

OLAP 기술을 이용한 학업성취도 분석 시스템(SAAS)의 설계 및 구현

(Design and Implementation of a Students' Achievement Analysis System using OLAP Technology)

박 미 현[†] 김 명^{††}
(Mihyeon Park) (Myung Kim)

요약 다각도로 분석된 학업 성취도를 신속하게 제공받는 학생들은 그렇지 않은 학생들에 비해 차 후에 더 나은 학업 성과를 보인다. 그러나 현재의 한국 학교 현장에서 시행되고 있는 평가 방법은 이를 반영하지 못하고 있다. 본 연구에서는 이러한 서비스를 온라인으로 교사와 학생에게 제공하는 '학업성취도 분석시스템'인 SAAS를 설계하고 구현하였다. 이 시스템은 비즈니스 데이터를 다차원적으로 분석하여 부가가치를 창출하는데 쓰이는 신기술인 OLAP을 이용한다. 학생, 과목, 기간 차원에 대한 종합적 성취도 분석 결과가 산출되고, 이는 편리한 인터페이스를 통해 제공된다. 시스템의 성능향상을 위해서 분석결과의 부분적 사전연산 방식과 큐브의 청크 단위 저장 방식이 사용되었다. SAAS는 전국 단위 학생 성적 분석을 위해서도 쓰일 수 있는 정도로 성능면에서 확장성을 지니며, 제 7차 교육과정 개혁에도 적용이 가능하다.

Abstract Students who are frequently and promptly informed of their current education achievement levels perform better than those who are not. However, current Korean school environments cannot afford to provide students with such services. In this study, we designed and implemented a student's achievement analysis system called SAAS. The system uses OLAP technology which is essential to extract desired data and to derive value-added information from an enterprise data warehouse. SAAS produces multidimensional analysis results on students, courses, and periods. Teachers and students can easily obtain desired analysis results through user-friendly interfaces. The performance of the system is enhanced by pre-computing some of the analysis results and using a chunk-based MOLAP cube storage scheme. The system is scalable to handle nationwide exams, and is extendable to embrace the 7th national curriculum revision.

1. 서론

교육평가의 주요 기능은 학생 개개인이 성취해야 할 교육 목표의 성취정도를 점검하고, 그 결과를 평가자와 평가대상자에게 제공해 줌으로써, 이들의 교육적인 노력 및 의사결정을 돕는 것이다[1]. 이 기능이 제대로 발휘

되기 위해서는 학업 성취도가 다각도로 분석되어 원하는 분석결과가 원하는 시점에 신속하게 제공될 수 있어야 한다.

최근까지 주로 사용되어온 평가 방식은 '상대평가'로써, 평가 대상자의 점수와 전체 석차가 제공된다. 이는 교육의 효과를 바탕으로 한 의사결정에 적절하고 충분한 자료가 되지 못한다. 예를 들어, 학생의 수학성적이 60점이라면, 그 학생은 수학과목을 잘하지 못한다는 것만 알 수 있을 뿐, 수학의 어떤 부분에 취약한지는 알 수 없다. 이런 평가 방식은 교육평가의 본래 의도에 크게 미치지 못한다.

평가 분석 결과는 평가자나 평가대상자가 원하는 시점에 신속하게 제공되어야 제공 효과가 증대된다. 다양

· 본 연구는 2000년도 한국학술진흥재단의 유초중고 교육정책 및 교원 양성 연구(과제번호: 2000-00013)의 지원에 의해 수행되었음.

