

## 農業統計 데이터 베이스의 設計

朴 世 權\*  
徐 輔 環\*\*  
安 成 玉\*\*\*

- I. 序 論
- II. 統計데이터 베이스의 特性 및 現況分析
- III. 프로토타입의 통계 DBMS 具現
- IV. 農業데이터 베이스 構築을 위한 發展方向

### I. 序 論

최근에 들어 우리 社會는 각종 分野에서의 급진적인 發展에 따라 內外的으로 유통되는 情報・資料가 質的, 量的으로 급속하게 팽창하고 있으며 거의 모든 분야에서 적절한 情報를 신속히 수집하고 용도에 맞게 가공하여 적시에 제공해주는 소위, 情報管理의 시스템적 체계가 절실히 요구되고 있다. 이러한 時代的 추세는 우리의 農業分野에서도 예외가 아니다. 근년에 들어 政府 및 政策研究 기관들에 의하여 農業部分에 있어서 현안으로 부각되는 여러 문제에 대하여 政策方向이나 提案에 대한 토의가 빈번하게 이루어지고 있으나, 대체로 그 問題認識 및 提案의 結果에 대한 종합적 평가가 불충분한 상태에서

이루어지는 것이 대부분인바, 이는 주요 관심사에 대한 政策의 評價가 단순한 경험적 판단만으로는 불가능할 정도로 여러 부분이 상호 有機的이고 복잡하게 構成되어 있기 때문이다. 따라서 농업분야에 있어서 당면한 문제를 生産, 消費, 所得, 需給, 價格, 輸出入, 農業構造, 農村福祉, 財政, 金融 등의 여러 시각에서 종합적으로 검토 가능키 위하여는 무엇보다도 시스템적 의사결정 체제가 필요하며, 이를 위하여는 컴퓨터를 이용한 情報의 수집, 분석, 가공 및 분산의 체계구축이 시급한 선결과제라 할 수 있다. 이와 같은 시대적 요청에 따라 農業情報의 시스템적 管理를 위하여 농림수산 소프트웨어 하우스에서는 각종 農業關聯情報를 체계적으로 수립, 저장, 검색하는 데이터 베이스 시스템과 이를 이용하여 2차, 3차 加工情報를 算出해내는 意思決定 모델 베이스시스템으로 代別되는 統合農業情報 시스템의 基礎研究를 수행한 바 있다.

本研究는 기초연구에서 설정된 設計方向에 따라 농업데이터 베이스를 구체화하기 위하여, 農業데이터 베이스가 주로 統計分析을 위해 사용되는 점을 감안하여, 統計데이터의 특성 즉 數

\*研究委員.  
\*\*責任研究員.  
\*\*\*安陽工業專門大學 教授.

值性(numericalness)과 大量性(largeness)을 효과적으로 活用할 수 있는 데이터 모델을 설정하고 이를 우리의 農業 데이터 베이스에 적용 가능성을 제시하였다. 즉, 아직까지 우리 국내에 실용화 된 바 없으나 일반적인 데이터 管理機能 뿐 아니라, 統計分析, 要約情報管理 등의 기능을 갖는 통계데이터 베이스의 구축이 가능성을 프로토타입의 농업통계 데이터 베이스의 구축을 통하여 실험하고 이에 근거한 구체적인 發展方案을 提示코자 한다.

## Ⅱ. 統計데이터 베이스의 特性 및 現況分析

### 1. 데이터 베이스와 데이터 베이스 관리 시스템

데이터 베이스는 원래 같은 데이터가 여러 응용에 중복되어 사용될 수 있다는 多目的 共用的 개념에서 출발하였다. 따라서 데이터 베이스란 어느 한 組織에서 多數의 應用시스템이 공동으로 사용하기 위하여 저장된 運營데이터의 集合이라고 정의할 수 있다. 이러한 데이터 베이스의 特徵으로서는 실시간(realtime) 접근(access), 계속적인 변화, 同時共有, 內容에 의한 参照 등을 들 수 있다. 전통적 화일시스템에서는 論理的 화일구조와 저장된 物理的 화일構造간에 1:1의 대응을(mapping) 요구하게 되고, 또한 데이터의 物理的 構造에 대한 가정과 그에 대한 接近方法이 데이터를 使用하는 응용프로그램에 具現되어 있어야 하므로 결국 데이터의 중복성과 데이터의 종속성, 데이터의 非共有性 등의 문제점이 있다. 이러한 전통적인 화일시스템의

短點을 보완하고 使用者 입장에서 보다 편리하게 資料處理가 가능한 중앙집중식 관리체제의 要求에서 출발한 방법이 데이터 베이스 관리시스템(DBMS)이며 이러한 DBMS는 應用프로그램과 데이터의 중재자로서 모든 應用프로그램들이 데이터 베이스를 共用할 수 있도록 管理한다. DBMS를 이용하는 應用프로그램은 데이터 베이스 전체의 生成, 접근방법, 조작절차, 保安, 物理的 構造 등에 관계없이 원하는 데이터와 處理作業만을 DBMS에 요구하면 된다. DBMS는 데이터 구조의 정의기능(definition facility), 데이터의 조작기능(manipulation facility)과 제어기능(control facility) 등을 수행한다.

데이터 베이스 관리시스템은 資料를 관리하기 편하도록 데이터 모델이라는 데이터를 저장하는 概念的인 도구를 제공한다. 데이터 모델은 데이터가 論理的으로 조직될 수 있는 樣式이라고 말할 수 있으며 代表的인 데이터 모델로는 관계 데이터 모델(relational data model), 계층 데이터 모델(hierarchical data model), 네트워크 데이터 모델(network data model) 등의 3가지가 있다. 일반적으로 한 데이터 베이스 관리시스템에서는 어떤 現實 世界를 表現하기 위해 하나의 데이터 모델을 지원하게 되며 이러한 3가지 모델간의 주요 차이점은 데이터 要素간의 관계(relation)를 표현하는 방식에 있다.

관계 데이터 모델(relational data model)은 使用者에게 데이터 베이스 構造를 테이블 형태로 기술해 주며, <그림 1, 2, 3>은 식량작물의 植付面積과 生産量에 대한 情報가 이 모델에서 표현되는 例를 보여 준다. 그림에서 보는 바와 같이 관계 데이터 모델에서는 모든 데이터들 사이에 存在하는 관계(relation)를 테이블 형태로 표현한다. 이러한 관계 데이터 모델은 論理的

관계데이터 모델의 예<그림 1, 2, 3>

그림 1 생 산 지

| 생산지번호 | 생 산 지 | 식량작물<br>식부면적 |
|-------|-------|--------------|
| S 1   | 경 기 도 | 208398       |
| S 2   | 강 원 도 | 114095       |
| S 3   | 충 청 도 | 321673       |
| S 4   | 전 라 도 | 574031       |
| S 5   | 경 상 도 | 507718       |

그림 2 작물별 식부면적

| 작물번호 | 작 물 명 | 식부면적    |
|------|-------|---------|
| A 1  | 미 곡   | 1262324 |
| A 2  | 맥 류   | 207635  |
| A 3  | 잡 곡   | 46085   |
| A 4  | 두 류   | 211912  |
| A 5  | 서 류   | 47376   |

그림 3 생산지별 작물별 식물작물 생산량

| 생산지번호 | 작 물 번 호 | 생 산 량  |
|-------|---------|--------|
| S 1   | A 1     | 767278 |
| S 1   | A 2     | 2391   |
| S 1   | A 3     | 4988   |
| S 2   | A 4     | 254878 |
| S 2   | A 5     | 45797  |

構造가 간단해서 이해하기 쉽고 데이터의 意味와 그것에 대한 操作이 明確한, 使用者 위주의 모델이다.

