

研究報告 148
1988. 12

統合農業情報시스템 ⑥

農業 데이터 베이스
構築의 發展方向

統計 데이터 베이스의 觀點에서

朴世權(首席研究員)

徐輔環(責任研究員)

安成玉(研究員)

韓國農村經濟研究院

빈 면

研究報告 148

統合農業情報시스템 ㉔

農業 데이터 베이스 構築의 發展方向 - 統計 데이터 베이스의 觀點에서 -

要 約

컴퓨터 시스템을 이용한 농업관련정보의 축적·분석·가공과 이를 각종 정책의사결정시 상황지표로 활용함으로써 농업의 효율성 및 경쟁력을 높이는 것이 매우 시급한 실정이다. 이러한 농업정보 관리문제를 시스템적 접근으로 해결하기 위하여 본 소프트웨어 하우스에서는 각종 농업관련정보를 체계적으로 수집·저장하는 데이터 베이스 시스템과 이로부터 2차·3차 加工情報를 산출해 내는 意思決定 시스템으로 大別되는 統合 農業情報 시스템의 구축을 위한 기초연구로써 현재 우리나라의 농림수산 정보 시스템의 현황과 문제점을 파악하고 이러한 문제해결의 관점에서 「통합농업 정보시스템의 분석 및 설계」 연구를 수행한 바 있다(연구보고 148 [I]).

본연구는 이의 계속된 보완·구체화 작업으로 농업 데이터 베이스의 특성과 문제점을 분석하고 농업 데이터 베이스의 효율적인 구축을 위한 방향 제시의 필요성에서 전개되었다. 농업 데이터 베이스는 통계분석을 위해 주로 사용되는 일종의 통계 데이터 베이스라 할 수 있으며 농업 데이터 베이스의 구축시 이러한 통계 데이터 베이스의 특성을 잘 고려해야 한다. 일반적으로 통계 데이터 베이스를 관리하기 위하여 기존의 범용 데이터 베이스 관리 시스템을 이용하나, 이 경우 데이터 저장과 액세스의 비효율성, 개념적 모델링도구와 질의어 등의 부족등으로 사용자의 요구를 충족시키지 못

II

하기 때문에 특별한 고려가 요구된다. 그리하여 본연구에서는 통계 데이터 베이스의 특성을 고려하여 데이터에 대한 일반적인 관리 기능뿐 아니라 통계 분석, 요약 정보의 관리 등 통계적 기능까지 포함한 통계 데이터 베이스 관리시스템의 관점에서 농업 데이터 베이스의 효율적 구축을 위한 발전 방향을 제시하였다.

1. 統計 데이터 베이스의 特性

統計 데이터 베이스는 데이터 베이스의 데이터가 단순 統計量의 계산 뿐만 아니라 일반적인 통계 응용에서 필요한 통계분석(statistical analysis)을 위해 주로 사용되는 데이터 베이스를 말한다. 이러한 統計 데이터 베이스의 特性은 다음과 같다. 첫째, 수록되는 데이터의 수치성과 대량성이다. 둘째, 데이터는 統計分析을 위해 사용되는 측정 데이터(measured data)와 측정 데이터의 의미를 부여하고 分類基準이 되는 매개데이터(parameter data)로 구성된다. 셋째, 산재 데이터(sparse data)가 일반적이다. 넷째, 分析目的으로 다수의 요약 부분집합들(summary subsets)이 생성되고 관리되어야 한다. 다섯째, 갱신이 거의 없는 안정적(static) 데이터이다. 여섯째, 데이터의 사용이 범주 속성(column-wise) 중심이다.

이와 같은 통계 데이터 베이스를 이용하는 통계분석가나 일반 사용자는 데이터에 대한 일반적인 관리기능 뿐 아니라 통계분석, 요약 정보의 관리 등 통계적 기능까지 요구한다. 따라서 데이터 베이스 관리 시스템은 데이터를 모델링하고 저장·조작하는 기능 및 데이터 베이스에 統計的 분석기법을 적용시킬 수 있는 기능도 포함해야 한다. 이러한 시스템을 統計 데이터 베이스 관리 시스템이라 한다. 이는 지금까지의 범용 데이터 베이스 관리 시스템에 서술된 통계 데이터 베이스의 특성을 반영한 시스템인 것이다.

2. Prototyp의 統計 데이터 베이스 관리 시스템(S-M 시스템)의 編造

통계 데이터 베이스의 효율적인 처리를 위하여 요약 데이터 베이스의 관리와 메타 데이터 베이스의 관리를 위해 데이터 사전/디렉토리(Data Dictionary / Directory)가 첨부된 통계 데이터 베이스 관리 시스템(이하 이

를 S-M 시스템이라 칭한다. S-M은 통계 데이터 베이스 관리(Statistical Data Base Management)의 약자임과 동시에 요약(Summary) 과 메타(Meta) 데이터 베이스의 약자인 중복의 의미를 갖는다)을 프로토타입(Prototype)으로 設計 및 具現한 실험적 시스템을 제시한다.

제안된 S-M시스템이 관리하는 원시 데이터 베이스, 요약 데이터 베이스 등은 제시된 그래프형태의 논리 모델(Logical model)에 의해 VSAM(virtual storage access method)화일의 저장구조를 이용하여 저장하고 이용된다. S-M시스템은 統計 데이터 베이스 구축 모듈, 메타 데이터 베이스 생성 및 관리모듈, 자동 요약 데이터 세트 생성 모듈, 요약 데이터 세트 갱신 모듈 및 S-M시스템 경로 배정 모듈의 5개의 모듈로 구성되어 있다 (구성도는 본문 <圖 3-1> 참조).

3. 農業 데이터 베이스 構築을 위한 發展方向

지금까지의 연구를 토대로 農業 데이터 베이스 구축을 위한 소프트웨어에 대한 몇 가지 안을 제시해 보겠다. 첫째, 제 1안으로 統計 데이터 베이스 전용의 목적으로 독자적으로 개발된 시스템의 도입이다. 둘째, 제 2안으로 상업용 데이터 베이스 관리 시스템으로 계층형 데이터 베이스 관리 시스템을 도입하고 統計 데이터 베이스의 특성을 보조할 소프트웨어를 도입하거나 개발하는 것이다. 셋째, 제 3안으로 상업용 데이터 베이스 관리 시스템으로 관계형 데이터 베이스 관리 시스템을 도입하고 統計 데이터 베이스 特性을 보조할 소프트웨어들을 도입 또는 개발을 통해 전위 시스템으로 운영하는 것이다.

이상의 세가지 중 본연구에서 제안하고자 하는 방안은 제 3안이다. 제 1안과 제 2안은 限界點이 너무 많아 統計 데이터 베이스의 特性 때문에 생기는 데이터 저장과 액세스의 비효율성이 있다고 하더라도 보조 소프트웨어로서 보완을 계속한다는 방침 아래 제 3안을 선택하는 것이 바람직한 것 같다. 그러나 제 3안을 선택한다고 하더라도 관계형 데이터 베이스 관리 시스템의 도입만으로는 統計 데이터 베이스의 特性을 살릴 수 없어, 부가적인 소프트웨어의 도입과 개발이 반드시 필요한데 이를 위한 지속적인 연

IV

구와 지원이 필요하다.

효율적 統計 데이터 베이스 구축을 위한 전위시스템은 기존의 統計 데이터 베이스의 運營을 지원할 수 있는 소프트웨어들의 도입과 함께 보완적인 성격을 띤 소프트웨어의 開發을 통해 완성시켜 나가는 것이 바람직하다. 그리하여 Prototype으로 개발한 S-M시스템의 경험을 살려 이 시스템에서 제시한 메타 데이터 베이스와 요약 데이터 베이스가 具現된 시스템으로의 완성을 위하여 도입과 더불어 지속적인 연구로 보조 소프트웨어의 개발을 통해 農業 데이터 베이스의 구축시 효율적으로 운영될 수 있는 전위시스템을 완성시켜 나가야겠다.

머 리 말

本稿는 1988년 農林水産 소프트웨어 하우스 事業의 일환으로 農業 데이터 베이스의 구축을 위한 프로토타입의 農業統計 데이터 베이스를 설계·구현하고 그 發展方向에 관한 研究內容을 수록한 報告書이다.

오늘날 고도의 經濟成長과 各種分野의 급진적인 發展에 따라 情報가 質的·量的으로 팽창하고 있으며 모든 分野에서 研究 및 意思決定에 必要한 적절한 情報를 適時에 提供해 주는 시스템체계가 絶실히 要求된다. 이러한 時代的 추세와 수입개방압력 등 우리의 農業分野에서의 環境變化에 따라 컴퓨터를 利用한 情報의 蒐集·加工·分析과 이를 各種政策 決定時 狀況 指標로 活用함으로써 農業의 경쟁력 열세를 만회하고 生産性 向上을 기해야 할 것이다. 本研究은 이러한 事業의 具體化 作業의 일환으로 農業 데이터 베이스가 주로 統計分析을 위해 使用되는 점을 감안하여, 기존의 統計 데이터 베이스의 現況과 問題點을 파악하고, 이를 農業 데이터 베이스에 適用함으로써 一般的인 데이터 管理機能 뿐 아니라 統計分析과 要約情報의 管理 등 統計分析的 기능을 포함한 農業統計 데이터 베이스 시스템의 구축을 위한 프로토타입(prototype)의 統計 데이터 베이스의 구축과 이에 근거한 具體的인 發展方向을 提示하였다.

그러나 統計 데이터 베이스의 역사가 워낙 짧아 보편화되어 있지 않고 참고문헌이나 관련 시스템의 미비로 인한 부족한 점이 많으리라고 생각된다. 이러한 점은 차후 소프트웨어 하우스의 계속사업으로 보충해 나가려고 한다.

끝으로 本研究가 이루어질 수 있도록 아낌없는 협조를 하여 주신 農林水産部 관련 직원 여러분께 감사드리는 바이다.

1988. 12.

韓國農村經濟研究院長 金 榮 鎮

目 次

第1章 序 論

- 1. 研究의 目的 및 必要性 1
- 2. 研究의 方法과 範圍 2

第2章 統計 데이터 베이스의 特性 및 現況 分析

- 1. 데이터 베이스와 데이터 베이스 관리 시스템 4
- 2. 統計 데이터 베이스와 統計 데이터 베이스 관리시스템 11
- 3. 統計 데이터 베이스 관리를 위한 기존 연구의 現況 分析 14

第3章 프로토타입(prototype)의 統計 데이터 베이스 관리시스템(S-M 시스템)의 設計 및 具現

- 1. S-M 시스템의 基本構造 18
- 2. 논리 모델의 그래프 표현 20
- 3. 메타 데이터 베이스의 設計 25
- 4. 요약 데이터 베이스의 設計 34
- 5. 具現結果 및 시스템의 文書化 38

第4章 農業 데이터 베이스 構築을 위한 發展方向

- 1. 연구 결과 요약 46
- 2. 發展方向 47

表 目 次

第 3 章

表 3-1	농업 조수입 데이터의 코드화일	29
-------	------------------------	----

圖 目 次

圖 2-1	화일 중심의 자료처리 시스템에서 응용프로그램과 화일 과의 관계	5
圖 2-2	데이터 베이스 관리 시스템 역할의 기본구조	7
圖 2-3	관계 데이터 모델의 예	9
圖 2-4	계층 데이터 모델의 예	9
圖 2-5	네트워크 데이터 모델의 예	10
圖 2-6	범주속성과 요약속성의 예	12

第 2 章

圖 3-1	S-M 시스템의 기본 구조	19
圖 3-2	농업 조수입 원시 데이터	21
圖 3-3	농업 조수입을 위한 행렬 표시(matrix representation) ..	22
圖 3-4	농업 조수입의 Semantics를 위한 그래프 표현	24
圖 3-5	계층구조의 데이터 사전/디렉토리의 구성	28

圖 3- 6	경로 결정 N-S 도표	30
圖 3- 7	A경로 운영 과정	31
圖 3- 8	B경로 운영 과정	32
圖 3- 9	C경로 운영 과정	33
圖 3-10	요약 데이터 베이스의 기술통계량	35
圖 3-11	자동 요약 데이터 세트 구성 과정	35
圖 3-12	자동 요약 데이터 세트 생성 N-S 도표	36
圖 3-13	사용자 요구 요약 데이터 세트 생성 N-S 도표	37
圖 3-14	A경로 출력	38
圖 3-15	B경로 출력	39
圖 3-16	C경로 출력	39

第 1 章

序 論

1. 研究의 目的 및 必要性

인류는 원시사회, 농경사회, 산업사회를 거쳐 이제 情報化社會로 진입하였다. 이러한 社會에서는 컴퓨터를 이용한 정보축적·분석·가공 처리의 양과 질의 정도가 그 社會의 경쟁력과 발전의 정도를 좌우한다고 말할 수 있다. 이러한 시대적 추세와 수입개방 압력 등 우리의 農業分野에서의 환경변화에 따라 컴퓨터시스템을 이용한 농업관련정보의 축적·분석·가공과 이를 각종 정책 의사결정시 상황지표로 활용함으로써 농업의 효율성 및 경쟁력을 높이는 것이 매우 시급한 실정이다.

이러한 問題를 농업정보관리의 시스템적 접근으로 해결하기 위하여 본 소프트웨어 하우스에서는 각종 농업관련정보를 체계적으로 수집·저장하는 데이터 베이스 시스템과 이로부터 2차·3차 加工情報를 산출해 내는 意思決定 시스템으로 大別되는 統合 農業情報 시스템의 구축을 위한 기초 연구으로써 현재 우리 나라의 농림수산 정보 시스템의 현황과 문제점을 파악하고 이러한 문제해결의 관점에서 「통합농업 정보시스템의 분석 및 설계」 연구를 수행한 바 있다(연구보고 148 [1]).

본 연구는 이의 계속된 보완·구체화 작업으로 농업 데이터 베이스의 특성과 문제점을 분석하고 농업데이터 베이스의 효율적인 구축을 위한 방향 제시의 필요성에서 전개되었다. 농업 데이터 베이스는 통계 분석을 위해 주로 사용되는 일종의 통계 데이터 베이스라 할 수 있으며 농업 데이터 베이스의 구축시 이러한 통계 데이터 베이스의 특성을 잘 고려해야 한다. 일반적으로 통계 데이터 베이스를 관리하기 위하여 기존의 범용 데이터 베이스 관리 시스템을 이용하나, 이 경우 데이터 저장과 액세스의 비효율성 개념적 모델링 도구와 질의어 등의 부족 등으로 사용자의 요구를 충족시키지 못하기 때문에 특별한 고려가 요구된다.

그리하여 본 연구에서는 통계 데이터 베이스의 특성을 고려하여 데이터에 대한 일반적인 관리 기능뿐 아니라 통계 분석, 요약 정보의 관리등 통계적 기능까지 포함한 통계 데이터 베이스 관리시스템의 관점에서 농업데이터 베이스의 효율적 구축을 위한 발전 방향을 제시하고자 한다.

2. 研究의 方法과 範圍

研究의 方法과 範圍는 다음과 같다.

첫째, 문헌을 통해 수집한 기술현황 자료를 분석하였다. 즉, 통계 데이터 베이스의 특성을 고려하여 기존의 범용 데이터 베이스 관리시스템을 확장하거나 독자 개발을 통하여 통계 데이터 베이스 관리시스템을 구축하고자 하는 기존 연구의 現況을 分析하였다.

둘째, 실제로 S-M 시스템이라고 命名지은 Prototype의 통계 데이터 베이스 관리 시스템을 설계하고 구현하였다. 이러한 S-M 시스템은 실제 농업 데이터 베이스 구축을 위해 본 소프트웨어 하우스에서 독자적인 통계 데이터 베이스 관리 시스템을 개발하기 위한 목적 보다는 기존의 연구·개발된 통계 데이터 베이스 관리 시스템의 도입과 확장·보완을 위한 방향 설정의 목적으로 개발되었다. 즉, Prototype의 시스템의 설계·구현 작업을 통하여 이론적으로 분석한 내용을 실험함으로써 보다 구체적이고 실제

적인 경험을 통해, 새로운 각도로 문제점과 나아갈 방향을 제시해 보고자 한다.

