병원에서의 의사 결정지원을 위한
데이터 웨어하우스 구축에 대한 연구

- 고혈압 관리 및 의료보험 사전심사를 중심으로

연세대학교 보건대학원

보건정보관리학과

조 경 원

병원에서의 의사 결정지원을 위한 데이터 웨어하우스 구축에 대한 연구

- 고혈압 관리 및 의료보험 사전심사를 중심으로

지도 채 영 문 교수

이 논문을 보건학 석사학위 논문으로 제출함

2000年 6月 30日

연세대학교 보건대학원

보건정보관리학과

조 경 원

조경원의 보건학 석사학위논문을 인준함

심사위원 _____ 인

심사위원 _____ 인

심사위원 _____ 인

연세대학교 보건대학원

2000 년 6 월 30 일

감사의 글

나의 작은 신음에도 응답하시는 나의 주인되신 하나님께 감사드립니다.

부족하지만 지금까지의 결실이 완성될 수 있도록 따뜻한 가르침으로 조언해 주신 김남현선생님과 박충식선생님, 그리고 2 년반의 시간이 절대 헛되지 않도록 항상 도전과 용기를 주시며 세심한 지도와 배려로 이끌어 주신 채영문 선생님께 깊은 감사를 드립니다.

오늘이 있기까지 늘 격려를 아끼지 않으신 남은우 교수님, 류황건 교수님, 박창식 교수님과 기도로 늘 힘이 되어 주신 김정태목사님, 바쁜 업무중에도 결실을 맺을 수 있도록 배려해주신 하바스 메디메디아 가족들께도 감사를 드립니다.

대학원 생활 내내 함께 지내며 많은 도움을 준 송미라 선생님, 승희, 미영, 은경언니, 기호, 경회, 규형선배, 연구실 동료, 친구 희경이와 마음으로 큰 힘이 되어준 소연, 윤희언니, 문선향사모님과 모르는 것 많은 후배를 힘들때마다 가르쳐주고 도와준 이동하박사님, 이영빈선생님께도 감사의 마음을 전합니다. 멀리 떨어져 있지만 존재만으로도 위로가 되어준 양원, 수봉, 대학친구들에게도 미안함과 고마움을 느낍니다. 지금까지 지내오면서 많은 분들의 도움을 받았으나 일일이 감사를 드리지 못함을 죄송하게 생각합니다.

저의 배움이 단지 지식이 아니라 삶에서의 지혜, 삶이 있는 신앙, 사유하는 지성을 가진 사람으로 더욱 굵은 매듭을 지어가도록 하겠습니다.

무엇보다도 작은 화초에 물을 주듯 항상 믿음과 사랑으로 지켜봐주신 부모님과 가족들에게 감사드리고 언제나 사랑으로 격려해준 사랑하는 사람에게 고마운 마음을 전합니다.

1999 년 7 월

조 경 원

목 차

표 목 차

그림목차

국문요약

급변하는 병원의 경영 환경에 능동적으로 대처하여 질적으로 차별화된 의료 서비스를 제공하기 위해서는 축적된 정보를 효과적으로 활용할 수 있는 시스템의 구축이 요구되어진다. 체계적인 진료 및 경영 전략을 세우기 위하여 병원의 가장 기본 정보인 진료정보를 체계적으로 수집하고 효율적으로 활용할 수 있는 방안을 마련하는 연구 또한 필요하다.

본 연구에서는 의료 분야 데이터 웨어하우스(data warehouse)를 임상과 비임상으로 구분하여 고혈압 관리 데이터 마트(data mart)와 사전 심사 데이터 마트를 구성한 후 이를 토대로 정보 검색 기법 및 OLAP(OnLine Analytical Processing : 온라인 분석 처리)을 이용한 사용자 용 프로그램을 개발함으로써 병원에서의 고혈압 관리 및 의료 보험 사전 심사, 의사 결정 지원을 위한 유연한 질의를 실행하도록 하였다. 개발된 프로그램의 사용자인 심혈관 분야 전문의와 의료 보험 사전 심사 요원의 요구사항 파악을 통하여 데이터 웨어하우스로 구축해야 하는 데이터베이스의 범위를 정하고 이에 따라 데이터를 추출, 정제, 변환함으로써 데이터 웨어하우스를 구축하였으며, microsoft 사의 SQL Server Enterprise Manager 를 이용하여 star 스키마와 snowflake 스키마의 데이터 모델링 기법을 활용하였다. 또, 임상 및 비임상 두 분야의 데이터 마트를 토대로 의료보험 사전 심사 시스템과 고혈압환자 분석 시스템으로 구성하였다. 의료 보험 사전 심사 시스템과 고혈압 환자 분석 시스템은 의약품 검색 모듈과 분석 모듈로 구성되어 있는데, 의약품 검색을 위하여 AID 검색 엔진을 이용하여 Microsoft 사의 Visual C++로 구현하였고, 분석 모듈의 OLAP(On-Line Analytical Processing)을 위한 GUI(Graphic User Interface)는

Microsoft 사의 VBA(Visual Basic for Application)로 개발하였다.

본 연구의 특징으로는 첫째, 병원의 각 서버에 흩어져 있던 복잡한 데이터들을 수집하여 분석에 활용할 수 있도록 표준화하고 변환 작업을 거침으로써 병원의 다양한 데이터의 파악이 가능하도록 하였다. 둘째, 병원의 기존 시스템 데이터 뿐 아니라 외부 데이터로서 KIMS 의약품 데이터 베이스를 데이터 웨어하우스와 통합하고 검색기법을 활용하여 약물 정보를 검색하게 하였다. 따라서 데이터 웨어하우스의 구축 및 활용을 통해 의사가 진료시 약물 정보를 짧은 시간에 최대한 이용함으로써 의료 보험 사전 심사 및 고혈압 관리를 위한 의사 결정에 의약품 정보를 최대한 활용할 수 있도록 하였다.

향후 본 시스템을 실제 병원에서 사용중인 OCS(Order Communication System) 데이터베이스 및 보험 청구 데이터베이스, 퇴원 요약 데이터베이스와 연동하고 지식 경영시스템과 통합함으로써 복잡한 병원의 데이터에 쉽게 접근하여 다양한 사용자의 요구 사항에 따라 다차원 분석을 용이하게 할 수 있을 것이다. 이와 같은 축적된 정보의 활용을 통해 병원의 경영 환경에 능동적으로 대처할 수 있는 질적으로 차별화된 의료 서비스를 제공할 수 있을 것이며, 병원에서의 전사적인 지식 경영의 비전 및 이의 전략 수립을 체계화할 수 있을 것이다.

Ⅰ. 서 론

1. 연구 배경 및 필요성

병원의 전산화는 전국민 의료 보험 이후 대폭 증가한 보험 청구 관련 업무처리를 지원하는 것을 기점으로 확산되기 시작하였으나 그 이후 수년간 뚜렷한 발전의 전환점을 갖지 못하였다. 여러 가지 경제 여건과 환자 서비스 증대, 업무 효율성 제고의 요구를 수용하기 위해 OCS(Order Communication System), PACS(Picture Archiving Communication System), CIS(Clinical Information System), EMR(Electronic Medical Record)등 다양한 범위에서의 업무 처리와 데이터 보관을 위해 발전되어 왔다. 그러나 이 과정에서 축적된 데이터는 정보로서의 가치를 지녔음에도 불구하고 사장되어 활용성이 미약한 현실이다. 급변하는 병원의 경영 환경에 능동적으로 대처하여 질적으로 차별화된 의료 서비스를 제공하기 위해서는 축적된 정보를 효과적으로 활용할 수 있는 시스템의 구축이 필요하다. 세계화, 전문화, 정보화의 시대라 불리우는 현대 시대에 정보의 활용은 개인, 기업, 국가에 이르기까지 매우 중요한 문제이며 정보화 사회를 살아가는 현대인에게 원하는 정보에 대한 정확하고 빠른 접근은 성공의 여부를 결정짓는 중요한 요소라 할 수 있다. 또한, 의료 기관의 양적 증가와 경쟁심화, 진료 수가 지불 방법의 변화 가능성, 의료 시장의 개방, 국민의 의료 이용 형태 및 의식의 변화 등 내외적 의료환경이 변화하는 실정에서 의료 서비스 목표의 정립과 질 향상을 위한 병원 혁신의 노력이 필요하다. 병원들이 여러 가지 형태의 정보 시스템 개발을 추진하고 있지만 대부분의 경우 병원이 처해있는 현안들을 해

결하고 경쟁력을 키우는데 도움을 주기에는 미흡한 실정이다.

특히 21 세기 정보 지식 사회에 대비하여 병원에 산재되어 있는 다양한 형태의 정보를 통합하여 가치있는 정보로 재창조하여 활용이 가능하도록 함으로써 교육, 연구, 진료 부문에서 세계 최고 수준을 유지하고, 나아가 국민 건강 증진에 기여를 할 수 있는 토대가 된다. 정보 시스템의 급격한 발전과 더불어 많은 양의 데이터가 축적되기 시작하면서, 필요한 정보를 찾아내어 가치 있는 지식으로 승화시키는 것에 대한 필요성이 야기되고 있다(이순철, 1999). 즉 대량의 데이터를 지식으로 효과적으로 저장, 관리, 활용할 수 있는 지식 기반 경영시스템의 필요성이 대두되고 있다(조성원, 1998). 단순히 데이터나 정보를 입력 · 조회 · 출력하는 역할이 아니라 지식을 바탕으로 지속적인 업무지식을 창출하고, 효과적으로 지식 기반 경영시스템을 운용할 수 있는 전사적인 지식 경영(knowledge management)의 비전 및 이의 전략 수립이 요구되고 있다(Daelim, 1998). 또한 최근 몇 년에 걸쳐 대기업을 중심으로 대량의 데이터를 포함하고 있는 정보 시스템을 구축, 관리할 수 있는 데이터 웨어하우스(data warehouse)가 새로운 아키텍쳐로 각광받고 있으며, 이와 함께 방대한 양의 데이터로부터 함축적이며 잠재적 유용성이 있는 정보를 발견할 수 있는 지식 탐사 방법인 데이터 마이닝(data mining)에 대한 중요도가 또한 증대되고 있다(조남철 등, 1999). 의료 분야에서도 다른 분야와 마찬가지로 의료 정보 자원의 효율적 활용을 위한 지식 기반 경영이 요구되고 있는 실정이다. 의료 분야의 지식 경영은 전국민 의료관련 정보의 많은 부분을 차지하고 국민의료복지에 관한 정책 수립 시 결정적 자료가 되는 의료 보험에서의 지식 경영과 환자 개개인의 임상적 정보를 포함하고 연구, 교육, 진료에 활용할 수 있는 자료가 되는 병원에서의 지식 경영으로 나눌 수 있다. 과거와 다르게 최근의 병원들은 정보화로 인해

서 상당한 양의 의료 데이터가 저장되어 있어서 이의 효과적인 이용에 관심을 가지게 되었다. 그러나 기존 통합 병원 정보 시스템은 일반 관리와 원무 관리 중심에서 벗어나지 못하고 있다. 품질 좋은 의료 서비스를 제공하기 위해서 환자 중심의 진료 및 진료 지원, 임상 연구 등을 종합적으로 지원하기 위한 데이터 웨어하우스의 필요성이 대두되기 시작했다(김종호 등, 1999).

대부분의 기업이나 단체의 업무 구성은 계층적인 구조로 구성할 수 있지만, 의료 분야는 여러 측면이 겹치는 특수한 구조이다. 이러한 복잡한 구조를 관리하기 위해 데이터를 의료 서비스별로 그룹화할 필요가 있으며, 단위 의료 서비스는 여러 명의 의료 제공자를 포함한다. 데이터 웨어하우스의 이용자는 전체적인 처리 과정에서 산출되는 데이터에 관심을 가질 수가 있다.

체계적인 진료 및 경영 전략을 세우기 위해서는 병원의 가장 기본 정보인 진료정보를 체계적으로 수집하고 효율적으로 활용하여 이를 기반으로 한 효율적인 경영 전략의 수립이 중요하나 매우 미흡한 실정이므로 진료 데이터를 효율적으로 활용할 수 있는 방안을 마련하는 연구가 필요하다.

본 연구는 병원의 기본 정보인 진료정보 및 진료지원정보, 원무행정정보를 대상으로 병원에서의 진료 및 경영 전략 수립과 관련하여 고혈압 관리 및 의료보험 사전심사의 의사결정지원을 위한 데이터 웨어하우스를 구축하고자 하였다.

2. 연구의 목적

본 연구는 병원의 의무 기록 정보 및 원무 행정 정보, 의약품 정보를 대상으로 하여 정형 데이터(structured data) 및 비정형 데이터(unstructured data)에 대한 가공 및 추출, 요약을 수행함으로써 환자에 대한 정보를 관리하고 이로써 병원 경영 및 진단에 활용할 수 있는 데이터 웨어하우스를 구축하고자 한다. 구체적인 목적은 다음과 같다.

첫째, 병원의 의무 기록 정보 및 원무 행정 정보, 의약품 정보에 대하여 병원 데이터 웨어하우스를 구축한다.

둘째, 병원 데이터 웨어하우스를 기반으로 의료 보험 사전 심사 데이터 마트와 고혈압 관리 데이터 마트를 구축한다.

세째, 구축된 데이터 마트로부터 사용자의 다차원 접근 및 활용을 위한 정보 검색 기법(information retrieval method)과 OLAP(On-Line Analytical Processing)등의 사용자의 데이터 접근을 프로그램을 설계 및 구현한다.

II 이론적 배경

1. 데이터 웨어하우스의 정의와 구조

가. 정의

데이터 웨어하우스는 1980 년대 중반 IBM 이 자신의 하드웨어를 판매하기 위해 처음으로 도입했던 개념으로, IBM 은 인포메이션 웨어하우스(information warehouse)라는 용어를 사용하였다. 이후 이 개념은 많은 하드웨어, 소프트웨어 및 툴 공급 업체들에 의해 이론적, 현실적으로 성장하였으며, 1980 년대 후반 Inmon 이 데이터 접근 전략으로 데이터 웨어하우스 개념을 제시(Inmon, 1993)한 이래 여러 가지 정의가 발표되었다. 데이터 웨어하우스라는 용어를 처음으로 제시한 Inmon 은 ‘ 의사 결정을 지원하기 위한 주제 지향적으로 통합된 여러 시간대의 갱신 삭제되지 않는 데이터의 집합’ 이라 정의하였다. Poe 는 ‘ 의사 결정 지원 시스템의 기초로 사용되는 읽기 전용의 분석적 데이터베이스’ 라고 정의하고 있다(Poe, 1996). 데이터 웨어하우스는 단순히 하나의 데이터베이스가 아니라 데이터가 수집되고 처리되고 활용되는 하나의 환경으로 이해되어야 하며, 이런 의미에서 데이터 웨어하우징(warehousing)이라는 단어가 사용되기도 한다. Gardner 는 ‘ 데이터 웨어하우징은 기업의 독립 및 상세정보를 모으기 위하여 다양한 원본으로부터 데이터를 수집하고 관리하는 과정’ 이라고 정의한다(Gardner, 1998).

데이터 웨어하우스는 이와 같이 여러가지 의미로 받아들여지고 있으며, 따라서 한마디로 정의하기는 결코 쉬운 일이 아니다. 가장 일반적으로 받아들여지고

있는 데이터 웨어하우스의 정의는 Inmon 이 기술한 것으로서 다음의 성격을 지닌 데이터의 집합체를 의미한다:

- **주제 지향성** (subject oriented)

데이터 웨어하우스내의 데이터는 일상적인 트랜잭션을 처리하는 프로세스(process) 중심 시스템의 데이터와 달리 일정한 주제별 구성을 필요로 한다. 예를 들어 보험 회사의 경우 프로세스 중심의 시스템으로는 '자동차 보험', '생명 보험', '개인 연금 보험' 등이 해당되지만, 이들의 주제 영역을 보면 '고객', '약관', '청구' 등이 될 수 있다.

- **통합성** (integrated)

데이터 웨어하우스내의 데이터는 고도로 통합되어야만 한다. 예를 들어, 기존의 환경에서는 남자와 여자를 남/여, Male/Female, 1/0 등으로 다양하게 표현될 수 있으나 데이터 웨어하우스에서는 이들을 통합할 필요가 있다. 예를 들면, 남자와 여자는 '남'과 '여'로 통합된다.

- **비휘발성** (non-volatile)

데이터 웨어하우스는 오직 두 가지 작업(operation)을 갖게 되는데, 하나는 데이터를 적재(loading)하는 것이고, 다른 하나는 데이터를 읽는 것, 즉 접근(access)하는 것이다. 이를 달리 표현하면 데이터 웨어하우스에 일단 데이터가 적재되면 읽기전용으로 존재한다는 것이다. 따라서, 데이터 웨어하우스의 데이터는 운영 시스템(operational system)에서 수시 발생되는 갱신이나 삭제 등이 적용되지 않으므로 수시로 변한다는 의미의 '휘발성'을 갖지 않게 된다.

- **시계열성** (time variant)

운영 시스템의 데이터는 접근하는 순간에 정확해야만 의미가 있게 된다. 그러나, 데이터 웨어하우스의 데이터는 일정한 시간 동안의 데이터를 대변하는 것으

로 스냅 샷(snap shot)과 같다고 할 수 있다. 따라서 데이터 구조상에 시간이 아주 중요한 요소로서 작용한다. 이와 같은 이유에서라도 데이터 웨어하우스의 데이터에는 수시적인 갱신이나 변경이 발생할 수 없다.

본 연구에서는 의료 분야의 데이터 웨어하우스를 ‘의료 정보 시스템 및 기타 부서 자체 내에서 축적된 데이터를 경영 분석이나 각 부서에서 분석 및 통계용 정보들로 가공하여 생성하고 이것을 필요로 하는 부서들이 공유하여 병원 경쟁력을 제고하고 새로운 정보 및 병원 지식을 생성해 주는 기반 시스템’이라고 정의하도록 하겠다.

나. 기존의 데이터 웨어하우스 아키텍처

데이터 웨어하우스 아키텍쳐는 정보데이터의 중앙 저장소로서의 역할을 하는 관계형 데이터베이스 관리 시스템 서버에 기반을 둔다. 데이터 웨어하우스에서 운영 데이터와 그 과정은 데이터 웨어하우스와는 완전히 분리되어 있다. 이러한 중앙 정보 저장소는 데이터 웨어하우스 정보의 원본 데이터인 운영 시스템과 최종 사용자 질의 및 분석 툴에 의해 접근, 관리되는 환경을 만들기 위해 설계되어지고, 많은 주요 요소(key component)들이 그 주위에 위치해 있다(Berson 등, 1999).