† 정 회 원 : 이화여자대학교 교육대학원 컴퓨터교육전공
parkmihyeon@hotmail.com

†† 중신회원 : 이화여자대학교 컴퓨터학과 교수
mkim@ewha.ac.kr

논문접수 : 2000년 9월 8일

심사완료 : 2001년 4월 9일

한 평가 결과를 피드백 받은 피험자들이 차후에 더 나은 성과를 얻는다는 연구결과들[2]이 있으나, 실제로 학교 현장에서 시행되고 있는 평가보고는 적절하다고 볼 수 없는 수준이다. 또한 평가 결과에 대해 다양한 정보를 제공해 주고 있는 시스템조차 전문한 실정이다. 대부분의 CAT(Computer-Assisted Test)[3]들조차도 테스트 후에 제시하는 결과는 정답을 맞춘 문항 수와 오답을 낸 문항에 대한 답을 보여줄 뿐이다. 이는 교육에 대한 투자가 부족한 여건에서 학생 성취도 분석에 대한 연구도 부족할 뿐 아니라, 현재 학교 실정에서는 평가 작업, 평가결과의 관리/분석/제공 작업 전체가 전문 관리/분석 능력을 갖추었다고 볼 수 없는 담당 교사가 맡고 있고, 그 과정이 수작업을 통해 이루어진다는데 더 큰 원인이 있다.

본 연구에서는 학생의 학업 성취도를 다차원적으로 분석하여 그 결과를 신속하게 제공하는 학업 성취도 분석 시스템인 SAAS (Student's Achievement Analysis System)를 설계하고 구현하였다. 이 시스템은 다음과 같은 세 가지 특징을 갖는다. 첫째, 학습평가 전체 과정을 전산화하고, 그 과정에서 산출되는 모든 정보를 데이터베이스화하여 여러 사용자가 정보를 공유, 검색, 관리할 수 있도록 한다. 학습평가 전체 과정이란 시험의 채점으로부터 평가 결과의 분석과 그 결과의 제시를 말한다. 둘째, 학업 성취도의 분석결과는 학년이 지나면서 계속 누적되도록 하고, 이로부터 다차원적 분석결과를 산출하도록 한다. 셋째, 평가결과의 분석은 편리한 인터페이스를 통해 사용자에게 온라인으로 신속하게 제공된다. 이 시스템은 최근들어 대용량 데이터를 다차원적으로 분석하기 위해 쓰이는 OLAP (On-line Analytical Processing)[4, 5, 6] 기술을 적용한다. 따라서, 분석용 데이터를 흔히 쓰이는 RDB (Relational Database)에 저장해두고 SQL로 사용자의 질의에 응답하는 것보다 훨씬 신속하게 작업을 처리할 수 있고 여러 사용자가 동시에 사용할 수 있도록 고성능이 보장된다. 이 시스템은 한 학교 전체 학생들의 학업 성취도 분석을 목표로 설계되었으나, 전국 규모의 모의고사 결과의 평가/분석에도 충분히 사용될 수 있는 확장성을 갖는다. 또한 본

연구는 교육 평가시스템에 최초로 OLAP 기술을 도입한 것이며, 한국교육과정평가원에서 현재 분석, 연구중인 성취도 분석 시스템을 기반으로 하였고 때문에 본 연구를 통해 설계 구현된 학업성취도 분석시스템의 확장성은 더욱 커질 수 있다.

본 논문은 다음과 같이 구성된다. 2절에서는 SAAS의 설계 근거 자료인, 한국 교육과정 평가원의 '국가 교육 과정에 근거한 평가기준 및 도구개발 연구'에 대하여 설명하고, 각 교육평가 자료가 SAAS에 어떻게 적용되는지 살펴본다. 3절에서는 이러한 평가 요소들을 기준으로 하여 학업성취도 분석시스템을 설계한다. 그리고, 시스템 효율을 증진시키기 위해 사용한 여러 가지 방법을 살펴보고, 성능을 비교한다. 4절에서는 SAAS의 기능을 설명하고, 5절에서 결론과 향후 연구과제에 관해 설명한다.