세째, 앞의 두 연구방법을 토대로 농업 데이터 베이스 구축을 위한 발전방향을 제시해 보았다. 즉, 먼저 몇 개의 안을 제시하여 각 대안들의 장점과 한계점을 分析한뒤, 지금까지의 연구의 결과와 현실성을 고려하여 가장 바람직한 방안을 결론적으로 제시하였다.

第2章

統計 데이터 베이스의 特性 및 現況 分析

1. 데이터 베이스와 데이터 베이스 관리 시스템

가. 데이터 베이스 概念

산업사회가 고도로 발달해감에 따라 하나의 조직에서 사용되는 데이터의 양은 상당히 방대해지며, 또한 여러 응용들에 중복 사용되어지게 된다. 이와 같이 데이터 베이스는 원래 같은 데이터가 여러 응용에 중복되어 사용될 수 있다는 다목적성에서 발전된 데이터 共用的 概念에서 출발하였다. 따라서 데이터 베이스란 어느 한 조직의 다수 응용 시스템들이 사용하기 위해 統合 貯藏된 운영 데이터의 집합이라고 정의할 수 있다. 이를 몇 개의 항목으로 나누어 설명하면 다음과 같다.

첫째, 데이터 베이스는 최소의 중복 (minimal redundancy)만을 허용하는 통합된 데이터 (integrated data)이다.

둘째, 데이터 베이스는 테이프나 디스크와 같이 컴퓨터가 액세스하여 처리할 수 있는 저장장치에 수록된 저장된 데이터 (stored data)이다.

셋째, 데이터 베이스에 있는 데이터는 어느 조직이나 기관이건 그 기능을 수행하는데 반드시 유지해야할 데이터인 운영 데이터 (operational data)이다.

내췁, 데이터 베이스는 한 조직에 있는 여러 응용 시스템들이 공동으로 소유하고 유지하며, 이용하는 공용 데이터(shared data)이다.

나. 데이터 베이스 관리 시스템(Data Base Management System : DBMS)

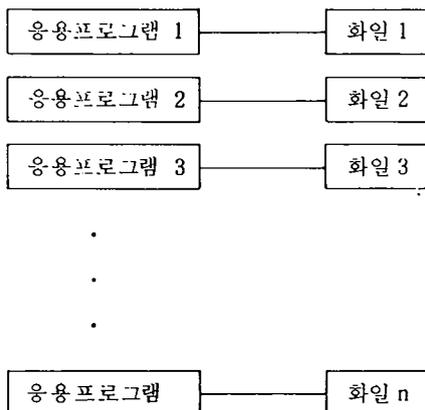
1. 데이터 베이스 관리 시스템의 배경 및 정의

먼저 화일 중심의 자료 처리 시스템에서 데이터 베이스 관리 시스템으로의 발전과정을 살펴보겠다.

화일을 중심으로 한 종래의 자료처리 시스템에서는 <圖 2 - 1 > 과 같이 각 응용 프로그램은 個別的으로 자기자신의 화일을 관리 유지해야 한다. 또한 각 응용 프로그램은 자기 데이터 화일을 검색하고 삽입, 삭제 및 갱신할 수 있는 루틴을 포함해야 한다.

이러한 시스템에서의 응용 프로그램은 논리적 화일구조와 저장된 물리적 화일구조간의 1 대 1의 대응을 요구하게 되고, 데이터의 물리적 構造에 대한 가정과 그에 대한 액세스 방법이 이 데이터를 사용하는 응용 프로그램에 구현되어 있어야 하며, 또한 데이터 화일의 共有가 어렵게 되어 결국 단독 사용만을 허용하게 된다. 여기서 볼 수 있는 문제점은 크게 데이터의 중복성과 데이터의 중복성과 데이터 共有의 非用利性을 이야기 할 수 있다.

圖 2 - 1 화일 중심의 자료처리 시스템에서 응용프로그램과 화일과의 관계



첫째, 데이터 종속성 (data dependence)이란 데이터의 구성 방법이나 구성형식 액세스방법이 변경되면 이에 관련된 응용 프로그램도 같이 변경되어야 하는 응용 프로그램과 데이터 간의 相互依存關係를 말한다.

예를 들어 인덱스된 순차 화일구조로 구성된 데이터 화일을 순차화일구조로 바꾸면 이 화일을 액세스하는 모든 응용 프로그램은 이에 따라 수정되어야 한다. 이러한 문제는 응용 프로그램이 많고, 그 화일 구성방법이나 액세스방법을 변경해야되는 상황에서는 아주 중요한 문제가 된다.

둘째, 응용 프로그램들은 같은 內容의 데이터를 같은 樣式에 같은 구조로 요구하는 경우도 있고, 또 어떤 때는 같은 內容의 데이터를 양식과 구조만 다르게 요구하는 때도 있다. 화일 시스템에서는 같은 데이터를 필요로 하는 경우이거나 일부분만 같은 데이터를 요구하는 경우 構造가 다르면 별도의 중복된 화일을 유지해야 한다. 이것을 데이터 중복성 (data redundancy)이라 한다. 이러한 데이터 중복성은 데이터의 일관성 (consistency) 유지의 어려움, 중복된 동일한 데이터에 대해 동등한 보안유지의 어려움, 중복된 데이터를 저장하기 위한 추가저장공간의 비용과 갱신작업시 중복된 모든 데이터를 전부 수행해야 되므로 過多 갱신비용의 지출 등의 問題點을 야기시킨다.

세째, 데이터를 共有하는 면에서도 한계점을 가지고 있다. 대부분의 화일 중심의 데이터 처리 시스템에서는 어느 한 프로그램이 하나의 화일을 OPEN하여 사용하면 그 프로그램의 수행이 모두 끝날 때까지 다른 프로그램은 그 화일을 사용할 수가 없어 데이터의 동시공유성 (concurrent sharing)을 지원하지 못한다.

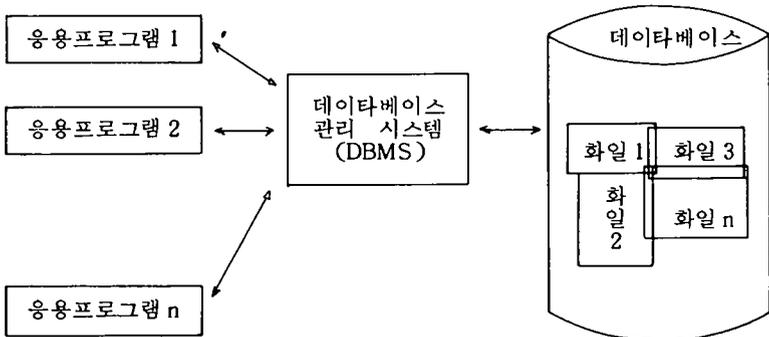
따라서 이러한 화일 시스템의 短點을 보완하고 사용자 입장에서 보다 편리하게 자료처리를 수행하는 중앙집중식 관리체제의 요구에서 출발한 새로운 방법이 데이터 베이스 관리 시스템 (data base management system : DBMS)이다. 즉, 데이터 베이스 관리 시스템은 화일 시스템에서 야기되는 데이터의 종속성, 중복성과 데이터 共有의 어려움 등의 문제를 해결하기 위한 방법으로 제안되었다.

여기서 데이터 베이스 관리 시스템이라 하면 응용 프로그램과 데이터의

중재자로서 모든 응용 프로그램들이 데이터 베이스를 공유할 수 있도록 관리해 주는 소프트웨어(software)로 정의할 수 있다. 이러한 시스템하에서 데이터 베이스를 이용하려는 응용 프로그램들은 데이터 베이스 관리 시스템을 통해서만 이용이 가능하다. 이것은 데이터 베이스 관리 시스템이 데이터 베이스의 구성, 액세스 방법, 管理維持에 대한 모든 책임을 지고 있다는 것을 의미한다.

이 데이터 베이스 관리 시스템을 통하여 데이터 베이스를 이용하는 방법의 기본적인 구조는 <圖 2 - 2 >에 表現되어 있다. 데이터 베이스 관리 시스템을 이용하는 응용 프로그램은 데이터 베이스 전체의 생성, 액세스 방법, 조작 절차, 보안, 물리적 구조 등에 대해서 관련할 필요없이 원하는 데이터와 처리 작업만을 데이터 베이스 관리 시스템에 요구하면 된다. 반면에 데이터 베이스 관리 시스템은 데이터 베이스를 종합적으로 조직하고 액세스하며 전체적으로 통제할 수 있는 프로그램들로 구성되어 있다. 이러한 데이터 베이스 관리 시스템이 데이터 베이스를 여러 응용 프로그램이 共有할 수 있도록 관리하기 위해서는 데이터 구조의 정의기능 (defenition facility), 데이터의 조작기능 (manipulation facility) 과 데이터의 제어기능(control facility) 등의 필수기능을 가지고 있어야 한다.

圖 2 - 2 데이터 베이스관리 시스템 역할의 기본구조



② 데이터 모델(Data Model)

앞서 설명된 데이터 베이스 관리 시스템은 자신의 자료를 관리하기 편하

도록 데이터 모델이라는 데이터를 저장하는 概念的인 도구를 提供한다.

즉, 데이터 모델 (Data Model)이라 함은 개략적으로 데이터가 논리적으로 조직될 수 있는 양식이라고 말할 수 있다.

데이터 모델에는 상업용 데이터 베이스 관리 시스템이 지원할 수 있는 대표적인 형태로서 관계 데이터 모델 (relational data model), 계층 데이터 모델 (hierarchical data model) 과 네트워크 데이터 모델 (network data model) 등의 3가지가 있다. 그러나 한 데이터 베이스 관리 시스템에서는 어떤 현실세계를 표현하기 위해 하나의 데이터 모델만이 사용할 수 있는데 그 이유는 데이터 베이스 관리 시스템이 하나의 데이터 모델만을 지원할 수 있기 때문이다.

이 3개의 데이터 모델간의 주요 차이점은 데이터 요소간의 관계 (relationship)를 표현하는 방식에 있다. 이제 이 3개의 데이터 모델을 각각 개략적으로 설명하고 특징을 비교해 보겠다.

가) 관계 데이터 모델 (relational data model)

관계 데이터 모델은 사용자에게 데이터 베이스 構造를 테이블 형태로 기술해 주는 것이다. <圖 2-3>은 식량작물의 식부면적과 생산량에 대한情報가 이 모델에서 표현되는 예를 보여준다. <圖 2-3>에서 보는 바와 같이 관계 모델에서는 모든 데이터들과 그 데이터 사이에 존재하는 관계 (relationship)를 테이블 형태로 표현한다. 이 관계 데이터 모델은 논리적 구조가 간단해서 이해 하기 쉽고 데이터의 의미와 그것에 대한 조작이 명확하다는 長點이 있다.

나) 계층 데이터 모델 (hierarchical data model)

계층 데이터 모델은 데이터 베이스의 논리적 구조가 계층 (hierarchy)性質을 갖는 트리 (tree) 형태로 表現되는 것을 말한다. 계층 데이터 모델의 예로서 <圖 2-4>와 같은 계층 레코드 타입 구조로 농가경제를 표현할 수 있다. 이 계층 데이터 모델은 사용자가 데이터 베이스의 계층 구조를 알아야 된다는 단점을 가지고 있지만 컴퓨터의 중앙처리 장치의 부담을 적게하는 장점을 가지고 있다.

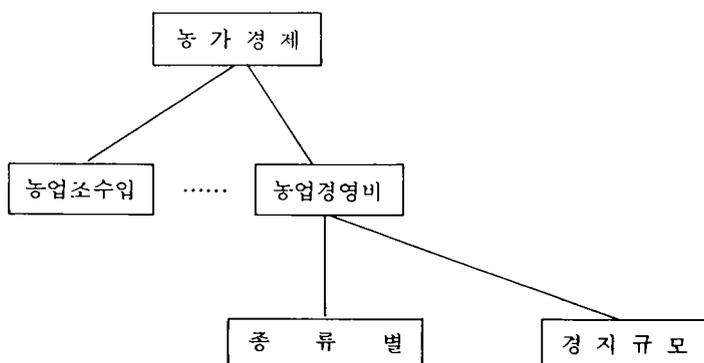
圖 2 - 3 관계 데이터 모델의 예

생산지			작물별 식부면적		
생산지번호	생 산 지	식량작물 식부면적	작물번호	작 물 명	식부면적
S 1	경 기 도	208398	A 1	미 곡	1262324
S 2	강 원 도	114095	A 2	맥 류	207635
S 3	충 청 도	321673	A 3	잡 곡	46085
S 4	전 라 도	574031	A 4	두 류	211912
S 5	경 상 도	507718	A 5	서 류	47376

생산지 작물별 식량작물 생산량

생산지번호	작 물 번 호	생 산 량
S1	A 1	767278
S1	A 2	2391
S1	A 3	4988
S2	A 4	254878
S2	A 5	45797

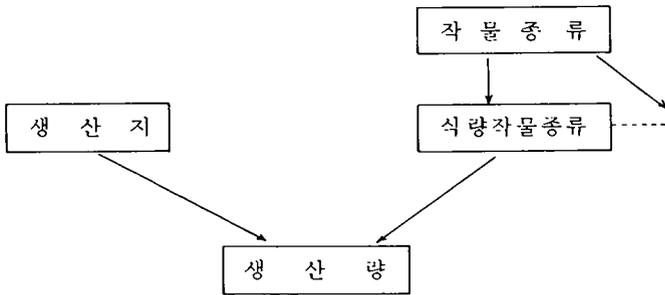
圖 2 - 4 계층 데이터 모델의 예



다) 네트워크 데이터 모델

네트워크 데이터 모델은 데이터 베이스의 論理的 構造를 표현하는 자료 구조도가 트리라는 제약을 받지 않고 일반 그래프의 성질을 갖는 형태이다. <圖 2-5>는 식량작물의 생산지와 작물종류별 생산량이 네트워크 관계 모델로 어떻게 표현되는지를 보여준다. 네트워크 데이터 모델은 계층 데이터 모델보다 그 구조가 복잡하고 이들을 처리하는 조작어가 복잡해지나 트리는 구조상의 제약이 없기 때문에 관계를 표현하는데는 더 융통성이 있다.

圖 2-5 네트워크 데이터 모델의 예



다. 農林水産部 데이터 베이스의 現況

현행 농수산부의 정책시행에 필요한 각종 情報資料들은 경제, 생산, 양곡 유통업무로 나누어 IBM4341-K10 시스템을 사용하여 전산처리를 하고 있다. 이 중 일부 업무는 데이터 베이스화 되어 있으며, 계층형인 DL/I 데이터 베이스 관리 시스템을 사용하고 있다. 나머지 업무는 VSAM(virtual storage access method) 화일을 이용하여 구축되어 있다.

그러나 최근에 들어 사회의 발전과 함께 농수산분야에서도 자료의 양적 증가와 질적 다양화에 따른 기술적인 자료처리의 필요성이 급격히 증대되고 있는 실정이다. 이러한 추세에 발맞추어 각종 農業關聯情報를 體系的으로 蒐集·貯藏하는 농업 데이터 베이스 시스템의 발전 방향에 대해 「統合農業情報시스템의 分析 및 設計」(연구보고 148 統合農業情報시스템□)에 제

시한 바 있다. 이 보고서에서 제시된 20개로 구분된 데이터 베이스를 효율적으로 구축하기 위하여 본 연구보고에서는 농업 데이터 베이스의 특성과 문제점을 살피고 이를 고려한 데이터 베이스의 설계 및 데이터 베이스 관리 시스템의 발전방향 등에 관한 구체적인 연구를 수행하였다.

2. 統計 데이터 베이스와 統計데이터 베이스 관리시스템

가. 統計 데이터 베이스의 特性

統計 데이터 베이스는 데이터 베이스의 데이터가 단순 統計量의 계산 뿐만 아니라 일반적인 통계 응용에서 필요한 통계분석 (statistical analysis) 을 위해 주로 사용되는 데이터 베이스를 말한다. 統計 데이터 베이스의 特性은 다음과 같다.