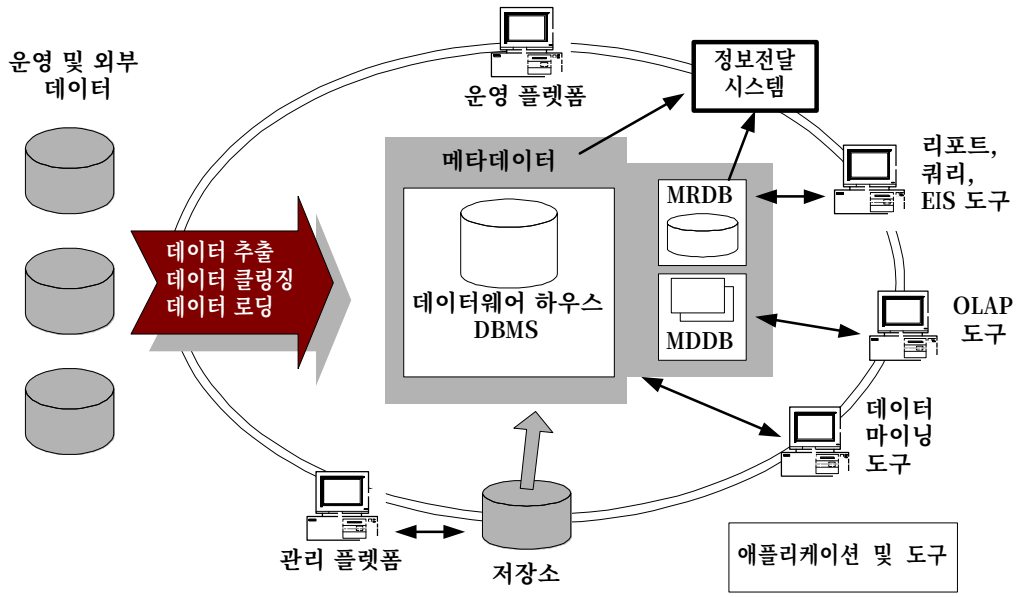


그림 1. 기존의 데이터 웨어하우스 구조

전통적으로 데이터 웨어하우스의 원본 데이터는 운영 데이터 저장소(ODS: operational data store) 또는 운영 응용 프로그램의 데이터들이 된다. 데이터가 데이터 웨어하우스에 입력되면서, 통합된 구조와 형태로 변형되는데, 변형 과정은 데이터의 변환, 요약, 필터링, 상세화등을 포함한다. 데이터 웨어하우스내의 데이터는 대량의 과거 데이터 콤포넌트를 포함하기 때문에 데이터 웨어하우스는 그 기간내의 동일한 데이터베이스에 대한 다른 데이터구조를 관리할 뿐 아니라, 대량의 데이터를 보관·관리할 수 있어야 한다.

1) **원본 데이터**

데이터 웨어하우스 환경으로 유입되는 데이터의 원본을 말한다. 조직내부의 OLTP(On-Line Transaction Processing)**시스템 및 과거 보관용 데이터와 조직 외부에서 들어오는 여러 자료 등의 외부 파일 및 각종 데이터베이스 정보 등이 있다.**

2) **데이터 추출 및 변환**

원본 시스템의 데이터를 데이터 웨어하우스로 이관하는 것을 말하며 데이터의 이관 주기 및 방법 등을 고려해야 한다. 또한 원본 데이터의 필드 값이 정확하지 않을 수 있기 때문에 필요한 경우 데이터의 정리작업이 필요하다. 데이터 웨어하우스에 저장되는 데이터는 원본 상태 그대로도 저장하지만 사용자들이 자주 보는 자료의 형태대로 요약하여 저장한다.

3) **저장 장소로서의 데이터베이스**

실제로 저장할 데이터의 종류와 데이터의 구조, 저장 기간 그리고 요약의 형태 등을 결정한다.

4) **메타**(meta) **데이터**

데이터의 구조, 데이터요약에 사용된 알고리즘, OLTP **와 데이터 웨어하우스 사이의 대응관계 등의 정보를 포함한다.**

5) 데이터 마트

데이터 웨어하우스의 데이터 중에서 일정한 주제의 데이터를 별도의 저장소에 중복 저장하여 사용자들의 사용 편의성을 제고하도록 한다.

6) 데이터 사용도구

데이터 웨어하우스에 있는 데이터를 분석하기 위한 도구로 조회 및 리포팅(reporting) 도구 및 OLAP 도구, 마이닝 도구 등이 있다(고재성, 1998).

2. 데이터 웨어하우스의 연구 및 적용현황

가. 국내 데이터 웨어하우스시스템

데이터 웨어하우스 적용분야는 제조업, 금융업, 유통업, 통신업, 무역업 등에 적용되고 있는데 의료 분야에서는 그 활용이 미진하다. 외국의 의료용 데이터 웨어하우스는 여러 주제(subject)에 적용 및 연구되고 있다. 국내 업체의 경우, 의료용 데이터 웨어하우스는 두개 병원에서 의무 기록관리 부분에 대한 데이터 마트 수준의 데이터 웨어하우스가 구축되고 있는 실정이다. 국내 연구의 경우에는 진료 지원에서 종합 검진 관리의 데이터 마트가 프로토타입(prototype)으로 개발되어 있으며 이대 목동 병원에서 1998 년 11 월에 의료 정보 데이터 마트를 구축하

여 병원 내부의 의료 서비스, 고객 관리 등 경영 자료로 활용하고 OLTP 를 데이터 마트로 구축하여 의무 기록 자료, 임상 병리, 의사 처방 등에 적용되고 있다. 의료 보험 관련 자료를 의료 수요자, 공급자에게 관련 정보를 통보하기 위한 OLTP 를 운영하고 있으며, 트랜잭션 데이터를 한곳에 모아 분석하는 데이터 웨어하우스 시스템은 개발되지 못한 실정이다. 보건 의료 기관 등 총 30 개 기관에서 구축·운용하고 있으나, 활용 기관은 많지 않은 것으로 나타난다.

나. 국외 데이터 웨어하우스 시스템

외국의 데이터 웨어하우스 구축사례는 1994년에 미국 의료 보험 회사중 테네스주 정부의 고용인을 포함한 270만 명의 의료 서비스를 제공하는 가장 큰 회사인 BlueCloss BlueShield에서 20여개의 시스템 제안서중 Sybase/Erisco를 선정하여 7개월만에 빠른 개발능력을 갖추었다. Adaptive Component Architecture의 초석이 되는 데이터베이스로써 일일 25,000건의 청구 업무를 수행하고 경영 분석 자료를 도출 활용하며 노인 의료 구성에 대한 인류 통계학자와 서비스 분석자들을 위해 이 시스템을 활용함으로써 의료수요자, 공급자의 정보제공을 위한 인프라 구축으로 의료 보험 정보 종합 관리를 하고 있다. 보험 회사로 출발했던 Anthem Blue Cross Blue Shield가 지금은 종합적인 의료 관리 회사(health care management company)가 되어 있는데, 이 회사는 NCR의 Teradata RDBMS를 도입했고, 이를 처리하기 위해 16개의 노드를 갖는 NCR WorldMark 고도 병렬 처리 서버를 도입했다. 여기에 저장된 데이터는 1.3 Tera Byte에 달한다. 이러한 시스템을 활용하여, 보험 사기를 줄이고, 의료 기관과의 지급할 보험금 협상, 정확한 위험 관리, 의료 서비스에 대한 통계적 지식 공유로 생존률 높이기에 도움이 되었다.

이 회사는 이러한 데이터 웨어하우스의 도입으로 분리되어 운영되던 회사를

통합하는 목적을 달성했을 뿐만 아니라, 결국 의료의 질과 비용 두 가지 측면에서 효과를 거두게 되었다. 예를 들면, 다음과 같은 분야에서 활용되었다.

- 보험 사기 : 특정한 의사의 진료 행위가 일반적이지 않은 패턴을 보이는 경우, 이를 분석해서, 하나의 의료 행위에 여러 번의 진료비 지급이 있던 것을 밝힐 수 있었다.
- 지식 공유 : 어떤 수술에 대해, 수술 성공율이 높았던 병원으로 환자를 보냄으로 해서, 사망률을 낮출 수 있었다.
- 보험금 협상 : 많은 병원에 대한 진료비에 관련된 정확한 정보를 보유함으로써, 보험금 협상에 유리해 질 수 있었다(Claudio, 1998).

또한 Christiana care에서 통합된 의료시스템에서의 데이터 웨어하우스를 도입하였는데, Christiana care 는 비영리 기관으로 개인들의 건강 증진을 목표로 한다. Christiana care health system 은 1997 년, 기존의 Delaware 의 Medical Center에서부터 IDS(integrated delivery system)으로 재구축하였다. Christiana care에서는 개개의 업무단위(business unit)내의 운용계 정보 시스템의 의사 결정지원응용프로그램은 기존에 존재하고 있다. 데이터 웨어하우스 도입 시, 기존 시스템이 충분하지 않은 이유로서 데이터 웨어하우스 도입에 대한 위험, 이익, 비용에 대해 분석한 것을 살펴보면, 위험으로는 적은 사람들만이 운영계 시스템과 의사 결정 지원 시스템(decision support system)과의 구별을 할 수 있고, 데이터 통합의 부가 이익을 구체화시켜 설명하기가 어렵고, 시스템을 이용할 사용자가 정적인 리포트에만 익숙해져 있어서 동적인 시스템에 익숙하지 않으며, 전체적인 데이터 웨어하우스 구축 프로젝트 수행 기간 동안 지원에 계속 유지될 지에 대한 우려가 있었다는 것이다. 잠재적인 이익으로 보았던 것은 IDS가 전체에 대해 유일한 결과를 내줄 수 있는 데이터를 통합할 수 있을 것이고, 분석을 위한 사전에 정리된 데이터의 준비로 동

적이고, 이용하기 쉬운 의사 결정도구로서의 OLAP을 통하여 해당 업무 담당자들이 신속하게 결과를 얻을 수 있고, 업무상의 문제를 빨리 해결할 수 있게 되었다.

대상이 되는 환자들에 대한 더 자세한 이해와, 이를 통한 의료 서비스 관리 시스템의 성능 개선에 대하여는 반복적인 리포트 작성시에 더 이상 정보 처리부서의 손이 필요하지 않고 손쉽게 작성이 가능하다는데 의의를 두었다. 이에 메타 데이터에 대해 사용자의 접근이 가능해짐은 물론이고 데이터의 질에 대한 신뢰가 높아졌으며 성능 평가가 가능해지고 의사들 사이에 정보를 중시하는 분위기가 형성될 것이라고 보았다.

마지막으로 데이터 웨어하우스 구축 비용은 적어도 백만 달러부터이고, 보통 5백만 달러는 드는데, 컨설턴트 비용으로 50만 달러, 하드웨어, 소프트웨어 각15만 달러, 기타사무집기 2만 달러 정도였으며, 투자한 비용을 고려한 ROI(Return On Investment)를 생각했을 때, 도입한 이후 4-5년째부터 투자한 비용을 회수한다고 보았다.

데이터 웨어하우스 구축 팀을 구성할 때는 의료 서비스를 잘 이해하는 것이 중요하므로 데이터 웨어하우스 시스템의 프로젝트 관리자는 정보통신분야를 잘 아는 전문의였고, 프로젝트 관리자는 간호사나 시스템 분석가였다고 한다. 이 외에 데이터베이스, 웹에 경험이 있는 간호사나 시스템 분석가, 두 명의 보건의료 시스템 개발 경력이 있는 컨설턴트를 활용했고 이들 외에 데이터베이스관리자, 시스템 분석가들을 활용했다.

데이터 웨어하우스를 구축할 때는 최소 9개월에서 1년이 걸리며, 성공할 수 있는 목표를 만들기 위한 제안을 한다. 모든 가정과 가능한 문제점을 명확히 기술해서 주관 부서가 올바른 결정을 하도록 하였는데, 우선은 계속할 지에 대한 결정을 내릴 수 있는 시점을 제공하고, 위험과 중요한 방향을 결정을 짓는 매주 미팅을 가

지며, 어려운 문제를 가장 일찍 다뤄서 그 영향을 최소화하도록 하였고, 데이터 웨어하우스의 이용을 결정하기 위한 요구사항을 미리 결정하고, 상세화하도록 하였다(Ewen 등, 1999).

3. 데이터 웨어하우스 관련연구분야

가. 관계형 데이터베이스(Relational Database)

사회 각 부분의 자동화(automation)가 진행되고, 많은 양의 정보가 축적됨에 따라, 생성되는 데이터를 체계적으로 저장하려는 시도가 있어왔다. 데이터베이스란 정보저장소로서, 데이터베이스를 관리하고 저장하는 DBMS(database management system) 또는 데이터베이스 시스템에 의해 관리되는 데이터의 모음이라고 받아들여지고 있다.

DBMS 란 일종의 소프트웨어로서, 일반적으로 다음의 4 가지 기능을 가지고 있다.

① 데이터베이스의 논리적 구조라고 볼 수 있는 스키마(schema)를 정의할 수 있고,

② 사용자가 데이터베이스에서 일정부분을 간단히 가공해서 볼 수 있는 질의 언어(query language)를 제공하며,

③ 정전등의 사고나, 허가되지 않은 사용자의 접근 등으로부터 데이터베이스를 안전하게 보관하고,

④ 여러 명의 사용자가 동시에 상호 간섭을 받지 않고 저장된 데이터베이스

를 이용하면서도, 하나의 데이터에 동시에 접근해서 데이터를 망가뜨리는 일이 없도록 해주는 기능이다.

1960 년대 말 DBMS 의 출범 이후 여러 역사를 거쳐, 1970 년대 말 이후부터 지금까지 가장 널리 쓰이고 있는 DBMS 의 종류는 관계형(relational) 데이터베이스 관리 시스템이다.

관계형 데이터베이스 시스템(또는 관계형 시스템)은 다음과 같은 성질이 있어야 한다(Date, 1993).

① 사용자에게는 데이터가 테이블로 인식되어야 한다.

② 적어도 select, project, join 의 연산을 제공해야 하며, 제공되는 연산으로 기존의 테이블에서 새로운 테이블을 생성하는 연산이어야 한다.

관계형 시스템은 이런 성질로 인해 다양한 장점이 있다. 사용자에게 인식하기 쉬운 형태이면서, 또한 단순하기 때문에 이론적, 수학적인 정립이 가능했고, 이론적 정립을 바탕으로 체계적인 성능 향상 기법을 도입할 수 있었다. 반면, 사용자가 쉽게 이용할 수 있는 연산자를 통해 다양한 방법으로 관계형 시스템 내의 데이터를 활용하는 기법을 제공했던 것이다. 그러나 관계형 시스템의 약점도 또한 이런 성질에서 기인한다.

관계형 시스템에서는 데이터를 테이블로만 구성하고, 테이블의 관리 측면으로 인해 여러 군데로 관련된 데이터가 쪼개어져 저장되었었다. 이런 기법으로는 복잡한 형태의 데이터를 표현하기 어려울 뿐만 아니라, 이러한 데이터를 활용하기 위해서는 join 연산을 수행해야 하는데, join 연산은 일반적으로 수행시간이 오래 걸리는 연산이다. 또한 제공되는 연산자를 수행하는 과정을 사용자가 완전히 이해하지 못한 경우에 원하는 정보를 얻을 수 없었다. 즉, 사용자는 정보를 얻기 위해서는 관계형 시스템내의 데이터에 대해 이미 완전히 파악하고, 얻으려는 정

보의 형태를 미리 알고 있어야만 한다는 것이다.

이러한 약점을 보완하기 위해 다양한 방법이 제시되었다. 여러 테이블에 흩어진 데이터를 한곳에 모아서 이용하자는 방법이 데이터 웨어하우스, 또는 데이터 마트 방법론이고, 얻을 수 있는 정보의 형태를 모르거나, 시스템 내의 데이터를 완전히 파악하지 않은 상태에서 정보를 얻어보자는 방법이 데이터 마이닝 방법론이다.

나. 정보 검색(Information Retrieval)

방대한 데이터로부터 주어진 시간 내에 원하는 정보를 발견하는 것은 매우 어려운 일이다. 매일 쏟아지는 방대한 정보량 중에서 꼭 필요한 것을 찾는 것은 힘든 작업이며 막대한 비용 발생이 있다. 이런 관계형 시스템이 가지고 있는 구조적인 문제점을 해결하기 위해 많은 연구가 행해졌고 이런 노력의 대가로 많은 솔루션이 나오게 되었는데 정보 검색 시스템 또한 그들 중의 하나이다. 정보 검색 기법은 RDBM 와는 달리 B-Tree 방식과 역화일(inverted file)을 이용한 빠른 색인어 접근이 가능한 구조를 갖고 있어 비정형 데이터의 처리나 놀라운 검색 속도 등의 장점을 보유함으로써 RDBMS 가 가지고 있는 구조적인 문제점들을 해결하고자 한 것이다. 누구든지 정보를 매우 쉽게 찾을 수 있도록 만들어져 다양한 컴퓨팅 환경에서 대량의 정보를 수집, 검색, 가공하여 이용자에게 이용자가 원하는 정보를 편하고 빠르며 정확하게 가져다 주는 도구로써 정보 검색 기법은 원하는 정보에 대한 접근을 효율적으로 지원함으로써 정보 수집을 위해 소비되는 시간과 노력을 크게 단축시켜 줌에 따라 기하급수적으로 증가하고 있는 전산화되는 데이터에 적절하게 대응하는 방법이라 할 수 있다. 이 기법은 자료의 내용을 분석 가공하여 특정한 개념을 포함하는 자료를 검색할 수 있는 시스템의 요구가 증가하

므로 본문에 나타난 개념을 검색할 수 있는 전문 검색 기능을 하며 멀티미디어 자료와 전문 데이터의 요구 증가에 의한 비정형 데이터의 처리가 가능하며 하부 네트워크의 종류에 구애 받지 않고 동일한 검색 속도를 유지할 수 있고 논리 연산자, 절단 연산자, 인접 연산자, 검색 제한자 사용, 색인어 검색, 히스토리 검색이 가능한 방법이다. 정보 검색 엔진은 자연어 검색을 지원하는데, 이는 문헌 내용이 주요 개념을 나타내는 단어나 문구 등을 문헌에 나타난 그대로 검색어로 채택하는 경우 언어의 통제 과정이 불필요하기 때문에 별도의 색인 작업이 필요없고 같은 주제가 여러 가지 용어로 표현되는 것을 방지하며 같은 개념이 다른 여러 개의 용어에 의해 표현될 수 있어서 심층정보의 검색이 어렵거나 포괄적인 개념의 정보검색이 어려운데 자연어와 통제어를 병행하여 사용하는 추세이므로 시소러스의 역할이 증대되고 있다. 시소러스는 문헌 중에 자연어를 정보검색의 색인어 번역할 때 쓰이는 용어 통제 장치로 동의어, 유사어, 관련어의 분류와 구분을 나타내는 통제 언어로 검색의 효율을 높이기 위해 시소러스를 활용하여 다양한 자연어를 일정하게 약속된 언어로 변환함으로서 일관성을 유지하고 용어와 용어의 관계를 나타내 줌으로서 전체적인 체계를 조직화하는 것이 필요하고 용어의 부정확성이나 용어의 변화 및 다양성에 대처할 수 있고 본문중의 용어와 검색어의 용어를 조정하여 검색의 정확률을 향상시키고자 한 것이다. 자연어 검색 기법은 크게 세 가지로 나눌 수 있는데, 질의어에 포함된 탐색어들의 정확한 조합을 통해 이와 완전히 일치하는 문헌을 검색하는 불리안(boolean) 논리 검색과 문헌의 색인어와 질의어에 용어의 상대적 중요성을 나타내는 0 과 1 사이의 가중치를 부여함으로서 불리안 검색기법을 확장 일반화하기 위한 퍼지(fuzzy) 검색과 주어진 질의어에 대한 적절성은 그 문서가 이용자에게 적절한 조건 확률로서 표현되어야 하고 그 계산은 데이터베이스의 통계적인 특성과 표본을 통해 추정되어지는데 이

를 매칭(matching) 함수를 이용해 질의어에 가중치를 구함으로서 적합성 확률의 순으로 문헌을 순위화하는 확률검색이 있다. 한국어의 경우 용언의 활용성이 다양하므로 불용어 리스트만으로 불용어를 처리하기는 무리이기 때문에 어형 통제보다 색인어 선별을 위한 형태소 분석이 필수적이고 한국어에서 색인어 선별이 중요한 이유는 복합명사와 고유명사, 외래어 등 사전 미등록어를 추정할 때 단어가 사전 미등록어 인지를 알 수 없기 때문이다.