계층데이터 모델(hierarchical data model)은 데이터 베이스의 論理的 構造가 계층(hierarchy)性質을 갖는 트리(tree) 형태로 表現되는 것을 말하며, 예로서 <그림 4>와 같은 계층구조로써 農家經濟데이터를 表現할 수 있다. 이 계층 데이터 모델은 使用者가 데이터 베이스의 계층구조를 알아야 된다는 短點을 가지고 있지만 컴퓨터의 中央處理 장치의 부담을 적게 하는 長點을 가지고 있다.

네트워크 데이터 모델(network data model)은 데이터 베이스의 論理的 構造를 表現하는 資料構造가 制限을 받지 않고 一般的인 그래프의

그림 4 계층 데이터 모델의 예

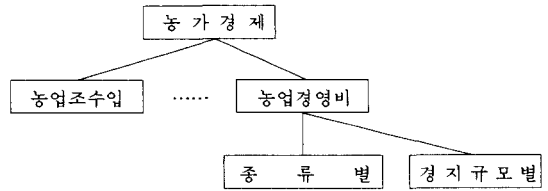
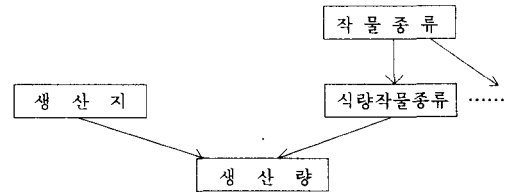


그림 5 네트워크 데이터 모델의 예



형태를 갖는다. <그림 5>는 食糧作物의 生産地와 作物種類別 生産量이 네트워크 관계모델로 어떻게 表現되는지를 보여준다. 네트워크 데이터 모델은 계층데이터 모델보다 그 구조가 복잡하고 이들을 처리하는 조작어가 복잡해지나, 트리라는 構造上的 制約이 없기 때문에 데이터의 관계를 表現하는데 融通성이 있다.

## 2. 統計데이터 베이스

데이터 베이스의 일반적인 性格外에도 統計데이터 베이스는 추가적으로 다음과 같은 特徵들이 부가됨으로써 設計 및 具現上的 어려움을 더해 주고 있다.

첫째, 論理的 接近에 어려움을 주고 저장, 처리의 效率性 提高가 특별히 요구되는 데이터의 數值性和 大量性.

둘째, 통계 데이터의 특성상 數值데이터가 어떤 實體에 대한 측정데이터와 이에 意味를 부여하는 메타(meta)데이터로 구성되는 分離性. 예로서 <그림 6>은 테이블 형태로 나타낸 作物生

그림 6 범주속성과 요약속성의 예

| 매개 데이터(parmeta) |     |    |     | 측정데이터<br>(measured data) |       |
|-----------------|-----|----|-----|--------------------------|-------|
| 품 목             | 도   | 년  | 계 절 | 생 산                      | 소 비   |
| 쌀               | 경기도 | 80 | 봄   | —                        | 700   |
|                 |     |    | 여름  | —                        | 900   |
|                 |     |    | 가을  | 1,200                    | 1,000 |
|                 |     |    | 겨울  | —                        | 900   |
| 보리              | 강원도 | 81 | 봄   | .....                    | ..... |
|                 |     |    | 여름  | .....                    | ..... |
|                 |     |    | 가을  | .....                    | ..... |
|                 |     |    | 겨울  | .....                    | ..... |

범주속성(category attribute) ← → 요약속성(summary attribute)

産/消費데이터이다.

처음 4개의 속성들(品目, 道, 年, 季節)은 매개 데이터이고 나머지 2개의 속성들(生産, 消費)은 測定 데이터이다. 매개데이터를 위한 속성들은 그것들이 測定데이터의 범주를 포함하기 때문에 범주속성(category attributes)이라 하고 測定 데이터를 위한 속성들은 그것들이 統計의 要約/分析을 위해 使用되기 때문에 요약속성(summary attribute)이라 한다.

셋째, 공백값에 대한 效果的인 저장과 액세스를 위한 압축기법이 요구되는 데이터의 散在性.

넷째, 데이터의 大量性으로 인한 원시데이터의 接近時 時間的, 經濟的 非效率性 제거를 위한 要約데이터의 必要性.

다섯째, 다행스럼제도 初期作業後, 갱신사항이 거의 없는 시계열 데이터상의 安定性 등을 들 수 있다.

### 3. 統計데이터 베이스 관리시스템

이와 같은 特性을 갖는 統計데이터 베이스의 管理시스템을 구축하기 위해서는 첫째, 기존의 統計패키지에서 데이터 조작 및 화일조작 기능을 보강 둘째, 기존의 범용 데이터 베이스 관리 시스템에서 데이터조작, 수치처리, 分析등의 機能 補強을 통한 統計데이터 베이스 管理시스템으로의 擴張 셋째, 전적으로 새로운 소프트웨어로써 독자적인 統計데이터 베이스 管理시스템을 개발하는 方法을 고려할 수 있다. 統計데이터 베이스를 이용하는 使用者는 일반적으로 데이터에 대한 管理機能뿐 아니라, 統計分析, 要約情報의 管理 등 統計的 機能까지 요구한다. 따라서 통계 DBMS는 데이터를 모델링하고 저장, 조작하는 機能 뿐 아니라 統計的 分析方法을 적용시킬 수 있는 機能도 포함하도록 요구되며 이러한 시스템을 일반 데이터 베이스 관리시스템과 구별하여 統計데이터 베이스 管理시스템(Statistical Database Management System:SDBMS)이라 한다.

### 4. 統計데이터 베이스의 現況分析

#### 가. 범용데이터 베이스의 擴張 補完 研究

기존의 상용 데이터 모델을 사용하여 확장된 시스템으로 데이터 베이스를 구축한 대표적인 시스템의 特徵은 다음과 같다.

① TPLS(Table Producing Language Statistics)

질의어인 TPL의 要約데이터를 표현하는 基本 機能은 TREE構造의 화일에서부터 要約데이

을 정의하는 것이다. 이러한 機能은 TREE構造의 복잡한 接近경로를 미리 뷰(view)로서 정의하여 수행된다. 따라서 使用者는 TREE構造의 接近경로를 매 질의어마다 明示하지 않아도 되는 利點이 있으나, 뷰 정의의 自體가 상당히 복잡하여 理解하기 힘들고, 뷰를 정의하는 使用者는 데이터 화일의 TREE構造 및 接近 경로에 대하여 사전에 알고 있어야 하는 제약이 있다. TPL의 要約테이블 정의기능은 매우 다양하고 강력하나, 가장 큰 短點은 배치형태로 수행되는 단독시스템이라는 점이다.

[2] HSDB(Hierachical Summary Data Base)

HSDB는 원시데이터로부터 원시要約테이블을 정의할 수 있도록 구성되어 있다. EST(Elementary Summary Table)가 하나의 속성(attribute)에 대해서만 要約데이터를 정의하며, 간단한 분류화면만을 제공하므로 오퍼레이션이 단순하다. 또한 EST에 정의되지 않은 要約데이터는 데이터사전/디렉토리(data dictionary/directory)에 저장하는데 이는 要約데이터 管理의 이중구조를 가져와 使用者로 하여금 要約데이터의 管理에 일관성을 어렵게 하는 단점이 있다.

[3] SRT(Statistical Relational Table)

要約테이블인 SRT는 독자적인 저장구조를 통하여(K-map 형태로 명칭) 저장되어, 부가적인 저장기법 및 접근기법이 필요하며 SRT를 管理하는 QBSRT라는 질의언어를 제공한다.

[4] STRAND(Simple to Read and Understand)

STRAND는 ER모델링技法에 의한 要約데이터를 管理하는 질의언어로서 오퍼레이션 要約化

(summarization)연산을 이용하여 要約데이터를 저장하는 릴레이션을 만들 수 있다. 그러나 이러한 作業이 수행되기 위해서는 데이터 베이스 管理者(Data Base Administrator : DBA)가 미리 스키마에 각 統計量을 구하는 프로시유어를 정의해야만 하므로 使用者立場에서는 制限要素가 많다. 또한 이 시스템은 統計情報를 관리하는 基本構造로써 릴레이션을 사용하므로, 要約테이블을 利用하는 方法에 비하여 要約데이터 管理가 개념적 수준에서 直接的으로 이루어지지 못하고, 內部的 수준에서 이루어진다. 이 질의어는 INGRES 데이터 베이스 管理시스템에 전위시스템으로 具現되어 있는데 使用者가 接近경로를 명시해야 하는 계층적 構造의 言語라는 短點도 있다.

