

교육통계정보를 위한 웹기반 OLAP시스템의 구축

이기준⁰¹, 박재민¹, 강성국¹, 황수찬²

¹ 한국교육개발원 교육통계센터

{standard, mksmile, promaz}@kedi.re.kr

² 한국항공대학교 컴퓨터공학과 schwang@kau.ac.kr

Construction of Web based OLAP system for Education Statistics Information

Ki-Jun Lee⁰¹, Jae-Min Park¹, Seong-Guk Kang¹, Soo chan Hwang²

¹ Center for Educational Statistics, Korea Educational Development Institute

² Dept of Computer Engineering, Korea Aerospace University

요 약

최근 과학적 정책결정의 합리성을 뒷받침하기 위해서 통계에 대한 수요가 급증하고 있는 실정이다. 따라서 통계 수요자들은 사용자가 필요한 통계표를 직접 실시간적으로 산출하고자 하는 요구를 지속적으로 제기하고 있다. 본 연구에서는 의사결정 및 통계시스템에서 주로 사용하고 있는 데이터웨어하우스, OLAP 시스템 및 스타스키마에 관하여 알아보고 수요자들의 다양한 요구에 부응하기 위한 교육인적자원 통계정보 시스템의 구축에 대해 기술한다. 또한 교육인적자원 통계정보시스템의 핵심기능인 Web OLAP 시스템을 통해서 수요자가 원하는 교육통계정보를 실시간적으로 산출하기 위한 시스템의 구축방안에 대하여 기술하고자 한다.

1. 서 론

다가오는 21세기는 지식정보화 사회로서 그 나라가 보유한 지식 정보의 양과 질이 국가 경쟁력을 좌우하는 매우 중요한 요소가 될 것이며, 교육은 미래의 정보화 시대를 주도해 나갈 수 있는 국가 경쟁력의 핵심적인 관건으로서 그 비중이 강조되고 있는 것이 현실이다[1,2]. 국가의 교육력과 교육 여건을 향상시키기 위해서는 교육통계 정보 체계가 과학적이고 선진적으로 정비되어 정책결정 과정에 합리성을 뒷받침해야 한다[3].

교육통계정보는 1965년부터 조사·수합된 사회분야의 핵심 통계정보로서, 우리나라 교육현황 전반에 대한 과학적·종합적 진단은 물론, 관련 정책 수립을 비롯한 제반 교육 기획·시행·평가·연구 등을 위해 정보를 제공하고 있으며, 교육통계사업은 1997년에 기존의 수작업에서 탈피, 데이터베이스를 기반으로 하는 자동화 시스템을 구축하여 시스템을 통한 통계자료의 조사 및 검증, 서비스 체계를 구축하였다[4].

이와 같이 구축된 교육통계 데이터베이스를 바탕으로 각종 통계자료집, 인터넷을 통한 정형 장표 서비스 및 데이터베이스 운영센터를 통한 통계 분석 서비스를 수행하고 있다. 그러나 대부분의 통계자료를 서비스하고 있는 기관 또는 사이트에서는 통계자료집의 원본을 스캐닝하거나 워드프로세서로 제작된 원본 전자파일을 직접 서비스하는 경우가 대부분이다. 또한 한국교육개발원, 통계청 등 일부 사이트에서 통계정보를 DB화하여 서비스하고는 있으나 통계표에 포함되는 구분항목에 대한 선택정

도의 기초적인 시스템만을 제공하고 있는 상황이다.

그러나 최근 과학적 정책 결정 및 주요 의사결정의 기본이 되는 통계에 대한 수요 증가에 의하여 수요자들이 요구하는 통계의 종류 및 조건이 다양하고 통계 산출에 소요되는 시간적 비용에 대하여 불만이 제기되고 있는 실정이다.

이와 같은 요구에 부응하기 위하여 인터넷을 기반으로 하여 사용자가 원하는 통계자료를 산출하기 위한 시스템을 구축하여야 하는 필요성이 대두되었다.

따라서 본 논문에서는 교육통계 데이터베이스를 사용하여 급증하는 교육통계의 수요에 대처하기 위한 의사결정 시스템에서 대표적으로 사용되는 OLAP시스템의 웹기반 설계 및 구축에 대하여 기술하고자 한다.