다. OLAP

OLAP 이란 On-Line Analytical Processing 의 약자로, 1993 년에 관계형 데이터베이스로 명성을 얻은 E.F.Codd 박사가 만든 것이다. OLAP 는 OLTP(On-Line Transaction Processing)와는 대조를 이루며, OLTP 는 다양한 활동의 데이터를 기록하는 반면, OLAP 는 대화식으로 지금까지 기록된 데이터를 분석하거나 미래 데이터를 투사하는 일을 다룬다. 한 가지 다차원 개념 관점은 OLAP 는 중요 이론 중 하나이다(조재희 등, 1996). OLAP 와 관련된 다차원 모델은 사용자가 각기 다른 차원 또는 키를 컴퓨터 화면의 행, 열 또는 페이지와 자유롭게 연계시키도록 허용한다. 컴퓨터 화면의 회전과 연속적 세부 수준까지의 탐색은 OLAP 와 관련하여 가장 흔하고 가장 기본적인 기능이다. 사용자들이 데이터에서 잘못된 내용이나 추세를 강조해서 보여주는 방향을 선택할 수 있다면, 그들이 데이터를 더 효율적으로 분석할 수 있는 가능성이 많아진다. OLAP 의 대표적인 특징은 다음과 같다.

① 데이터를 다차원 형태로 구축한다.

실제 기업에서 사용되고 있는 정보는 기간, 제품, 부서, 지역등 매우 다양하게 발생하는 다차원 정보이다. 예를 들면 기간만 하여도 분석자는

년 단위로 분석하다가 분기, 월, 주, 일 단위 등의 차원으로 볼 수도 있다.

② 최종 사용자가 필요한 정보를 직접 얻는다.

예전부터 제공되어온 EIS(Executive Information System)/DSS 도구도 전산 담당자의 개입을 최소로 하여 왔으나 최종 사용자가 출력된 리포트를 통하거나 전산 담당자의 도움을 받지 않고 직접 필요한 정보를 얻기 위하여 데이터베이스를 접근하도록 지원하는 것은 사용자의 변화하는 욕구를 충족시키기 위한 중요한 요소이다.

③ 사용자가 화면상에서 대화식으로 여러 계층사이를 drill-down, drill-up 하면서 필요한 정보를 얻는다.

OLAP 유형은 별도로 구축된 다차원 데이터베이스의 정보를 접근하는 다차원 OLAP 즉, MOLAP(Multi dimensional OLAP)과 기존의 관계형 데이터베이스를 기반으로 하여 사용자에게 다차원으로 접근할 수 있도록 하는 관계형 OLAP 즉, ROLAP(Relational OLAP)으로 구분한다. 사용자 인터페이스는 양자가 동일하나 데이터가 물리적으로 저장되는 방식은 전혀 다르다.

OLAP 도구는 업무의 계층을 모델화하고 구현하며, crosstab 요약으로 통합한다. 데이터 browser 는 사용자가 crosstab 의 컬럼이나 열로 나타나는 여러 개의 카테고리를 합치거나 재조합하고 그 중에서 원하는 부분을 선택하는 기능을 제공한다. 또, 사용자는 drill-up 이나 drill-down 을 통해, 그러한 카테고리 계층을 확장하거나 요약할 수 있다. 이러한 강력한 방법은 사용자가 간단한 통계 도구나 질의 도구를 통해 데이터를 활용할 수 있게 해준다.

OLAP crosstab 구조를 구현할 수 있는 몇 가지 방법이 있는데, 관계형 테이블이나, 업계 표준의 다차원 데이터 구조나, 그때그때 생성되는 가상의 crosstab 의 구조를 구현하는 방법이다. 상용소프트웨어는 클라이언트(client) 소프트웨어 또

는 서버(server) 소프트웨어로 구현되고, 현재 대부분은 웹에서의 접근을 제공하고 있다(Ewen 등, 1999).

라. 데이터 마이닝(data minig)

80 년대 이후 다양한 분야에 데이터베이스가 적용되기 시작하면서 축적되기 시작한 데이터의 양은 폭발적으로 증가하고 있다. 생산품에 붙어있는 바코드(bar code)로 관리되는 판매 자료, 기상 예측이나 지질 연구를 위해 수집되는 과학적 관측 자료, 세금이나 연금 관리를 위한 정부의 자료 등과 같이 여러 분야에서 생성되는 데이터는 점점 값싸지는 저장 매체와 강력해진 컴퓨터의 처리 능력, POS(point of sale)나 각종 센서 등의 장치를 사용한 데이터 수집 능력 향상을 통해 데이터베이스로 유입된다. 이렇게 모아진 데이터에는 중요한 내용이 포함되어 있다고 볼 수 있다. 과학적 관측 자료에는 어떤 현상을 규명하기 위한 증거가 되는 정보가 포함될 수 있고, 판매 자료에는 중요한 시장, 고객의 구매 추이 등의 정보가 있어서, 새로운 제품 판매 전략을 세우는데 활용될 수 있는 자료가 분명히 포함되어 있는 것이다.

하지만 데이터의 수집 능력에 비하면, 데이터의 분석 능력은 한참 떨어지는 지식 병목(knowledge bottleneck)이라는 문제를 발생시키고 있다. 데이터는 많지만, 의사 결정에 도움을 줄(decision-support) 유용한 지식은 부족하다는 것이다. 그 동안 데이터의 분석을 통해 유용한 지식을 추출하는 것은 그 분야에 능통한 사람이 통계적 도구나, 스프레드 시트(spread sheet) 또는 데이터베이스 언어(database language)를 활용함으로써 행해졌다. 하지만, 분석자의 지식에 의존적이라는 문제점과 데이터의 증가 속도를 고려할 때 이러한 방식의 데이터 분석은 한계가 있다고 여겨지게 되었다.

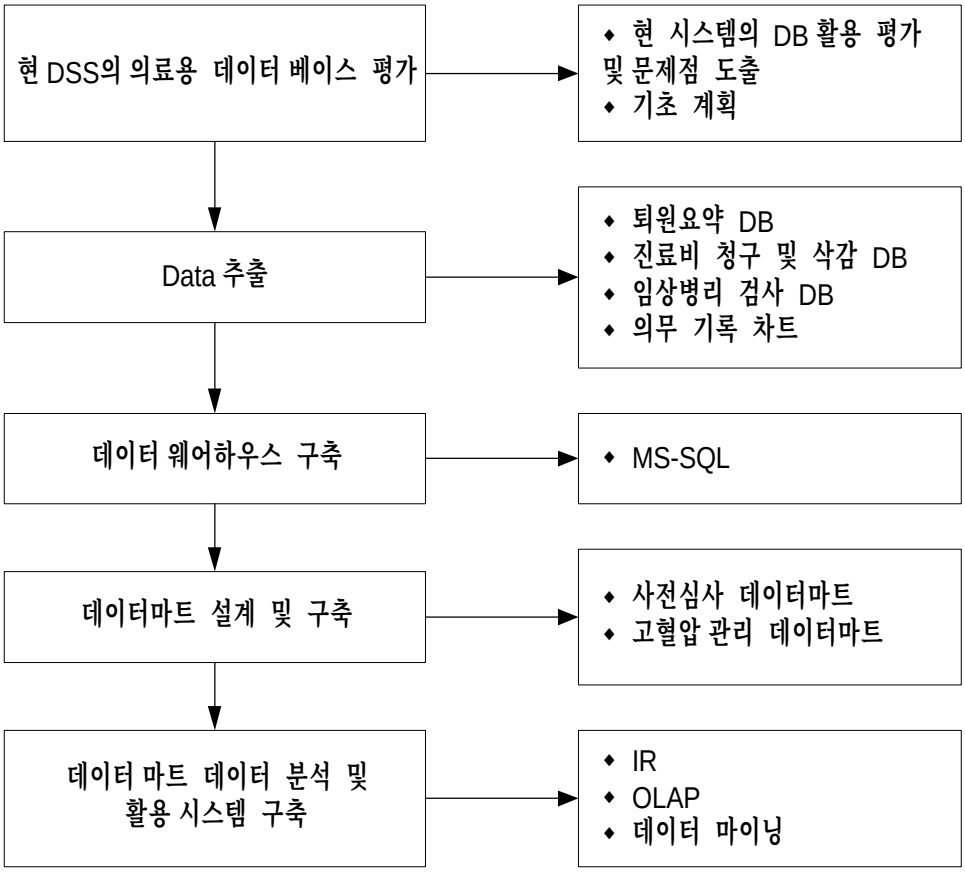
데이터 마이닝은 이러한 문제를 해결하기 위해 90 년대 초에 생겨난 연구 분야이다. 아직 일반적으로 받아들여지는 데이터 마이닝에 대한 정의는 존재하지 않지만, ‘대용량의 데이터 베이스에서 유용한 정보를 추출하고 이용하는 일련의 정보를 찾아내고 이를 바탕으로 데이터에 대한 통찰(insight)을 얻자는 것’으로 정의된다. 숨겨진 형태의 유용한 정보의 예로는 “40 대의 흡연하는 남성이 고혈압인 경향이 있다”와 같은 형태이다.

데이터베이스로부터 분석 대상이 되는 데이터를 선별하는 과정, 선별된 데이터를 적절한 형태로 가공하고 정제(cleaning)하는 과정, 변환된 데이터에 패턴(pattern)추출 알고리즘을 수행하는 과정, 알고리즘의 결과를 재가공하고 기존의 지식과 통합하는 과정 등의 전체 수행 과정을 데이터 마이닝이라고 지칭하기도 하고, 이중에서 마이닝 알고리즘을 적용하는 단계만을 따로 구분하여 데이터 마이닝이라고 하기도 한다. 데이터 마이닝 결과로 얻을 수 있는 지식의 종류에 따라 연관규칙(association rule), 순차 패턴(sequential pattern), 분류(classification), 클러스터링(clustering), 예외 탐사(outlier discovery)등으로 데이터 마이닝의 작업(task)을 구분하기도 한다. 이중에는 데이터 마이닝 분야에 고유한 것도 있고, 기존의 기계학습(machine learning)이나 통계에서 다루어 왔던 형태도 포함되어 있다. 그러나 데이터 마이닝은 이러한 패턴을 실세계(real-world)의 데이터베이스에서 추출하는 것으로써 반드시 적당한(tractable) 계산량으로 이루어져야 한다는 것이 데이터 마이닝의 고유 관심 분야라고 할 수 있다. 데이터 마이닝은 데이터베이스, 인공지능, 기계학습, 통계학 등 기존의 여러 연구분야를 기반으로 출발했으나 이제는 독자적인 연구분야로 구축되고 있다고 할 수 있으며, 데이터 마이닝의 독자적인 시장을 구축하기 시작한 현재에도 기존 연구분야와의 관계에 대해서는 아직 많은 논의가 계속 되고 있다. 아직까지는 데이터 마이닝이 거대한 크기의 데이터베이

스에 적용되기에는 제한점이 있지만, 다루는 데이터의 크기로 인해 기존 연구와 구별된다고 일반적으로 받아들여지고 있다.

Ⅲ. 연구 방법

1. 연구의 틀



연구의 틀

2. 연구 대상 및 범위

가. 연구 대상

본 연구는 병원의 의무 기록 및 처방 전달 시스템 운영 데이터 중 임상병리 검사 결과, 방사선 검사 결과, 퇴원환자 수술 내역, 입원약국 자료와 기타검사결과 및 외부자료로서 KIMS 의약품 데이터에 대하여 데이터 웨어하우스를 설계하고 구축하고자 한다.

진료데이터의 가장 기본이 되는 것은 의무 기록실의 퇴원 요약 정보라 할 수 있다. 이는 환자에 대한 기본 정보가 퇴원 요약지에 포함되어 있기 때문이다. 특히, 미국에서는 국가 차원에서 Uniform Hospital Discharge Data Set 을 만들어서 관리하면서 진료 정보 데이터베이스의 가장 기본적인 항목을 설정할 때 이 내용을 중심으로 하고 있다. 본 연구에서는 여기에 근거하여 심혈관 센터의 환자 중 고혈압에 관련된 심혈관 계 환자의 퇴원 요약 정보를 근간으로 환자의 검사 결과 데이터 및 진료 데이터를 대상으로 하였다. 데이터 웨어하우스는 사용자에 따라 임상 데이터 웨어하우스, 비임상 데이터 웨어하우스로 구분할 수 있다. 병원의 데이터가 무한하여 본 연구에서는 임상에 관련하여 사전 심사 마트를 구축하고자 하였다.

이에 따라 1999 년 5, 6, 7 월 동안 Y 병원 심혈관 센터를 방문한 환자 중에서 주진단 또는 부진단 중에 고혈압이 있는 환자에 대한 퇴원 요약 데이터베이스, 외래 처방 데이터베이스, 임상 병리 검사 데이터베이스, 기타 검사 데이터베이스, 약국 데이터베이스로 고혈압 관리 데이터 마트를 구축하였다.

그리고, 1998 년 12 월부터 1999 년 5 월까지 Y 병원의 해당 기간에 퇴원한 입원 환자(매월 약 2,000cases)에 대한 OCS 처방 데이터베이스, 의료 보험 진료비 청구 데이터베이스, 임상 병리 검사 데이터베이스, 의료 보험 진료비 삭감 및 이의 신청 데이터베이스로 사전 심사 데이터 마트를 구축하였다.

나. 연구의 범위

병원의 의무 기록 정보 및 원무 행정 정보, 의약품 정보의 정형 데이터 및 비정형 데이터에 대한 처리, 가공 및 추출을 수행함으로써 병원 데이터 웨어하우스를 구축하고 병원 데이터 웨어하우스를 기반으로 의료 보험 사전 심사 데이터 마트와 고혈압 관리 데이터 마트를 설계한다. 구성된 데이터 마트에 OLAP 및 IR 기법을 적용하여 기존 운영 정보의 table 목록, table layout, 코드내역을 수집하여 데이터 분석을 하고 이러한 구조화문서에 대한 검색을 지원하며 데이터 마이닝기법을 이용하여 각 데이터 마트에 대해 새로운 지식을 추출한다. 이를 통해 병원 경영 및 진단을 위한 의사 결정 지원 시스템 개발에 활용하였다.

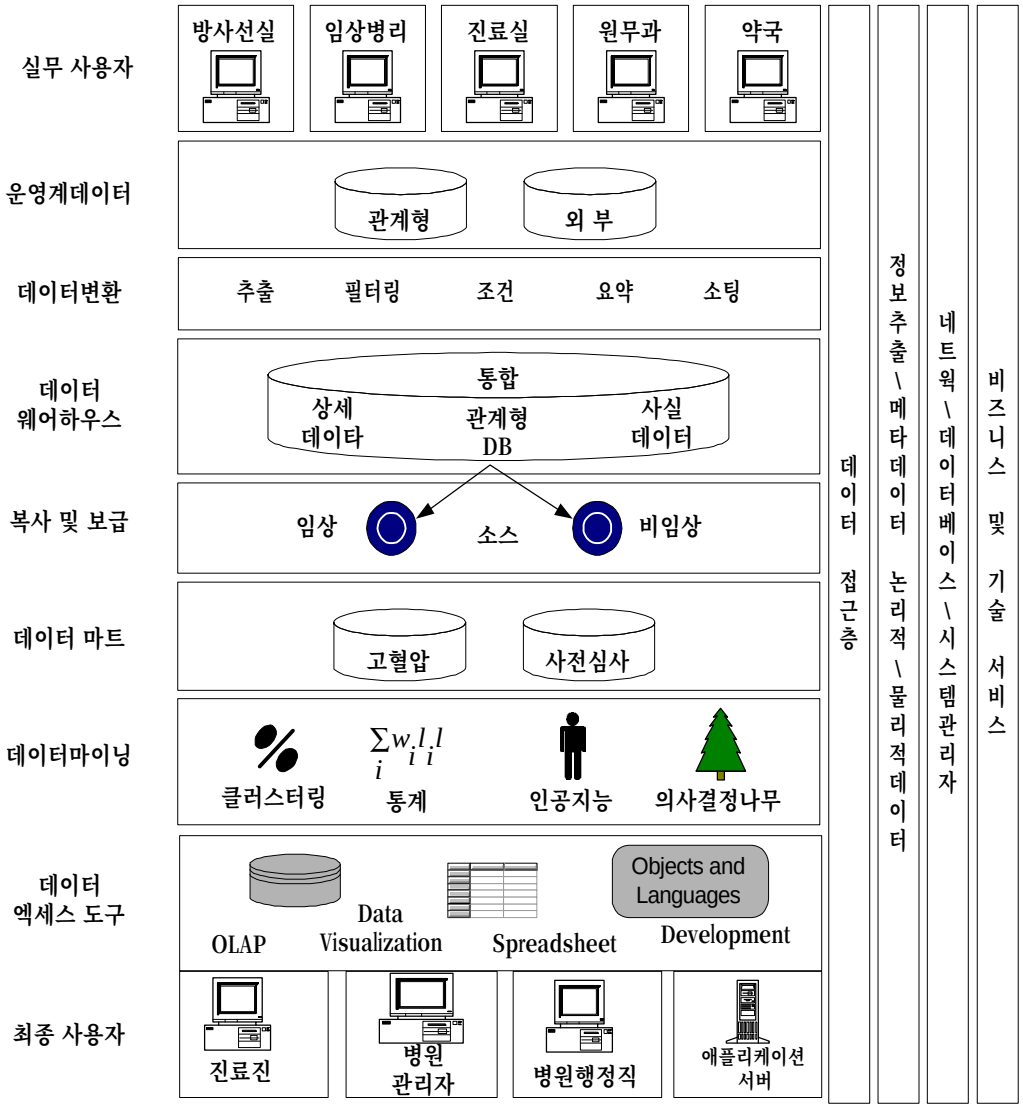


그림 2. **데이터 웨어하우스** framework

분석업무를 효과적으로 지원하기 위해 여러 가지 계층을 구현한다. 현재 방사선검사, 임상병리검사, 약국, 원무계, 보험 심사계 등 실무로부터 OCS **를 통하여**

존재하는 데이터베이스, 의약품 데이터베이스, 환자 기본 정보 데이터베이스 등 운영 데이터가 존재하고 이들을 추출, 변환하여 데이터 웨어하우스를 구성하였다. 의료용 데이터 웨어하우스는 사용자에 따라, 진료와 진료 지원을 기반으로 하는 임상 데이터 웨어하우스(clinical data warehouse)와 원무관리와 일반관리를 기반으로 하는 비임상 데이터 웨어하우스(non-clinical data warehouse)로 구분할 수 있다. 비임상 데이터 웨어하우스는 경영층의 의사 결정을 위함이고 임상 데이터 웨어하우스는 환자에 대한 정보를 저장하여 의료 서비스와 의료 연구를 주목적으로 한다(Pedersen 등, 1998). 따라서 임상과 비임상으로 구분하여 고혈압 관리 데이터 마트와 사전 심사 데이터 마트 형태로 구성한다. 이러한 데이터마트 토대 위에 OLAP 도구와 같은 다차원분석에 기반을 두고서 의사 결정을 위한 유연한 질의를 실행하는 지식정보계층을 두었다. 스프레트쉬트와 다양한 분석을 통해서 최종 사용자의 전략적 의사 결정에 도움을 줄 수 있도록 하였다.

