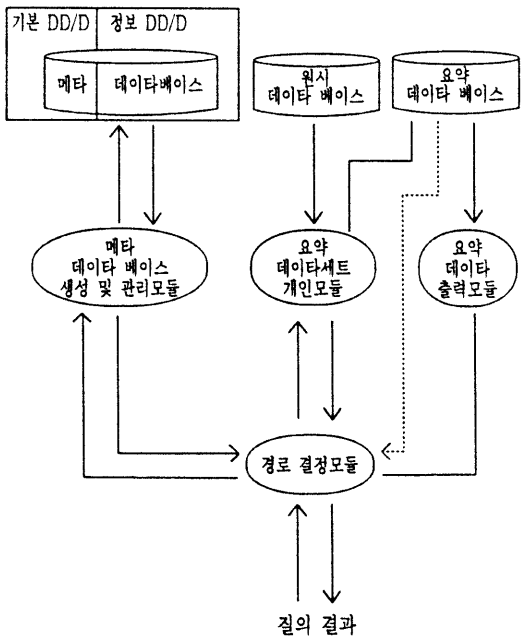

Fig. 5 Path of CASE 3

IV. 결 론

통계 데이타베이스의 효율적 이용을 위해 통계 분석을 위한 사용자 요구에 의해 요약정보를 제공하기 위하여 방대한 양의 원시 데이타베이스에 접근하는 부담을 줄이는 노력이 필요하다. 이를 위해서는 데이타베이스에 통계량을 미리 계산하여 저장함으로써 사용자에게 빠른 응답시간내에 통계 정보를 제공하려는 요약 데이타베이스와 이의 효율적인 관리와 사용의 편리를 위한 메타 데이타베이스가 생성되고 관리 되어야만 한다.

본 논문에서는 요약 데이타베이스를 효율적으로 이용한 통계 분석작업의 환경과 사용자의 편이성을 지원하기 위하여, 계층구조 형태인 데이타 사전/디렉토리의 형태로 독립적으로 운영되는 메타 데이타베이스와 메타 데이타의 한글처리와 메뉴 형태의 입출력이 가능한 관리기의 설계 및 구현작업과 이를 이용한 운영 방법등이 제시 되었다.

그러나 이 시스템은 메타 데이타베이스의 보안을 위한 보안 메카니즘과 제공하고자 하는 통계 정보를 테이블 형태에서 벗어나 그래프 형태로 제공 할 수 있는 사용자 그래프 인터페이스에 관한 보완 작업이 요구되어 이에 관한 연구를 진행하고자 한다.

참 고 문 헌

1. Chin, F.Y., Ozsoyoglu, G., "Auditing and Inference Control in Statistical Databases", IEEE Tran. on Software Enginnering, Vol.8, Nov6, Nov. 1982.
2. Wertz, C.J.,"The Data Dictionary : Concepts and Uses". North-Holand QED information Sciences, Inc., 1986.
3. Su, Y.W.S., "Modeling Integrated Manufacturing Data with SAM*", IEEE Database Engineering 1986, pp. 34 - 49.
4. 안성옥, 서보환, 박세권, "농업 데이타베이스 구축의 효율적 방향 통계 데이타베이스적 관점에서", 한국농촌경제 연구소, 연구보고, 148 - 6, 1988.12.
5. 안성옥, 황종선, "통계 데이타베이스의 효율적 처리를 위한 DSM 시스템의 설계 및 구현에 관한 연구", 한국정보과학회, '89년 봄 학술발표논문집, 1989.
6. Kacmar, C., Leggett, J., Schnase, J. L. and C. Boyle, ''Data Management Facilities of Existing Hypertext Systems", Hypertext Research Lab. Texas A & M Univ., TAMU 88-018, Sept.1988
7. Waterworth, J., ''Multimedia Technology and Applications", Ellis Horwood Ltd., 1991, pp 33-59
8. Mahapara, P.k. and J.F.Courtney, "Research Issues in Hypertext & Hypermedia for Business Application", ACM SIGBIT DATABASE 23.4, Fall 1992, pp 10-18