2. 관련 연구

2.1 데이터 웨어하우스

데이터웨어하우스는 1980년대 중반 IBM이 자신이 하드웨어를 판매하기 위해 처음으로 도입했던 개념으로, IBM은 인포메이션 웨어하우스(Information Warehouse)라는 용어를 사용하였다. 이후 이 개념은 많은 하드웨어, 소프트웨어 및 툴 공급 업체들에 의해 이론적, 현실적으로 성장하였으며, 1980년대 후반 Inmon이 데이터 접근 전략으로 데이터웨어하우스 개념을 사용함으로써 많은 관심과 집중을 받게 되었다.

가장 대표적인 정의는 Inmon의 것으로 그는 데이터웨어하우스를 '기억의 의사 결정 과정을 지원하기 위한 주

제 중심 적이고 통합적이며 시간성을 가지는 비휘발성 자료의 집합'으로 정의하고 있다[5].

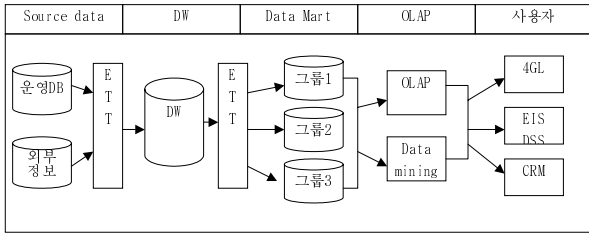


그림 1. 데이터웨어하우스 개념도

2.2 OLAP

데이터 웨어하우스는 의사 결정에 필요한 정보를 사용자에게 효율적으로 제공하기 위해 구축된 거대한 데이터 저장소이다. 데이터 웨어하우스에 근간을 두고 있는 OLAP은 기업의 의사 결정 지원을 위하여 다차원적인 데이터 분석 및 복잡 질의 처리를 제공하는데 효율적이다 [6].

OLAP 시스템은 사용자 의사결정을 위하여 주요 비즈니스 항목들을 다양한 각도에서 자유롭게 비교할 수 있는데 이러한 조회작업을 다차원질의라 한다. 다차원질의는 그림 2에서 보이듯이 사용자가 큐브의 일부분을 자신이 원하는 형태로 절단하여 살펴보는 것에 비유할 수 있다. 이러한 이유로 다차원질의를 슬라이싱과 다이싱(Slicing & dicing)이라고 한다.

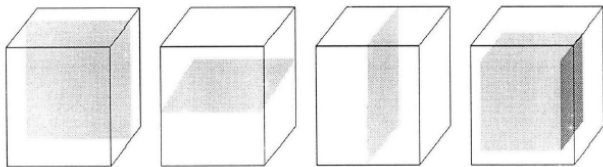


그림 2. 슬라이싱과 다이싱

최근 OLAP의 형태는 클라이언트-서버 형태에서 웹기반으로 변화하고 있다. 웹 OLAP은 클라이언트 서버형 OLAP에 비해 소프트웨어 설치비의 저렴함으로 인해 구축비를 절감할 수 있고, 사용자는 인트라넷이나 인터넷상에서 웹 브라우저를 사용하여 질의를 수행 할 수 있다. 또한 메타데이터를 OLAP서버, DW서버, 웹서버등에서 생성할 수 있어 시스템 다운으로 인한 메타데이터 재구성 시간을 줄이는 장점을 보유하고 있다. 초기의 웹OLAP은 OLAP 도구에 의해 사전에 작성된 보고서를 HTML의 형태로 저장하는 방식을 취했다. 이 경우 사전에 정의된 내용의 보고서만 볼 수 있다는 한계를 지닌다.

다음으로 나타난 인터페이스는 사용자 요청에 따라 HTML 문서를 자동적으로 생성하는 방식이다. 웹브라우저, 웹서버, 웹OLAP 게이트웨이, OLAP 엔진의 4-Tier의 형태를 취한다. 이 경우 사용자는 개발자가 구축한 파라메타 값을 변경하여 다양한 형태의 비정형 보고서를 실시간으로 생성할 수 있다. 사용자는 다양한 인터페이스

를 사용할 수 있으며, 클라이언트 이동된 데이터에 대해 추가적인 프로세싱이 수행될 수 있어 HTML의 많은 제약 사항을 극복할 수 있다. 최근 클라이언트와 서버의 모든 프로그램이 자바나 ActiveX로 작성되고 있는 추세이다 [7].

2.3 스타스키마

OLTP 환경에서는 데이터베이스를 설계하는데 E-R 다이어그램(Entity-Relationship Diagram)이나 정규화(Normalize) 기법이 널리 사용되었다. 그러나 ER 다이어그램에 의해서 설계된 데이터베이스는 의사결정시스템 및 통계시스템에서는 적합하지 않다[8].