본 연구에서는 데이터 마트로부터 새로운 정보를 추출하고자 하는 데이터 마이닝 계층을 제외하였다.

3. 연구 분석 및 방법

가. 데이터 웨어하우스에서의 정보 흐름

데이터 웨어하우스는 단순히 하나의 데이터베이스가 아니라 데이터가 수집되고 처리되고 활용되는 하나의 환경으로 이해되어야 하며, 이런 의미에서 데이터 웨어하우징이라는 용어가 더욱 타당하다는 주장도 있다. 데이터 웨어하우스의

진정한 가치는 데이터를 수집하는 것보다는 정보의 흐름을 관리하는 것이다(hackathorn, 1995)(조재희 등, 1996). 전체적인 데이터 웨어하우스의 흐름은 데이터 흐름(dataflow)과 메타 흐름(metaflow)으로 구성된다. 데이터 흐름은 내부 흐름(inflow), 상향 흐름(upflow), 하향 흐름(downflow), 출력 흐름(outflow)로 구성된다.

표 1. 데이터 웨어하우스내의 자료흐름

흐 름			설 명
데이터 흐름 (dataflow)	입력흐름(inflow)		운영시스템과 기타 데이터 원본으로부터 변환단계를 거쳐 데이터 웨어하우스 내부로 유입되는 과정임
	자체흐름 (ownflow)	상향흐름 (upflow)	데이터가 다양한 정도로 상세하게 집약되어 요약되어짐
		하향흐름 (downflow)	오래된 데이터는 저장매체로 이동하거나 정화기준에 따라 폐기함
	출력흐름(outflow)		데이터 마트로 복사되거나 쿼리/리포팅 도구, OLAP 도구, 데이터 마이닝등의 다양한 어플리케이션에서 사용됨
메타흐름(metaflow)			데이터 흐름과 관련한 메타데이터의 흐름

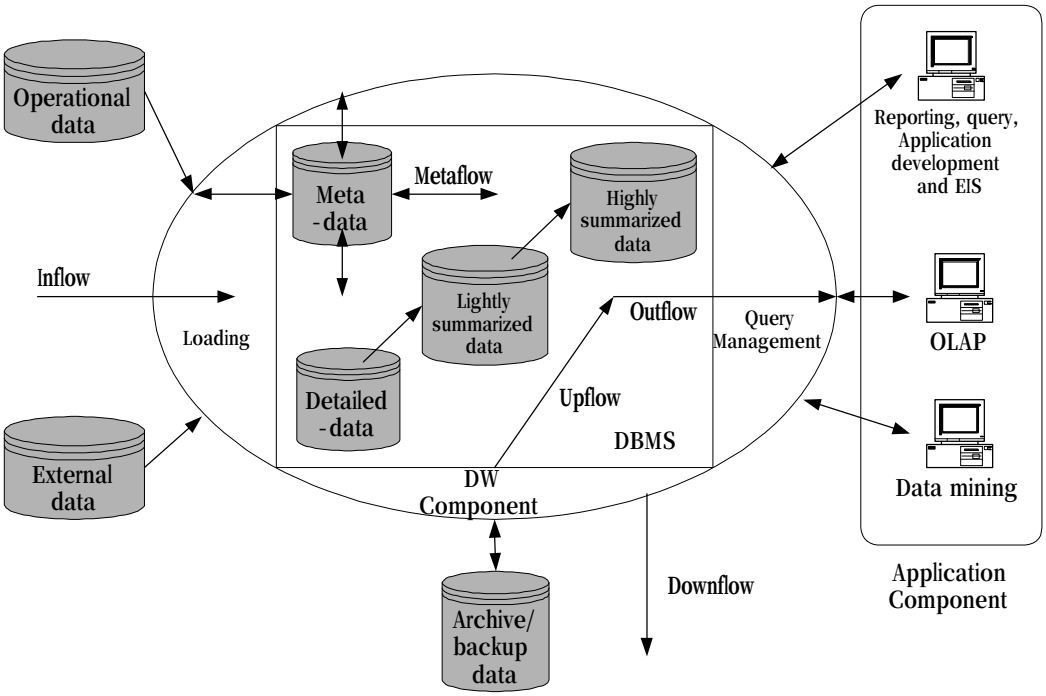


그림 3. 데이터 웨어하우스 내의 데이터 흐름도

1) 입력 흐름

데이터 웨어하우징 환경에서 첫번째 흐름은 데이터가 운영시스템과 기타 데이터 원본으로부터 변환 단계를 거쳐 데이터 웨어하우스 내부로 유입되는 과정이다. 이때, 운영시스템의 성능에 가능한 영향을 미치지 않아야 하며, 운영시스템과 관련하여 데이터의 표준화와 정확성 확보, 웨어하우스 데이터와의 대응관계 설정과 같은 요소들을 고려하여야 한다. 상이한 운영시스템상의 데이터의 질은 차이가 많이 나므로, 이러한 데이터가 데이터 웨어하우스 환경으로 이동되기 전에 운영시스템상의 데이터를 표준화 할 필요가 있다.

데이터 표준화, 데이터의 정확성 검토, 운영 데이터와 웨어하우스 데이터의 대응 관계 설정등의 문제를 해결해야 하는 단계라고 할 수 있다. 운영시스템상의 데이터는 의사 결정 지원형태와는 다르기 때문에 데이터 변형과정이 필요한데, 데이터 변형은 변환(conversion), 보완(enrichment), 요약(summarization)이라는 3 가지의 단계를 구성하게 된다. ‘변환’은 운영시스템상이 데이터 값과 포맷상의 불일치를 해결하는 것을 목적으로 하고, ‘보완’은 데이터 웨어하우스 내에서 분석시 사용할 새로운 개체를 만들기 위해서 다양한 운영시스템 원본으로부터 오는 상이한 항목을 결합하는 것을 말하며, 추출된 원시 데이터는 분석가에게 추세를 제공하기 위해 ‘요약’되는 과정을 거칠 필요도 있다. 또한 웨어하우스 데이터는 기업 내부의 데이터뿐만 아니라 외부의 다양한 자료원으로부터도 데이터를 흡수하여 보강되어야 한다(조재희 등, 1996).

2) 상향흐름

데이터가 운영시스템으로부터 웨어하우스로 이동되는 과정에서 많은 양의 데이터 변형이 이루어진다. 상향흐름은 웨어하우스 내로 유입된 데이터가 부가가치를 더하는 과정으로 다양한 정도로 요약될 수 있다.

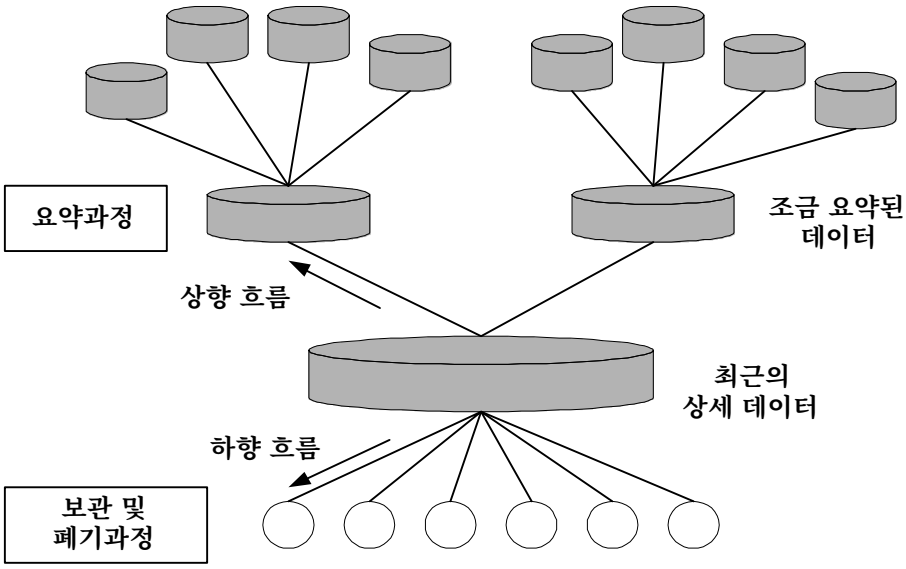


그림 4. 상향 흐름과 하향 흐름

3) 하향 흐름

데이터 웨어하우스 내의 데이터는 시간이 지남에 따라 그 가치가 감소하게 된다. 이러한 데이터는 저장되거나 정화 기준에 따라 제거되어진다.

4) 출력 흐름

비즈니스 사용자들이 웨어하우스 내의 데이터를 이용할 수 있게 된다. 웨어하우스 내의 데이터는 사용자에게 이용되기 전에 데이터 마트로 복사되기도 한다. 이러한 데이터는 쿼리/리포팅 도구, OLAP 도구, DSS, 데이터 마이닝등이 다양한 어플리케이션에서 사용된다.

5) 메타 흐름

메타 데이터는 데이터를 기술하는 정보로서 데이터에 대한 데이터라 할 수 있다. 또한 메타-모델은 어디에 메타 데이터를 저장할 것인가를 나타내는

것으로 메타 정보가 많을수록 데이터에 대한 정보를 잘 표현한 것이라 할 수 있다. 데이터 원본, 구조 및 관계성과 웨어하우스가 운영 원본으로부터 유도되는 방법이나 의사 결정 지원을 위한 업무 규칙등과 같은 데이터의 상태를 표현하기 위해서 메타 데이터는 필요하다. 메타 데이터는 웨어하우스 구축 도구 및 모델링, 구축, 로딩 및 데이터 웨어하우스 관리, 뿐만 아니라 의사 결정 시스템, 운영계와 웨어하우스 데이터베이스를 질의 및 분석할 때도 사용되어진다. 데이터 웨어하우스 구축도구에서 메타 데이터는 데이터 웨어하우스 관리자에 의해 사용되는 기술적 데이터와 데이터 웨어하우스 사용자에 의해 사용되는 비즈니스 데이터로 구분된다. 여기서 기술적 데이터는 원본 데이터와 데이터 웨어하우스의 구성을 설명하는 데이터이며, 비즈니스 데이터는 데이터 웨어하우스에 적재된 시점이나 신뢰성 등을 비즈니스 관점에서 설명한다.

이들 기술적 또는 비즈니스 메타 데이터들 중, 사용자들의 일차적인 관심은 데이터의 변환과정에 대한 정보라고 할 수 있다. 즉, 데이터의 형식, 길이, 값의 범위 등과 같은 데이터 자체에 대한 정보들보다는 데이터 웨어하우스에 적재된 자료가 어떤 변환과정을 거쳐 생성된 자료인지를 알려주는 변환 프로세스에 관한 정보가 더욱 중요하다고 할 수 있다.

나. 데이터 웨어하우스 구축 방법

1) 데이터 변환 및 클렌징 방법

데이터가 여러 개의 운영계/기존 시스템으로 존재하는 경우에는 어떤 형태든 통합과 변환이 필요하다. 데이터 변환(transformation)은 데이터의 재구

조화(restructuring), 재정의(redefining), 재연산(recalculation), 합계계산(summarizing) 등으로 포함하는 넓은 개념이다. 그 내용으로는 split, merge, truncate fields 와 같은 재구조화(restructuring), 불필요한 필드들의 제거, 데이터의 형태와 길이들의 변환, 날짜· 전화번호 등의 서로 다른 형태들의 통일, drived 필드의 계산 및 제품코드·조직코드와 같은 코드 변환 등을 포함한다. 그 세부 변환 기법을 살펴보면 다음과 같다(Informix, 1998).

- 통합(Integration)

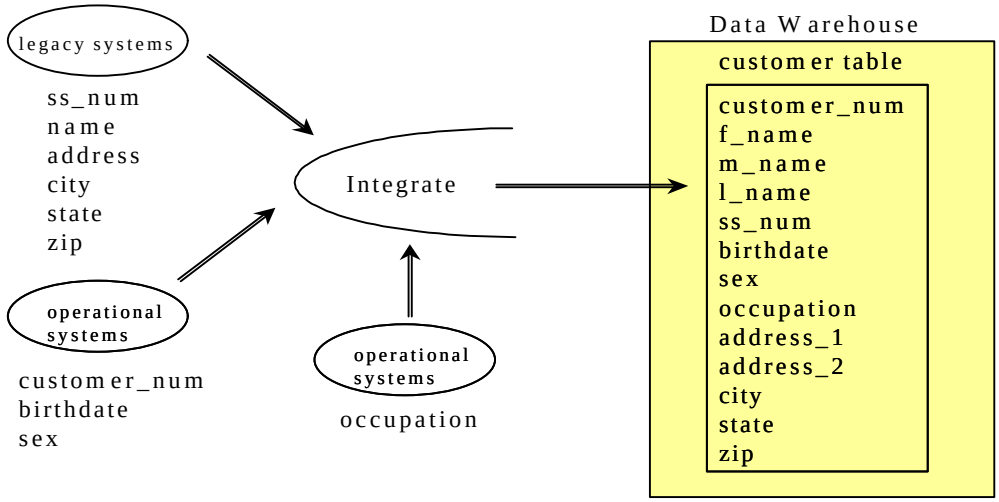


그림 5. 데이터베이스 통합

- 분리(Separation)

분리는 하나의 데이터를 여러 개의 분리된 필드로 나누는 작업으로, 여러 개의 데이터의 하나의 필드로 묶는 concatenation 과 상대적 개념이다.

- 정규화(Normalization)

데이터 웨어하우스의 데이터 변환 과정에서 정규화(normalization)는 하나의 레코드(record)를 여러 개의 레코드로 나누는 작업이다.

- 일치(Reconciliation)

일치는 데이터의 내용상의 차이들을 줄이고 데이터의 일관성을 향상시키기 위한 방법이다.

- 클렌징(cleansing)

space 문제와 같은 필드손실, 띄워 쓰기, 오타, 필드 추가로 인한 Null 값에 대하여 오류를 수정하는 방법이다. 데이터 오류를 해결하는 방법으로는 일반적으로 수작업을 하며, 클렌징 도구를 사용하기도 한다.

2) 데이터 모델링

데이터베이스에 포함된 데이터 항목을 효율적이고 체계적으로 이용하기 위해서는 데이터에 대한 모델링을 한 후에 이에 근거하여 데이터베이스를 구축하여야 한다. 데이터 모델링이란 병원이나 기업의 정보구조를 실체(entity)와 관계(relationship)를 중심으로 명확하게 체계적으로 표현하고 문서화 하는 기법이다. 데이터 모델링은 연관 조직의 정보 요구에 대한 정확한 이해를 제공하고 분석과 개발과 사용자간의 의사 소통 수단으로 활용되며 시스템 개발의 기초 자료의 제공을 그 목적으로 한다. 데이터 모델링 단계는 크게 논리적 데이터 모델링과 물리적 데이터 모델링으로 구분한다. 논리적 데이터 모델링이란 데이터 구조에 대한 논리적인 설계를 하는 단계로써 정확한 업무 중심으로 자료 항목을 추출된 실체와 속성들의 관계를 구조적으로 설계하는 단계이다(오해석, 1996).

데이터 웨어하우스내의 모델링 측면을 살펴보면, 두가지 핵심 요소인 통합성과 다차원성을 고려해야 한다. 통합성이란 다양한 데이터 원본으로부터 입수되어 서로 다른 구조와 형식을 가지고 있는 데이터들이 데이터 웨어하우스 내에서 하나의 구조와 형식을 가지게 됨을 의미한다. 다차원성이란 데이터가 분석을 위한 여러 가지 차원을 가지고 있음을 의미한다. 이와 같은 다차원적인 데이터 구조를 표현하기 위한 방법이 다차원적 모델링(multidimensional modeling)이다. 다차원 모델링은 ROLAP 과 MOLAP 모두에서 수행할 수 있으나, MOLAP 에서는 다차원 분석을 위해 특별한 모델링이 필요하지 않다. 다차원 데이데이터베이스의 구조 자체에서 이를 지원하기 때문이며, 결국 다차원 모델링은 ROLAP 에 필요한 모델링 기법이다. 다차원 모델링은 사용자의 다양한 요구사항을 충족시키는 복잡한 질의를 신속하게 처리하는 데 목적이 있다. 따라서 다차원 모델링의 내용은 사용자의 요구사항, RDBMS 의 기능, ROLAP 도구 등 세가지 요소에 따라서 변하게 된다.

데이터 웨어하우스를 위하여 현재까지 제시된 모델표현방식으로는 사실(fact)과 차원(dimension)만을 사용하는 star 스키마 구조와 star 스키마에서 차원을 정규화(normalization)시킨 snowflake 스키마가 존재한다. 두 기법을 혼용해서 사용하기도 한다.

데이터의 빠른 접근을 중요하게 여기는 데이터 웨어하우스의 특성상 star 스키마에서는 데이터의 무결성보다 정보의 무결성(information integrity)을 중요시하게 되고 경우에 따라서는 데이터의 중복도 허용하는 반(reverse) 정규화를 기피하지 않으며 테이블 수를 절대적으로 줄이려고 한다. 차원 테이블들은 대량의 레코드를 갖는 사실 테이블 속에서 찾고자 하는 레코드를 join 을 통하여 빠르게 접근하게 하기 위한 수단으로서 활용된다(고재성, 1998).

하나의 사실 테이블에는 이런 차원이 하나 이상 존재할 수 있으며 이러한 이유로 데이터 웨어하우스에서의 데이터 모델링은 다차원 모델링으로 불리기도 한다(Ralph Kimball, 1996). star 스키마는 사실테이블을 중심으로 차원 테이블에 대해서는 전혀 정규화가 일어나지 않은 형태이다. 차원 테이블의 속성들(attribute)에 대해서 다시 정규화 하게 되면 이를 snowflake 스키마라고 부른다.

- star 스키마

  중앙에 사실테이블을 가지면서 주위에 reference 데이터를 포함하는 차원 테이블을 갖는 논리적 구조이다.

- snowflake 스키마

  차원요소(dimension Attribute)마다 1 개의 차원 테이블을 가지고 있고, 이 차원 테이블에는 차원요소에 대한 정보와 상위 차원 테이블로 join 하기 위한 키가 들어 있다. snowflake 스키마의 차원 테이블은 정규화가 잘 되어 있어서 중복이 안 되며, 차원 테이블의 크기도 작다.