첫째, 수록되는 데이터의 수치성 (numericness) 과 대량성 (largeness) 이다. 이 수치성과 대량성은 통계 데이터 베이스의 논리적 접근에 어려움을 주고 저장·처리의 效率性을 위해 특별한 고려를 필요하게 한다.

둘째, 데이터는 統計分析을 위해 사용되는 측정 데이터 (measured data) 와 측정 데이터의 의미를 부여하고 分類基準이 되는 매개 데이터 (parameta data) 로 구성된다. 매개 데이터와 측정 데이터는 속성 (attribute) 들의 형태로 묘사된다. <圖 2-6>은 테이블 형태로 나타난 작물생산/소비 데이터이다. 처음의 4개의 속성들 (품목, 도, 년, 계절)은 매개 데이터이고 나머지 2개의 속성들 (생산, 소비)은 측정 데이터이다. 매개 데이터를 위한 속성들은 그것들이 측정 데이터의 범주를 포함하기 때문에 범주속성 (category attributes) ”이라 하고 측정 데이터를 위한 속성들은 그것들이 통계의 요약/분석을 위해 사용되기 때문에 “ 요약속성 (summary attribute) ”이라 한다.

- 범주속성들의 값의 조합 (combination)은 요약속성을 위한 합성키 (composite key)와 같은 役割로 쓰인다.
- 범주속성들은 높은 중복성 (redundancy)을 가지고 있어 범주속성을 위

한 효과적인 저장과 액세스를 위한 고려가 필요하다.

- 범주속성의 범위가 일반적으로 작은 반면 요약속성의 범위는 대체적으로 크며 수치 데이터이다.

圖 2 - 6 범주속성과 요약속성의 예

매개 데이터 (Parameta data)			측정데이터 (measured data)			
품 목	도	년	계 절	생	산	소 비
쌀 ⋮ 보리	경 기도	80	봄	—	700	
			여름	—	900	
			가을	1,200	1,000	
			겨울	—	900	
	강원도	81	봄	⋮	⋮	
			여름	⋮	⋮	
			가을	⋮	⋮	
			겨울	⋮	⋮	
			⋮	⋮	⋮	
			⋮	⋮	⋮	
			⋮	⋮	⋮	
			⋮	⋮	⋮	

범주속성 (category attribute)
요약속성 (summary attribute)

세째, 산재 데이터 (sparse data)가 일반적이다. 이의 효과적인 저장과 액세스를 위한 데이터 압축기법 (compression technique)들이 사용되어야 한다.

네째, 分析目的으로 다수의 요약 부분집합들 (summary subsets)이 생성되고 관리되어야 한다.

다섯째, 갱신이 거의 없는 안정적 (static) 데이터이며, 시계열 데이터인 경우가 많다.

여섯째, 데이터의 사용이 범용 데이터 베이스에서 처럼 레코드 중심(row wise)이 아니라 범주속성(column-wise) 중심이다. 이를 위해 轉置 파일(transposed file) 기법이 유용하게 사용될 수 있다.

일곱째, 메타 데이터를 취급한다.

이와 같은 통계 데이터 베이스의 특성에 따라 주요 응용 분야로는 소득, 고용, 생산, 수출입, 인구센서스 등을 수록하는 사회경제 데이터, 재무제표 재고, 증권거래상황 등을 수록하는 기업거래 데이터, 에너지, 기상과학 실험결과를 수록하는 과학실험 데이터 등이 대표적이다. 농업 데이터 베이스도 농업 센서스, 생산량, 재고량, 소비량, 生産費, 식부면적 등의 사회경제 데이터로 통계분석을 위해 주로 사용되는 통계 데이터 베이스이다. 그러므로 농업 데이터 베이스의 구축시 이러한 통계 데이터 베이스의 특성을 고려해야 할 필요가 있다.

나. 統計 데이터 베이스 관리 시스템(SDBMS)

이와 같은 특성을 갖는 統計 데이터 베이스의 관리를 위해서는 統計 패키지 데이터의 조작 및 화일 조작기능을 보강할 수도 있고, 기존의 범용 데이터 베이스 관리 시스템에 데이터 조작, 수치처리, 분석기능의 보강 등을 통한 統計 데이터 베이스 관리 시스템으로의 확장, 또는 전적으로 새로운 소프트웨어로서 독자적인 통계 데이터 베이스 관리 시스템을 개발할 수도 있다.

統計 데이터 베이스를 이용하는 통계분석가나 일반 사용자는 데이터에 대한 일반적인 관리기능 뿐 아니라 통계분석, 요약 정보의 관리 등 통계적 기능까지 요구한다. 따라서 데이터 베이스 관리 시스템은 데이터를 모델링하고 저장·조작하는 기능 및 데이터 베이스에 統計的 분석기법을 적용시킬 수 있는 기능도 포함해야 한다. 이러한 시스템을 統計 데이터 베이스 관리 시스템(Statistical Database Management System: SDBMS)이라 한다. 이는 지금까지의 범용 데이터 베이스 관리 시스템에 앞에서 서술한 통계 데이터 베이스의 특성을 반영한 시스템인 것이다.

統計 데이터 베이스 관리를 위한 첫번째 방법인 統計 패키지의 활용방법

은 통계 패키지들이 데이터에 대한 적합한 저장 구조 및 접근경로의 부재와 제공되는 데이터 관리 기능의 한정성으로 인해, 부분적인 도움을 줄 수는 있으나 전체적인 문제해결을 위해서는 부적합하다고 생각되어 연구범위에서 제외시켰다.

두번째 方法인 범용 데이터 베이스 관리 시스템의 확장 또는 독자 개발에 의한, 통계 데이터 베이스 관리 시스템의 구축에 대하여 살펴 보겠다.

이와 같이 통계 데이터 베이스 관리 시스템이 필요한 이유는, 통계 데이터 베이스를 관리하기 위해 기존의 범용 데이터 베이스 관리 시스템을 그대로 이용하기에는 데이터 저장과 액세스의 비효율성, 함수성 (functionality) 과 사용의 편의성의 부족, 統計 데이터의 特性에 적합한 개념적 모델링 도구 (conceptual modeling tool), 내부적 모델링 도구 (internal modeling tool), 질의어 등의 부족으로 인해 사용자의 요구를 충족시키지 못하기 때문이다.

3. 統計 데이터 베이스 관리를 위한 기존연구의 現況 分析

앞에서 논한 기존의 범용 데이터 베이스 관리 시스템을 확장하거나 독자 개발을 통하여, 통계 데이터 베이스 관리 시스템을 구축하여 통계 데이터 베이스를 관리하고자 하는 연구가 많이 있어 왔다. 연구에 의해 개발된 시스템과 질의어 또는 제안에 대해 간략히 살펴보면 다음과 같다.

가. 범용데이터 베이스의 확장보완 연구

기존의 상용 데이터 모델을 사용하여 확장된 시스템으로 統計 데이터 베이스 관리 시스템을 구축한 연구는 다음과 같다.

GENISYS [ManD 81], RAPID [TuHc 79], 미국 노동통계부 (Bureau of Labor Statistics) 의 Wiss 등에 의해 구현된 Table Producing Language System [TPLS :UsIb80], Tsukuba 대학의 Ikeda 와 Hiroshima 대학의 kobayashi 에 의해 Model 204 데이터 베이스 관리시스

템에 구현된 HSDB[Ikek 81], IBM San Jose Lab의 Ghosh와 Codd에 의한 릴레이서널 모델을 확장한 Statistical Relational Teble(SRT)와 질의어인 QBSRT(Query by statistical Relational Teble)의 제안, Berkely 대학의 Johnson에 의한 Simple to Read and Linderstand(STRAND) 질의어의 제안, Kohji shibano와 Hideto Sato에 의한 CAS SDB, Wisconsin대학의 Klug에 의한 질의어인 Abe[Klug81]와 Ozsoyolu가 제안한 STBE[Ozso 856] 질의어 등이다.

이들의 중요한 특징을 간략히 살펴보면 아래와 같다.

① TPLS의 질의어인 TPL의 요약 테이블 표현하는 기본기능은 TREE 구조화일에서부터 요약 테이블을 정의하는 것이다. 이러한 기능은 TREE구조의 복잡한 접근 경로를 미리 뷰(View)로서 정의하여 수행된다. 따라서 사용자는 TREE 構造의 접근 경로를 매 질의어마다 명시하지 않아도되는 이점이 있으나, 뷰 정의 자체가 상당히 복잡하여 이해하기 힘들고, 뷰를 정의하는 사용자는 데이터 화일의 TREE 구조 접근 경로에 대하여 잘 알고 있어야 하는 제약이 있다. TPL의 요약 테이블 정의 기능은 매우 다양하고 강력하나 가장 큰 단점은 배치(batch) 형태로 수행되는 단독(stard-alone) 시스템이라는 것이다.

② HSDB는 원시 데이터로부터 원시 요약 테이블(Elementary Summary Table:EST)을 정의할 수 있도록 하였는데 EST가 하나의 속성(attribute)에 대해서만 요약 데이터를 정의하며, EST를 관리하는 기능 중에서 분류화만을 제공하므로 오퍼레이션이 단순하다. 또한 EST에 정의되지 않은 요약 데이터는 데이터 사전 / 디렉토리(Data Dictionary/Directory)에 저장하는데 이는 요약 데이터 관리의 이중구조를 가져와 사용자로 하여금 요약 데이터의 관리에 일관성을 이루지 못하게 한다.

③ Ghosh에 의한 요약 테이블인 SRT는 독자적인 저장구조를 통하여 K-map 형태로 저장되어, 부가적인 저장기법 및 접근기법이 필요하며 QBSRT는 SRT를 관리하는 화면을 통한 질의어이다.

④ STRAND는 ER 모델링기법에 의한 요약 테이블 관리하는 질의어로서 오퍼레이션 요약화(summarization)를 이용하는 요약 데이터를 저장하는

릴레이션을 만들 수 있다. 그러나 이러한 작업이 수행되기 위해서는 데이터 베이스 관리자(Database Administrator: DBA)가 미리 스키마에 각 통계량을 구하는 프로시저를 정의해야만 하므로 사용자 입장에서는 제한요소가 많으며, 또한 이 시스템은 통계정보를 관리하는 기본구조로서 릴레이션을 사용하므로, 요약 테이블을 이용하는 방법에 비하여 요약 데이터 관리가 개념적 수준에서 직접적으로 이루어지지 못하고 내부적 수준에서 이루어진다. 이 질의어는 현재 INGRES 데이터 베이스 관리 시스템에 전위 시스템으로 구현되어 있는데 사용자가 접근 경로를 명시해야 하는 계층적 구조의 언어라는 단점도 있다.

⑤ Abe(Aggregates-by-Example)는 Zloof[Zloof75]가 제안한 QBE(Query-by-Example)와 유사한 질의어로서 화면을 이용하여 릴레이셔널 시스템을 기반으로 복잡한 통계 질의어를 처리할 수 있도록 설계되었다.

⑥ 릴레이셔널 모델을 근간으로 하고 있으나 요약 데이터를 관리하는 독자적 기본구조를 가지고 있어(Case Western Reserve 대학의 Ozsoyoguel, 가 제안한 요약 데이터를 관리하는 Primitire Summary Table(PST)이라는 요약 테이블과 질의어인 STBE(Summary Table by Example)는 기존 시스템의 확장 또는 독자개발의 연구분류의 중간위치에 있다고 볼 수 있다. STBE는 화면을 이용하여 질의어를 처리하는 Abe와 유사하나, PST라 불리는 요약 테이블을 요약 데이터를 관리하는 기본구조로서 사용하는 점이 Abe와의 가장 큰 차이점이다. 다른 시스템과 비교해서 STBE의 가장 큰 특징은 사용자에게 정의된 요약 테이블을 릴레이션 형태로 변환하여 저장하는 점이다. 따라서 데이터 베이스 관리 시스템은 요약 테이블을 위한 부가적 저장 구조가 필요 없으며, 질의어 처리를 더욱 효과적으로 수행할 수 있게 된다. 그러나 STBE는 요약테이블이 관리하는 요약 데이터를 반드시 하나의 어트리뷰트에 한정하여야 하는 단점이 있다.

나. 독자 개발 연구

독자적인 통계 데이터 베이스 관리 시스템의 개발에 관한 연구는 다음과 같다.

미국의 Berkeley 대학의 LBL에서 개발한 SUBJECT[Chan81], SE-EDIS[McCa81], GUIDE[Won82]와 SAM*(Semantic Association Model)[Su82], RAM[Sund78]과 Ozsoyoglu에 의한 SSDB 등이 있다. 이들 시스템들의 중요한 특징은 아래와 같다.

① Chan과 Shoshani에 의한 SUBJECT 시스템은 Cluster 노드와 Cross Product 노드들의 계층구조에 의해 요약 세트의 개념을 세우고, Semantic modelling과 요약과 통계정보의 단순한 질의를 위해 범주와 요약 속성으로 데이터를 구분하며, 요약 데이터와 원시 데이터의 행렬표시를 위해 cluster abstraction과 cross product abstraction을 사용한다. cross product abstraction은 n차원 행렬의 데이터형을 사용하여 요약 데이터를 나타내는 새로운 데이터 모델링 도구(tool)이며, cluster abstraction은 복잡한 범주속성의 계층구조와 요약속성의 수집(collection)을 위해 사용된다. 또한 이 시스템은 사용자가 情報의 위치를 알아내고 검색을 돕기 위해 cluster와 cross product의 계층구조에서 browsing기능을 포함한 7개의 메뉴화된 코멘트(command)의 세트를 사용한다.

② GUIDE(Graphical User Interface for Databass Exploration)는 SUBJECT와 SEEDIS의 사용자 인터페이스(Interface)로 사용되는데 SUBJECT의 디렉토리 역할, 도움 메시지(help message), 데이터 베이스 스키마의 적절한 위치의 zooming기능, 중간결과를 가진 부분적 질의기능 등의 特徵을 포함한 그래픽 도구이다.

③ SAM*는 7개의 다른 연관(association)타입을 가지고 세트, 행렬, 벡터, 시계열 등과 같은 복잡한 데이터 타입들을 지원하며 실제구현은 G(generalized:일반화) - relation을 사용하였다.

④ RAM의 infological framework은 대상(object)들의 그룹에 관한 초기 메시지들을 사용하며, 통계 조사들을 나타내기 위해 n-cube의 박스구조가 사용된다.

⑤ SSDB(System for Statistical Database)는 개념모델로서 DA(Data Abstraction)모델을 변형한 HODM(Heterogeneous Operational Data Model)을 설계했으며, 질의어로는 STBE를 사용하였다.