2. 학업성취도 분석 시스템 SAAS의 이론적 배경

SAAS는 한국 교육 과정 평가원에서 1998년 3월에 고시한 '국가 교육과정에 근거한 평가기준 및 도구 개발 연구' [1] 보고서를 토대로 설계되었다. 이 연구는 제6차 국가 교육과정에 근거하여, 고등학교의 특정 학년 학생이 특정 과목으로부터 최소한 어느 정도의 수준을 성취해야 할 것인지에 대한 기준을 개발하고, 평가 도구위 개발, 평가 도구의 활용과 적용 방안 등을 주된 내용으로 하고 있으며, 2002년부터 초/중학교에 확대 적용되는 제 7차 교육과정과의 연계도 고려되고 있다. 여기서, 이 보고서의 내용이 SAAS에 어떻게 적용되었는지, 그리고 향후 어떻게 확장될 수 있는지를 검토해 보고자 한다.

이 보고서는 고등학교 공통필수 10개 과목(윤리, 국어, 공통수학, 공통사회, 국사, 공통과학, 체육 I, 음악 I, 미술 I, 공통영어)에 대해, 고등학교 1학년생들이 성취해야 할 범위와 정도를 설정하고, 평가 기준으로는 (1) 성취기준, (2) 평가기준, (3) 등급화 방안을 제시한다.

첫째, 성취기준은 각 교과와 단원을 대영역, 중영역, 소영역으로 나누고, 소영역별로 학생들이 성취해야 하는 능력 또는 특성에 대한 기준이다. 이는 SAAS에서 교과 정보 데이터의 과목과 단원을 나누는 분류 기준으로 사

대영역	I			II			III			IV			V			
중영역	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
성취결과	상	중	하	중	중	상	상	중	상	중	상	중	상	상	중	중

그림 1 A 학생의 특정과목 학년말 성적 (중영역 개수 = 16)

용된다.

둘째, 평가기준은 성취기준에 준하여 평가되는 활동의 수준이다. 학생들의 학업 성취도는 상/중/하 세 수준으로 나눈다. '중수준'은 '고등학교 1학년 학생이 충실한 교수, 학습과정을 거치면 누구나 성취하리라 기대되는 수준'이고, '상수준'은 그 이상을, '하수준'은 그 이하를 성취한 것을 뜻한다.

셋째, 등급화 방안은 과목 전체에 대한 성취도를 파악하기 위한 방안으로써, 학년말에 각 과목별로 수/우/미/양/가의 평어로 기록하게 되어 있다. 학기말에 각 평가영역에서 받은 '상/중/하' 점수가 '수/우/미/양/가' 평어로 환산된다. '수'는 성취기준 '상'에 해당하고, '우'는 '상'과 '중'사이에 속한다. '미'는 '중'수준에 해당하고, '양'은 '중'과 '하' 사이에 속하고, '가'는 '하'수준에 속한다. 학업성취 결과는 증명영역별로 그림 1과 같이 표기된다.

3. 학업성취도 분석 시스템 설계

SAAS는 학생과 교과목 시험에 관한 다량의 정보로부터 데이터 웨어하우스 (data warehouse, 분석용 데이터베이스) [4, 5] 를 구축하여 검색과 관리가 용이하도록 하고, 평가 정보를 다차원적으로 분석하여 신속하게 사용자에게 제공하는 시스템이다. 이 시스템은 최근 정보기술(IT) 시장에서 비즈니스 데이터의 분석에 활발하게 사용되기 시작한 OLAP 기술[6, 7]을 사용하여, 처

리속도나 저장공간 효율 면에서 고성능을 보장하기 때문에, 전국 규모의 모의고사와 같은 대단위 시험의 분석 시스템으로도 확장이 가능하고, 사용하기 쉬운 사용자 인터페이스를 제공한다.

이제, SAAS에 관해 구체적으로 설명하기로 한다. 첫째, SAAS가 기본적으로 제공하는 질의문의 종류를 살펴보고, 이들을 분석하여 공통인자들을 추출하고 각 인자들이 갖는 특성에 맞춰 효율적인 분석이 가능하도록 데이터 웨어하우스의 스키마를 정의한다. 둘째, SAAS의 효율적인 저장구조를 정의한다. 셋째, 시스템의 성능을 높이기 위해 적용한 여러 가지 방법들을 설명한다.