표 2. star 스키마와 snowflake 스키마의 비교

	star 스키마	Snowflake 스키마와
장점	- join 의 수가 적음 - SQL 이 간단 - join 을 줄여줌으로써 성능 향상 - 복잡한 구조가 쉽게 모델링 됨 - 숫자 형태의 단위는 fact 에만 존재	- 한 dimension 내의 여러 개의 계층 지원 - cross dimensional modeling - dimension table 이 많은 경우도 지원 - 데이터베이스 마케팅 지원
단점	- 비교적 저장 공간이 많이 듬(데이터의 중복성 허용) - cross dimensional modeling 이 안 됨	- join 의 수가 많음(작은 dimension table 간) - SQL 문장이 복잡

3) 사용자를 위한 데이터 접근 방법

의사 결정권자가 이용하는 프로그램을 분류하자면, 리포트 생성, 데이터 마이닝/통계, 데이터 브라우징이 있다. 리포트 생성 프로그램은 잘 알려진 업무 처리를 요약해준다. 마이닝/ 통계 프로그램은 대용량의 데이터를 분석해서, 업무의 동향이나 새로운 전환점을 찾게 해준다. 브라우징은 OLAP 라고도 불리는데, 데이터를 적당한 통합 단계에서 빨리 처리하는 방법을 제공

해준다. OLAP 는 이해하기 쉽고, 자주 이용되는 테이블 내의 요약을 기본으로 하고 있다. 테이블 내에서 두개의 OLAP cube 를 대용량의 데이터베이스로부터 구축할 때는 많은 계산량을 필요로 한다.

IV. 연구 결과

1. 의료 데이터 웨어하우스 구조

가. 데이터 웨어하우스 구조

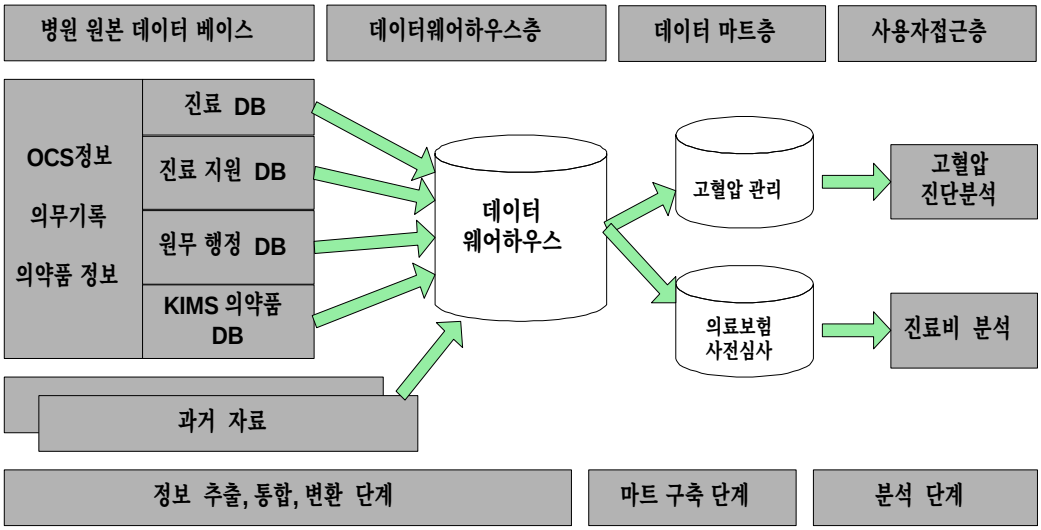


그림 6. 데이터 웨어하우스 구조

데이터 웨어하우스내의 데이터흐름을 염두해 두고 의료용 데이터 웨어하우스 구조를 그림 6 과 같이 도출하였다.

나. 데이터 웨어하우스내의 데이터 흐름

전체적인 병원에서의 데이터 웨어하우스의 흐름은 데이터 흐름과 메타 흐름으로 구성된다. 데이터 흐름은 입력 흐름, 상향 흐름, 하향 흐름, 출력 흐름으로 구성된다. 상향 흐름과 하향 흐름은 데이터베이스 자체 흐름이다. 본 연구에서는 개발 환경의 특수성(자료 양의 소규모, 하향 흐름 불필요 등)을 고려하여 하향 흐름은 구축하지 않았다.

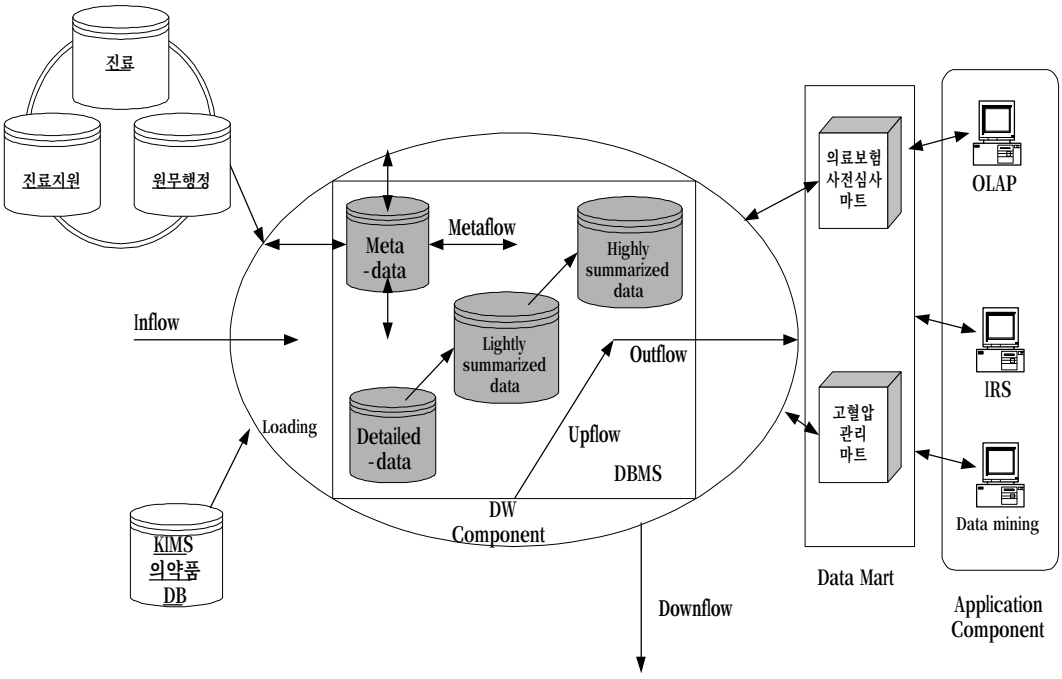


그림 7. 병원 데이터 웨어하우스 내의 데이터 흐름

1) 입력 흐름의 개발

기존의 운영 데이터베이스로부터 데이터 웨어하우스에 데이터를 적재하기 위해서, 임상병리 및 방사선실 데이터베이스와 OCS 데이터베이스와 같은 운영계 데이터베이스로부터 데이터를 추출하고 의무 기록 데이터와 심혈관 센터의 고혈압 환자에 대한 설문 데이터, 의약품정보 데이터와 같은 외부 데이터를 수집하여 변환 및 통합 과정을 거쳤다. 사전 심사 및 고혈압 전문가와 인터뷰를 통하여 데이터의 표준화 및 변환 작업을 하였으며, 운영 데이터베이스의 대상 필드를 SQL 로 변환함으로써 데이터의 입력 흐름을 완성하였다.

2) 상향 흐름의 개발

데이터의 요약과정으로 차원 내에 정의된 속성들 간에는 일반적으로 잘 정의된 속성 계층이 존재한다. 기간 차원 내에는 속성으로 년, 월, 일이 존재하는데, 계층적으로 년 아래에 월이, 월 아래에 일이 들어간다. 따라서 속성 계층은 하향식 데이터 구조를 제공하며, 이를 토대로 사용자는 한 단계 아래의 상세 정보를 검색할 수 있고 한 단계 위의 정보 또한 검색할 수 있다. 예를 들면, M_fact 에서 퇴원일을 기준으로 월별, 분기별, 년도별로 요약이 된다.

3) 출력 흐름의 개발

의료용 의사 결정 지원 시스템 사용자에게 궁극적인 다차원 분석 정보를 제공해 주는 출력 흐름은 사용자들의 다차원 질의를 시각적으로 구성하여

다차원 분석에 질의(query)형태로 변환하여 전달하도록 구현하였다. 웨어하우스 내의 데이터를 의료진 또는 의사 결정자에게 이용되는 전단계로서 의료 보험 사전 심사 데이터 마트와 고혈압 관리 데이터 마트를 구성하였다. 이러한 데이터는 OLAP, DSS, 데이터 마이닝을 통해 의료진에게 의사 결정 지원을 하도록 하였다.

4) 메타 데이터

메타 데이터의 테이블은 데이터 웨어하우스를 구성하고 있는 병원 운영계 시스템의 테이블에 대한 정보로서 각 운영 시스템의 테이블/파일의 정보, 그리고 각 테이블/파일의 컬럼/필드에 대한 정보이다. 데이터의 변환에 대한 정보도 각각의 작업을 구성하고 있는 과제들에 대한 정보로 구성되어 있다. 데이터 웨어하우스를 구성하는 원본 시스템 개체, 원본 테이블(source table) 개체, 그리고 원본 요소(source data)개체는 운영시스템으로부터 추출한 원본 데이터에 대한 정보를 관리한다. 또한, 매핑(mapping)개체는 원본 데이터와 데이터 웨어하우스에 저장되어 있는 데이터간의 관계에 대한 정보를 저장하고, 마트를 구성하는 테이블을 관리하게 된다.

2. 병원의 의사 결정지원을 위한 데이터 웨어하우스 구축

가. 원본 데이터

원무과의 운영시스템으로부터 받은 데이터는 다음과 같다. 99 년 1-6 월 퇴원한 보험환자의 급여 현황에 대한 급여 데이터가 처방 건수 기준으로 2,032,016 건이고, 동일 기간 퇴원한 보험 환자 중 비급여 항목에 관한 비급여 데이터가 처방 건수 기준으로 677,193 건, 지정 진료비 항목에 대한 지정 진료비 데이터가 처방 건수 기준으로 220, 541 건이었다. 그 외에 이의 신청건에 대한 데이터로 보험청구데이터베이스와 임상병리검사 데이터베이스를 추출하였다. 또한 심혈관 센터의 운영 시스템으로부터 받은 데이터는 1999 년 5, 6, 7 월 동안 Y 병원 심혈관 센터를 방문한 입원환자 중에서 주진단 또는 부진단중에 고혈압이 있는 환자에 대한 퇴원 요약 데이터베이스, 외래 처방 데이터베이스, 임상 병리 검사 데이터베이스, 기타 검사 데이터베이스, 약국 데이터베이스를 추출하였다.

표 3. 진료 부문 데이터베이스

구분	테이블명	내용	DB형태	레코드수
진료정보	Chief_complaint	주진단명리스트	mdb	8,022
	Out_dx	퇴원환자진단명정보	rpt	8,048
	Hyper	문진설문지	dbf	944
수술정보	Out_oper	퇴원환자수술정보	rpt	597

표 4. 진료 지원 부문 데이터베이스

구분	테이블명	내용	DB형태	레코드수
임상병리검사 (사전 심사)	Lab9902	임상병리검사9902	mdb	95,465
	Lab9903	임상병리검사9903	mdb	110,023
	Lab9904	임상병리검사9904	mdb	100,982
	Lab9905	임상병리검사9905	mdb	100,866
	Lab9906	임상병리검사9906	mdb	87,496
임상병리검사	입원일반화학검사1	입원일반화학검사	rpt	51,695
	입원일반화학검사2	입원일반화학검사	rpt	56,613
	입원일반화학검사3	입원반화학검사	rpt	46,990
	입원일반화학검사4	입원일반화학검사	rpt	33,400
	입원일반화학검사5	입원일반화학검사	rpt	24,822
	입원혈액학검사1	입원일반화학검사	rpt	58,579
	입원혈액학검사2	입원일반화학검사	rpt	68,127
	입원혈액학검사3	입원일반화학검사	rpt	60,217
	입원혈액학검사4	입원일반화학검사	rpt	53,774
	입원혈액학검사5	입원일반화학검사	rpt	50,363
	입원기타검사1	입원기타검사	rpt	2,055
	입원기타검사2	입원기타검사	rpt	1,891
	입원기타검사3	입원기타검사	rpt	1,148
	외래기타검사1	외래기타검사	rpt	2,935

	외래일반화학검사1	외래일반화학검사	rpt	47,858
	외래일반화학검사2	외래일반화학검사	rpt	48,362
	외래일반화학검사3	외래일반화학검사	rpt	44,656
	외래혈액학검사1	외래혈액학검사	rpt	60,662
	외래혈액학검사2	외래혈액학검사	rpt	59,999
	외래혈액학검사3	외래혈액학검사	rpt	53,425
방사선검사	Chest	방사선검사	rpt	2,761
약국처방	Druglist	입원약국자료	rpt	146,969
	DrugEDI	의약품 EDI 코드	mdb	1,419
의무 기록	Pat_list	외래환자명단	rpt	11,506
	Hyper11	퇴원요약지	dbf	967
	Out_pat	퇴원환자기본정보	rpt	1,868

표 5. 원무 행정부문 데이터베이스

구분	테이블명	내용	DB형태	레코드수
지정진료비내역	Money1	지정진료비	mdb	103,828
비급여내역	Nopay	비급여	mdb	208,086
보험청구내역	Song_9901	보험청구9901	mdb	142,195
	Song_9902	보험청구9902	mdb	118,220
	Song_9903	보험청구9903	mdb	140,066
	Song_9904	보험청구9904	mdb	132,911
	Song_9905	보험청구9905	mdb	128,455
	Song_9906	보험청구9906	mdb	138,468

표 6. 외부 데이터베이스

구분	테이블명	내용	DB형태	레코드수
의약품정보	KIMS	KIMS의약품정보1	mdb	5,876
	KIMSi	KIMS의약품정보2	mdb	8,246

나. 데이터 변환

대개의 경우, 원본 시스템에서 추출한 데이터를 그대로 데이터 웨어하우스의

사실테이블(fact table)로 만들 수 없으므로 데이터 변환을 시켜야 한다. 데이터의 무결성은 데이터 웨어하우스의 중요한 요소이며 원본 시스템 경우보다 데이터의 각종 무결성을 체크하여야 한다.

의무 기록 데이터, OCS 운영 데이터, 의약품 데이터 등의 원본 데이터를 추출·변환·통합한다. 데이터 웨어하우스 데이터베이스를 구축함에 있어서 각 데이터의 형태 및 크기를 구분한다. 원본 데이터로부터 변환 개체의 생성은 세부변환프로세스에 의하여 구체화된다. 이들 각각 프로세스에 대해 설명하면 다음과 같다.

1) 통합

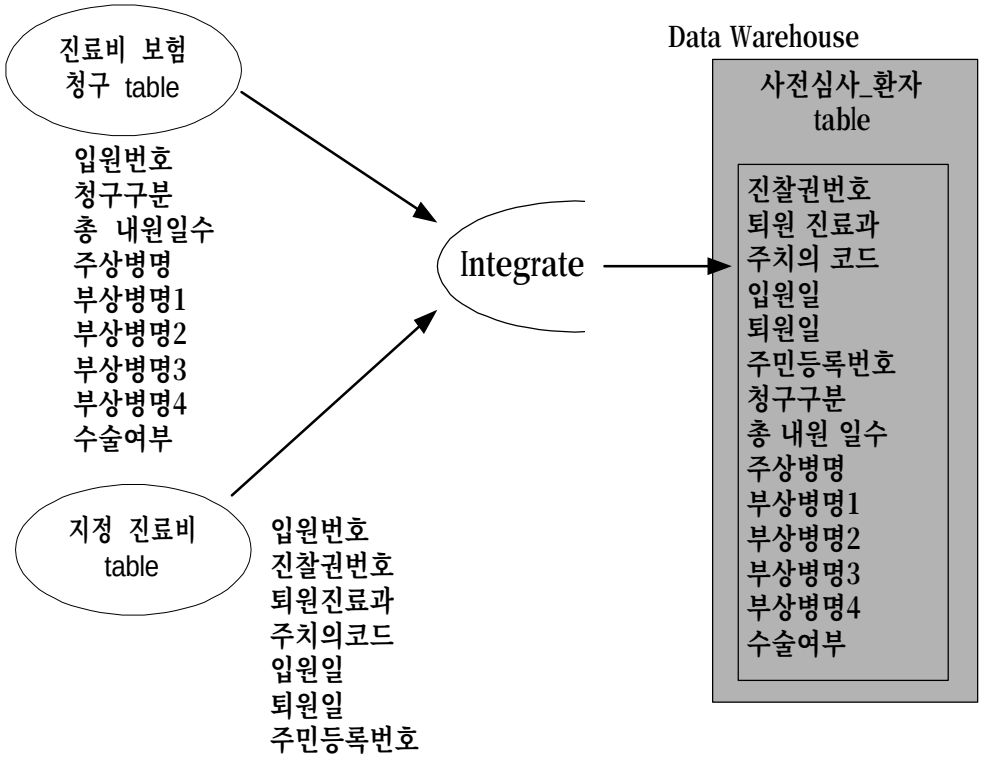


그림 8. 데이터 변환 – 통합

한 환자에 관한 정보들이 여러 테이블로 나누어져 있다. 따라서 사전 심사를 위한 환자의 기본정보테이블을 구성하기 위하여 원무계의 지정 진료비 테이블과 사전 심사계의 진료비 보험 청구비 테이블의 필드를 통합하였다.