[5] PST(Primitive Summary Table) 및

STBE(Summary Table by Example)

릴레이셔널 모델을 근간으로 하고 있으나 要約데이터를 管理하는 獨自的 基本構造를 가지고 있어 기존시스템의 擴張 또는 獨自開發의 研究分類의 중간위치에 있다고 볼 수 있다. STBE는 화면을 이용하여 질의어를 처리하는 ABE(Aggregate—By—Example)와 유사하나, PST라 불리는 要約테이블을 要約데이터를 管理하는 基本構造로써 사용하는 점이 다르다. 따라서 DBMS는 要約테이블을 위한 부가적 저장구조가 필요없으며, 질의어 處理를 더욱 效果的으로 수행할 수 있게 한다. 그러나 STBE는 要約테이블이 管理하는 要約데이터를 반드시 하나의 어트리뷰트에 限定하여야 하는 短點을 갖고 있다.

### 나. 統計패키지의 擴張 補充 研究

기존의 統計패키지에 데이터의 조작 및 화일 조작 기능을 보강하는 方法이 있으며 현재 많이 쓰고 있는 統計패키지들은 SAS, Minitab, SPSS, BMDP 등인데, 이들 統計패키지들의 統計分析 機能은 광범위하고 매우 강력하나 대부분의 패키지들은 데이터 관리기능이 극히 制限的으로 提供되고 있으며 특히 센서스 데이터 베이스와 같은 방대한 規模의 데이터세트에 대한 分析作業을 수행할 경우 統計패키지들은 매우 저하된 性能을 나타내고 있다. 이를 保完하는 具體적인 研究에 대한 진행은 아직 없으며, 다만 SAS/DB2와 같이 統計패키지와 관계형데이터 베이스간에 상호 데이터를 공유하는 式의 接近方法은 現在 연구되고 있는 중이다.

### 다. 獨自 開發 研究

獨自의인 統計데이터 베이스 管理시스템의 開發에 관한 研究는 美國의 버클리 大學의 LBL에서 개발한 SUBJECT, SEEDIS, GUIDE와 SAM\*, RAM과 SSDB 등이 있으며 이들 시스템의 주요 특징은 다음과 같다.

#### [1] SUBJECT

SUBJECT시스템은 Cluster노드와 Cross-product노드들의 계층 구조에 의해 要約세트의 概念을 세우고, 범주와 요약속성으로 데이터를 구분한다. 要約데이터와 原始데이터의 行列表示를 위해 Cluster abstraction과 Cross-product abstraction 등의 연산을 사용한다. Cross-product abstraction은 3次元行列의 데이터를 사용하여 要約데이터를 나타내는 데이터 모델링 도

구이며, Cluster abstraction은 복잡한 범주 속성의 계층구조와 요약속성의 수집을 위해 使用되는 도구이다.

#### [2] GUIDE(Graphical User Interface for D/B Exploration)

SUBJECT와 SEEDIS의 使用者 인터페이스로 사용되는 그래픽 도구이며 SUBJECT의 디렉토리 역할, 도움메세지, 위치의 Zooming기능 및 중간결과를 이용한 部分的 질의기능 등의 특징을 포함한다.

#### [3] SAM\*

7개의 연관(association)형태를 가지고 세트, 行例, 벡터, 시계열 등과 같은 복잡한 데이터 타입들을 지원하며, 실제 具現은 G(generalized) 관계를 使用한다.

#### [4] RAM

Infological framework로 對象(object)에 관한 초기 메시지들을 사용하며, 統計調查 등에서와 같이 복수의 대상을 구현하기 위하여 n-cube의 박스構造가 使用된다.

#### [5] SSDB(Systems for Statistical Database)

SSDB는 개념모델로써 DA(Data Abstraction)모델을 변형한 HODM(Heterogenous Operational Data Model)을 利用하며 질의어로는 STBE를 使用한다.

### Ⅲ. 프로토타입의 통계 DBMS 具現

본절에서는 要約데이터 베이스와 메타 데이터 베이스의 효율적 管理를 위하여 데이터사전/디렉토리가 첨부된 통계데이터 베이스 관리시스템 《이하 S-M 시스템이라 칭하고, S-M은 통계데이터 베이스관리(Statistical Data Base Management)의 약자임과 동시에, 要約(Summary)과 메타(Meta) 데이터 베이스의 약자임, 중복의 意味를 가진다》을 프로토타입으로 설계 및 구현된 실험적 시스템을 提示한다.

#### 1. S-M 시스템의 基本構造

提案된 S-M 시스템이 관리하는 원시 데이터

베이스와 要約데이터 베이스는 提示된 그래픽 형태의 理論모델에 의해 VSAM(Virtual storage Access Method) 화일의 저장구조를 이용하여 설계 구현되었다. S-M 시스템은 <그림 7>과 같이 5개의 모듈로 구성되는데 각 모듈이 수행하는 기본적인 기능은 다음과 같다.

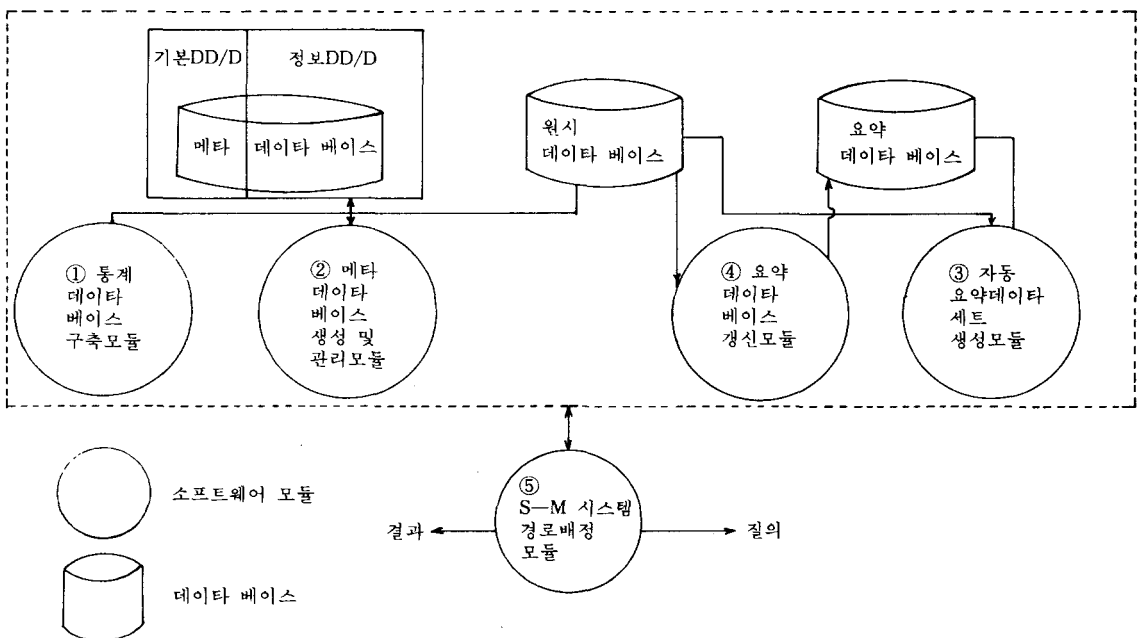
#### ① 統計데이터 베이스 구축모듈

원시 수치데이터를 論理모델의 계층구조에 따라서 2次元行列로 VSAM화일에 저장하여 원시 데이터 베이스를 구축한다.