3.1 질의문과 데이터 웨어하우스의 스키마

SAAS의 성능이 어떤 연산들을 대상으로 최적화 되어야 하는가를 알아보기 위해, 평가자료 분석에 주로 쓰이는 질의의 종류를 분류/분석하는 작업이 선행되어야 한다. 학업 성취도 분석에 주로 쓰이는 질의문들은 다음과 같다:

- ① E 시험에서 A 학생의 특정 과목 평가영역별 성취도는?
- ② E 시험에서 A 학생 성적을 반 평균과 비교하면?
- ③ E 시험에서 A 학생성적을 학년 평균과 비교하면?
- ④ E 시험에서 X 반의 평균 성적과 학년 전체 평균과 비교하면?
- ⑤ E 시험에서 A 학생성적과 B 학생성적을 영역별로

기간 (3×2×4=24)

기간 키	년도	학기	시험일자
T01	1998	1	1998-3-23
T02	1998	1	1998-4-30
T03	1998	2	1998-9-30

학생 (10×40=400)

학생 키	학년	반	번호	이름
SA001	3	1	1	김병철
SA002	3	1	2	김병환
SA030	3	2	1	김순국

사실테이블 (평가분석)

기간 키	시험영역 키	학생 키	점수
T01	TA001	SA001	1
T02	TA002	SA002	0
T03	TA003	SA003	2
T04	TA004	SA004	2
T05	TA005	SA005	2

사실테이블의 튜플 수
= 24×400×400 = 3,840,000

과목별평가영역 (10×40=400)

시험영역 키	과목	대단원	중단원	평가영역
TA001	영어	1	1	1
TA002	영어	1	2	2
TA003	영어	2	1	1

그림 2 SAAS 데이터 웨어하우스의 스타스키마

비교해보면?

- ⑥ A 학생의 국어과목 성적을 지난 번 시험과 비교해보면?
- ⑦ X 반의 평균 성적을 지난 번 시험과 비교해 보면?
- ⑧ 학년 전체의 성적을 지난 번 시험과 비교해 보면?
- ⑨ 특정 과목에서의 지난 해 학생 성적과 올해 학생성적을 비교해 보면?
- ⑩ A 학생의 올 1학기 해당과목에 대한 성적 추이는 ?
- ⑪ 학년 전체의 올 한해 해당과목에 대한 성적 추이는?
- ⑫ 특정 과목에 대해 P 교사가 지도한 반과 Q 교사가 지도한 반의 성적을 비교해보면?
- ⑬ 우수반과 열등반의 성적 추이는?

(단, '성적비교'는 '영역별 성취도 비교'를 의미한다.)

이 리스트를 자세히 살펴보면 ①학생, ②과목, ③기간이라는 요소들로 나누어진다 것을 알 수 있다. '학생' 요소는 '개별 학생', '학생이 속한 반'과 '학년' 등을 포함한다. '과목' 요소는 평가되는 과목인데, 이는 평가 영역인 '대단원'에서 '중단원'으로 또는 '최하위' 평가영역으로 세분화 되어질 수 있다. '기간' 요소는 평가 시행일자를 의미하며, '평가 일', '학기', '년' 단위로 그룹화 되어진다. SAAS의 데이터 웨어하우스는 이 세 가지 요소들을 기준으로 하여 구축된다.

SAAS가 분석할 평가 데이터는 학생, 과목, 기간에 의해 정의되는 3차원 데이터이다. 즉, 위에 열거된 학업성취도 분석 질의문은 물론이고, 학생, 과목, 기간에 대한 모든 조합이 가능하도록 설계되어야 한다. 그림 2에 SAAS의 데이터웨어하우스 스키마가 있다. 분석 데이터

가 있는 사실 테이블(fact table)의 각 튜플(tuple)은 4개의 애트리뷰트를 갖는데, 그 중에서 4번째 애트리뷰트인 점수(score)는 분석할 데이터이고, 남은 3개는 그 점수의 특성을 나타내는 3개 차원 정보이다. 사실테이블의 튜플에는 각 차원의 멤버를 나타내는 키(key)가 저장되어 있고, 그 키를 갖는 각 차원 멤버들은 해당 차원 테이블 (dimension table)에 하나의 튜플로 저장되어 있다.