2) **정규화**

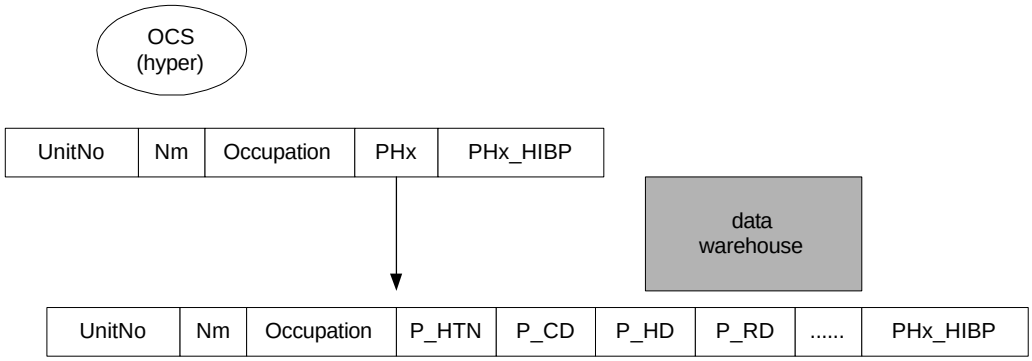


그림 9. **데이터 변환 – 정규화**

정규화는 하나의 필드가 여러 개의 필드를 생성하는 것을 말하는데, 이 연구의 데이터에서는 과거력이나 가족력에 관한 데이터에 이러한 방법들이 필요하였다. 과거력이라고 하면, 한 환자의 과거 병력을 의미하는데, 질병이 다양하므로 과거력 또한 수없이 많은 경우를 가지게 된다. 그러나 구축하고자 하는 것은 심혈관 센터의 고혈압 환자에 대한 데이터 웨어하우스를 구축하는 것이므로 고혈압과 관련있는 질환 12 **가지를 기준으로 하여** 12 **가지 과거력의 여부를 가지도록 하였다. 위의 그림에서 보듯이,** PHx(**과거력**)**이** index **를 이용하여** P_HTN(**고혈압**), P_CD(**뇌혈관 질환**), P_HD(**심장 질환**), P_RD(**신장 질환**), P_VD(**혈관 질환**), P_retina(**고혈압성 망막증**), P_BM(**당뇨병**), P_lipid(**고지혈증**), P_gont(**통풍**), P_broncho(**기관지수축**), P_sexual(**성기능 장애**)**등 여러 필드의 값들을 채우는 형식으로 변환하였다.**

3) 분리

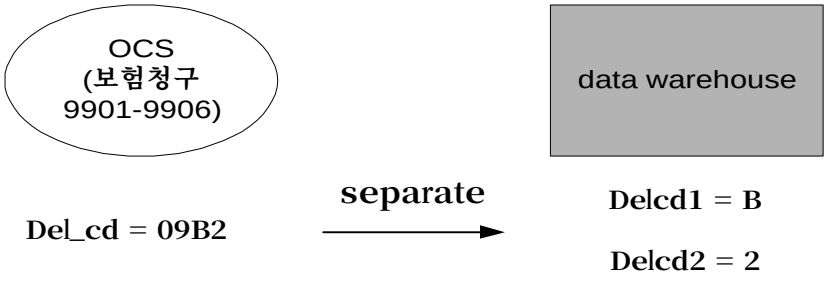


그림 10. 데이터 변환 – 분리

분리는 하나의 데이터 값을 여러 필드로 구분하는 것을 의미한다. OCS 의 보험청구테이블에는 진료비 삭감에 관한 필드로 Del_cd 가 있는데, 이것은 Delcd1(삭감사유)와 Delcd2(재료/행위구분)으로 구분할 수 있다. Del_cd 중에서 가운데 알파벳값이 Delcd1(삭감사유)에 해당하고 Delcd2(재료/행위구분)에 해당하는 것이다. 또한, 의약품 정보 테이블의 경우에도 Moh 라는 필드의 값을 DrugNo(복지부 분류 번호), kimscode(KIMS 분류 번호), kimsclass(KIMS 분류 영문 명칭)으로 구분하였다.

4) 일치

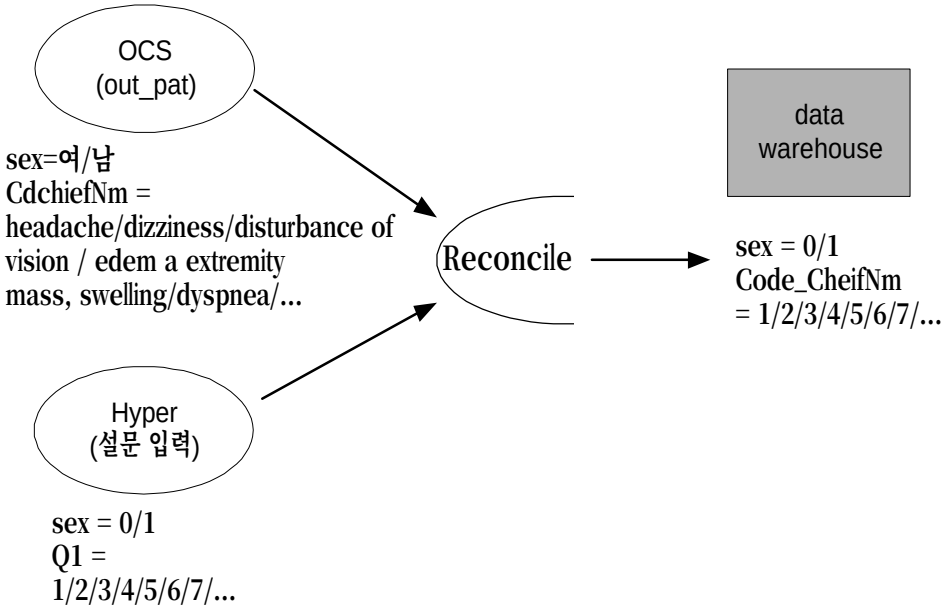


그림 11. 데이터 변환 – 일치

OCS 로부터 받은 퇴원 요약 테이블에는 성별이 남/여로 구분이 되어있으며, 심혈관센터에서 시행하고 있는 환자설문 입력 테이블에는 성별이 1/0 으로 입력되어 있으므로, 이것을 일치시킬 필요가 있다. 따라서 데이터 웨어하우스에서는 성별인 sex 필드에서 남자는 1, 여자는 0 으로 일치시키도록 하였다.

5) 조합(combination of column)

중요한 유도된 변수들은 여러 필드에 존재하는 값들을 사용하여 만들어

진다. 이것은 레코드에 존재하는 데이터를 이용하여 새로운 변수를 만들게 된다. 고혈압 분석에 필요한 고혈압 관리 마트 구성할 때, 실제로 키나 몸무게에 관련된 필드를 사용하는 경우보다는 비만도에 대한 데이터가 필요하다. 따라서 비만도의 계산을 위해 함수를 사용하여 관련된 두개의 필드를 새로운 하나의 필드로 생성하여 분석에 사용하였다. 비만도(BMI)는 (weight)/키(height)^2 로 계산이 된다. 먼저 BMI 라는 필드를 생성하고 다음의 SQL 질의를 수행하였다.

SQL Query :

UPDATE mart_hyper

SET mart_hyper.BMI = mart_hyper.Weight/(mart_hyper.Height/100)^2

6) index 및 join

실제로 병원에서 환자의 합병증과 관련된 필드들에 표준화된 Index 가 부여되어 있지 않고 병원 자체에서 사용하는 Index 가 따로 존재 했다. 따라서 분석에 필요한 표준화 작업에 ICD10 코드를 활용했고, 그 ICD10 코드에서도 고혈압분석에 필요한 질병의 분류가 현재 없는 상태이기에 고혈압과 연관되어 있는 질환들에 대한 질병군들을 11 단계로 분류하여 Index 를 생성할 필요가 있으므로 다음과 같이 index 를 통해 join 하였다. 합병증과 관련된 필드 생성작업(11 개의 필드: com_HTN com_Sexual)을 설명하면 다음과 같다. 먼저, 각 질환별로 고혈압과 관련된 질환을 기준으로 ICD10 코드를 사용하여 Index 를 부여하는데, ICD Table 에 C_Class 를 추가하여 고혈압과 관련된 12 가지 질병에 관한 Index 를 부여하고 ICD Table 의 ICD 필드와 Mart_Hyper Table 의 MainDiseCd DiseCd20 필드들은 같은 ICD10 코드이

므로 각각 새로운 필드 20 **개**(C_Code1 C_Code20)**를 만들어** Left Join **을 이용하여** ICD Table **의** C_Class **의** Index **가** C_Code1 C_Code20 **까지 삽입되도록 하였다**.

SQL Coding :

Update mart_hyper set mart_hyper.C_Code = ICD.C_Class

where mart_hyper.MainDiseCd = ICD.ICD

….

Update mart_hyper set mart_hyper.C_Code20 = ICD.C_Class

where mart_hyper.DiseCd20 = ICD.ICD

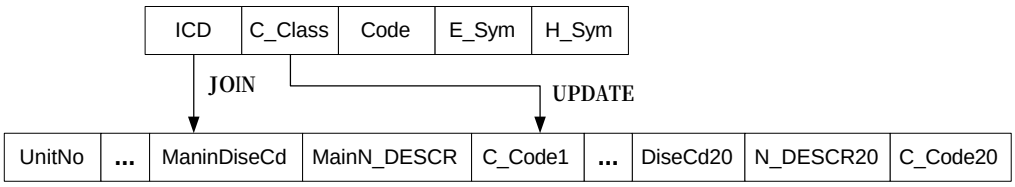


그림 12. **데이터 변환** – index **부여 및** join1

이제 합병증에 관한 새로운 필드들(11 **개의 필드** : com_HTN, Com_CD, …, Com_Sexual)**을 추가하여 각 질병의 여부를 확인하게 되는데**, **일단** 11 **개의 필드를 추가한 뒤 다음의 코딩을 질의로 작성 후 실행하였다**.

SQL Coding :

Update mart_hyper set mart_hyper.com_HTN = 1

where mart_hyper.C_Code1 = 1

….

Update mart_hyper set mart_hyper.com_Sexual = 1

where mart_hyper.C_Code20 = 11

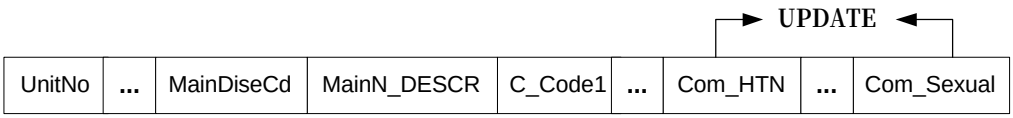


그림 13. **데이터 변환** – index **부여 및** join2

또한 병원에서 사용되고 있는 데이터의 종류들은 다수의 Text **형태의 데이터들로 구성되어 있다. 이러한** Text **의 데이터들은 분석할 때 검색이나** SQL **질의등에 상당히 많은 제약을 받기 때문에 이러한 데이터들의** index **부여에 의한 코드화는 필수적이다. 방사선 검사**(chest PA) **및 심전도 검사**(EKG), **안저촬영**(Retina) **등이 이에 해당한다**.

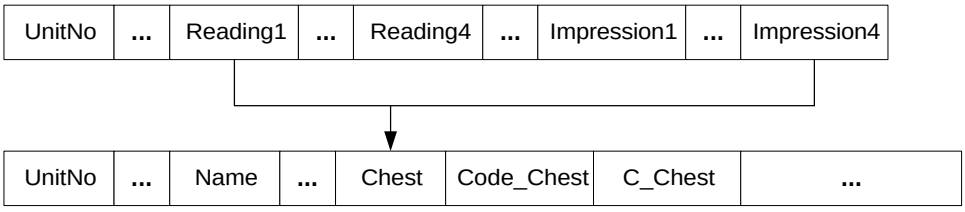


그림 14. **데이터 변환** – index **부여 및** join3

기존의 시스템의 Chest Table **의 데이터를 모두 합쳐서 데이터 웨어하우스** Chest **필드로 삽입한 뒤** Chest **필드의 내용을 참조하여** Code_chest **필드와** C_Chest **필드에 필요한** index **를 부여하였다**.

code_Chest Index – 0: **정상**, 1: Cardiomegaly, 2:Pulmonary congestion, 3:1+2

C_Chest Index – 0: **정상**, 1: **비정상**(code_Chest **가** 1,2,3 **인 경우**)

또한, 심전도 검사의 경우에도 EKG 필드의 내용에 따라 Code_EKG 필드와 C_EKG 필드에 필요한 index 를 부여하였다.

Code_EKG Index – 0: ischemia pattern(LVH), 2:arrythmia pattern, 3:1+2

C_EKG – 0: 정상, 1: 비정상(Code_EKG 가 1, 2, 3 인 경우)

병원에서 사용하고 있는 데이터를 살펴보면, 실제 검사 테이블 형태가 검사항목이라는 필드아래 모든 검사명이 입력되고 그에 따른 결과 필드가 있으나, 실제로 분석 시에는 시행되었던 검사명을 하나의 변수로서 사용하게 되므로 한 필드의 값들을 또다른 새루운 필드로 구성하였다.

다. 데이터 마트 모델링

데이터 웨어하우스 데이터베이스를 기반으로 데이터 분석에 필요한 항목을 정의하여 사전 심사 데이터 마트와 고혈압 데이터 마트를 설계한다. Access 데이터베이스로부터 요구 항목을 설정하기 위하여 MS-SQL 을 사용하여 질의하여 데이터 마트를 구성한다. 고혈압 마트는 심혈관센터 환자를 대상으로 주진단 또는 부진단 중에 고혈압이 있는 환자에 대한 퇴원 요약 데이터베이스, 외래 처방 데이터베이스, 임상 병리 검사 데이터베이스, 기타검사 데이터베이스, 약국 데이터베이스로부터 ICD 10 코드 중 I1_._ 인 환자만 추출하고, 비전산화 항목은 의무 기록으로부터 입력받은 데이터로 구성되었다.

고혈압 데이터 마트는 환자의 특성에 따라 다음과 같이 구분 될 수 있다. 고혈압 진단 및 치료에 고려해야 할 환자의 특성은 증상, 위험 요인, 이학적 검사, 검사실 검사 결과 및 표적 장기 손상과 같은 동반 질환, 질병의 중증도, 약물 치료에 대

한 반응(순응도, 부작용), 혈압의 변이 등이므로 이러한 환자의 특성들을 반영하는 변수들로 스키마를 구성하였다.

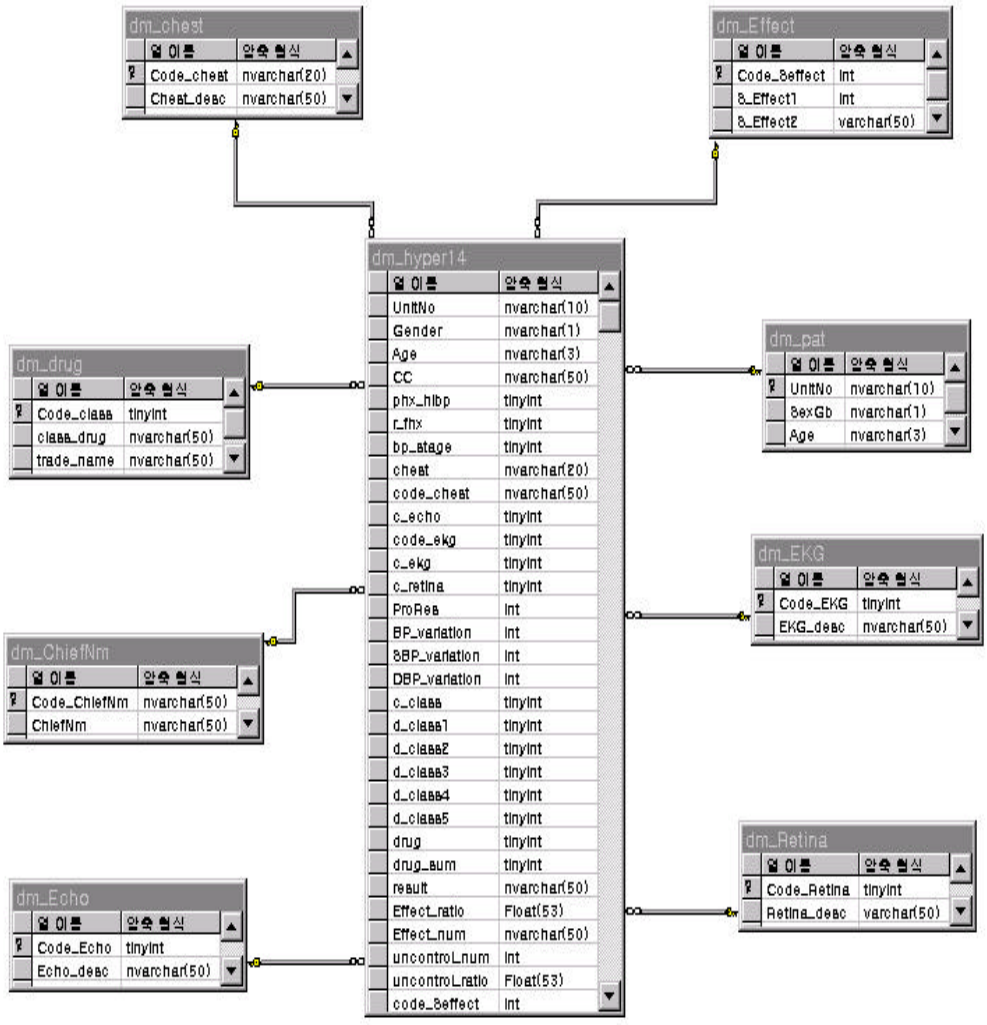


그림 15. 고혈압 관리 데이터 마트를 위한 star 스키마 일부

그림 15 는 의료용 데이터 웨어하우스에서 고혈압 관리 마트를 구성하는 고혈

압 관리에 대한 star 스키마의 일부를 보여주고 있다. 그림 15 에서 차원 테이블(dimension table)의 키에는 키표시가 있고, 사실 테이블(fact table)의 키는 차원테이블들의 키들로 구성되어진다. 이 스키마는 한 개의 사실 테이블과 여덟 개의 차원테이블로 구성되어있다(dm_echo, dm_retina, dm_pat, dm_EKG, dm_drug, dm_ChiefNm, dm_Chest, dm_Effect). Dm_hyper14 인 사실은 위의 여덟 개의 차원에 의해 요약되어진다.

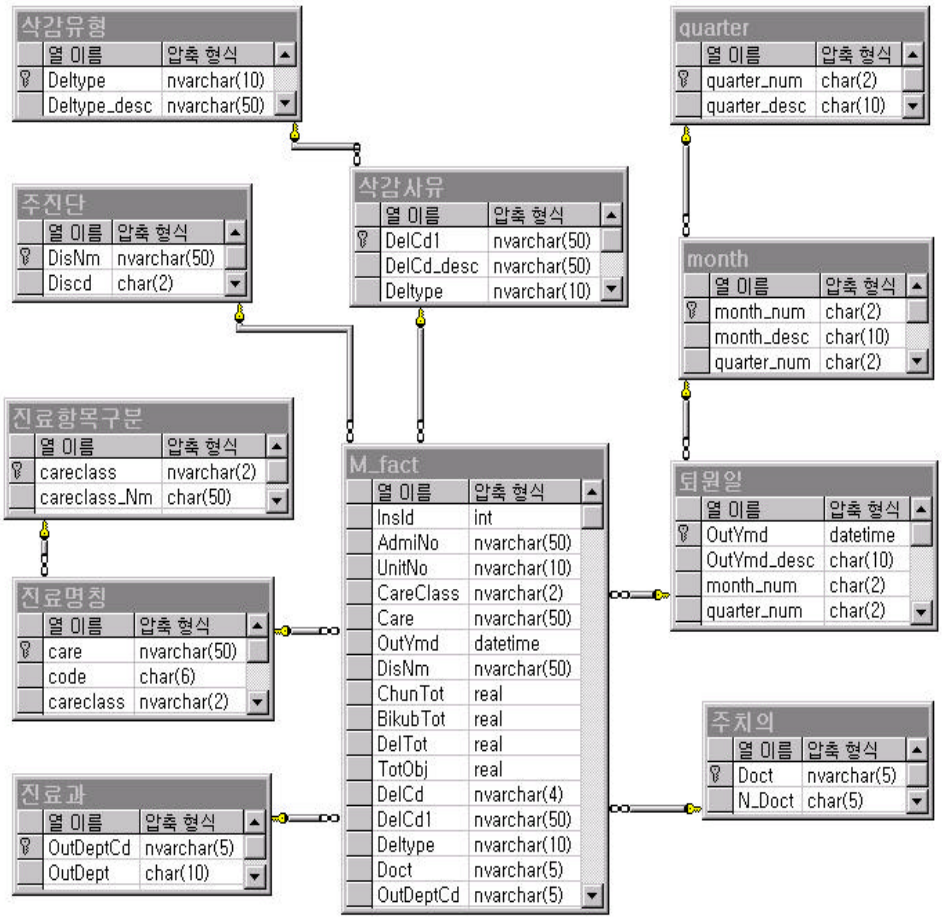


그림 16. 사전 심사 데이터 마트를 위한 snowflake 스키마 일부

의료 보험 사전 심사 데이터 마트는 퇴원한 입원 환자에 대한 OCS 처방데이터베이스, 의료 보험 진료비 청구 데이터베이스, 임상 병리 검사 데이터베이스, 의료 보험 진료비 삭감 및 이의 신청 데이터베이스로 구성하였다. 의료 보험 사전 심사 데이터 마트는 환자의 특성에 따라 다음과 같이 구분 될 수 있다. 의료 보험 진료비 청구 환자를 대상으로 청구 및 삭감, 이의 신청에 고려해야 할 사항은 진료비 삭감 유형, 진료 항목, 진료 명칭, 진료과, 삭감 사유, 주치의, 검사실 검사 결과 및 약물 치료 등이므로 이러한 환자의 특성들을 반영하는 변수들로 스키마를 구성하였다.