#### ② 메타데이터 베이스 生成 및 管理모듈

메타데이터를 效率的으로 처리할 수 있도록 데이터사전/디렉토리(DD/D) 형태의 메타情報를 저장하고 使用者의 요구에 따라 필요한 메타 데이터情報를 出力한다.

그림 7 S-M 시스템의 基本구조



[3] 自動要約데이터세트生成모듈

原始데이터 베이스로부터 論理모델의 계층구조에 따라서 要約데이터 베이스를 自動生成한다.

[4] 要約데이터세트갱신모듈

시스템에 의하여 自動적으로 생성된 要約데이터 세트와는 다른 特定構造의 要約세트를 요구하는 使用者에게 原始데이터 베이스로부터 새로운 要約세트를 형성하여 提供한다.

[5] S-M시스템 경로배정모듈

使用者의 질의에 따라 原始데이터 베이스 또는 要約데이터 베이스로부터의 검색경로를 決定한다.

2. 論理모델의 그래프表現  
(Graphic Representation)

가. 데이터 테이블

統計데이터 베이스의 特性과 問題點에서 본바와 같이 統計데이터인 農業組收入데이터를 그대로 D/B에 저장하기에는 <그림 8>과 같이 범주 속성의 반복성 問題로 기억장소와 액세스의 非效率性을 수반하게 된다. 이를 解決키 위하여 <그림 9>와 같은 2次元行列로 추정 데이터를 저장하고, 매개데이터는 메타데이터 베이스에 단 1번만 저장하면 배열 線型化 方法에 의해 빠른 속도로 액세스가 가능하다. 여기서 각 行과 列의 데이터構成은 각각 가장 적절한 계층구조에 의해 순서가 決定되어 표현된다.

나. 시맨틱스(semantics)를 위한 그래프 表現

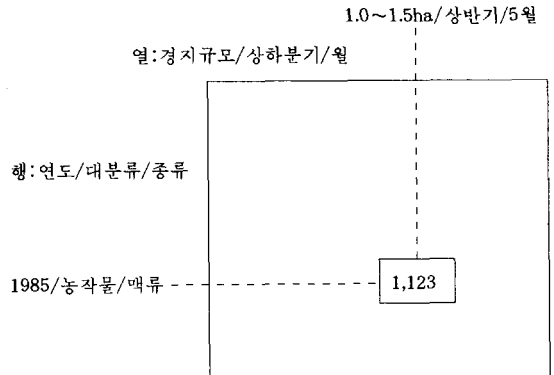
의미개념을 그래프 형태로 表現키 위하여

그림 8 농업 조수입 원시 데이터

| 연도   | 맥 류  | 종 류 | 경지규모    | 상하반기 | 월 별 | 조수입   |
|------|------|-----|---------|------|-----|-------|
| 1984 | 농작물  | 미 곡 | 0.5ha미만 | 상반기  | 1   | 91786 |
| "    | "    | "   | "       | "    | 6   | ⋮     |
| "    | "    | "   | "       | 하반기  | 7   | ⋮     |
| "    | "    | "   | ⋮       | "    | ⋮   | ⋮     |
| "    | "    | "   | 0.5~1.0 | 상반기  | 1   | 70042 |
| "    | "    | "   | ⋮       | "    | 6   | 93636 |
| "    | "    | "   | ⋮       | 하반기  | 7   | ⋮     |
| "    | ⋮    | 맥 류 | ⋮       | ⋮    | ⋮   | ⋮     |
| "    | ⋮    | ⋮   | ⋮       | ⋮    | 12  | ⋮     |
| ⋮    | 비농작물 | 축산물 | 0.5ha미만 | ⋮    | ⋮   | ⋮     |
| ⋮    | ⋮    | ⋮   | ⋮       | ⋮    | ⋮   | ⋮     |
| 1985 | 농작물  | 미 곡 | ⋮       | ⋮    | ⋮   | ⋮     |
| ⋮    | ⋮    | ⋮   | ⋮       | ⋮    | ⋮   | ⋮     |

← 매개데이터 (범주속성) →      → 추정데이터 (요약속성) ←

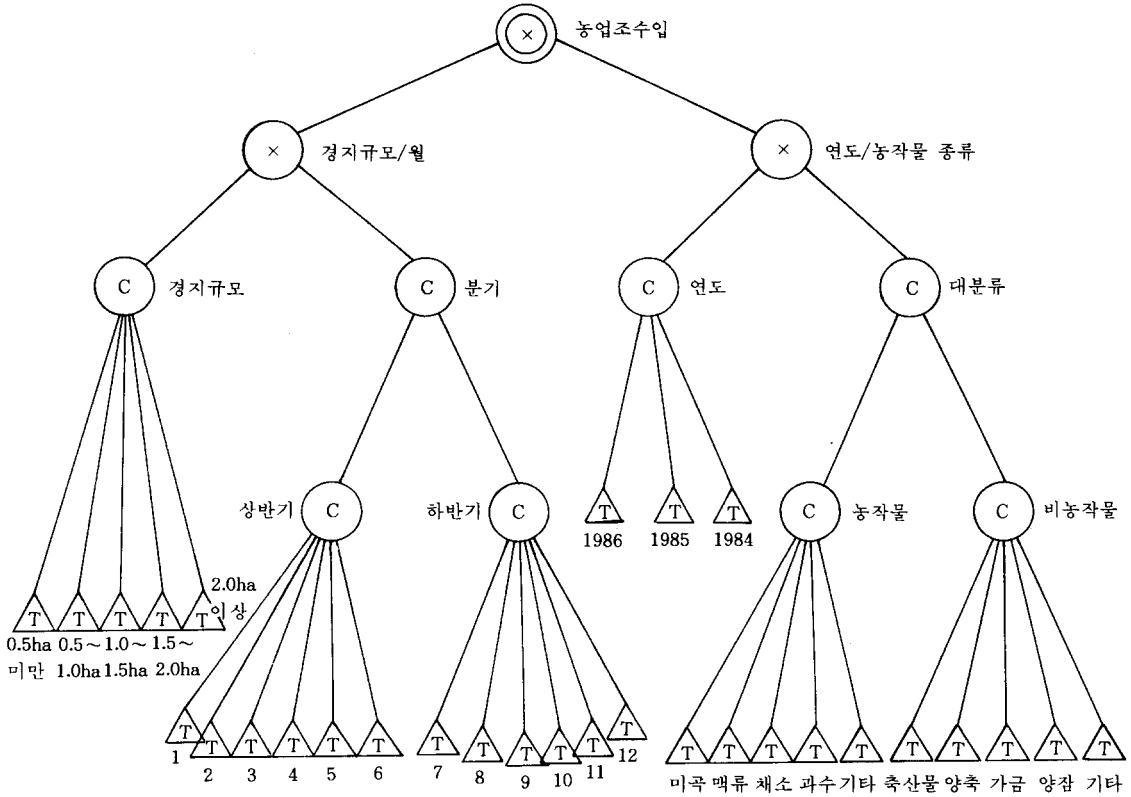
그림 9 농업조수입을 위한 행렬표시(matrix representation)



LBL研究所에서 개발한 SUBJECT시스템의 方法論을 적용하여 理論모델을 계층구조의 그래프表現으로 제시해 보았다. 노드에는 크로스 프로덕트(cross-product)노드와 클러스터(cluster)노드의 2종류가 있으며, 이 노드들이 연결되어 Directed acyclic graph를 구성하고 있다. <그림



그림 10 농업조수입의 시맨틱스를 위한 그래프 表現



8)과 <그림 9)의 農業組收入 데이터가 <그림 10)과 같은 그래프 형태로 表現된다.

들의 모든 값들과의 조합에 의하여 그 값을 形成한다.