第 3 章

프로토타입 (prototype)의 統計 데이터 베이스 관리시스템 (S-M 시스템) 의 設計 및 具現

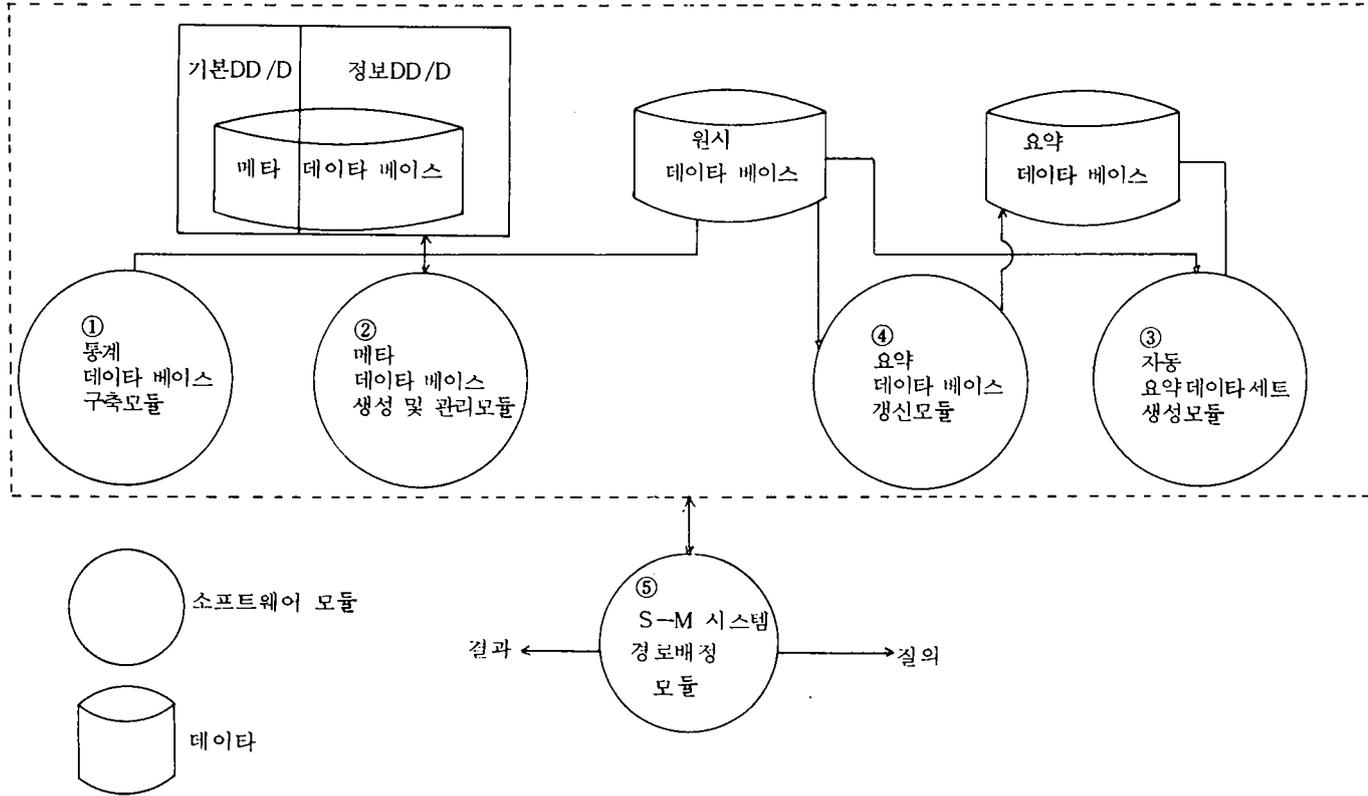
통계 데이터 베이스의 효율적인 처리를 위하여 요약 데이터 베이스의 관리와 메타 데이터 베이스의 관리를 위해 데이터 사전 / 디렉토리 (Data Dictionary/Directory)가 첨부된 통계 데이터 베이스 관리 시스템 (이하 이를 S-M 시스템이라 칭한다. S-M은 통계 데이터 베이스 관리 (Statistical Data Bass Management)의 약자임과 동시에 요약 (Summary) 과 메타 (Meta) 데이터 베이스의 약자인 중복의 의미를 갖는다.)을 프로토타입 (Prototype)으로 設計 및 具現된 실험적 시스템을 제시한다.

1. S-M 시스템의 基本構造

제안된 S-M시스템이 관리하는 원시 데이터 베이스, 요약 데이터 베이스 등은 제시된 그래프형태의 논리모델 (Logical model)에 의해 VSAM (vertual storage access method)화인의 저장구조를 이용하여 저장하고 이용된다. 統計 데이터 베이스의 효율적인 처리를 위한 統計 데이터 베이스 관리 시스템인 S-M시스템의 구성은 <圖 3-1>과 같다.

S-M시스템은 5개의 모듈로 구성하는데, 각 모듈이 수행하는 기본적인 기능은 다음과 같다.

圖 3 - 1 S-M시스템의 기본구조



① 統計 데이터 베이스 구축 모듈 : 원시 데이터의 수치 데이터 (measured Data)를 논리모델의 계층구조에 의한 2 차원 행렬로 VSAM 화일에 저장하여 원시 데이터 베이스를 구축한다.

② 메타 데이터 베이스 생성 및 관리모듈 : 메타 데이터를 효율적으로 처리할 수 있는 데이터 사전 / 디렉토리 (Data Dictionary/Directory) 형태의 메타정보의 저장 및 사용자의 요구에 따라 필요한 메타데이터정보를 출력한다.

③ 자동 요약 데이터 세트 생성 모듈 : 원시 데이터 베이스로부터 논리모델의 계층구조에 의해 요약 데이터 베이스를 자동 생성한다.

④ 요약 데이터 세트 갱신 모듈 : 자동적으로 시스템에 의하여 생성된 요약 데이터 세트와는 다른 構造의 요약세트를 요구하는 사용자에게 원시 데이터 베이스로부터 새로운 요약세트를 생성하여 제공하며, 동시에 요약 데이터 베이스를 갱신한다.

⑤ S-M시스템 경로 배정 모듈 : 사용자의 질의에 따라 원시 데이터 베이스 또는 요약 데이터 베이스로부터의 검색 및 요약 데이터 베이스의 경로를 결정하여 그 경로를 통하여 해당 모듈을 운영하여 결과를 출력시킨다.

프로토타입의 S-M시스템의 설계 및 구현시 농업조수입 (농림수산부 농가경제조사결과보고 참조) 데이터를 실례로 하여 수행한다.

2. 논리 모델의 그래프 표현 (Graph Representation)

가. 문제제기

앞에서 설명한 統計 데이터 베이스의 특징과 문제점에서 본바와 같이, 통계 데이터인 농업 조수입 데이터의 범주속성 (category attribute)을 그대로의 형태로 데이터 베이스에 저장하기에는 <圖 3-2>에서 보듯이 반복성의 문제와 대량성의 문제로 기억장소와 액세스의 비효율성을 수반하게 된다. 이와 같은 문제의 해결방법으로 <圖 3-3>과 같은 2 차원 행렬로 축

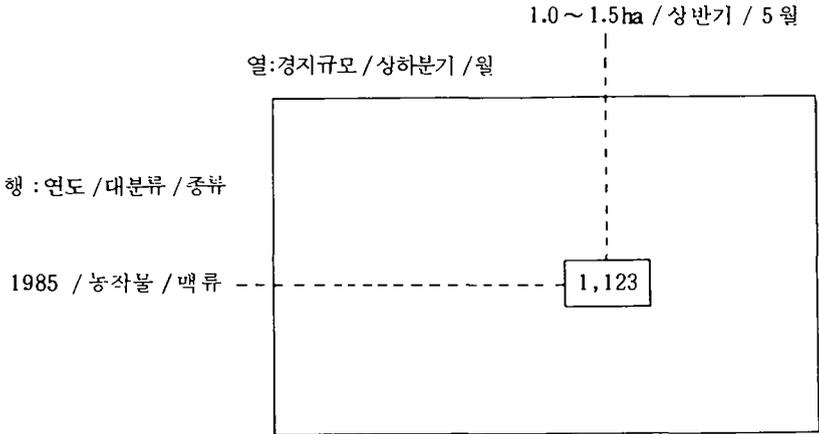
정데이터를 저장하며 매개 데이터는 메타 데이터 베이스에 단 1번만 저장하며 배열 선형화(array linearization)방법에 의해 빠른 속도로 액세스한다. 여기서 각 행과 열의 데이터의 구성은 각각이 가장 적절한 계층 구조에 의해 순서가 결정되어 表現된다. 즉, < 圖 3 - 4 > 와 같은 그래프형태로 모델링을 하여 자동에그리게이션(automatic aggregation)과 메타 데이터를 데이터처럼 運營하며, 또한 요약 데이터 베이스의 쉬운 運營을 가능하게 한다.

圖 3 - 2 농업 조수입 원시 데이터

연 도	매 개 류	종 류	경지규모	상하분기	월 별	조 수 입
1984	농 작 물	비 곡	0.5 ha미만	상반기	1	91786
"	"	"	"	"	∴	∴
"	"	"	"	"	6	∴
"	"	"	"	하반기	7	∴
"	"	"	∴	"	∴	∴
"	"	"	∴	"	1	70042
"	"	"	0.5~1.0	상반기	∴	∴
"	"	"	∴	"	6	93636
"	"	"	∴	하반기	7	∴
"	"	매 류	∴	∴	∴	∴
"	∴	∴	∴	∴	12	∴
∴	∴	∴	∴	∴	∴	∴
∴	비농작물	축 산 물	0.5ha미만	∴	∴	∴
∴	∴	∴	∴	∴	∴	∴
1985	농 작 물	미 곡	∴	∴	∴	∴
∴	∴	∴	∴	∴	∴	∴

← 매개데이터 (범주속성) → 측정데이터 (요약속성) *

圖 3 - 3 농업 조수입을 위한 행렬표시(matrix representation)



나. Semantics를 위한 그래프표현(Graph representation)

의미개념을 그래프형태로 표현한 LBL연구소에서 개발한 SUBJECT시스템의 방법론을 적용하여 논리모델을 계층구조의 그래프표현으로 제시해 보겠다.

노드(node)에는 크로스프로덕트(cross product)노드와 클로스터(cluster)노드의 2종류가 있으며, 이 노드들은 연결되어 directed acyclic graph를 구성하고 있다. 즉, <圖 3-2>와 <圖 3-3>의 농업조수입 데이터가 <圖 3-4>와 같은 그래프형태로 표현된다. “X”로 표기된 노드들은 크로스프로덕트 노드이며, “C”로 표기된 노드들은 클로스터 노드이다.

X-노드로 표기된 크로스프로덕트 노드는 統計 데이터 베이스의 범주속성의 다차원 성질을 표기한 것이다. 이러한 노드는 또한 큰 크로스프로덕트의 복잡성을 감소시키기 위한 다계층을 表示하기도 한다. 예를 들어 7원소의 한 개의 크로스프로덕트 노드가 3개와 4개의 원소를 가진 2개의 크로스프로덕트 노드로 각각 표현된 뒤, 2개의 크로스프로덕트 노드가 상위수준의 크로스프로덕트 노드에 의해 연결되어 나타내진다. 질의를 위한 크로스

프로덕트 노드의 semantics 는 다음과 같다. 즉, 선택되지 않은 크로스프로덕트 노드의 모든 가지들이 에그리게이션 (aggregation) 된다. 또한 크로스프로덕트 노드 그 자체의 선택은 모든 가지들을 에그리게이션 한다. 결국 에그리게이션을 위해 선택된 크로스 프로덕트 노드들은 선택된 값들과 선택되지 않은 가지들의 모든 값들과의 조합에 의하여 그 값을 形成한다.

C-노드로 표현된 클로스터 노드들은 데이터 공간의 주어진 차원에서 원소의 갯수가 너무 많을 때 그것을 복수의 수준으로 표기하여 한 번에 원소의 작은 부분만을 처리하려는 속성을 표기한 것이다. 질의를 위한 클로스터 노드의 의미는 크로스 프로덕트 노드와는 달리 질의에 포함된 선택된 가지들만 에그리게이션되고 선택되지 않은 가지들은 에그리게이션되지 않는다. 한편 클로스터 노드 그 자체의 선택은 모든 가지들을 에그리게이션 한다.

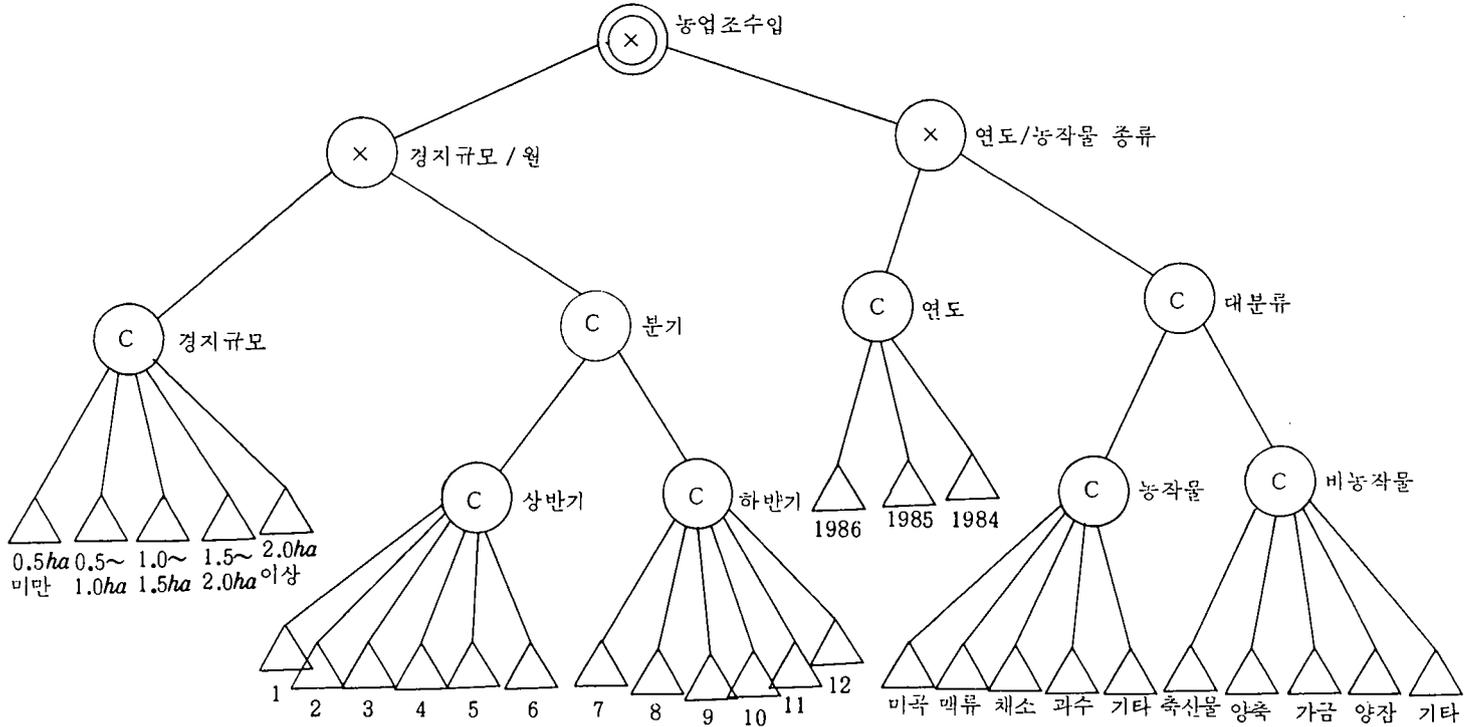
실제적으로 클로스터 노드들은 복잡한 범주속성의 계층구조와 요약속성의 수집 (collection) 을 위해 사용되며, 크로스 프로덕트 노드들은 범주속성의 합성키 (composite key) 들을 표현하기 위해 사용된다.

다. 特 徵

위에서 논한 방법에 의한 논리모덴의 그래프 표현의 特徵은 다음과 같다. 첫째, 그래프 구조가 시각적이고 의미묘사가 정확하다.

둘째, 사용자는 노드들의 타입을 알 필요 없이 각 노드들의 의미규칙 (semantic rule) 들을 적용하여 요약속성들의 자동 에그리게이션 (automatic aggregation) 을 제공한다. 셋째, 이러한 자동 에그리게이션에 의하여 요약 데이터 베이스 구성을 쉽고도 빠르게 할 수 있다. 넷째, 범주 데이터의 반복성 문제를 해결할 수 있다. 다섯째, 사용자에게 데이터 베이스 내용을 계층메뉴형태로 제공함으로써 사용자가 이름이나 약어들을 기억할 필요가 없는 편의성을 지원한다.

圖 3 - 4 농업 조수입의 Semantics를 위한 그래프 표현



3. 메타 데이터 베이스의 設計

가. 메타 데이터의 定義 및 特性

① 定 義

메타 데이터란 데이터에 관한 情報을 총칭하는 것으로 다양한 방법으로 검색되고 운영되고 표현되는 데이터의 내용과 조직에 관한 체계적이고 설명적인 정보의 집합이다. 수치 데이터를 기준으로 볼 때는 범주데이터도 개별 수치 데이터의 성격을 규정하는 메타 데이터로 취급하기도 한다. 이러한 메타 데이터가 포함하고 있는 정보를 살펴보면 다음과 같다.