대부분의 데이터 웨어하우스는 다차원 모델을 표현하기 위하여 스타스키마 모델을 사용한다[9].

데이터 웨어하우스를 위하여 현재까지 제시된 모델 표현방식으로는 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 join을 통하여 빠르게 접근하게 하기 위한 수단으로서 활용된다[10].

하나의 사실 테이블에는 이런 차원이 하나 이상 존재할 수 있으며 이러한 이유로 데이터 웨어하우스에서의 데이터 모델링은 다차원 모델링으로 불리기도 한다[11]. star 스키마는 사실테이블을 중심으로 차원테이블에 대해서는 전혀 정규화가 일어나지 않은 형태이다. 차원 테이블의 속성들(attribute)에 대해서 다시 정규화 하게 되면 이를 snowflake 스키마라고 부른다[12].

3. 교육통계 웹 OLAP 시스템의 구현

본 장에서는 교육통계 분석서비스를 위한 사이트인 교육인적자원 통계정보시스템의 구성에 대해 설명하고 이중 맞춤형 분석 시스템인 다차원 웹 OLAP시스템의 구조와 웹 OLAP 시스템을 위한 다차원 DB의 설계에 대해서 기술한다.

3.1 교육인적자원 통계정보 시스템 구성

교육통계정보의 서비스는 교육인적자원 통계정보 시스템(<http://cesi.kedi.re.kr>)을 통하여 2장에서 설명한 바와 같이 정형장표와 비정형장표의 방법을 모두 제공하고 있다. 정형장표 서비스는 경험적으로 많이 사용이 되는 주요한 통계표에 대해서 개발자가 미리 제작해 놓은 통계표에 대해서 서비스를 받게 된다. 그림 3은 교육인

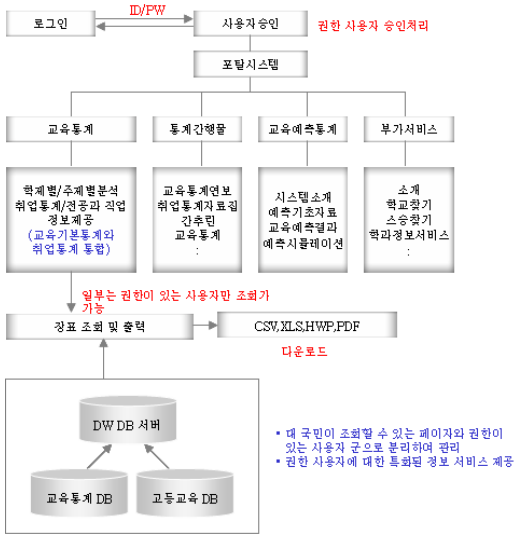


그림 3. 교육인적자원 통계정보시스템 구성도

적자원 통계정보 시스템의 구성도를 도표화한 것이다.

3.2 Web OLAP 시스템 구성

2장에서 설명했듯이 웹 OLAP의 방법을 하는 방법은 온라인을 통한 방법과 오프라인을 통한 방법의 2가지로 나뉜다. 이중 오프라인을 통한 본 논문에서 사용하고 있는 웹 OLAP 시스템의 구성도는 그림 4와 같다.

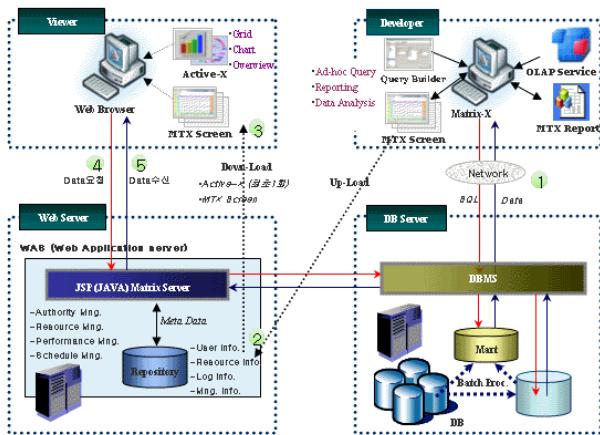


그림 4. 웹 OLAP 시스템 구성도

데이터 웨어하우스 설계자는 통합된 데이터 소스와 사용자의 요구사항을 분석하여 데이터 웨어하우스를 위한 다차원 분석 DB 및 정형 통계 장표를 생성하게 된다. 생성된 DB를 기반으로 제작된 정형통계표는 데이터를 포함 시킨 파일로 전환되어 웹서버에 저장되고 저장된 파일을 웹브라우저에 설치된 ActiveX 파일에서 읽어 들여 서비스하게 된다.