1] 크로스 프로덕트 노드(X-노드)

統計데이터 베이스의 범주속성의 多次元性質을 表記한 것으로 복잡성을 감소시키기 위하여 다계층으로 表示한다. 예를 들면, 7원소를 가진 한 개의 크로스 프로덕트 노드가 3개와 4개의 원소를 가진 2개의 크로스 프로덕트 노드로 각각 表現한 뒤, 2개의 크로스 프로덕트노드에 의해 연결되어 나타내진다.C-노드와 달리 選擇되지 않은 X-노드의 모든 가지들은 에그리게이션된다. 결국, 에그리게이션을 위하여 선택된 X-노드들은 선택된 값들과 選擇되지 않은 가지

2] 클러스터 노드(C-노드)

데이터 공간의 차원에서 원소의 갯수가 너무 많을 때 그것을 복수의 수준으로 表記하여 한번에 源素의 작은 부분만을 처리하려는 속성을 表記한 것이다. 질의를 위한 C-노드의 意味는 X-노드와는 달리 질의에 포함된 가지들만 에그리게이션 되고 選擇되지 않은 가지들은 에그리게이션 되지 않는다. 실제적으로 C-노드들은 복잡한 범주 속성의 계층구조와 要約속성의 수집을 위해 사용되며, X-노드들은 범주속성의 合成키들을 表現하기 위해 사용된다.

### 3. 메타데이터 베이스의 設計

#### 가. 메타데이터의 定義 및 特性

메타데이터란 데이터에 관한 情報을 총칭하는 것으로 데이터의 內容과 組織에 관한 체계적이고 說明的인 情報의 집합이다. 수치데이터를 기준으로 볼 때는 범주데이터도 個別수치 데이터의 性格을 규정하는 메타데이터로 취급할 수 있다. 메타데이터가 포함하고 있는 情報는 명칭(name), 약어(alias), 분류(label), 說明(discription), 保安事項(security specification), 論理構造(logical structure), 접근경로(access path), 연결명세(linkage specification), 處理節次(processing procedure) 등이다. 메타데이터의 特性은 다음과 같다.

○ 메타데이터의 데이터 타입: 메타데이터는 標準的이고 固定된 길이의 데이터 타입들 뿐만 아니라, 가변길이 텍스트와 같은 擴張되고 복잡한 데이터 타입(예: 도서목록)들도 포함한다.

○ 메타데이터 構造: 융통성있는 확장가능 계층 데이터構造(open-ended hierachical data structure)이며 각 노드들은 메타데이터 엔티티들의 모임(cluster)을 表示하고, 이 노드들은 연결(linkage)이나 연합(association)을 表現하는 선(line)에 의하여 構造를 형성한다.

○ 메타데이터의 액세스형태: 대부분의 데이터 베이스에서 使用者는 데이터의 選擇을 위해 메타데이터의 일부분만을 자주 사용한다.

○ 메타데이터 갱신: 메타데이터값의 갱신은 새로운 데이터 베이스가 生成될 때 그에 따른 메타데이터 값들이 첨가되며, 또한 存在하는 데이터 베이스의 構造, 화일 또는 略語, 첨가 등의 변화에 따라 갱신된다.

#### 나. 메타데이터 베이스의 管理와 S—M시스템의 運營

##### ① 使用者 入力

메타데이터 베이스를 통하여 해당 화일을 찾아 그래프 형태의 構造를 보고 검색하고자 한 데이터에 관련된 명칭들의 코드(表 1참조)를 찾아 입력시킨다.

〈예제1〉

1984年, 非農作物 가금을 사용하는 경지구모 1.5~2.0ha의 농가가 6월 중에 얻은 조수입의 값을 원하면 메타데이터 베이스를 통하여 다음과 같이 해당코드를 찾아 入力시킨다.

|    |     |    |      |      |   |
|----|-----|----|------|------|---|
| 1  | 2   | 3  | 4    | 1    | 6 |
| 연도 | 대분류 | 종류 | 경지구모 | 상하반기 | 월 |

##### ② 시스템의 運營

메타데이터 베이스를 활용하여 使用者가 질의 코드를 入力시키면 우선 S—M시스템의 경로배정 모듈에서 경로결정을 위해 플래그값을 生成한다. 이 모듈에 의해 결정된 플래그 값에 따른 S—M시스템의 경로를 살펴보면

○ A경로: 플래그값이 0인 경우로 원시데이터 베이스에서 직접 검색한다.

○ B경로: 플래그값이 1인 경우로 시스템에 의해 미리 제공된 계층구조에 의해 自動生成된 要約데이터 베이스로부터 검색한다.

○ C경로: 플래그값이 -1인 경우로 使用者가 自動生成된 要約데이터 베이스와는 다른 構造의 要約세트를 요구할 때 원시데이터 베이스로부터 새로운 要約세트를 生成하여 제공하며, 이 경우 要約데이터 베이스를 새롭게 갱신한다. 단, 다른 使用者에 의해 이미 저장된 값은 要約데이터 베이스에서 검색만 한다. 각 경로에 대한 運營은 〈그림 11〉과 같다.

表 1 농업 조수입 데이터의 코드화일

| 범주 데이터의 코드화 순서 및 범위 |         |         |         |         |         |         |
|---------------------|---------|---------|---------|---------|---------|---------|
| 코드화순서               | CODE(1) | CODE(2) | CODE(3) | CODE(4) | CODE(5) | CODE(6) |
| 해당데이터               | 연 도     | 대 분류    | 종 류     | 경지구모    | 상하분기    | 월       |
| 범 위                 | 0~3     | 0~2     | 0~5     | 0~5     | 0~2     | 0~6     |

| 범주 데이터의 코드값                |                       |         |            |           |             |         |
|----------------------------|-----------------------|---------|------------|-----------|-------------|---------|
| 범 주 데 이 타                  | 0                     |         | 1          |           | 2           |         |
| 연 도                        | 0                     |         | 1          |           | 2           |         |
| CODE(1)                    | 1984~1986의 요약값        |         | 1984       |           | 1985        |         |
| 대 분 류                      | 0                     |         | 1          |           | 2           |         |
| CODE(2)                    | 농작물과 농작물이외 수입을 합한 요약값 |         | 농작물        |           | 비농작물        |         |
| CODE(2)=1경우<br>종류, CODE(2) | 0                     | 1       | 2          | 3         | 4           | 5       |
|                            | 농작물요약값                | 미 곡     | 맥 류        | 채 소       | 과 수         | 기 타     |
| CODE(4)=2경우<br>종류 CODE(3)  | 0                     | 1       | 2          | 3         | 4           | 5       |
|                            | 비농작물요약값               | 축산물     | 양 축        | 가 금       | 양 잠         | 기 타     |
| CODE(4)<br>경 지 구 모         | 0                     | 1       | 2          | 3         | 4           | 5       |
|                            | 경지구모요약값               | 0.5ha미만 | 0.5~1.0ha  | 1.0~1.5ha | 1.5~2.0ha   | 2.0ha이상 |
| CODE(5)<br>상 하 분 기         | 0                     |         | 1          |           | 2           |         |
|                            | 상하분기를 합친 그 해의 요약 값    |         | 상반기(1월~6월) |           | 하반기(7월~12월) |         |
| CODE(5)=2경우<br>월, CODE(6)  | 0                     | 1       | 2          | 3         | 4           | 5       |
|                            | 상반기요약값                | 1월      | 2월         | 3월        | 4월          | 5월      |
| CODE(5)=2경우<br>월, CODE(6)  | 0                     | 1       | 2          | 3         | 4           | 5       |
|                            | 하반기요약값                | 7월      | 8월         | 9월        | 10월         | 11월     |