- 명칭과 약어 (Names 와 Aliases) : 명칭은 메타 데이터 엔티티들을 위한 유일한 식별자 (identifier)이며, 약어나 동의어 등은 명칭을 대체할 수 있는 식별자이다.
- 분류와 설명정보 (Labeling 과 Descriptive Information) : 명칭과 약어를 보충하는 분류, 온라인 타이틀, 주석, 주제 색인 용어, 코드화되어 있는 정보들을 포함한다.
- Data Derivation 과 Quality : 특수한 필드나 어트리뷰트에 관한 Derivation 과 Quality 정보는 작성과 수정의 절차와 역사, 소오스 인용 및 신뢰성평가 등을 포함한다.
- 보안사항 (Security Specification) : 보안사항은 입력, 삭제, 첨가, 그리고 수정 등의 기능을 보유한 프로그램이나 사용자의 종류와 범위, 이용한계 등을 지적한다.
- 논리구조 (Logical Structure) : 구조정보는 데이터의 논리적모델의 구조 뿐 아니라 다른 사용자들에 의해 다양한 뷰 (views)들과 배열과 같은 복잡한 데이터 타입들을 묘사한다.
- 접근경로와 연결명세 (Access Path 와 Linkage Specification): 메타 데이터는 다른 엔티티들이 어떻게 연결되어 있는지 등을 설명한다.

- 처리절차 (Processing Procedures) : 데이터의 변환, 오류시 표기, 구간화 과정 등을 표기한다.

2 특 성

메타 데이터의 사용을 기능별로 살펴보면 데이터 定義, 문서화 (documentation), 데이터 선택, 데이터 운영 (Data Manipulation)과 데이터 디스플레이 (Data Display)로 설명할 수 있다. 이러한 메타 데이터의 특성은 다음과 같다.

- 메타 데이터의 데이터 타입 : 메타 데이터는 주로 도서목록과 같은 문서화 (textual) 형태로서 효과적인 사용형태는 표준적이고 고정된 길이의 데이터 타입들 뿐만 아니라 가변 길이 텍스트 같은 확장되고 복잡한 데이터 타입들도 포함한다.
- 메타 데이터 構造 : 융통성 있는 확장 가능 계층 데이터 구조 (open-ended hierarchical data structure)이다. 각 노드들은 메타 데이터 엔티티들의 다른 타입들의 정보의 계층 모임을 표시하고, 이 노드들은 명백한 연결 (linkage)이나 연합 (association)을 나타내는 선 (line)에 의하여 구조를 형성한다.
- 메타 데이터에서의 논리연결 (Logical Linkage) : 메타 데이터는 보다 더 확장된 정보나 다른 영역의 키나 포인터로 제공될 다른 메타 데이터의 어트리뷰트들의 이름을 포함하여 연결시키며, 또한 데이터 베이스로부터 데이터 원소 수준으로 어떤 총체적 메타 데이터 어트리뷰트들의 자동계승 (automatic inheritance)의 기능을 가진다.
- 메타 데이터의 액세스형태 : 대부분의 데이터 베이스에서 사용자는 데이터의 선택을 위해 메타 데이터의 일부분만이 자주 사용된다.
- 메타 데이터 갱신 : 메타 데이터 값의 갱신은 새로운 데이터 베이스가 생성될 때 그에 따른 메타 데이터 값들이 첨가되며, 또한 존재하는 데이터 베이스의 구조, 화일 또는 약어, 첨가 등의 변화에 따라 변경된다.

나. 메타 데이터 베이스의 構成

메타 데이터 베이스는 S-M 시스템의 실제운영을 위한 모든 정보 특히

요약 데이터 베이스를 효율적으로 운영하기 위한 정보들을 저장하며, 데이터를 정의하고 서술하는 데이터 사전과 데이터 베이스 각 성분의 위치와 구조적 정보를 제공하는 데이터 디렉토리가 결합된 데이터 사전/디렉토리(Data Dictionary/Directory : DD/D)의 형태로 구성된다. 이러한 데이터 사전/디렉토리에서 사용하는 메타 정보는 기본 메타 데이터와 정보 메타 데이터로 나누어진다.

기본 메타 데이터는 화일의 이름 및 범주 데이터의 이름 및 코드화된 범주 데이터의 값과 레코드의 갯수 각 필드의 크기 등이 저장되며, 정보 메타 데이터는 그래프형태의 구조를 명시하고 각 기본 메타 데이터의 구간화, 확장된 필드 속성, 통계적으로 유도된 데이터, 수치 분류화 정보 등이 저장된다. 현 시스템에서는 기본 메타 데이터와 일부 정보 메타 데이터만이 저장되었다. 앞으로 요약 데이터 베이스의 효율적 이용을 위한 확장되고 계층화된 형태의 메타 데이터 베이스에 대한 연구가 보완되어야겠다. 이를 위해 실시간 응답이 가능하고 통계 데이터의 분석에 적합한 <圖 3-5>와 같은 계층구조형태인 데이터 사전/디렉토리 시스템의 연구를 진행시켜야겠다.

다. 메타 데이터 베이스의 管理와 S-M 시스템의 運營

1 사용자 入力

메타 데이터 베이스를 통하여 해당화일을 찾아 그래프 형태의 구조를 보고 검색 하고자 하는 데이터에 관련된 명칭들의 코드(<表 3-1> 참조)를 찾아 입력시킨다.

(예제 1)

1984년 비농작물 가금의 경지규모 1.5~2.0 ha의 6월달의 조수입의 값을 원하면 메타 데이터 베이스를 통하여 해당코드를 찾아 入力시킨다.

1	2	3	4	1	6
연 도	대분류	종류	경지규모	상하분기	월

圖 3 - 5 계층구조의 데이터 사전/디렉토리의 구성

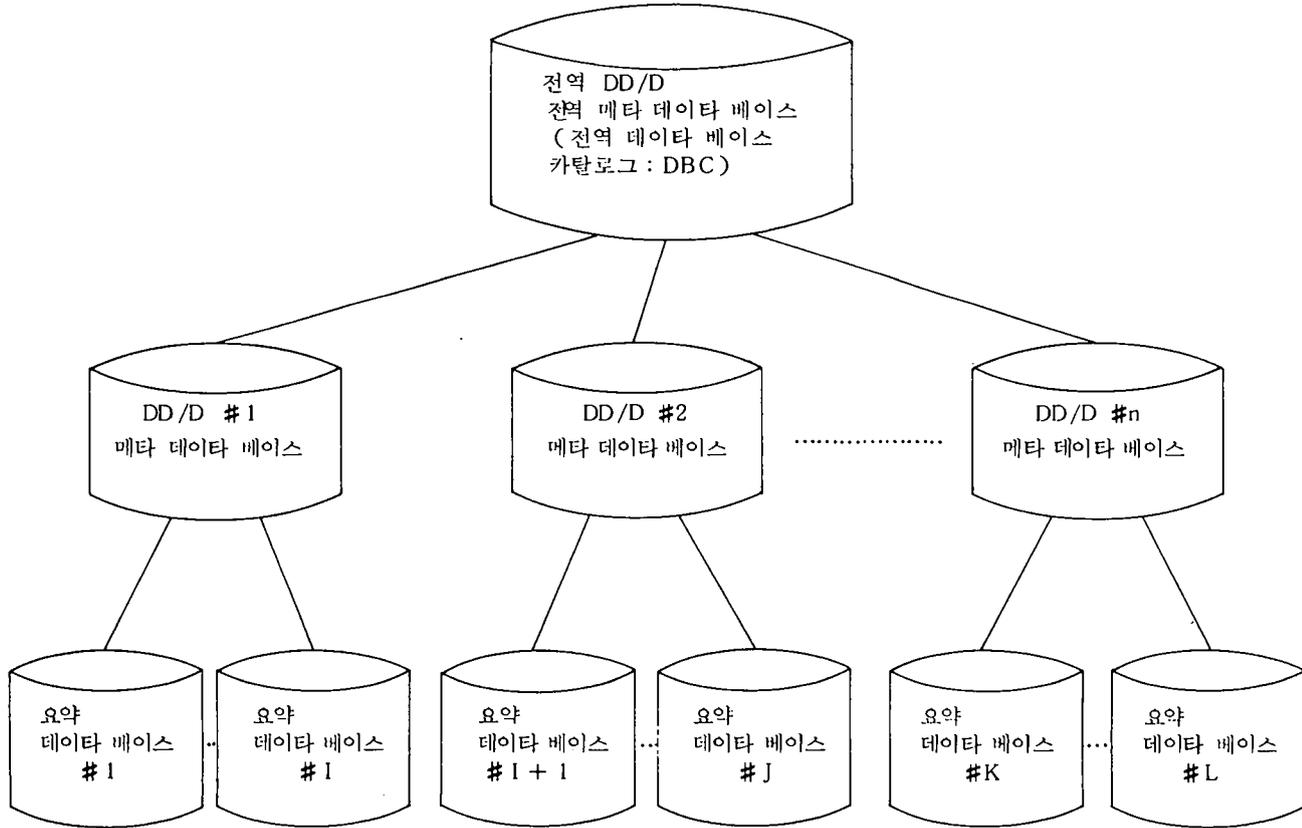


表3 - 1 농업 조수입 데이터의 코드화일

범주 데이터의 코드화 순서 및 범위							
코드화 순서	CODE (1)	CODE (2)	CODE (3)	CODE (4)	CODE (5)	CODE (6)	
해당 데이터	연도	대분류	종류	경지규모	상하분기	월	
범위	0 ~ 3	0 ~ 2	0 ~ 5	0 ~ 5	0 ~ 2	0 ~ 6	
범주 데이터의 코드값							
범주 데이터							
연도 CODE (1)	0 1984 ~ 1986의 요약값	1 1984	2 1985	3 1986			
대분류 CODE (2)	0 농작물과 농작물 이외의 수입을 합한 요약값	1 농작물	2 비농작물				
CODE(2) =1경우 종류 CODE(3)	0 농작물 요약값	1 미곡	2 맥류	3 채소	4 과수	5 기타	
CODE(2) =2경우 종류 CODE(3)	0 비농작물 요약값	1 축산물	2 양축	3 가금	4 양잠	5 기타	
CODE(4) 경지 규모	0 경지규모 요약값	1 0.5 ha 미만	2 0.5 ~ 1.0 ha	3 1.0 ~ 1.5 ha	4 1.5 ~ 2.0 ha	5 2.0 ha 이상	
CODE(5) 상하 분기	0 상하분기를 합친 그 해의 요약값	1 상반기 (1월 ~ 6월)		2 하반기 (7월 ~ 12월)			
CODE(5) =2경우 월 CODE(6)	0 상반기 요약값	1 1월	2 2월	3 3월	4 4월	5 5월	6 6월
CODE(5) =2경우 월 CODE(6)	0 하반기 요약값	1 7월	2 8월	3 9월	4 10월	5 11월	6 12월

(예제 2)

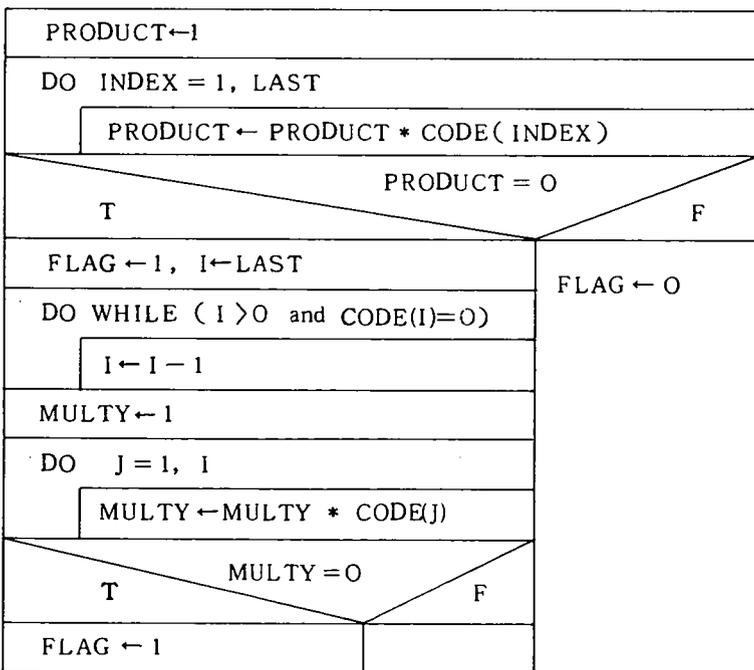
요약 데이터 베이스에서 검색하기를 원하면 그 부분의 코드값에 0을 입력시키면 된다. 만약 1985년 농작물의 값을 원하면 나머지 데이터의 코드값을 0으로 하면 된다.

1	1	0	0	0	0
---	---	---	---	---	---

② 시스템의 운영

메타 데이터 베이스의 활용을 통한 사용자 질의에 해당하는 코드가 입력되면 우선 S-M 시스템의 경로 배정모듈에서 경로결정을 위한 플래그(flag) 값을 생성한다. 이러한 경로결정 알고리즘을 나씨-슈나이더만 (Nassi-sunaideman : N-S) 도표로 나타내면 <圖 3-6> 과 같다.

圖 3 - 6 경로 결정 N-S도표

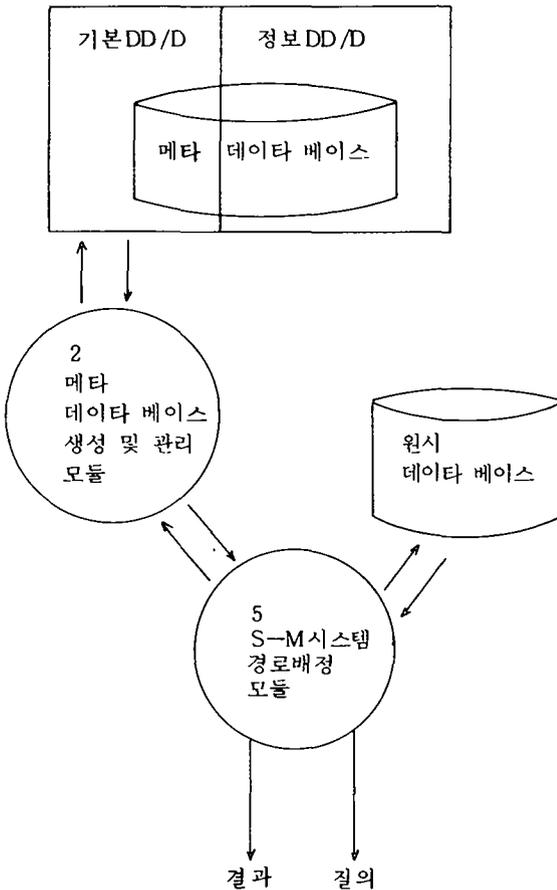


이 모듈에 의해 결정된 플래그 값에 의한 S-M 시스템의 경로를 살펴보면 아래와 같다.

가) A 경로

플래그 값이 0인 경우로 원시 데이터 베이스에서 직접 검색한다. <圖3-7>과 같은 경로로 운영된다.

圖 3 - 7 A경로 운영 과정

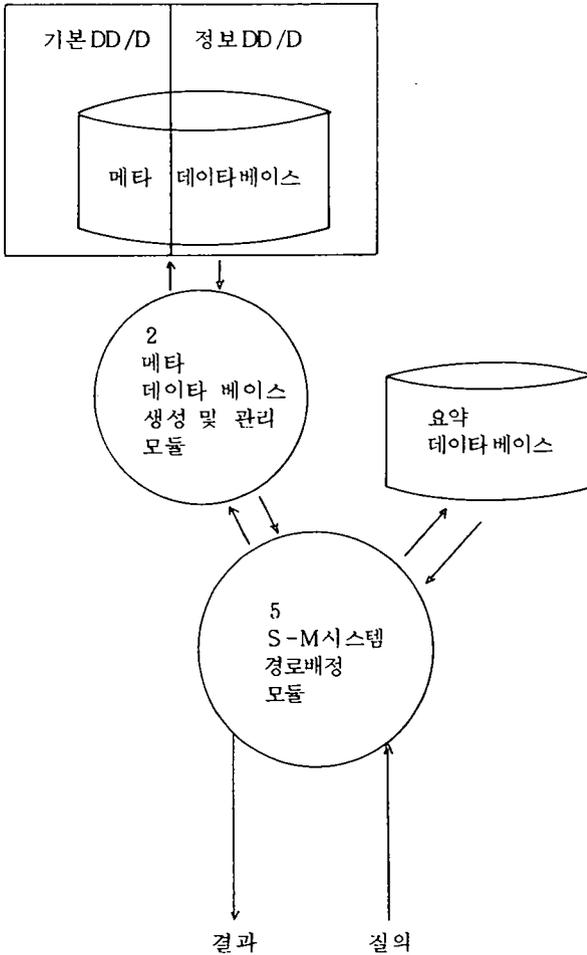


나) B 경로

플래그 값이 1인 경우로 시스템에 의해 미리 제공된 계층 구조에 의해 자

동 생성된 요약 데이터 베이스로부터 검색한다. <圖 3-8>과 같은 경로로 運營된다.