반면 비정형 통계표는 사용자가 웹서버를 통해서 비정형통계분석에 필요한 데이터를 요청하고 해당데이터가 사용자의 Local PC에 완전히 전송이 되면 Web OLAP툴을 사용하여 해당 데이터의 분석을 수행하게 된다.

이 작업을 위해 사용하고 있는 웹 OLAP 시스템은

MS-Excel의 피벗테이블을 기반으로 하고 있다. 일반적 Web OLAP툴은 사용자 인터페이스가 각 시스템별로 독자적으로 구축이 되어 툴을 사용하기 위해서는 툴에 따른 교육을 습득해야하나 본 시스템에서 적용하고 있는 Excel 기반 Web OLAP은 일반적으로 사용하고 있는 Excel을 사용할 수 있는 사용자라면 특별한 교육을 받지 않아도 사용할 수 있다는 장점을 지닌다.

또한 일반적인 Web OLAP 툴의 경우 자료의 분석을 위해서 분석을 위한 쿼리 및 레포트를 작성하고 이를 실행할 때 DB에서 해당 쿼리에 맞는 결과만을 전송받는 방법을 사용한다. 이 경우 사용자의 요구에 따라 표의 피벗은 가능하나 추가적인 데이터의 추가나 변경이 어렵다. 그러나 본 시스템에서 사용하는 MS-Excel 엔진 기반의 Web OLAP은 분석의 대상이 되는 데이터를 사용자 Local PC로 전송이 되어 메인메모리상에 로드되고 로드된 데이터에 한해서는 사용자의 요구에 맞게 통계표를 작성할 수 있다.



그림 5. Excel을 활용한 Web OLAP 화면

3.3 Web OLAP을 위한 다차원 DB의 구축

교육인적자원 통계정보시스템에서 구현한 Web OLAP에서 구축한 다차원 DB는 유초중등, 고등, 취업통계와 관련하여 총 16개를 구축하였으며 세부 내역은 표 1과 같다.

표 1. 교육인적자원 통계시스템 다차원 DB 현황

분야	세부분야	차원 수	사실 값수	분석가능 통계수	분석DB구축건수
유초중등	학교현황등 9개	98	28	8,589,432,403	524,785
고등	학교현황등 6개	72	13	224,635	351,352
취업	취업현황	27	27	2,097,151	122,187
합계	16개	197	68	8,591,754,189	998,324

표 1에서 보이는 분석가능 통계수는 각 세부분야별 생성가능한 통계표의 수를 나타낸 것이다. 생성 가능 통계수는 각 세부분야별로 조합이 가능한 차원수와 조합이 가능한 사실값 수의 곱으로써 산출된다.

일반적 Web OLAP툴에서는 스타스키마의 형태로 구축되

어 있는 다차원 DB의 사실테이블과 차원테이블의 조인을 통하여 분석이 끝난 데이터를 전송하여 결과값만을 전송한다. 반면 교육인적자원 통계서비스에서 사용하고 있는 Web OLAP시스템은 데이터의 전송이 완료된 후 사용자 PC의 Excel 엔진을 사용하여 분석을 시도한다. 따라서 일반적 Web OLAP툴에서는 차원테이블과 사실테이블과의 조인 성능이 시스템 반응속도를 결정하는 중요한 요소이다. 그러나 교육인적자원 통계서비스에서는 DB엔진이 아니고 Excel엔진을 사용하기 때문에 사용자의 Local PC의 Web OLAP툴에서 차원테이블과 사실테이블을 조인할 수 있는 기능이 없다. 따라서 데이터베이스에서 전송시에 차원테이블이 조인된 데이터가 전송되어야 한다. 즉, 그림 6과 같이 사용자가 세부분야에 해당하는 다차원DB를 선택하면 다차원DB에 관련된 사실테이블과 차원테이블을 조건에 맞게 조인한 후, 조인된 결과로 발생하는 최종 테이블을 사용자 Local PC로 모두 전송하게 되고 사용자 PC에서는 전송받은 데이터를 Web OLAP ActiveX 프로그램을 사용하여 피보팅, 필터링, 그룹핑등의 기능을 사용하여 사용자가 원하는 형태의 통계표를 작성하게 된다.