그림 16 은 의료용 데이터 웨어하우스에서 의료 보험 사전 심사 마트를 구성하는 의료 보험 사전 심사에 대한 snowflake 스키마를 보여주고 있다. 그림 16 에서 차원 테이블의 키에는 키 표가 있고, 사실 테이블의 키는 차원 테이블들의 키들로 구성되어진다. 이 스키마는 한 개의 사실 테이블과 열 개의 차원 테이블로 구성되었다(DM_Delcd, DM_carelass, DM_care, DM_outdept, DM_doct, DM_discd, DM_admi, DM_unit, DM_out, DM_Deltype). M_fact 사실 테이블의 ChunTot(총 청구횟수), bikubTot(총 비급여횟수), DelTot(총 삭감횟수), TotObj(총 이의 신청횟수)는 차원테이블에 의해서 요약되어진다.

라. 사용자의 데이터 접근을 위한 어플리케이션 개발

운영 데이터베이스는 심혈관 센터, 의무 기록과, 원무과, 임상 병리과, 방사선과 등에서 받는 데이터베이스이며, 외부 데이터베이스는 의약품 데이터베이스, 의무 기록 차트 입력 데이터베이스 등 몇 개로 구성되어있다. 어플리케이션 컴포넌트(application component)는 의료 보험 심사 시스템과 고혈압 환자 분석 시스템으로 구성되어 있다. 의료 보험 심사 시스템과 고혈압 환자 분석 시스템은 의약품

검색을 위하여 AID Search 엔진을 이용하여 Visual C++로 구현하였다. 의약품 검색 모듈과 분석모듈을 개발하였는데, OLAP 을 위한 GUI(Graphical User Interface)는 VBA(Visual Basic for Application)로 개발하였다.

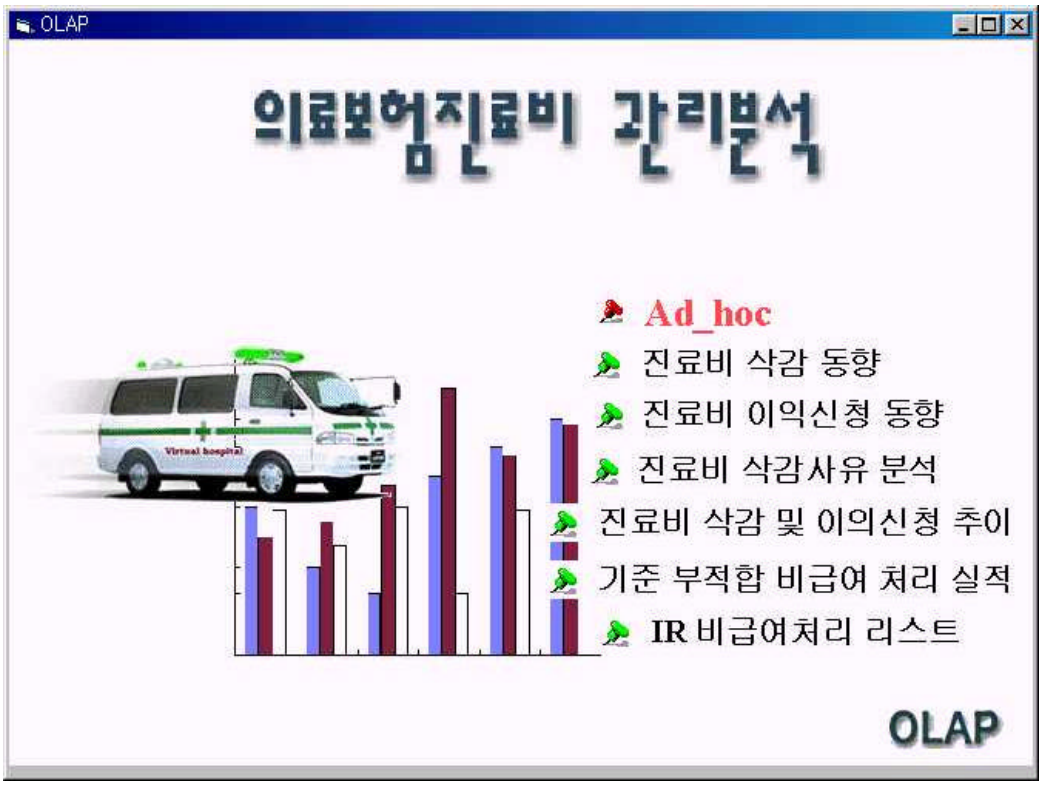


그림 17. 사전 심사 OLAP 초기화면

그림 17 은 의료 보험 사전 심사 데이터를 이용하여 OLAP 분석을 할 수 있는 초기화면이다. 데이터 웨어하우스 설계 시 주제 영역별로 나뉘어진 데이터 원본을 사용자는 버튼 클릭으로 OLAP 분석을 수행할 수 있게 했다. 예를 들면, Ad-hoc 버튼을 클릭할 경우 사용자가 임시적으로 원하는 영역에 대하여 항목들의 분석은 Ad-hoc 버튼 클릭으로 각 주제 영역의 측정치들을 모두 선택적으로 분석하여

사용자 관점의 의사 결정을 수행하게 할 수 있다.

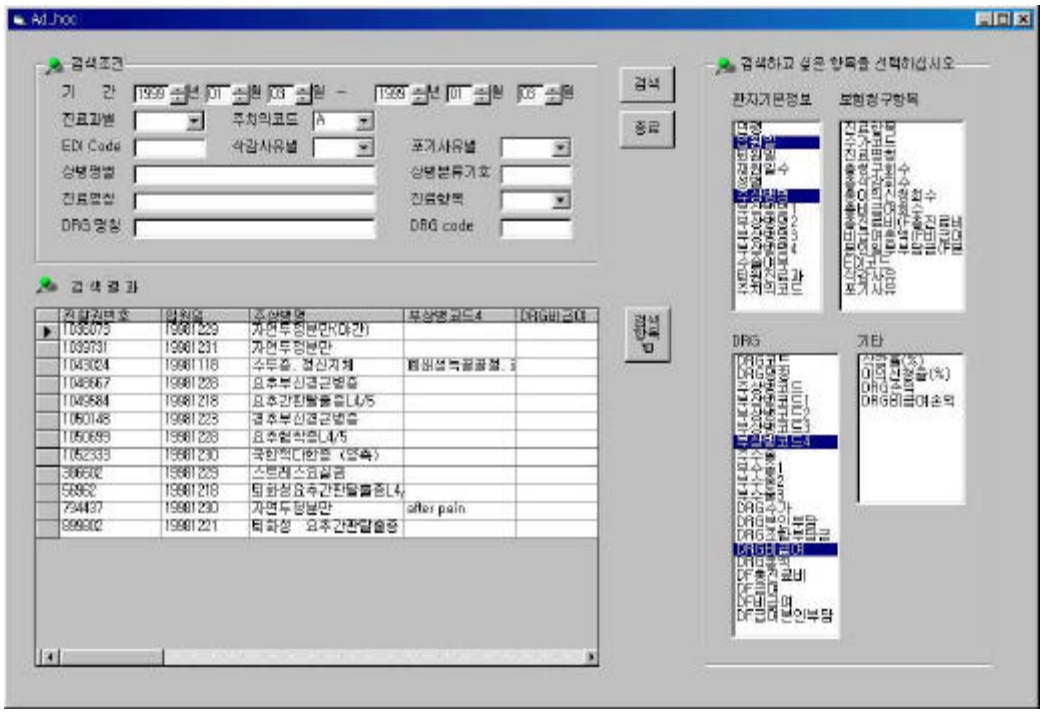


그림 18. 의료 보험 사전 심사 조건 분석 화면

그림 18 은 사용자가 필요한 목적에 따라 다각적으로 분석이 가능하도록 검색 조건과 검색 항목을 선택하여 자료를 산출하도록 설계하였다. 원하는 검색 조건별로 선택가능하며, 화면 오른쪽의 구분별 속성 선택 화면이 나타나고, 화면 하단 왼쪽에는 사용자가 차원별 속성들과 측정속성별로 원하는 정보를 얻을 수 있게 하였다. 또한 DRG 전면 실시에 대비하여 DRG 와 현행 행위별 수가제 하에서의 검색 항목을 넣어 DRG 시행에 대비하여 진료비 관련 정보를 손쉽게 사용자들이 획득하여 진료비 관리 활동을 위한 자료를 산출하는데 소요되는 시간을 줄이고자 하였다.

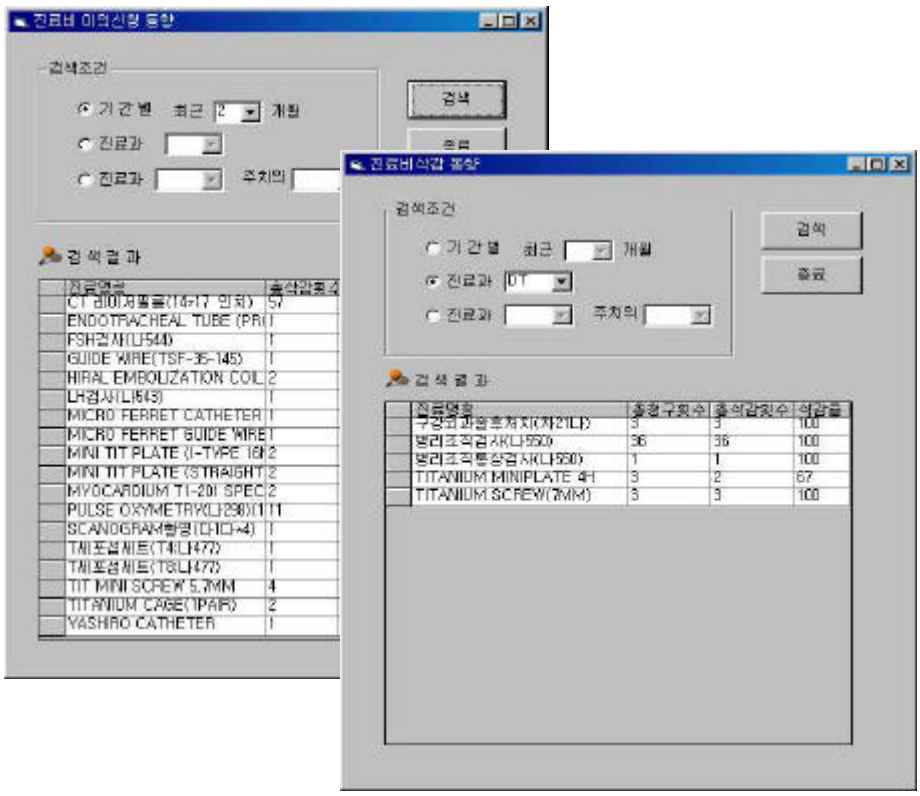


그림 19. 진료비 삭감 및 이의 신청 동향 화면

그림 19는 차원 속성 선택 영역에서 화면 상단에 기간별, 진료과별, 진료과의 주치의 별의 차원속성이 위치하고 하단 컬럼에는 진료 명칭 별로 총청구횟수, 총 비급여 횟수 등의 측정량을 나타낸 화면이다. 진료 명칭은 행위별 진료명칭으로 써 최근의 주요 삭감되는 진료항목 및 이의 신청 항목에 대하여 원하는 기간별, 진료과 별, 주치의 별로 대처 방안을 세울 수 있도록 삭감 동향을 파악할 수 있게 하였다.

그림 20. 진료비 삭감 및 이의 신청 추이 분석 화면

그림 20 은 최근 3 개월간의 진료 항목 별로 진료비 삭감율, 이의 신청율, 삭감 사유의 추이를 파악하기 용이하도록 그래프를 이용하여 보여줌으로써 정보 분석이 용이하도록 하였다. 진료항목이란 기본 진료료(초재진료, 응급 의료 관리료 등), 입원료, 투약 및 조제료, 주사 및 주사료, 마취료, 이학요법료, 정신요법료, 수술 및 처치료, 검사료, 방사선료, CT 료 등이다.

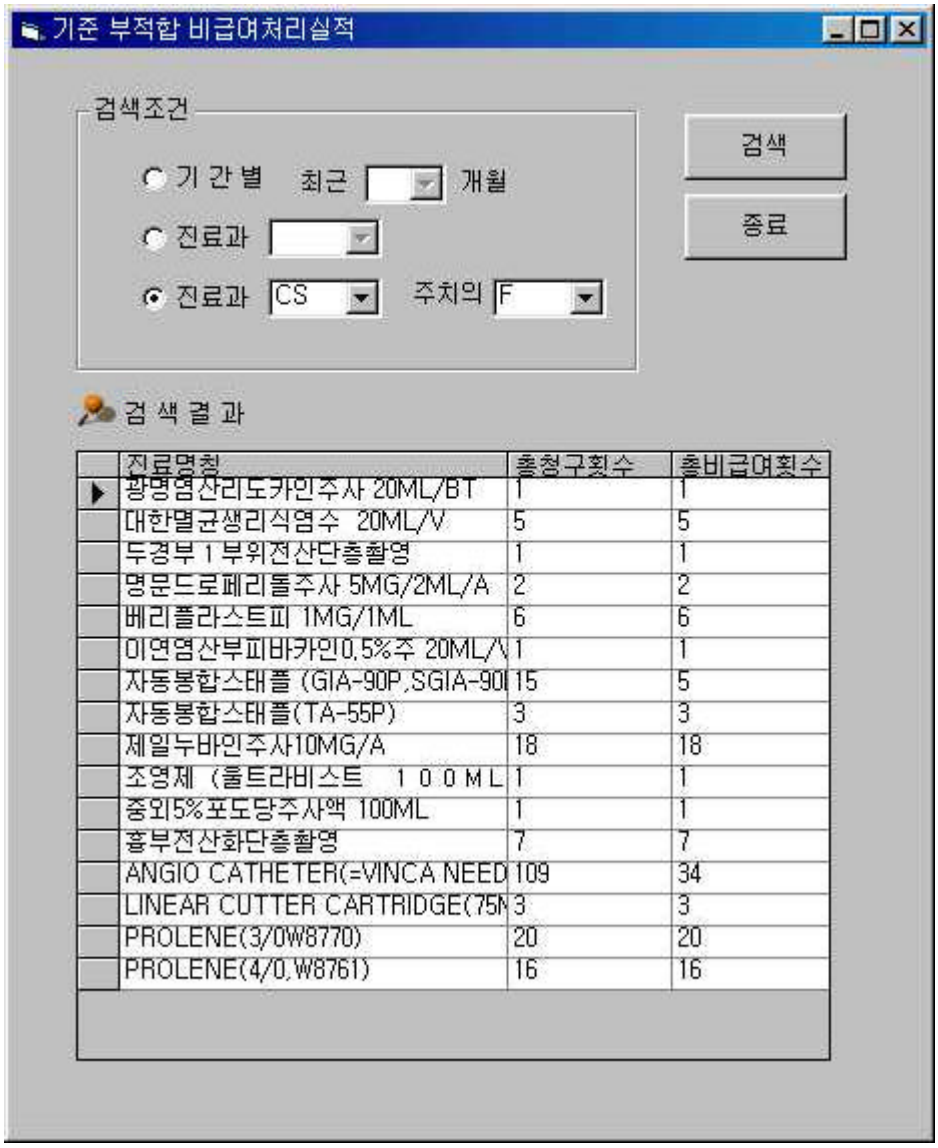


그림 21. 기준 부적합 비급여 처리 실적 분석 화면

보험 기준이 부적합하여 급여 청구가 불가능하다고 판단되어 비급여 처리할 수가 있으나, 비급여의 남발과 비급여로 인한 진료비의 민원이 야기될 수가 있다.

따라서 그림 21 은 진료처방의 임의 비급여 처리 행위에 대한 적정성 관리를 위한 비급여 처리 정보를 제공하고자 한 것이다.

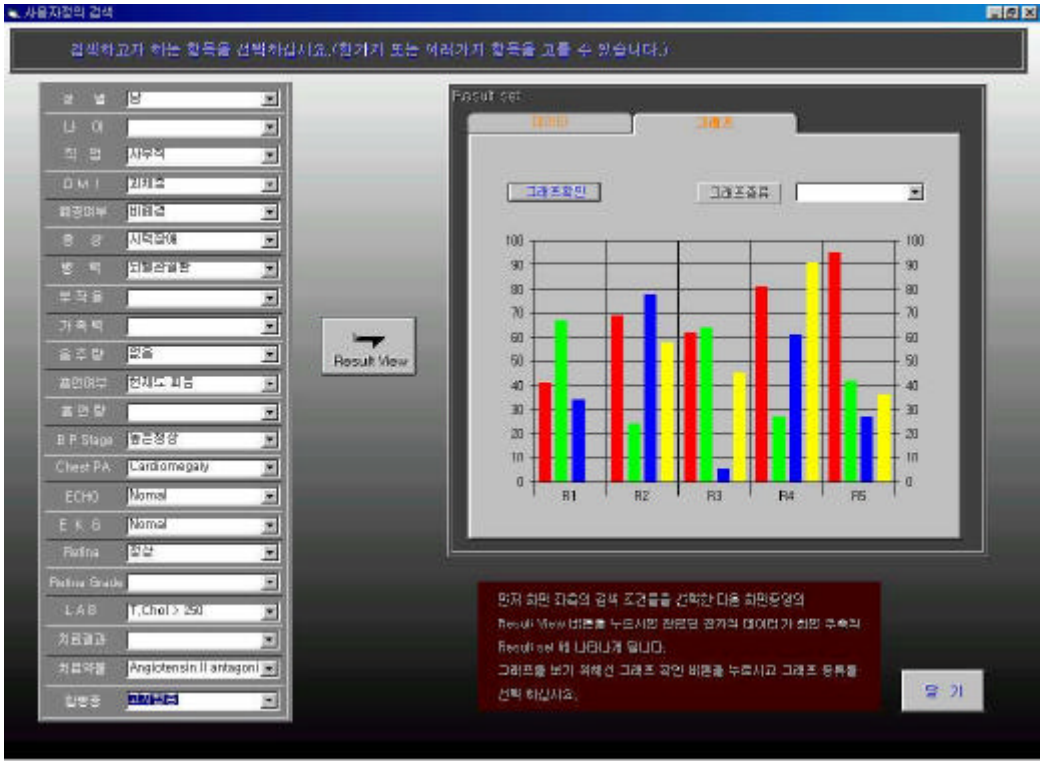


그림 22. 고혈압 관리 사용자 조건 분석 화면

그림 22 는 고혈압 관리를 위한 OLAP 화면중의 하나이다. 사용자가 필요한 목적에 따라 다각적으로 분석이 가능하도록 검색 조건과 검색 항목을 선택하여 자료를 산출하도록 설계하였다. 원하는 검색 조건별로 선택가능하며, 화면 왼쪽의 속성 선택화면이 나타나고, 화면 오른쪽에는 선택 속성에 따른 혈압의 변이를 나타내어준다. 조건으로 한 속성들은 고혈압 진단 및 치료에 영향을 미치는 환자의 특성으로 위험 요인, 표적 장기 손상과 같은 동반 질환, 질병의 중증도, 약물 치료

에 대한 반응, 약물 치료 등이며, 검색 항목에 대해 혈압의 변이를 의료진들이 쉽게 확인할 수 있도록 하여 진료 지원을 위한 자료를 산출하는데 소요되는 시간을 줄이고자 하였다.