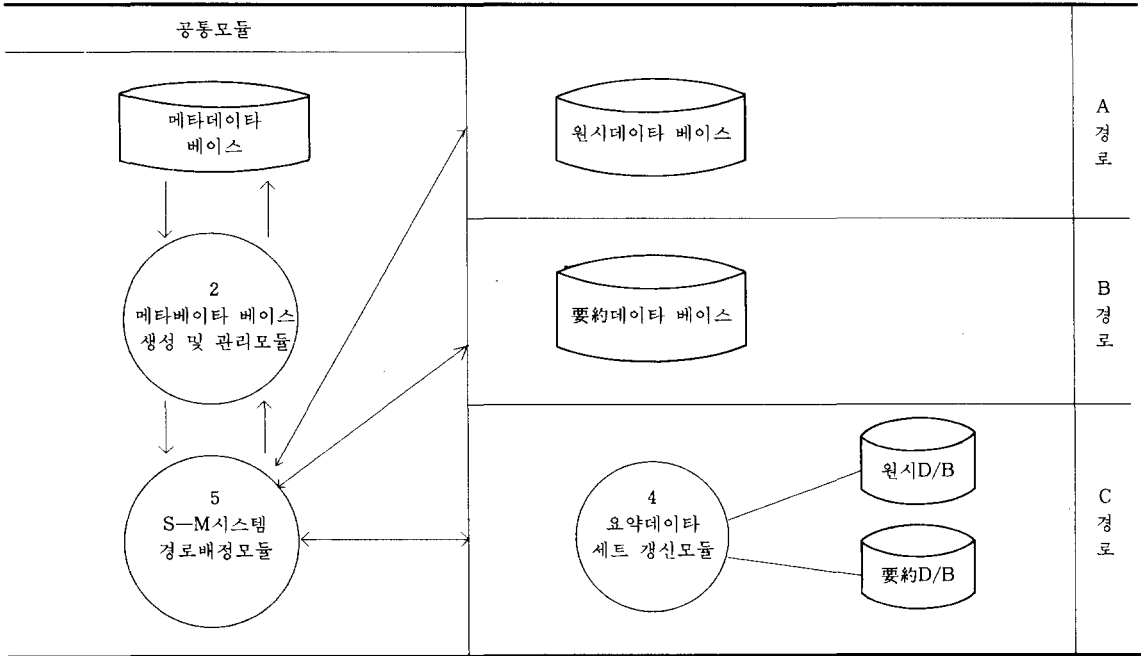
4. 요약데이터 베이스의 設計

가. 요약데이터 베이스의 必要性

일반적인 범용데이터 베이스 管理시스템은 使用者에게 저장된 데이터에 대한 풍부한 데이터 管理기능을 提供하는 반면 統計的 要約情報은 제공되지 않는다. 그러나 統計데이터 베이스에

서는 부분적으로 要約된 것만이 統計分析을 위 해 사용되는 경우가 대부분이다. 一般的으로 統計응용에 이용되는 데이터 베이스가 방대한 量의 資料를 저장하고 있고 統計分析作業이 반복적인 수행절차로 수행된다는 사실을 고려할 때 이러한 방대한 양의 원시데이터 베이스를 여러 번 반복적으로 접근하는 것은 막대한 費用을 가져오며, 응답성도 보장되지 않는다. 따라서, 아이

그림 11 S-M시스템의 경로



템의 수나 범주속성들을 제한함으로써 얻어지는 요약데이터 세트들의 모임인 요약데이터 베이스를 생성하여 한번 정의된 요약데이터에 대해서는 그 요약데이터를 접근하는 질의들이 원시 데이터를 접근하지 않고 결과를 바로 사용자에게 제공할 수 있도록 하는 기능을 통계 데이터 베이스의 효율적 이용이란 측면에서 매우 중요하게 평가된다. 그리하여 본시스템에서는 요약데이터 베이스를 생성하여 여러가지 통계량들을統合적으로 저장함으로써 이러한 통계량들을 요구하는 질의에 대해서는 요청 때마다 원시데이터 베이스를 접근하지 않고 직접 요약데이터 베이스만을 접근하므로 빠른 응답시간을 보장토록 하였다.

나. 요약데이터 베이스의 구성

우선 시스템에서 <그림 10>의 계층구조에 따라서 요약데이터를 자동 생성하여 요약데이터 베이스를 구성하는 과정으로 <그림 12>와 같은 경로가 운영된다. 이 과정은 원시데이터 베이스가 생성될 때 단, 한번만 원시데이터 베이스를 접근하여 구성에 맞는 통계량을 구하기 위하여 이용된다. 이러한 과정의 알고리즘을 나찌-슈나이더만 도표로 나타내면 <그림 13>과 같다. 또한 요약데이터 베이스가 생성될 때마다 그에 관한 정보가 메타데이터 베이스(데이터 사전/디렉토리의 형태)에 저장되어 요약데이터 베이스가 효율적으로 이용되게 해야 한다. 한편 S-M시스템의 운영시 사용자가 메타데이터 베이스의 검색 결과 필요한 요약

데이터 세트가 존재하지 않을 때 즉, 自動生成된 要約데이터 베이스와는 다른 構造의 要約세트를 요구할 때마다 원시데이터 베이스로부터 새로운 要約데이터 세트를 生成하여 제공하며, 이 경우 要約데이터 베이스를 갱신한다.

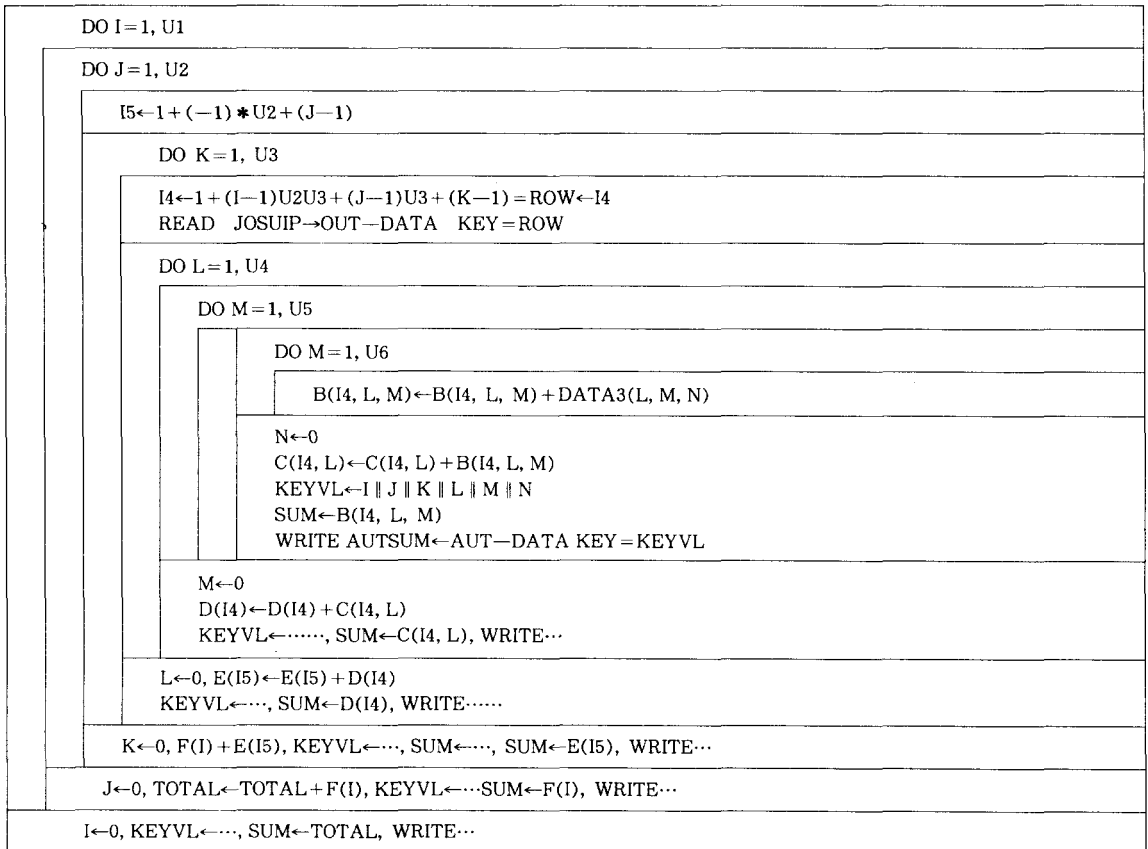
다. 要約데이터 베이스의 長點

要約데이터 베이스는 하나의 데이터分析 環境下에서 반복적으로 수행되는 統計分析을 지원하기 위하여 데이터 세트에 대해서 다양한 統計量을 미리 계산하여 저장한다. 그리하여 使用者는 원시데이터 베이스의 接近없이 分析作業을 수행할 수 있어 요구하는 統計量의 要約情報를 빠른 응답시간 내에 제공받을 수 있다. 이러한 要約데이터 베이스를 효율적으로 사용하기 위해서는 우선 가장 사용빈도가 높은 要約데이터 세트가 自動 生成되도록 적절한 계층구조가 先行되어야

그림 12 자동요약 데이터 세트 구성과정



그림 13 자동요약 데이터 세트 생성 N-S도표(기술 통계량은 합안구함)



한다. 또한 使用者의 다양한 요구에 따라 情報를 알기쉽게 테이블 형태로 나타내기 위하여 各 要約 데이터 세트에 대응하는 要約데이터의 具現에 관한 研究가 補完되어야 한다.