圖 3-8 B경로 운영 과정

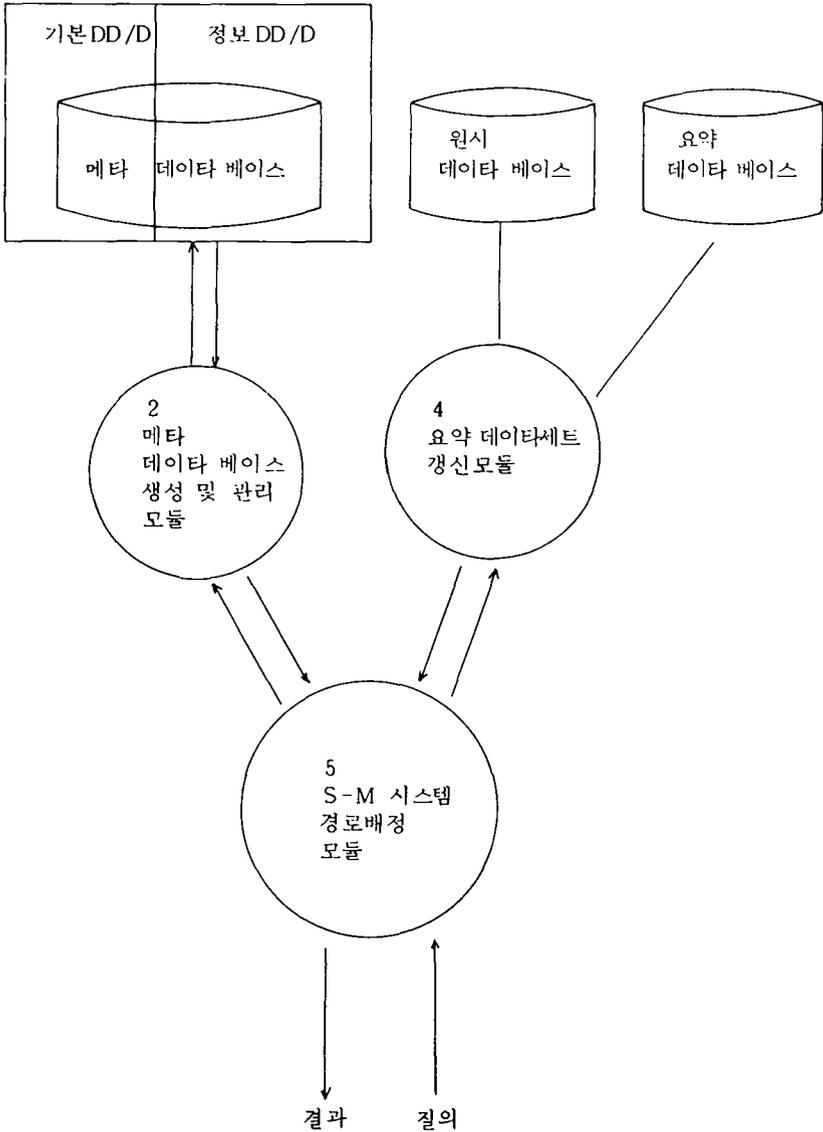


다) C-경로

플래그 값이 -1인 경우로 사용자가 자동생성된 요약 데이터 베이스와는 다른 구조의 요약세트를 요구할 때 원시 데이터 베이스로부터 새로운 요약

세트를 생성하여 제공하며, 또한 요약 데이터 베이스를 갱신한다. 단 다른 사용자에게 의해 이미 저장된 값은 요약 데이터 베이스에서 검색만 한다.(〈圖 3-9의 -----표시〉). 이 경로의 運營은 〈圖 3-9〉와 같다.

圖 3-9 C경로 운영 과정



4. 요약 데이터 베이스의 設計

가. 요약 데이터 베이스의 필요성

일반적인 범용 데이터 베이스 관리 시스템은 사용자에게 저장된 데이터에 대한 풍부한 데이터 관리 기능을 제공하는 반면 통계적 요약정보는 제공하지 않는다. 그러나 統計 데이터 베이스에서는 실제적으로 주로 일부분이나 요약된 것만이 통계 분석을 위해 사용된다. 일반적으로 통계 응용에 이용되는 데이터 베이스가 방대한 양의 자료를 저장하고 있고 통계분석작업이 반복적인 수행절차로 수행된다는 사실을 고려할 때 이러한 방대한 양의 원시 데이터 베이스를 여러번 접근하는 것은 막대한 비용을 가져오며, 응답성도 보장되지 않는다.

그리하여 아이템의 수나 범주속성들을 제한함으로써 얻어지는 요약 데이터 세트들(summary data sets)의 다수의 모임인 요약 데이터 베이스를 생성하여 한번 정의된 요약 데이터에 대해서는 그 요약 데이터를 접근하는 질의들은 원시 데이터를 접근하지 않고 질의의 결과를 바로 사용자에게 제공할 수 있도록 하는 기능은 통계 데이터 베이스의 효율적 이용이란 측면에서 중요하게 평가된다.

그리하여 본 시스템에서는 요약 데이터 베이스를 생성하여 여러 가지 통계량들을 통합적으로 저장함으로써 이러한 통계량들을 요구하는 질의에 대해서는 요구시마다 원시 데이터 베이스를 접근하지 않고 직접 요약 데이터 베이스만을 접근하므로 빠른 응답시간을 보장한다.

본 요약 데이터 베이스에서 구하는 수치 속성에 대한 기술 統計量은 최대값, 최소값, 합, 평균, 표준편차 등이다 (〈圖 3-10〉참조).

圖 3 -10 요약 데이터 베이스의 기술통계량

<u>KEY</u>	<u>MIN</u>	<u>MAX</u>	<u>SUM</u>	<u>MEAN</u>	<u>SD</u>
120000	.	.	1940291	.	.
.
.
.

나. 요약 데이터 베이스의 構成

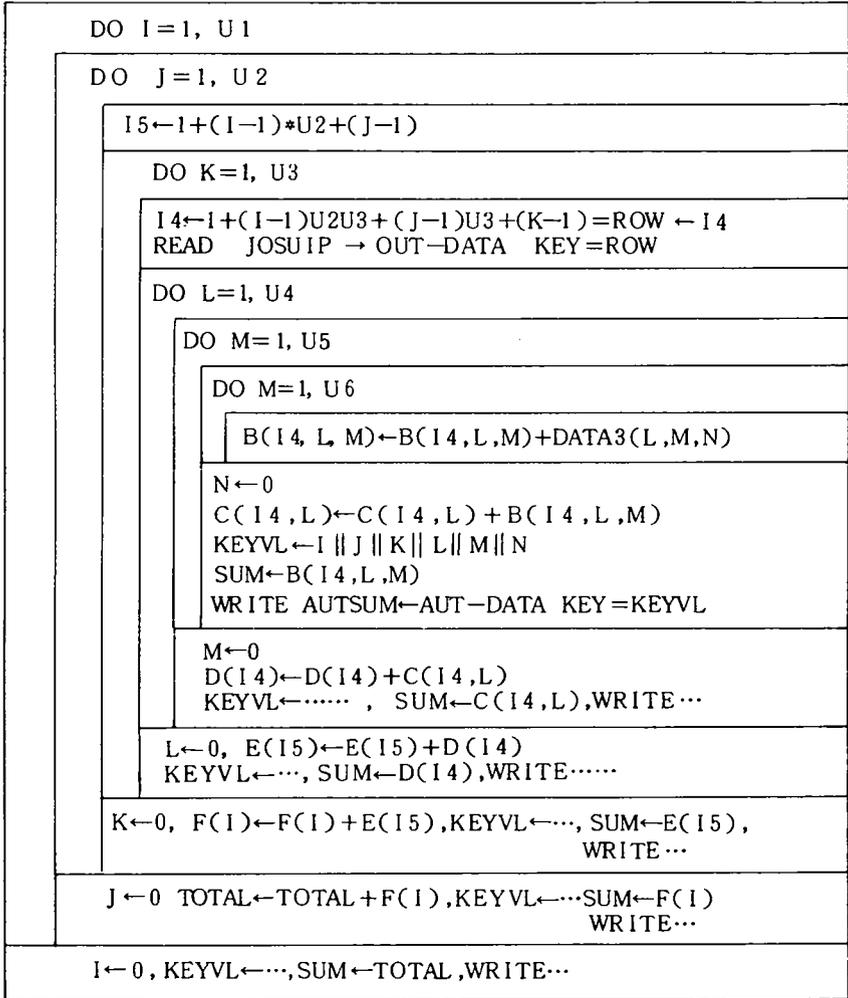
우선 시스템에서 <圖 3 - 4 > 의 화일구성의 계층구조에 의해 요약 데이터를 자동 생성하여 요약 데이터 베이스를 구성하는 과정으로 <圖 3 -11 > 과 같은 경로가 운영한다. 이 과정은 원시 데이터 베이스가 생성될 때 단 한번만 원시 데이터 베이스를 접근하여 구조에 맞는 통계량을 구하기 위하여 이용된다. 이러한 과정의 알고리즘을 나찌-슈나이더만 圖表로 나타내면 <圖 3 -12 > 와 같다. 또한 요약 데이터 베이스가 생성될 때마다 그에 관한 정보가 메타 데이터 베이스(데이터 사전 / 디렉토리의 형태)에 저장되어 요약 데이터 베이스가 효율적으로 이용되게 해야 한다.

한편 S-M 시스템의 실제 運營時 사용자가 메타 데이터 베이스의 검색 결과 필요한 요약 데이터 세트가 존재하지 않을 때 즉, 자동생성된 요약 데

圖 3 -11 자동 요약 데이터 세트 구성 과정

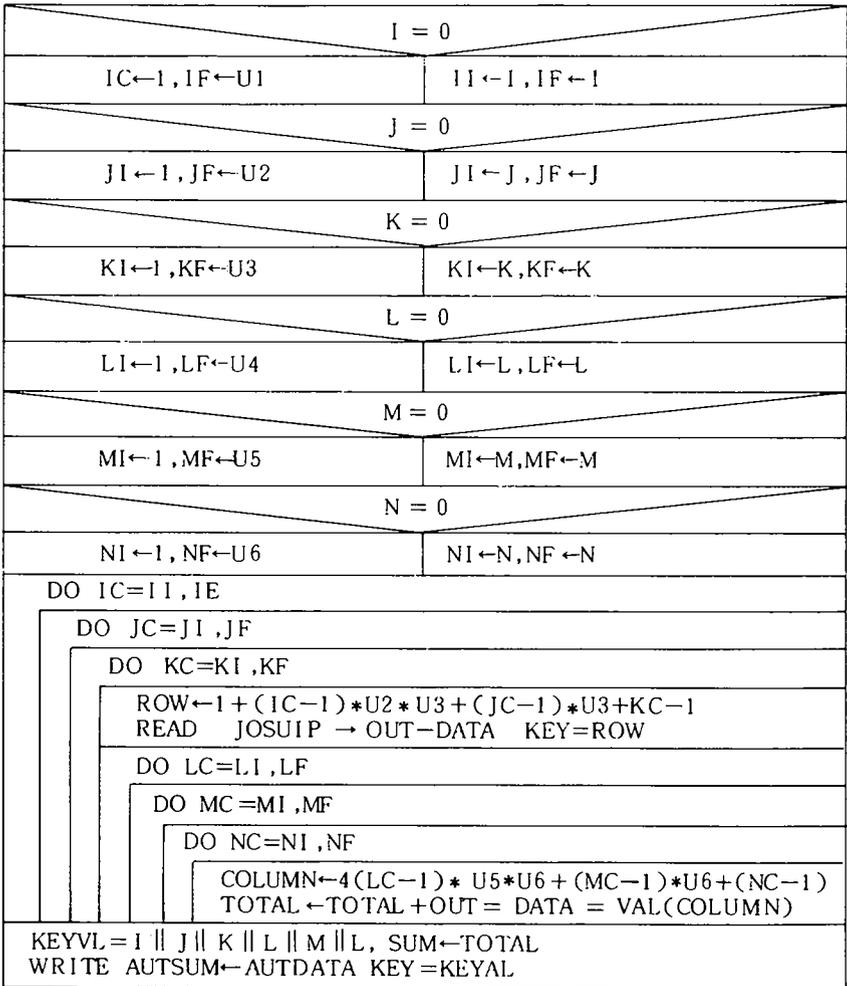


圖 3 - 12 자동 요약 데이터 세트 생성 N-S도표
(기술 통계량은 합안구함)



이타 베이스와는 다른 구조의 요약세트를 요구할 때마다 원시 데이터베이스로부터 새로운 요약 데이터 세트를 생성하여 제공하며, 또한 요약데이터 베이스를 갱신한다. 이 과정은 <圖 3 - 9>에서 이미 설명하였다. 이러한 과정의 알고리즘을 나쉴-슈나이더만 도표로 나타내면 <圖 3 - 13>과 같다.

圖 3 - 13 사용자 요구 요약 데이터 세트 생성 N-S도표



다. 요약 데이터 베이스의 長點과 고려사항

요약 데이터 베이스는 하나의 데이터분석 환경하에서 반복적으로 수행되는 통계 분석을 지원하기 위하여 데이터 세트에 대해서 다양한 통계량을 미리 계산하여 저장한다. 그리하여 사용자는 원시 데이터 베이스의 접근없이 분석작업을 수행할 수 있어 요구하는 統計量의 요약정보를 빠른 응답시간내에 제공받을 수 있다.

이러한 요약 데이터 베이스를 효율적으로 사용하기 위해서는 우선 가장 사용빈도가 높은 요약 데이터 세트가 자동생산되도록 적절한 계층구조가 선행되어야 한다. 또한 사용자의 다양한 요구에 따라 정보를 알기 쉽게 테이블 형태로 디스플레이하기 위하여 각 요약 데이터 세트에 대응하는 요약테이블(summary table)의 구현에 관한 研究가 보완되어야겠다.

5. 具現 結果 및 시스템의 文書化(Documentation)

가. 具現 結果

S-M 시스템을 VSAM 화일과 PL/I 언어를 이용하여 IBM -4341에 실제구현한 결과는 아래와 같다. 단 조수입 데이터 베이스를 예로 하여 출력한 형태는 메타 데이터 베이스에서 기본 메타 데이터와 요약 데이터베이스의 통계량중 합의 값만 인쇄한 경우이다.

① 사용자 질의가 A경로인 경우

원시 데이터 베이스에서 직접 검색을 원하는 경우로 메타 데이터 베이스에서 해당코드 값을 찾은 후 경로배정 모듈에서 플래그값을 0으로 세트한 뒤 A 경로를 통하여 運營되는데, 그 출력 값은 아래 <圖 3 - 14 >와 같다.

圖 3 - 14 A경로 출력

```

##### INPUT DATA #####
2      1      3      4      2      5
# 1985  # 농작물 # 채소   # 1.5-2.0ha # 하반기 # 11-월

SEARCH SELECT 0

FLAG(0) OUTPUT 213425      135701
    
```

② 사용자 질의가 B경로인 경우

자동생성된 요약 데이터 베이스로부터 검색하는 경우로 메타 데이터 베

스의 코드 값으로부터 플래그 값을 1로 세트한 뒤 B 경로로 운영되는데, 그 출력 값은 아래 <圖 3-15>와 같다.