3.2 SAAS 데이터의 저장 구조

SAAS는 데이터를 저장하기 위해서 OLAP 기술이 사용하는 데이터 저장방식을 쓴다. 흔히 사용되는 RDB를 사용하지 않는 이유는 RDBMS가 다량의 데이터를 저장, 검색, 갱신하는 데에는 효율적이나, 데이터를 다차원적으로 분석하여 그 결과를 온라인으로 여러 사용자에게 제공하기에는 적합하지 않기 때문이다[8]. SAAS는 OLAP의 3가지 데이터 저장 방식 중에서 MOLAP (Multidimensional OLAP) [7, 9] 방식을 쓴다. 이는 데이터를 다차원 배열에 저장하는데, 배열이 고밀도인 경우 높은 성능을 보장하여, SAAS에 매우 적합하다.

OLAP 기술은 방대한 양의 데이터로부터 계산되어야 하는 분석결과를 온라인으로 사용자에게 제공하기 위해서 분석 대상 데이터 뿐 아니라, 분석결과의 전부 또는 일부를 미리 계산하여 저장하여 둔다. 이를 사전 연산 (precalculation)한다고 하고, 이를 통해 생성된 데이터는 OLAP 큐브(cube)에 저장한다고 한다. MOLAP은 큐브를 다차원 배열 형태로 저장하는데, 이 때 배열의 차원 축을 잘못 설정하면 배열이 희박(sparse)하게 되고 엄청난 공간의 낭비를 초래하면서 연산의 속도가 느려진다.

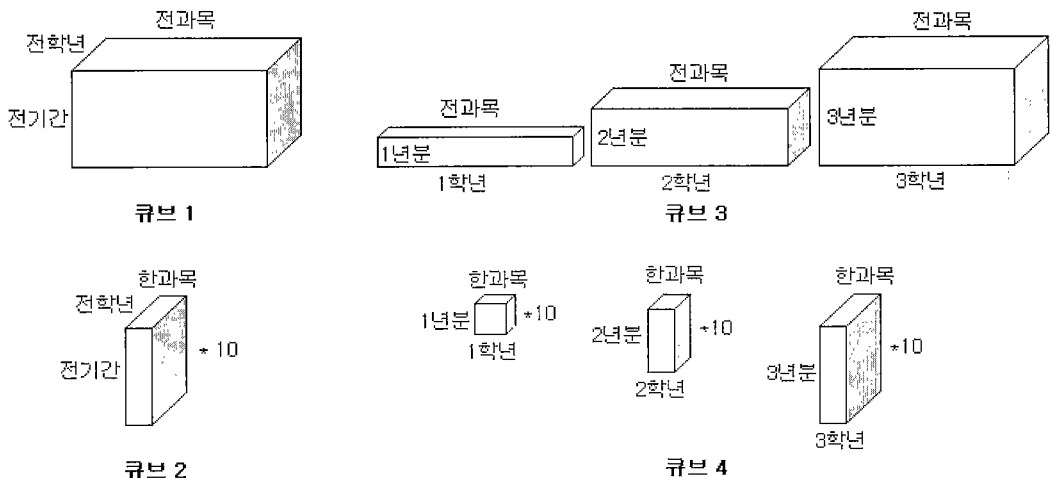


그림 3 차원항목을 달리하는 4가지 다차원 모형

이제 SAAS의 큐브 차원 축을 결정하기 위해 평가데이터의 특성을 분석해 보기로 한다.