### 5. 具現結果 및 시스템의 文書化

#### 가. 具現結果

S-M시스템을 VSAM파일과 PL/I言語를 이용하여 농림수산부 시스템(IBM4341)에 실제 구현한 結果는 다음과 같다. 단, 組收入데이터 베이스를 예로 하여 出力한 형태는 메타데이터 베이스에서 基本메타데이터와 要約데이터 베이스의 統計量중 畝의 값만 인쇄한 경우이다.

#### ① 使用者 질의가 A경로인 경우

원시데이터 베이스에서 직접 검색을 원하는 경우로 메타데이터 베이스에서 해당코드 값을 찾은 후 경로배정 모듈에서 플래그값을 0으로 세트한 뒤 A경로를 통하여 운영되는데, 그 出力값은 <그림 14>와 같다.

그림 14 A경로 出力結果

```

###검색코드 입력자료###
      2      1      3      4      2      5
#1985년 #농작물 #채소 #1.5~2.0ha #하반기 11월
검색결과 플래그값 0;A경로 선택
검색결과 코드 213425
      ##조수입 135,701원##
    
```

#### ② 使用者 질의가 B경로인 경우

自動生成된 要約데이터 베이스로부터 검색하는 경우로 메타데이터 베이스의 코드값으로부터

플래그 값을 1로 세트한 뒤 B경로로 運營되며 그 出力結果는 <그림 15>와 같다.

그림 15 B경로 出力結果

```

###검색코드 입력자료###
      2      1      3      4      2      5
#1984 #비농작물 #비농작 #경 지 #상하반 #1월~
          불합  규모합  기합  12월합
검색결과 플래그값 1;B경로 선택
검색결과 코드 120000
      ##조수입 1,940,291원##
    
```

#### ③ 使用者 질의가 C경로인 경우

自動生成된 要約데이터 베이스와는 다른 構造의 要約데이터 세트를 要求하는 경우로 메타데이터 베이스에서 해당 코드값을 찾은 후 경로배정 모듈에서 플래그 값을 -1로 세트한 뒤 C경로로 運營되는데, 그 出力값은 <그림 16>과 같다.

그림 16 C경로 出力結果

```

###검색코드 입력자료###
      1      0      0      3      0      0
#1984 #작 물 #미곡~ #1.0~ #상하반 #1월~
          전체합  기타합  1.5ha  기합  12월합
검색결과 플래그값 -1;C경로 선택
검색결과 코드 100300
      ##조수입 4,587,426원##
    
```

#### 나. S-M시스템 프로그램 概要

<表 2>는 S-M시스템의 基本構造에 따른 각 모듈별 프로그램 일람표를 要約하였으며 <表 3>은 이중 경로배정모듈의 IPO차트를 예시한 것이다. 즉 각 모듈별 프로그램, 使用言語, 관련화

表 2 프로그램 일람표

| 시스템名 | 통계 데이터 베이스 관리시스템                |       |      | 作成者                        |     | 確認  |   |
|------|---------------------------------|-------|------|----------------------------|-----|---|---|
| 段階名  |                                 |       |      | 作成日                        |     | PAGE  | 1 |
| 文書名  | 프로그램 일람표                        |       |      |                            |     | PAGE  | 1 |
| 번호   | 프로그램名                           | step수 | 사용언어 | 관련파일                       |     | 내   | 용 |
|      |                                 |       |      |                            | I/O |   |   |
|      | PL1SDB                          |       | PL/I | JOSUIP<br>(VSAM)           |     | • 원시 조수입 데이터를 VSAM File에 출력하여 원시 조수입 D/B구축  |   |
|      | PL100                           |       | PL/I | AUTSUM<br>(VSAM)           |     | • Summary D/B 구축을 위한 VSAM File의 정의 및 초기치 출력   |   |
|      | PLISUM1                         |       | PL/I | AUTSUM<br>(VSAM)           |     | File 구성의 계층 구조에 의해 요약 데이터를 자동으로 생성해서 VSAM File에 출력                                  |   |
|      | PL1IN                           |       | PL/I | JOSUIP<br>AUTSUM<br>(VSAM) |     | 사용자의 요구에 따라 원시 D/B 또는 요약 D/B로부터의 검색 및 요약 D/B의 갱신의 경로를 결정하여, 실제 통계 데이터 베이스 관리시스템을 운영 |   |
|      | PL1IN—<br>Subprogram<br>meta—DB |       | PL/I |                            |     | 메타 정보의 저장 및 사용자의 요구에 따라 필요한 메타 정보출력   |   |
|      | PL1IN—<br>Subprogram<br>USERSUM |       | PL/I | AUTSUM<br>(VSAM)           |     | 자동생성된 요약 D/B와는 다른 구조의 요약 세트들 요구하는 사용자에게 원시 D/B로부터 새로운 요약세트를 생성하여 요약 D/B를 갱신         |   |

일형태와 내용이 수록되어 있는 프로그램 일람표와 각 프로그램에 대한 설명과 入出力상황이 수록된 프로그램說明書를 작성하였는데, 이러한 S—M시스템의 文書化 作業은 실제시스템의 利用者가 시스템을 쉽게 理解하고 使用할 수 있도록 함과 동시에 S—M시스템의 유지보수 및 發展의시스템으로의 확장 研究를 지원토록 작성되었다.

#### IV. 農業데이터 베이스 構築을 위한 發展方向

##### 1. 研究結果 要約

지금까지 우리는 農業데이터 베이스가 統計데이터 베이스 性格을 가지고 있음을 說明한 뒤 이의 효율적인 구축을 위하여 統計데이터 베이스의 特性 및 現況을 分析하였다. 또한 農業데이터 베이스를 위한 統計데이터 베이스 管理시스템의 發展을 위한 具體的인 제안을 하고자,

表 3 플래그셋 프로그램 IPO차트

|      |                 |       |       |      |   |
|------|-----------------|-------|-------|------|---|
| 시스템명 | 통계데이터 베이스 관리시스템 | 作成者   |       | 確認   |   |
| 段階名  |                 | 作成日   |       |      |   |
| 文書名  | 프로그램 설명서        | 프로그램名 | PLIIN | PAGE | 3 |
| 개요   |                 |       |       |      |   |

사용자의 요구에 따라 원시 D/B 또는 요약 D/B로부터의 검색 및 요약 D/B의 갱신의 경로를 결정하여 실제 S-M 시스템을 운영함.