圖 3-15 B경로 출력

```
##### INPUT DATA #####
1      2      0      0      0      0
# 1984 #비농작물 #비농작물 합 #경지규모합 #상하분기 합 #1월-12월합
SUBPROGRAM SELECT 1
FLAG(1) OUTPUT 120000 1940291
```

③ 사용자 질의가 C경로인 경우

자동생성된 요약 데이터 베이스와는 다른 구조의 요약 데이터 세트를 요구하는 경우로 메타 데이터 베이스에서 해당코드 값을 찾은 후 경로배정 모듈에서 플래그 값을 -1로 세트한 뒤 C경로로 운영되는데, 그 출력 값은 아래 <圖 3-16>과 같다.

圖 3-16 C경로 출력

```
##### INPUT DATA #####
1      0      0      3      0      0
# 1984 #작물전체합 #미곡~기타합 #1.0-1.5ha #상하분기합 #1월-12월합
SUBPROGRAM SELECT -1
FLAG(-1) OUTPUT 100300 4587426
```

나. S-M 시스템 프로그램 概要

本節에서는 <圖 3-1>의 S-M 시스템의 기본구조에 따른 각 모듈별 프로그램의 개요가 설명되었다. 즉, 각 모듈별 프로그램名, 사용언어, 관련화일형태와 내용이 수록되어 있는 프로그램 일람표와 각 프로그램에 대한 설명과 入出力 상황이 수록된 프로그램 설명서를 작성하였다.

이러한 S-M 시스템의 文書化(Documentation)작업은 실제 시스템의 이용자가 시스템을 쉽게 이해하고 사용할 수 있도록 함과 동시에 S-M 시스

템의 유지 보수 (maintenance)와 확장을 가능케 하고 발전적 시스템으로의 연구를 용이하게 한다.

가. 프로그램 일람표

시스템名	통계 데이터베이스 관리시스템		作成者		確認	
段階名			作成日			
文 献 名	프로그램 일람표				PAGE	1
번호	프로그램名	step수	사용언어	관련화일		내 용
					I/O	
	PL1SDB		PL/I	JOSUIP (VSAM)	I	· 원시 조수입 데이터를 VSAM File에 출력하여 원시 조수입 D/B 구축
	PL100		PL/I	AUTSUM (VSAM)		· Summary D/B 구축을 위한 VSAM File의 정의 및 초기치 출력
	PL1SUM1		PL/I	AUTSUM (VSAM)		File 구성의 계층 구조에 의해 요약 데이터 베이스를 자동생성해서 VSAM File에 출력
	PL1IN		PL/I	JOSUIP AUTSUM (VSAM)		사용자의 요구에 따라 원시 D/B 또는 요약 D/B로부터의 검색 및 요약 D/B의 갱신의 경로를 결정하여, 실제 통계 데이터베이스 관리시스템을 운영
	PL1IN-Subprogram META-DB		PL/I			메타 정보의 저장 및 사용자의 요구에 따라 필요한 메타 정보출력
	PL1IN-Subprogram USERSUM		PL/I	AUTSUM (VSAM)		자동생성된 요약D/B와는 다른 구조의 요약세트를 요구하는 사용자에게 원시D/B로부터 새로운 요약세트를 생성하여 요약D/B를 갱신

시스템名	통계데이터베이스 관리시스템	作成者		確認																			
段階名		作成日																					
文 名	프로그램 설명서	프로그램名	PL1SDB	PAGE	1																		
개요																							
원시 조수입 데이터를 2 차원 행렬로 수치 데이터블 VASM File 에 저장하여 원시 조수입 D/B 구축																							
INPUT		치 리 과 정		OUTPUT																			
· 원시 조수입 데이터 INP-DATA 1		1. 아래와 같은 행렬구조를 가짐		· VASM File에 구축 OUT-DATA																			
KEY	VALUE(1)···VALUE(6)	연 : 경지규모 / 상하분기 / 월		ROW	COL(1)···COL(7)																		
99	9 (7) 9 (7)	행 30 60		99	9(7)···9 (7)																		
연도 / 대분류 / 종류																							
· INP -DATA 2																							
VALUE(1)···VALUE(6)		연도대분류 종류 경지규모 상하분기 월																					
9 (7) 9 (9)	99	<table border="1"> <tr> <td>I</td><td>J</td><td>K</td><td>L</td><td>M</td><td>N</td> </tr> <tr> <td>U1</td><td>U2</td><td>U3</td><td>U4</td><td>U5</td><td>U6</td> </tr> <tr> <td>(3)</td><td>(2)</td><td>(5)</td><td>(3)</td><td>(2)</td><td>(6)</td> </tr> </table>		I	J	K	L	M	N	U1	U2	U3	U4	U5	U6	(3)	(2)	(5)	(3)	(2)	(6)		
I	J	K	L	M	N																		
U1	U2	U3	U4	U5	U6																		
(3)	(2)	(5)	(3)	(2)	(6)																		
9 개		최대값																					
VALUE(6)···VALUE(6)		배열선형화에 의해																					
9 (7) 9 (7)	99																						
		2. KEY=ROW=1+(I-1) U2·U3+(J-1)·U3+(K-1)																					
		3. COL(PP)에서 P P = 1 + (L-1)·U5·U6+(M-1) U6+(M-1)																					
		4. ① COL(LL) = INP - DATA1 , VALUE(LL) ② LL=1+(IJ-1)·6+K K-1(JJ=2,10 and KK= 1,6), COL(LL)=INP - DATA .VALUE(KK)																					
보충설명																							
· 범주 데이터의 대량성과 반복성의 문제를 해결 · 행의 연도 / 대분류 / 종류를 배열 선형화 방법 (array linesization) 으로 계산하여 VSAM File 의 KEY 로 사용																							

시스템名	통계데이터베이스 관리시스템	作成者		確 認																													
段階名		作成日																															
文 書 名	프로그램 설명서	프로그램名	PLISUM1	PAGE	2																												
개 요																																	
<p>농업조수입 원시 D/B로부터 File 구성의 계층구조에 의해 요약 데이터 베이스를 자동생성</p>																																	
INPUT		처 리 과 정		OUTPUT																													
<p>• 농업조수입 원시D/B (JOSUIP)</p> <p>OUT-DATA</p> <table border="1"> <tr> <td>ROW</td> <td>COL(1)</td> <td>-----</td> <td>COL(6)</td> </tr> <tr> <td>99</td> <td>9 (7)</td> <td></td> <td>9 (7)</td> </tr> </table> <p>DATA3(5, 2, 6) 은 COL(1)~ COL(6)을 재정의</p> <p>• 각 코드별 최대값</p> <table border="1"> <tr> <td>U1</td> <td>U2</td> <td>U3</td> <td>U4</td> <td>U5</td> <td>U6</td> </tr> </table> <p>연 대 종 경 상 월 도 분 류 지 하 류 규 분 모 기</p>		ROW	COL(1)	-----	COL(6)	99	9 (7)		9 (7)	U1	U2	U3	U4	U5	U6	<ol style="list-style-type: none"> 주어진 계층구조에 의해 자동적으로 요약값이 생성되도록 다중 DO문을 구성 Key 값을 결정후 해당 농업조수입 데이터를 읽어들이 배열을 생성하여 해당 요약값을 위한 기억장소 확보 <ol style="list-style-type: none"> B(30,5,2) : 연도 / 대분류 / 종류 / 경지규모 / 상하 분기 C (30,5) : 연도 / 대분류 / 종류 / 경지규모 D(30) : 연도 / 대분류 / 종류 E(6) : 연도 / 대분류 F(3) : 연도 TOTAL : 농업조수입 전체 계산 후 요약 D/B에 출력 		<p>요약 D / B (AUTSUM)</p> <p>AUT-DATA</p> <table border="1"> <tr> <td>KEYVL</td> <td>SUM</td> <td>MEAM</td> </tr> <tr> <td>A(6)</td> <td>9(0)</td> <td>9 (8)</td> </tr> <tr> <td>MAX</td> <td>MIN</td> <td>SOV</td> <td style="background-color: #cccccc;"></td> </tr> <tr> <td>9 (0)</td> <td>9(8)</td> <td>9(8)</td> <td>A(0)</td> </tr> </table> <p>< 요약 D/B 기술 통계량 내용 ></p> <p>KEYVL : KEY 값 MIN : 어트리뷰트 값의 최소값 MAX : // 최대값 SUM : // 합 MEAN : // 평균 SDV : // 표준편차</p>		KEYVL	SUM	MEAM	A(6)	9(0)	9 (8)	MAX	MIN	SOV		9 (0)	9(8)	9(8)	A(0)
ROW	COL(1)	-----	COL(6)																														
99	9 (7)		9 (7)																														
U1	U2	U3	U4	U5	U6																												
KEYVL	SUM	MEAM																															
A(6)	9(0)	9 (8)																															
MAX	MIN	SOV																															
9 (0)	9(8)	9(8)	A(0)																														
보충설명																																	
<ul style="list-style-type: none"> 주어진 계층구조에 의해 요약 데이터 베이스를 자동 생성 프로그램의 변수나 배열명은 초기치를 0 로 세트 요약값을 원하는 부분의 코드는 0 값을 취함. 																																	

시스템명	통계데이터베이스 관리시스템	作成者		確 認												
段階名		作成日														
文 書 名	프로그램 설명서	프로그램명	PL11N	PAGE 3												
개 요																
<p>사용자의 요구에 따라 원시D/B 또는 요약D/B로부터의 검색 및 요약D/B의 생성의 경로를 결정하여 실제 S-M 시스템을 운영함.</p>																
INPUT		처 리 과 정		OUTPUT												
<p>사용자가 검색을 원하는 코드 CODE(6)</p> <table border="1" style="width: 100%;"> <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>연 도</td> <td>대 분 류</td> <td>종 류</td> <td>경 지 규 모</td> <td>상 하 분 기</td> <td>월</td> </tr> </table>		code (4)	code (2)	code (3)	code (4)	code (5)	code (6)	연 도	대 분 류	종 류	경 지 규 모	상 하 분 기	월	<ol style="list-style-type: none"> 1. 코드의 값중 요약데이터베이스의 통계량을 요구하는 O가 하나도 없는 경우는 'FLAG'의 값을 '0'으로 세트 2. 요약값을 요구하는 구조가 주어진 계층구조에 의할때, 즉 O코드값이 다른 숫자가 나오기 이전의 값으로 연속되어 있을 때 'FLAG'의 값을 "1"로 세트 3. 요약값을 요구하는 구조가 주어진 계층구조와 같지 않을때, 즉 O코드값이 순서없이 섞여 나올 때 'FLAG'의 값을 "-1"로 세트 4. 결정된 'FLAG' 값에 의 해 해당 경로를 걸쳐 운영 		<p>FLAG 값</p> <p>0 또는 1 또는 -1</p> <p>해당경로를 거침</p> <p>결 과</p>
code (4)	code (2)	code (3)	code (4)	code (5)	code (6)											
연 도	대 분 류	종 류	경 지 규 모	상 하 분 기	월											
보충설명																
<ul style="list-style-type: none"> • Prototype의 통계 데이터 베이스 관리 시스템을 농업조수입을 예제로 하여 구현시킴 • FLAG:0 → A경로 • FLAG:1 → B경로 • FLAG:-1 → C경로 <p style="text-align: right;">} 로 S-M 시스템을 운영</p>																

시스템名	통계데이터베이스 관리시스템	作成者		確認																															
段階名		作成日																																	
文書名	프로그램 설명서	프로그램名	PL11N-Sub-program METE-DB	PAGE	4																														
개요																																			
메타 정보의 저장 및 사용자에게 요구에 따라 필요한 메타 정보출력																																			
INPUT		처 리 과 정		OUTPUT																															
<p>· 메타 데이터 베이스를 통하여 해당 화일을 찾아그래프 형태의 구조를 보고 코드를 찾아 입력</p> <table border="1" style="margin-left: 20px;"> <tr> <td>I</td><td>J</td><td>K</td><td>L</td><td>M</td><td>N</td> </tr> <tr> <td>9</td><td>9</td><td>9</td><td>9</td><td>9</td><td>9</td> </tr> <tr> <td>연도</td><td>대분류</td><td>중류</td><td>경지규모</td><td>상하분기</td><td>월</td> </tr> </table>		I	J	K	L	M	N	9	9	9	9	9	9	연도	대분류	중류	경지규모	상하분기	월	<p>(Heta D/B)</p> <table border="1" style="margin-left: 20px;"> <tr> <td>YUNDO(0)</td><td>-----</td><td>YUNDO(3)</td> </tr> <tr> <td colspan="3" style="text-align: center;">A (5) A (5)</td> </tr> </table> <table border="1" style="margin-left: 20px;"> <tr> <td>HAWOL(0)</td><td>-----</td><td>HAWOL(6)</td> </tr> <tr> <td colspan="3" style="text-align: center;">A (6) A (6)</td> </tr> </table> <p>Ex) YUNDO(0) = 연도함 HAP' YUNDO(1) = '1984' YUNDO(2) = '1985' YUNDO(3) = '1986' 으로 저장되어 있어</p> <ul style="list-style-type: none"> · 1의 값에 있어 YUNDO (1)의 값을 출력시킴 · 나머지 코드도 전부 같은 방식으로 찾아 출력시킴 		YUNDO(0)	-----	YUNDO(3)	A (5) A (5)			HAWOL(0)	-----	HAWOL(6)	A (6) A (6)			<p>· 메타정보 출력</p> <p>① 기본메타 데이터 화일의 이름, 코드화 되어있는 범주데이터의 이름</p> <p>② 정보 메타 데이터 그래프 형태의 구조를 명시하고, 각 기본 메타 데이터의 구간화, 약어질, 단위</p>	
I	J	K	L	M	N																														
9	9	9	9	9	9																														
연도	대분류	중류	경지규모	상하분기	월																														
YUNDO(0)	-----	YUNDO(3)																																	
A (5) A (5)																																			
HAWOL(0)	-----	HAWOL(6)																																	
A (6) A (6)																																			
보충설명																																			
<ul style="list-style-type: none"> · 현재는 기본 메타 데이터만 구현되어 있음. · 메타정보의 효율적 이용을 위해 상하위 구조를 가진 데이터 사전 / 디렉토리의 구조로 구성되도록 보완적 연구가 필요함. 																																			

시스템名	통계데이터베이스 관리시스템	作成者		確認	
段階名		作成日			
文 書 名	프로그램 설명서	프로그램名	PL11N-Sub-program USERSUM	PAGE	5

개 요

자동생성된 요약D/B와는 다른 구조의 요약세트를 요구하는 사용자에게
 원시 D/B로부터 새로운 요약 세트를 생성하여 제공하며, 요약 D/B
 를 갱신.