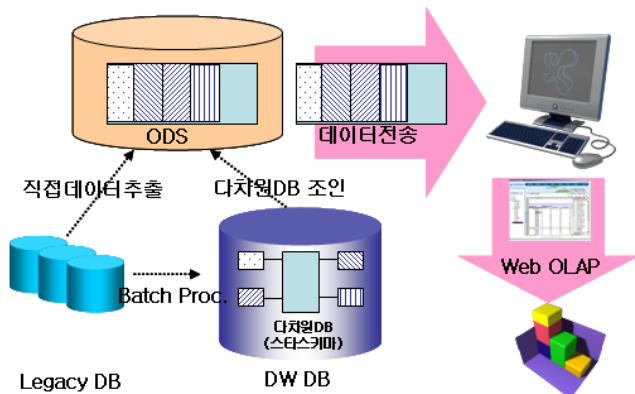


그림 6. 교육인적자원 시스템 Web OLAP 시스템 구성도

4. 결론 및 향후 과제

최근 통계에 관련된 관심 및 수요의 증대로 보다 다양한 교육통계 정보의 중요성이 대두되고 있는 상황이다. 이러한 상황하에 인터넷을 활용하여 수요자가 필요한 통계를 직접 손쉽게 실시간적으로 산출하고자 하는 요구가 꾸준히 제기되어왔다. 이러한 요구에 부응하기 위하여 교육인적자원 통계정보시스템을 구축하고, 특히 데이터웨어하우스기법을 도입한 Web OLAP시스템을 통하여 실시간적인 통계산출 시스템을 구축하였다. 이 시스템을 통하여 교육인적자원 통계를 일괄·수집 분석할 수 있어 통계분석 업무절차가 개선되어 통계관련 업무시간 및 효율성 증대에 많은 영향을 줄 것으로 보여진다.

본 시스템의 구축에는 먼저 수요자의 활용도를 높이기 위하여 일반 사용자에게 친숙한 MS-Excel 인터페이스를 사용하였으며, 시스템적인 응답속도를 높이기 위하여 데이터웨어하우스와 OLAP에서 사용하고 있는 스타스키마의 구조를 변형하여 시스템 응답속도를 증가시켰다.

교육인적자원 통계정보시스템에서 일반적인 시스템에서 제공하는 검색기능을 제공하고 있다. 그러나 통계청 및 통계 서비스 사이트에서는 본 시스템과 유사한 검색방법, 통계표의 제목 및 통계표의 구분항목에 대한 텍스트 서칭을 통한 검색에 그치고 있어, 통계수치에 대한 검색이 어려운 것이 현재 실정이다.

따라서 향후 연구과제로 고려해야할 사항으로는 교육 및 통계와 관련된 메타데이터에 대한 표준이 정의가 되어야 한다. 이 메타데이터의 표준이 제정되면 교육관련 통계의 검색이 수월해지고 정확성이 증대될 수 있을 것으로 보인다.

5. 참고문헌

- [1] 선우중호, 교육전산망과 대학도서관망의 통합 구성 및 운영에 관한 연구, 교육부, 1992
- [2] 엄장일, 교육 전산망에 대비한 대학 전산화 모형 개발에 관한 연구, 부산대학교 정보과학종합연구소, 1989
- [3] 정택희, 1998년도 교육기본통계조사 결과보고, 한국교육개발원, 1998
- [4] 이기준, 김정겸, "교육통계 데이터베이스를 이용한 시도 분산 데이터베이스의 구현", 한국정보과학회 춘계학술논문발표회, 1999
- [5] Inmon W. H: "Building the Data Warehouse" John Wiley & Sons, Inc. Canada 1996
- [6] 최은하, 김진호, 옥수호, "데이터웨어하우스의 개념적 설계를 위한 스타스키마에서 ER 도형으로의 변환기법", 한국정보과학회 춘계학술논문발표회, 2002
- [7] 박지나, "군인사 행정시스템을 위한 웹 OLAP 시스템의 구현", 이화여자대학교 정보과학대학원 석사학위논문, 2001
- [8] S.Charuhuri and Dayal, "An overview of Data warehouseing and OLAP Technology", SIGMOD record, March 1997.
- [9] Red Bric System, "Star Schema and STARjoin Technology", Red Brick Systems, White Paper, 1996
- [10] 고재성, "데이터웨어하우징과 데이터 마이닝 기법을 이용한 의사결정 사례에 관한 연구", 성균관대학교 경영대학원 석사학위논문, 1998
- [11] Kimball R, "The Data Warehouse Toolkit", John Wiley & Sons, Inc., 1996
- [12] Elzbieta Malinowski, Esteban Zimányi, "A conceptual solution for representing time in data warehouse dimensions", Proceedings of the 3rd Asia-Pacific conference on Conceptual modelling - Volume 53 APCCM '06, 2006