| INPUT   | 처리과정     | OUTPUT   |          |          |          |          |    |     |    |      |      |   |   |  |
|---|----------|----------|----------|----------|----------|----------|----|-----|----|------|------|---|---|--|
| 사용자가 검색을 원하는 코드 CODE(6) <table border="1" style="margin-top: 10px;"> <tr> <td>code (4)</td> <td>code (2)</td> <td>code (3)</td> <td>code (4)</td> <td>code (5)</td> <td>code (6)</td> </tr> <tr> <td>연도</td> <td>대분류</td> <td>종류</td> <td>경시규모</td> <td>상하분기</td> <td>월</td> </tr> </table> | code (4) | code (2) | code (3) | code (4) | code (5) | code (6) | 연도 | 대분류 | 종류 | 경시규모 | 상하분기 | 월 | <ol style="list-style-type: none"> <li>1. 코드의 값 중 요약데이터 베이스의 통계량을 요구하는 0가 하나도 없는 경우는 'FLAG'의 값을 '0'으로 세트</li> <li>2. 요약값을 요구하는 구조가 주어진 계층구조에 의할 때, 즉 0코드값이 다른 숫자가 나오기 이전의 값으로 연속되어 있을 때 'FLAG'를 "1"로 세트</li> <li>3. 요약값을 요구하는 구조가 주어진 계층구조와 같지 않을 때, 즉 0코드값이 순서없이 섞여 나올 때 'FLAG'의 값을 "-1"로 세트</li> <li>4. 결정된 'FLAG'값에 의해 해당 경로를 걸쳐 운영</li> </ol> | <p>FLAG값</p> <pre>                     graph TD                         A[0 또는 1 또는 -1] --&gt; B[해당경로를 거침]                         B --&gt; C[결과]                     </pre> |
| code (4)  | code (2) | code (3) | code (4) | code (5) | code (6) |          |    |     |    |      |      |   |   |  |
| 연도  | 대분류      | 종류       | 경시규모     | 상하분기     | 월        |          |    |     |    |      |      |   |   |  |

보충설명

- Prototype의 통계 데이터 베이스 관리 시스템을 농업조수입을 예제로 하여 구현시킴
- FLAG : 0 → A경로
- FLAG : 1 → B경로 로 S-M 시스템을 운영
- FLAG : -1 → C경로

실제로 S-M시스템이라는 統計데이터 베이스 管理시스템을 프로토타입으로 설계하고 具現하였다. 이러한 S-M시스템의 長點으로는 첫째, 범주속성의 반복성의 문제 해결로 인한 기억장소의 절약과 빠른 액세스를 가져오며 둘째, 要約데이터 세트가 같은 데이터 베이스의 다양한 뷰(View)들에 의해 쉽게 운영되며 自動에그리게이션이 용이하다. 한편 이 시스템의 限界點은 제한된 질의어를 가지고 있고, 그래픽형태의 構造를 가지고 있어 뷰 정의시 데이터 화일의 構

造점근경로를 알아야 하는 制約이 있다.

2. 發展方向

가. 發展方案

지금까지의 研究를 토대로 農業데이터 베이스의 發展方案을 현재의 상황으로 독자적인 開發이 어렵다는 관점에서 現存의 상용데이터 베이스 시스템 또는 이미 개발된 獨自的 데이터 베이스 시스템을 활용한다는 전제하에서 檢討하였다.



① 第1案

統計데이터 베이스 전용의 目的으로 기 개발된 시스템의 도입이다. 대표적인 시스템으로는 미국의 버클리大學의 LBL에서 개발된 SUBJECT, SEEDIS, GUIDE 등을 들 수 있으며 이 시스템의 長點은 기억장소와 액세스의 效率를 높이고, 使用者에게 使用의 편의성을 제공할 수 있다. 限界點은 범용(General Purpose) DBMS가 아니므로 인사, 급여, 기타 管理業務를 포함한 농림수산 관련정보를 총괄하기에는 制限點이 있으며 또 다른 범용 DBMS의 도입을 초래할 수 있으므로 낭비적인 要素와 두 개의 DBMS로 인한 데이터의 共有 및 디자인의 균형에 問題點을 야기시킬 수 있다.

② 第2案

범용 데이터 베이스 管理시스템으로 계층형 管理시스템을 도입하고, 統計데이터 베이스의 特性을 보조할 소프트웨어를 도입하거나 開發하는 것이다. 이 案의 長點으로는 現 농림수산부의 DBMS가 계층형이므로 技術人力과 프로그램의 친근성을 들 수 있으며 또한 이 시스템의 화 일관리체계가 CPU부담이 크지 않다는 점이다. 限界點은 계층데이터 베이스 自體의 단점으로 분산 데이터 베이스 구축의 어려움, 사용하는 言語의 복잡성, 應用프로그램의 開發에 많은 時間의 所要와 維持補修의 어려움 등을 들 수 있다.

③ 第3案

범용 DBMS중 관계형 DBMS를 도입하고 統計데이터 베이스 特性을 보조할 S/W들을 도입하거나 開發을 통해 전위시스템으로 운영하는 것이다. 이 案의 長點은 첫째, 데이터의 獨立性

으로, 저장된 데이터 베이스 資料의 物理的 構造 향상과 개발된 응용 S/W의 開發을 분리 수 행할 수 있다는 점이다. 둘째, 使用者의 관점에서 보는 데이터 베이스 스키마의 單純性으로 자료에 대한 검색이 資料의 접근경로에 의존하지 않고 보다 發展적인 형태의 內容으로 이루어지기 때문에 使用者의 부담을 줄일 수 있을 것이다. 限界點으로는 관계형 DBMS가 CPU의 부담이 크고, 統計데이터 베이스의 特性으로 인한 데이터 저장과 액세스의 非效率性을 들 수 있겠다.

나. 結 論

이상 3가지 案 중에서 1案과 2案은 현실적인 限界點이 너무 많아, 보조 S/W를 사용해서 데이터 저장과 액세스의 非效率性을 보완한다는 전제하에 第3案의 選擇을 제안한다. 즉, 세 가지 시스템의 접근방법에 대한 기존시스템의 現況을 分析하고, 초기형태의 S-M시스템을 구현한 結果, 비록 기억장소와 액세스 등의 效率性이 떨어지고 정규형 릴레이션 형태로의 변환 과정이 필요하나, 차기시스템 도입시 관계데이터 베이스 관리시스템의 運營이 확실시되므로 이를 이용할 수 있고 보조 S/W를 계속 개발함으로써 保完이 가능하므로, 기존의 관계 데이터 베이스 관리시스템의 확장을 통한 農業統計데이터 베이스를 구축하는 것이 바람직하다고 判斷된다.

參 考 文 獻

김상희, 박석, 「범주속성의 Regrouping 情報를 利用한 效率的인 統計的 질의처리」, 韓國情報科學會, '86年 가을 학술발표논문집, 1986.  
 노상욱, 김용철, 박석, 「통계적 응용을 위한 확장된 SOL인터페이스의 設計에 관한 研究」, 韓國情報

- 科學會, '88年 봄 학술발표 논문집, 1988.
- 박세권, 서보환, 안성옥, 「農業데이터 베이스 구축의 發展方向」, KREI연구보고 148-6, 1988. 12
- 조석봉, 전학성, 김영시, 김영찬, 「교환기 데이터 베이스를 爲한 메타정보 생성에 관한 研究」, 韓國 情報科學會, 1988年 봄 學術發表論文集, 1988.
- 최원석, 「統計分析 지원을 爲한 統合要約 데이터 테이블의 設計 및 具現」, KAIST, 電算學科 석사논문, 1986.
- Anthony Klug, 「Abe—A Query Language for Constructing Aggregates—by—Example」, First LBL Workshop on SDB, 1981.
- A. Shoshani, 「Statistical Databases: Characteristics, Problems and Some Solutions」, A LBL Perspective on SDB, 1982.
- Hideto Sato, Ryosuke Hotaka, 「For Large Meta Information of National Integrated Statistics」, First LBL Workshop on SDB, 1981.
- Maurizio Rafanelli, Fabrizio L. Ricci, 「Proposal of a Logical Model for Statistical Data Base」, Second International Workshop on SDB, 1983.
- Paul Chan, Aric Shoshani, 「SUBJECT: A Directory Driven System for Large SDB」, First LBL Workshop on SDB, 1981.
- Roy Hammond, 「Metadata in the RAPID DBMS」, First LBL Workshop on SDB, 1981.
- Yvonne M. Bishop, Stanley R. Freedman, 「Classification of Metadata」, Second International Workshop on SDB, 1983.