INPUT	처 리 과 정	OUTPUT																																																										
• 농업조수입 원시D/B (JOSUIP) OUT-DATA <table border="1"> <tr> <td>ROW</td> <td>COL(1)</td> <td>-----</td> <td>COL(10)</td> </tr> <tr> <td>99</td> <td>9(7)</td> <td></td> <td>9(7)</td> </tr> </table> • 코드값 <table border="1"> <tr> <td>I</td> <td>J</td> <td>K</td> <td>L</td> <td>M</td> <td>N</td> </tr> <tr> <td>9</td> <td>9</td> <td>9</td> <td>9</td> <td>9</td> <td>9</td> </tr> <tr> <td>연</td> <td>대</td> <td>종</td> <td>경</td> <td>상</td> <td>월</td> </tr> <tr> <td>도</td> <td>분</td> <td>류</td> <td>지</td> <td>하</td> <td></td> </tr> <tr> <td></td> <td>류</td> <td></td> <td>구</td> <td>분</td> <td></td> </tr> <tr> <td></td> <td></td> <td></td> <td>모</td> <td>기</td> <td></td> </tr> </table>	ROW	COL(1)	-----	COL(10)	99	9(7)		9(7)	I	J	K	L	M	N	9	9	9	9	9	9	연	대	종	경	상	월	도	분	류	지	하			류		구	분					모	기		1. 주어진 각 코드값에 의해 요약값을 요구하 는 범주데이터를 결정. 2. Key 값을 결정한 후 해당 농업 조수입 데 이타를 읽어들임. 3. 요약을 요구하는 부분 의 코드만 DO문을 이용하여 값을 계산. 4. 요약D/B에 출력하여 요약D/B를 갱신	요약D/B 갱신 (AUTSUM) AUT-DATA <table border="1"> <tr> <td>KEYVL</td> <td>SUM</td> <td>MEAN</td> </tr> <tr> <td>A(6)</td> <td>9(10)</td> <td>9(8)</td> </tr> </table> <table border="1"> <tr> <td>MAX</td> <td>MIN</td> <td>SOV</td> <td style="background-color: #cccccc;"></td> </tr> <tr> <td>9(10)</td> <td>9(8)</td> <td>9(8)</td> <td>A(10)</td> </tr> </table>	KEYVL	SUM	MEAN	A(6)	9(10)	9(8)	MAX	MIN	SOV		9(10)	9(8)	9(8)	A(10)
ROW	COL(1)	-----	COL(10)																																																									
99	9(7)		9(7)																																																									
I	J	K	L	M	N																																																							
9	9	9	9	9	9																																																							
연	대	종	경	상	월																																																							
도	분	류	지	하																																																								
	류		구	분																																																								
			모	기																																																								
KEYVL	SUM	MEAN																																																										
A(6)	9(10)	9(8)																																																										
MAX	MIN	SOV																																																										
9(10)	9(8)	9(8)	A(10)																																																									

보충설명

- 사용자에게 의해 자동생성된 요약 D/B와는 다른 구조의 요약세트를
 요구할때 처음은 이 Subprogram에 의해 계산하여 생성하여 요약
 D/B에 저장
- 같은 요약세트를 다시 요구할때는 이미 저장된 요약 D/B에서 단
 지 검색만함. (메타 데이터 베이스를 통하여 경로를 결정함)

第 4 章

農業 데이터 베이스 構築을 위한 發展方向

1. 연구결과 요약

지금까지 우리는 2 章에서 農業 데이터 베이스가 統計 데이터 베이스 성격을 가지고 있음을 설명한 후 이의 효율적인 구축을 위하여 統計 데이터 베이스의 특성 및 지금까지 국내외 연구의 現況을 分析하였다. 또한 3 章에서 농업데이터 베이스를 위한 통계 데이터 베이스 관리 시스템의 발전방향의 구체적 제안을 위해서 실제로 S-M 시스템이라는 統計 데이터 베이스 관리 시스템을 프로토타입으로 설계하고 구현하였다.

이러한 S-M 시스템은 실제 農業 데이터 베이스 構築을 위해 본 소프트웨어 하우스에서 독자적인 통계 데이터 베이스 관리 시스템을 개발하기 위한 기초연구로서의 목적 보다는 농업 데이터 베이스 구축을 위해 기존의 연구 개발된 통계 데이터 베이스 시스템을 도입하기 위해 실제 시스템의 구현을 통한 문제점을 살펴 본 뒤 합리적으로 방향을 제시하기 위한 목적으로 개발하였다.

이러한 S-M 시스템의 長點은 다음과 같다. 범주속성의 반복성의 문제해결로 인한 기억장소의 절약과 빠른 액세스를 가져온다. 또한 요약 데이터세

트가 같은 데이터 베이스의 다양한 뷰(multiple view)들에 의해 쉽게 運 轉되며 자동 에그리게이션이 용이하다.

한편 이 시스템의 한계점은 제한된 질의어를 가지고 있고 요약 데이터 베이스의 정의기능이 정해진 구간화에 의한 제한점이 있다. 또한 그래프형태의 構造를 가지고 있어 뷰(view)정의시 데이터 화일의 구조 접근 경로를 알아야 하는 제약이 있다.

2. 發展方向

가. 제 안

지금까지의 연구를 토대로 農業 데이터 베이스 구축을 위한 소프트웨어에 대한 몇 가지 안을 제시해 보겠다. 단 이 제안들은 농림수산부 컴퓨터 시스템이 24 MB 이상의 시스템(예를 들어 IBM에서는 4381/22 이상)으로 교체되며, 현재의 상황으로 독자적인 개발은 어렵다는 결론에서 현존의 상용 데이터 베이스 시스템 또는 이미 개발된 독자적인 統計 데이터 베이스 시스템을 활용한다는 전제하에서 검토하였다.

① 제 1 안

統計 데이터 베이스 전용의 목적으로 독자적으로 개발된 시스템의 도입이다. 대표적인 시스템으로는 미국의 Berkeley 대학의 LBL(Lawrence Berkeley Laboratory)에서 개발된 SUBJECT, SEEDIS, GUIDE 등을 들 수 있다.

이 안의 長點으로는 統計 데이터 베이스의 特徵을 잘 표현하는 시스템의 도입으로 인해 기억 장소와 액세스의 효율을 높이고 사용자에게 사용의 편의성을 제공할 수 있다.

限界點은 상업용 데이터 베이스 관리 시스템이 아니므로 인사관리, 급여계산 등을 포함한 농림수산부 전체업무를 총괄하기에는 제한점이 많아 또 다른 상업용 데이터 베이스 시스템의 도입을 초래하므로 생기는 낭비적인 요소, 양 데이터 베이스 관리 시스템과 관련된 데이터 공유 및 디자인의 균형에 어

려움을 배제할 수 없다.

② 제 2 안

상업용 데이터 베이스 관리 시스템으로 계층형 데이터 베이스 관리 시스템을 도입하고 統計 데이터 베이스의 특성을 보조할 소프트웨어를 도입하거나 개발하는 것이다. 대표적인 계층형 데이터 베이스 시스템으로는 IBM의 I-MS가 있다.

이 안의 長點으로는 현 농림수산부 데이터 베이스 관리 시스템이 계층형이므로 기술인력과 프로그램들의 친근성을 들 수 있으며, 또한 이 시스템이 제공하는 자료관리체제가 CPU 부담이 크지 않다는 것이다.

限界點은 계층형 데이터 베이스의 단점으로 분산 데이터 베이스 구축의 어려움, 사용하는 언어 DL/I 등의 복잡함, 응용 프로그램 개발에 많은 시간의 소요와 유지 보수의 어려움 등을 들 수 있다. 또한 계층형 데이터 베이스 시스템을 목적 프로그램으로 하고 개발된 통계 데이터 베이스를 위한 보조 소프트웨어가 거의 없어 자체 개발로만이 統計 데이터 베이스의 特性을 고려한 農業 데이터 베이스를 구축할 수 있다.

③ 제 3 안

상업용 데이터 베이스 관리 시스템으로 관계형 데이터 베이스 관리 시스템을 도입하고 統計 데이터 베이스 特性을 보조할 소프트웨어들을 도입 또는 개발을 통해 전위 시스템으로 운영하는 것이다. 대표적인 관계형 데이터 베이스 관리 시스템으로는 IBM의 DB2와 ORACLE이 있다. 이 두개의 시스템은 서로 장단점이 있다. IBM에서 개발된 DB2는 가격이 싼 반면 ORACLE은 가격이 비싼편이다. 개발면에서 보면 ORACLE이 DB2보다 쉽게 응용 소프트웨어를 개발할 수 있는 융통성이 있다. 선택의 여부는 하드웨어 시스템과 예산을 고려하여 충분한 검토후에 결정해야겠다. 또한 요약 테이블을 관리하여 統計 데이터 베이스의 운영을 지원할 수 있는 소프트웨어로서는 Ozsoyoglu에 의한 STBE와 STL, Ghosh와 Codd에 의한 SRT와 QBSRT, Ikeda와 Kobayashi에 의한 HSDB 등이 있다.

이 안의 長點은 다음과 같다. 먼저 관계형 데이터 베이스 시스템의 도입으로 얻을 수 있는 이점을 살펴 보겠다. 데이터의 독립성으로, 저장된 데이터 베이스 자료의 물리적 구조 향상과 개발된 응용 소프트웨어의 개발을 분리 수행할 수 있다는 것이다. 사용자의 관점에서 보는 데이터 베이스 스키마의 단순성으로 자료에 대한 검색이 자료의 접근 경로에 의존하지 않고 보다 발전적 형태인 內容으로 이루어지기 때문에 사용자들에게 그 부담을 줄일 수 있다. 또한 행정전산망에서는 이미 표준 데이터 베이스 관리 시스템으로 관계형 시스템을 채택하여 개발하고 있으므로 그 운영시에 프로그램과 데이터 베이스에서 질의의 호완성과 이식성을 높일 수 있다. 한편 統計 데이터 베이스의 특성을 고려하는 점에서도 관계형 데이터 베이스 관리 시스템을 목적 시스템으로 하고 개발된 소프트웨어들이 상당히 있으므로 도입하여 검토 후 활용함으로써 효율적인 農業 데이터 베이스를 구축할 수 있다.

限界點은 관계형 데이터 베이스 관리 시스템이 CPU의 부담이 크고, 통계 데이터 베이스의 특성 때문에 데이터 저장과 액세스가 비효율적이다. 또한 개념적 모델링 도구와 질의어도 만족시키지 못하나 이는 통계 데이터 베이스의 특성을 고려한 전위시스템의 도입 또는 개발을 통해서 보완이 가능하다.

나. 結 論

이상의 세 가지 중 본 연구에서 제안하고자 하는 방안은 제 3안이다. 제 1안과 제 2안은 限界點이 너무 많아 統計 데이터 베이스의 特性 때문에 생기는 데이터 저장과 액세스의 비효율성이 있다고 하더라도 보조 소프트웨어로서 보완을 계속 한다는 방침아래 제 3안을 선택하는 것이 바람직한 것 같다.

그러나 제 3안을 선택한다고 하더라도 관계형 데이터 베이스 관리 시스템의 도입만으로는 統計 데이터 베이스의 特性을 살릴 수 없어, 부가적인 소프트웨어의 도입과 개발이 반드시 필요한데 이를 위한 계속적인 연구와 지원이 필요하다. 지금까지 연구의 결과로 앞으로의 방향을 제시해 보면 다음과 같다.

효율적 統計 데이터 베이스 구축을 위한 전위시스템은 기존의 統計 데이터

베이스의 運營을 지원할 수 있는 소프트웨어들의 도입과 함께 보완적인 성격을 띤 소프트웨어의 開發을 통해 완성시켜 나가는 것이 바람직하다. 왜냐하면 기존의 개발의 시스템의 도입만으로는 2장 3절에서 설명한 바와 같이 분제성과 제한성이 있기 때문이다.

그리하여 3장에서 Prototype으로 개발한 S-M 시스템의 경험을 살려 이 시스템에서 제시한 메타 데이터 베이스와 요약 데이터 베이스가 具現된 시스템으로의 완성을 위하여 도입과 더불어 지속적인 연구로 보조 소프트웨어의 개발을 통해 農業 데이터 베이스의 구축시 효율적으로 운영될 수 있는 전위 시스템을 완성시켜 나아가야겠다.

이와 같이 統計 데이터 베이스의 관리를 위해서는 데이터 베이스 관리 시스템의 도입만으로는 문제가 해결되지 않는다. 즉, 이를 이용한 統計 데이터 베이스의 構築에서부터 전문기술인력이 필요하며, 統計 데이터 베이스 이용의 효율성을 위해 지속적인 전문인력에 의해 적극적으로 이용기법이 개발되어야겠다. 이를 위하여 연구인력의 적극적인 노력과 함께 지속적인 뒷받침이 요구된다. 그러므로써 효율적인 農業 데이터 베이스 構築을 통하여 양질의 자료를 요구자에게 쉽게 제공할 수 있으며, 또한 최종 목표인 정책입안과 관련된 의사결정 지원을 위한 농수산통계자료의 統合農業情報 시스템의 構築에 효율적으로 대처해 나가야겠다.

參 考 文 獻

- 김상희, 박석, 「범주속성의 Regrouping 정보를 이용한 효율적인 統計的 질의처리」, 韓國情報科學會, '86年 가을 학술발표논문집, 1986.
- 노상욱, 김용철, 박석, 「통계적 응용을 위한 확장된 SOL 인터페이스의 設計에 관한 研究」, 韓國情報科學會 '88年 봄 학술발표 논문집 1988.
- 노상욱, 박석, 「要約데이터의 效率的인 處理를 爲한 데이터 사전 / 디렉토리의 設計 및 具現 農林水産部, 「農家經濟 結果報告」, 1988
- 이석호, 「데이터 베이스론」, 正益社, 1985.
- 이애영, 박석, 「통계적 개체-관련성 모형의 응용」, '87 통계 데이터 베이스학술세미나 논문집 제1권 제1호, 1987.
- 조석봉, 전학성, 김영시, 김영찬, 「교환기 데이터 베이스를 爲한 메타 정보 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.
- 한상만, 이충일, 정찬진, 송영기, 「메타 시스템에서의 데이터 베이스自動生成研究」, '88 통계 데이터 베이스 학술 세미나 논문집 제2권 제1호, 1988.

- 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.
- Deane Merrill, [*Automatic Aggregation and Disaggregation of Data*], First LBL Workshop on SDB, 1981.
- Fredric Gey, [*Data Definition for Statistical Summary Data or Appearances Can Be Deceiving*], First LBL Workshop on SDB, 1981.
- Frank Olken, [*How Baroque Should a Statistical Database Management System Be?*], Second International Workshop on SDB, 1983.
- F. Gey, [*Data Definition for Statistical Summary Data or Appearances Can Be Deceiving*], A LBL Perspective on SDB, 1982.
- Hideto Ikeda and Yasuyuki Kobayashi, [*Additional Facilities of a Conventional DBMS to support Interactive Statistical Analysis*], First LBL Workshop on SDB, 1981.
- Hideto Sato and Ryosuke Hotaka, [*For large Meta Information of National Integrated Statistics*], First LBL Workshop on SDB, 1981.
- J. L. McCarty, [*Metadata Management for Large Statistical Data Bases*], A LBL Perspective on SDB, 1982.
- J. L. Mc Carthy, D. W. Merrill, [*SEEDIS: The Socio-Economic Environmental Demographic Information System*], A LBL Perspective on SDB, 1982.
- Maurizio Rafanelli, Fabrizio L. Ricci, [*Proposal of a Logical Model for Statistical Data Base*], Second International Workshop on SDB, 1983.

- Neil C. Rowe, [*Some Experiment in Evaluation of and Expert System for SDB*], Second International Workshop on SDB, 1983.
- Paul chan and Aric Shoshani, [*SUBJECT:A Directory Driven System for Large SDB*], First LBL Workshop on SDB, 1981.
- Robert A. Burnett and James J. Thomas, [*Data Management Support for Statistical Data Editing and Subset Selection*], First LBL Workshop on SDB, 1981.
- Robert A. Burnett, Paula J. Cowly, James J. Thomas, [*Management and Display of Data Analysis Environment for Large Data Sets*], Scnd International Workshop on SDB, 1983.
- Roger E. Cubitt, [*Meta Data :an Experience of its Uses and . Management*], Second International Workshop on SDB, 1983.
- Roy Hammond, [*Metadata in the RAPID DBMS*], First LBL Workshop on SDB, 1981.
- S. J. Eggers and A. Shoshani, [*Efficient Access of Compressed Data*], A LBL Perspective on SDB, 1982.
- Stanly Y. W. Su Sham B. Navathe, Don S. Batory, [*Logical and Physical Medelling of Statistical/Scientific DB*], Second International Workshop on SDB, 1983.
- Yvonne M. Bishop, Stanley R Freedman, [*Classification of Metadata*], Second International Workshop on SDB, 1983.

빈 면

研究報告 148
統合農業情報시스템 ⑥
農業 데이터 베이스 構築의 發展方向

1988년 12월

發行人 金 榮 鎮

發行處 韓國農村經濟研究院

1310-0510

서울특별시 동대문구 회기동 4-102

登錄 1979年 5月 25日 第 5-10號

電話 962-7311

印刷 協成社 文苑社

電話 739-3911~5

出處를 明示하는 한 자유로이 引用할 수 있으나 無斷轉載 및 複製는 禁함.