그림 23. 의약품 검색화면

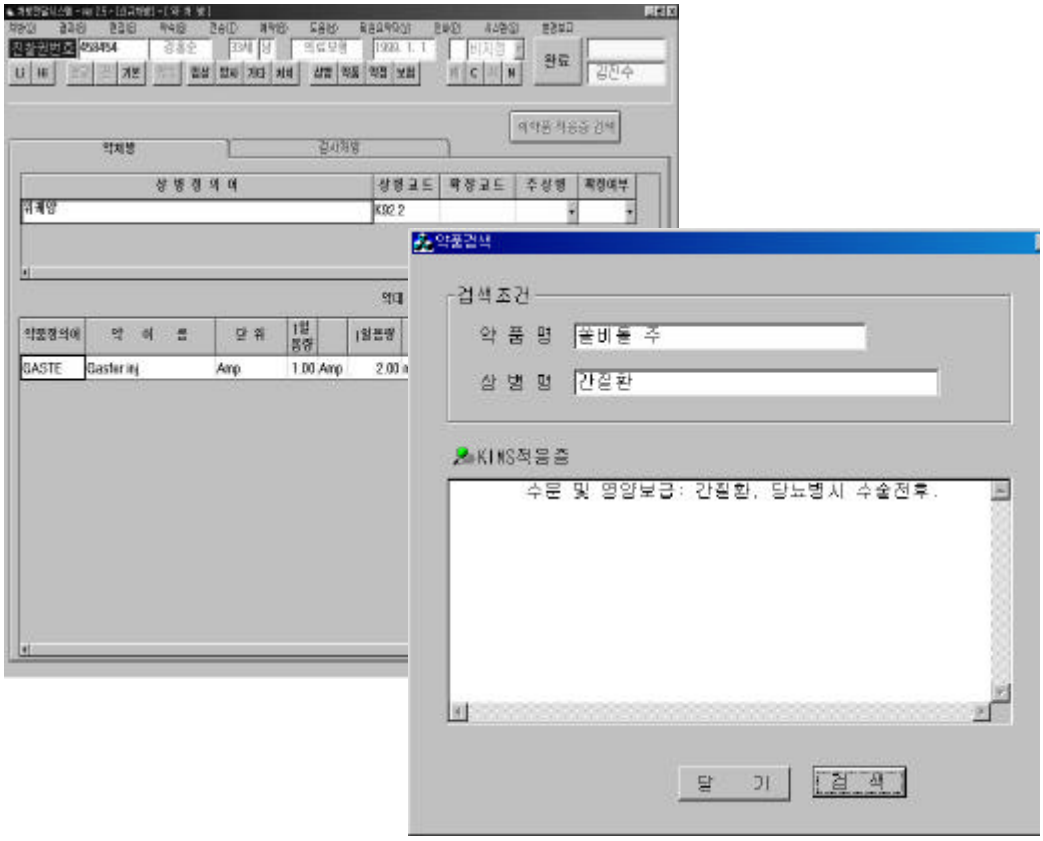


그림 24. KIMS 적응증 검색화면

의사가 처방을 내리는 시점에서 약제의 처방을 내고 해당 약제에 대하여 의약품의 적응증 범위내의 처방인지를 점검할 수 있도록 한 화면이다. 의약품 점검 버튼을 클릭하면 해당 환자의 상병 명과 검색 엔진이 수행되어 나온 KIMS 의약품의 적응증과의 매칭을 통해 적합 여부의 결과를 알려주어 약품 처방에 대한 의사결정을 지원하도록 하였다.

V. 고 찰

급변하는 병원의 경영 환경에 능동적으로 대처하여 질적으로 차별화된 의료 서비스를 제공하기 위해서는 축적된 정보를 효과적으로 활용할 수 있는 시스템의 구축이 필요하다. 단순히 데이터나 정보를 입력 · 조회 · 출력하는 역할이 아니라 지식을 바탕으로 지속적인 업무 지식을 창출하고, 효과적으로 지식 기반 경영시스템을 운용할 수 있는 전사적인 지식 경영의 비전 및 이의 전략 수립이 요구되고 있다. 또한 체계적인 진료 및 경영 전략을 세우기 위해서는 병원의 가장 기본 정보인 진료 정보를 체계적으로 수집하고 효율적으로 활용하여 이를 기반으로 한 효율적인 경영 전략의 수립이 중요하나 이에 대처할 만한 정책 및 개발이 매우 미흡한 실정이므로 진료 데이터를 효율적으로 활용할 수 있는 방안을 마련하는 연구가 필요하다. 본 연구에서는 병원의 진료 및 의무 기록 정보, 원무 행정 정보, KIMS 의약품 정보를 대상으로 하여 정형 데이터 및 비정형 데이터에 대한 가공 및 추출, 요약을 수행함으로써 환자에 대한 정보를 관리하고 이로써 병원 경영 및 진료에 활용할 수 있는 데이터 웨어하우스를 구축하였다.

여러 기업에서 데이터 웨어하우스를 개발하고 있으나, 의료 분야에서는 개발되어진 경우가 거의 없었다. 외국의 경우 의료용 데이터웨어하우스 구축에 있어서의 이슈에 관한 연구가 있었으며(Pedersen 등, 1998) 국내의 경우 의료용 데이터 웨어하우스 아키텍쳐 개발에 관한 연구가 있었다(김종호 등, 1999). 그러나 이러한 연구들은 데이터 웨어하우스를 구축하기 위하여 병원에서의 복잡한 데이터 수집하여 분석에 활용할 수 있도록 표준화하고 변환하는데 중점을 둔 본 연구와 구분된다. 의료 분야의 데이터 웨어하우스는 사용자에 따라, 진료와 진료 지원을

기반으로 하는 임상 데이터 웨어하우스와 원무 관리와 일반 관리를 기반으로 하는 비임상 데이터 웨어하우스로 구분할 수 있다. 비임상 데이터 웨어하우스는 경영층의 의사 결정을 주목적으로 하고, 임상 데이터 웨어하우스는 환자에 대한 정보의 저장 및 활용을 통한 의료 서비스와 의료연구를 주목적으로 한다(Pedersen 등, 1998). 따라서 본 연구에서는 임상과 비임상으로 구분하여 고혈압 관리 데이터 마트와 사전 심사 데이터 마트를 구성한 후 이를 토대로 정보검색기법 및 OLAP 을 이용한 사용자 접근 프로그램을 개발함으로써 병원에서의 고혈압 관리 및 의료 보험사전 심사 의사 결정 지원을 위한 유연한 질의를 실행하도록 하였다.

본 연구의 특징으로는 첫째, 병원의 각 서버에 흩어져 있던 복잡한 데이터들을 수집하여 분석에 활용할 수 있도록 표준화하고 변환 작업을 거침으로써 병원의 다양한 데이터의 파악이 가능하도록 하였다. 병원의 데이터는 기본적으로 환자의 기본인적정보의 관리를 위한 복잡한 프로세스에 의해 기본 데이터가 검사실마다 결과를 가지게 되고 이에 따라 여러 데이터베이스에 존재하게 되며, 또 다른 경우는 진료 처방 자료에 진료 의사 코드가 없는 경우 및 동일 환자, 동일 일자에 대하여 표준 등록시 진료과와 처방 발생시 진료과가 다른 경우가 발생하는 등 데이터의 유의하지 못한 경우가 많다. 본 연구에서는 이러한 병원 데이터의 특성을 파악하여 표준화하고 변환함으로써 데이터 웨어하우스를 구축하였다. 둘째, 병원의 기존시스템 데이터 뿐 아니라 외부데이터로서 KIMS 의약품 데이터 베이스를 데이터 웨어하우스와 통합하고 검색 기법을 활용하여 약물 정보를 검색하게 하였다. 지금까지의 병원 관리 시스템은 약물정보가 거의 배제된 상태이며, 약물처방시 과별 약물 정보 공유가 되지 않아 약물 부작용, 과다처방 등이 발생하고 획일적인 의사 결정에 의한 처방이 보편적으로 진행되고 있다. 따라서 데이터 웨어하우스의 구축 및 활용을 통해 의사가 진료시 약물정보를 최소한 짧은 시간에

최대한 이용하게 함으로써 의료 보험 사전 심사 및 고혈압 관리를 위한 의사 결정에 의약품 정보를 최대한 활용할 수 있도록 하였다.

본 연구에서는 병원에서의 의사결정지원을 위한 데이터 웨어하우스의 프로토타입을 구현하였으나 그에 대한 객관적인 성능 평가를 못한 점이 연구의 제한점으로 지적될 수 있다. 또한 병원의 업무에 지장이 없도록 데이터를 추출하여야 하기 때문에 최근의 충분한 양의 데이터를 추출하지 못했으며, 데이터 필드의 부재와 중복등으로 인하여 사용자 관점의 다차원 분석에 필요한 데이터 획득의 어려움이 있었다. 진료비에 관한 보다 구체적인 데이터를 수집한다면 데이터 웨어하우스의 구조에 상세한 속성들을 추가로 포함시킴으로써 상대적으로 중요한 속성을 가려내고, 데이터 웨어하우스의 양적, 질적으로 우수한 데이터에 비례하는 분석이 가능할 것이다. 또한 정보 검색 기법을 이용하여 데이터 웨어하우스를 구축하는 데 있어서 빠른 성능 및 비용 절감의 장점에 비하여 실제 구현 과정에서 병원 데이터의 복잡성 및 다양한 사용자 요구사항에 의한 다차원 분석상의 어려움이 있었다.

향후 본 연구에서 구현한 병원 데이터 웨어하우스의 프로토타입을 기반으로 실제 병원에서 사용 중인 OCS 데이터베이스 및 보험청구 데이터베이스, 퇴원요약 데이터베이스와 연동하고 지식 경영 시스템과 용이하게 통합할 수 있는 데이터 웨어하우스를 구축하고 실제로 시스템을 운영함으로써 성능을 평가할 필요성이 있다. 그리고 병원 데이터의 복잡성과 다양한 사용자 요구 사항에 의한 다차원 분석을 위하여 보다 안전성 있는 데이터웨어하우스 구축 도구를 활용할 필요가 있으며, 이를 통해 메타 데이터의 관리를 철저히 할 수 있을 것이다.

VI. 결 론

21 세기 정보지식사회에 대비하여 병원에 산재되어 있는 다양한 형태의 정보를 통합하고 가치있는 정보로 재창조하여 활용이 가능하도록 함으로써 급변하는 병원의 경영 환경에 능동적으로 대처하기 위한 전략적 정보체계가 요구된다. 현재 병원의 각 운영시스템을 중심으로 병원의 데이터 웨어하우스구축을 통해 현 운영시스템에 적재되어 있는 데이터를 분석하고 축적된 정보를 효과적으로 활용함으로써 질적으로 차별화된 의료 서비스를 제공하여야 한다.

이 연구는 병원의 진료 부문, 의무 기록 정보를 포함하는 진료지원부문, 원무 행정부문 정보 및 KIMS 의약품 정보를 대상으로 하여 정형 데이터 및 비정형 데이터에 대한 가공 및 추출, 요약을 수행함으로써 환자에 대한 정보를 관리하고 이로써 병원 경영 및 진단에 활용할 수 있는 데이터 웨어하우스를 구축하고, 이 데이터 웨어하우스로부터 진료 및 진료 지원 부문의 데이터베이스를 기반으로 하는 임상 데이터 마트로 고혈압 관리 데이터 마트와 원무 행정부문 데이터베이스를 기반으로 하는 비임상데이터 마트로 의료 보험사전 심사 데이터 마트를 구성하여, 정보검색기법 및 OLAP 을 이용한 사용자 접근 프로그램을 개발함을 목적으로 하였다. 이로써 단순히 데이터나 정보를 입력 · 조회 · 출력하는 역할이 아니라 지식을 바탕으로 지속적인 업무 지식을 창출하고, 효과적으로 지식 기반 경영시스템을 운용할 수 있는 전사적인 지식 경영의 비전 및 이의 전략을 수립하는 등 진단 및 병원 경영을 위한 의사 결정을 지원하도록 하였다.

시스템 사용자인 심혈관 센타 전문의와 의료 보험 사전 심사 요원의 요구 사항 파악을 통하여 데이터 웨어하우스로 구축해야 하는 데이터베이스의 범위를 정하

고 이에 따라 데이터를 추출, 정제, 변환함으로써 데이터 웨어하우스를 구축하였으며, star 스키마와 snowflake 스키마로 모델링하여 임상 및 비임상에 관련하여 데이터 마트를 토대로 의료 보험 심사 시스템과 고혈압환자분석시스템으로 구성하였다. 의료 보험 사전 심사 시스템과 고혈압환자분석시스템은 의약품 검색 모듈과 분석모듈로 구성되어 있는데, 의약품 검색모듈은 AID 검색 엔진을 이용하여 Visual C++로 구현하였고, 분석모듈에서의 OLAP 을 위한 GUI 는 VBA 로 개발하였다.

향후 본 시스템을 실제 병원에서 사용중인 OCS 데이터베이스 및 보험청구 데이터베이스, 퇴원요약 데이터베이스와 연동하고 지식 경영 시스템과 통합함으로써 복잡한 병원의 데이터에 쉽게 접근하여 다양한 사용자의 요구사항에 따라 다차원 분석을 용이하게 할 수 있을 것이다. 이와 같은 축적된 정보의 활용을 통해 병원의 경영 환경에 능동적으로 대처할 수 있는 질적으로 차별화된 의료 서비스를 제공할 수 있을 것이며, 병원에서의 전사적인 지식 경영의 비전 및 이의 전략 수립을 체계화 할 수 있을 것이다.

참 고 문 헌

고재성. 데이터웨어하우징과 데이터 마이닝기법을 이용한 의사 결정사례에 관한 연구. 성균관대학교 경영대학원 석사학위 논문, 1998

김종호, 김태훈, 민성우, 이희석. 의료용 데이터 웨어하우스 아키텍쳐의 설계 및 구현. 한국데이터베이스학회. ’99 춘계학술대회, 1999:393-394

오해석. 데이터베이스총론. 정익사, 1996

이순철. 지식경영의 이해. 삼성경제연구소, 1999

조남철, 손명호, 김태훈, 이희석. 데이터 웨어하우스 메타데이터 구축사례. 한국데이터베이스학회. ‘99 춘계학술대회, 1999:383-392

조성원. 컨설팅 방법론에 근거한 데이터 마이닝 접근 방법 및 사례 연구. 한국전문가시스템학회.’98 추계학술대회, 1998:130-138

조재희, 박성진. 데이터웨어하우징과 OLAP. 대청미디어, 1996

Berson A, Smith S, Thearling K. Building DataMining Applications for CRM. The McGraw-Hill Companies, 1999

Cass S. Data Extraction and Transformation for the Data Warehouse. Communication of The ACM, 1995:446-447

Claudio C. Health Care Management. Communication of The ACM, 1998;41(9):58-59

Date CJ. Data Introduction to Database Systems, Addison-Wesley Publishing Company, Inc., 1993

Daelim. Information & Telecommunication Co. Knowledge Management Committe. This

is Core of Knowledge Management. Korea: Changhae, 1998

Devlin B. Data Warehouse from Architecture to Implementation. Addison-wesley, 1997

Ewen EF, Medsker CE, Dusterhoft LE. Data Warehousing in an Integrated Health System; Building the Business Case. DOLAP ’98 Washington DC USA ACM, 1999 1-58113-120-8

Gardner SR. Building the data warehouse. Communications of the ACM, 1998; 41(9): 52-60

Inmon, W.H. Building the Data Warehouse(2nd Ed.). John Wiley & Sons, Inc., 1996

Kimball R. The Data Warehouse Toolkit. John Wiley & Sons, Inc., 1996

Barquin RC, Edelstein HA. Building, Using, and Managing the Data Warehouse. Pretice Hall PTR, 1997

Pedersen TB, Jensen CS. Research issues in clinical data warehousing. Proc. of the 10th IEEE Int’l Conference on Scientific and Statistical Database Management, 1998; 43-52

Pedersen TB, Jensen CS. Multidimensional Data Modeling for Complex Data. Proc. of the 10th IEEE Int’l Conference on Scientific and Statistical Database Management, 1999; 336-345

Abstract

Development of a Data Warehouse for Decision Support in Hospital: Hypertension Management and Medical Insurance Claims Review

Kyoung Won Cho

Graduate School of

Health Science and Management

Yonsei University

(Directed by Professor Young Moon Chae, Ph. D.)

There is a need for an information system that provides information to support decision making for clinicians as well as managers to help them cope with radically changing hospital environment. There is also a need for data warehouse that integrates relevant information from various databases to effectively support clinical as well as managerial decision-making.

In this study, data warehouse was constructed to support two domain specific data marts: hypertension management for clinical decision-making and insurance claims review for managerial decision-making. Microsoft SQL Server Enterprise Manager, which is a data warehouse development tool, was used to construct data warehouse by sampling, cleansing, and transforming of managerial and clinical data from various databases in hospital. Each data mart was constructed using star schema and snowflake schema. Clinical decision support system for hypertension management and decision

support system for claims review system were develop by AID search engine for information retrieval, Visual C++ for OLAP, and Visual Basic for GUI (Graphic user interface).

There were several significances in this study. First, variety of hospital data were standardized and integrated in a data warehousing architecture so that they can be easily extracted for various analysis. Second, KIMS drug database, which is an external database, was also integrated into the data warehouse so that doctors and managers can easily access drug information.

In the future, the data warehouse and two prototype decision support systems, developed in this study, should be integrated to the main hospital information system, especially the order communication system (OCS), for further testing and improvement at the actual sites. These systems can also be used in developing more comprehensive knowledge management system for hospitals.