먼저, 분석대상 데이터가 저장될 베이스 큐브(base cube)의 크기를 살펴보자. 큐브를 구성하는 셀(cell)의 개수는 각 차원의 멤버의 수를 곱한 값이다. 전형적인 고등학교를 기준으로 하면, 큐브의 각 차원 크기는 다음과 같다.

- 학생차원: 총 학생 수 1,200명 (반의 학생 수: 40명, 학년 당 반의 수: 10, 학년 수: 3)
- 기간차원: 총 평가횟수 24회(학기 당 평가 회수: 4, 연간 학기 수: 2, 학년 수:3)
- 과목차원: 총 평가영역 400개(과목당 대,중단원 수:각 5, 10, 과목 수: 10, 평가영역 수: 40)

위의 가정에 의하면 베이스 큐브의 셀 개수는 11,520,000 이고, 각 셀이 상/중/하를 의미하는 2/1/0의 값이므로 셀에 1 바이트를 할당하면, 큐브의 크기는 약 11.5 메가바이트가 되어 필요 이상으로 많은 저장공간을 차지하게 된다. 그 이유는 학생 차원에서 학년 전체를 하나의 축으로 구성했기 때문에, 1학년 학생들의 경우, 기간 차원의 셀들에는 1년치 자료만 들어 있고, 남은 2년치 셀들은 무효 값을 갖기 때문이다. SAAS를 구성하는 차원의 항목을 달리하여 그림 3과 같이 큐브 모형을 4가지로 나누어 보고 각 큐브들의 크기와 밀도를 비교해 볼으로써, 최적의 큐브 모형을 선정하였다.

각 큐브의 크기를 비교해 보자. 각 베이스 큐브를 도식화 한 것이 그림 3이다. 큐브 2는 과목별로 독립적인 큐브가 생성되므로, 10과목을 위해 10개의 서브큐브(subcube)로 구성된다. 큐브 3은 학년마다 1개씩 3개의 서브큐브로 구성되고, 큐브 2와 큐브 3의 혼합형인 큐브 4는 학년별로 각각 10개의 서브큐브가 만들어진다. 이 4가지 경우의 큐브의 크기와 밀도가 표 1, 2에 나타나 있다. 큐브 1과 큐브 2는 전체 학년을 하나의 큐브로 구성했기 때문에 1학년의 경우 2, 3학년을 위한 셀이 비어있다. 반면, 큐브 3과 큐브 4의 경우에는 학년별로 서브큐브가 구성되므로, 100% 밀도의 큐브가 생성된다.

따라서, 큐브 1과 큐브 2보다는 큐브 3, 큐브 4가 저장공간 측면에서 좀 더 효율적이라 할 수 있다. 또한 SAAS의 용도를 고려해 보면, 과목을 달리하여 성취도를 비교하는 것은 무의미하므로 전 과목을 하나의 큐브로 구성하는 큐브 3보다는 큐브 4가 보다 효율적인 모형이라 할 수 있다.

SAAS 분석데이터의 각 차원은 계층구조로 구성될 수 있다. 계층(hierarchy)을 두면 저장공간을 효율적으로 사용할 수 있을 뿐 아니라, 다차원적 연산도 매우 효율적으

로 신속하게 처리될 수 있다. SAAS 데이터의 각 차원에 그림 5와 같은 계층을 정의하였다. 그림 6은 계층구조를 반영한 SAAS 큐브의 논리적 구조이다. 음영이 있는 부분이 분석대상 데이터 (시험점수)가 있는 베이스 큐브(base cube)이다. 베이스 큐브는 384,000 (= 400 * 24 * 40) 셀로 구성되고, SAAS 큐브의 크기는 744,150 (= 410 * 33 * 55) 이 된다. 하나의 큐브가 1 메가 바이트를 넘지 않아 실용적으로 사용할 수 있다는 것을 알 수 있다.

큐브 종류	큐브 크기	큐브 밀도(%)
큐브 1	11,520,000	67%
큐브 2	1,152,000	67%
큐브 3	3,840,000	100%
큐브 4	384,000	100%

그림 4 큐브의 크기와 밀도

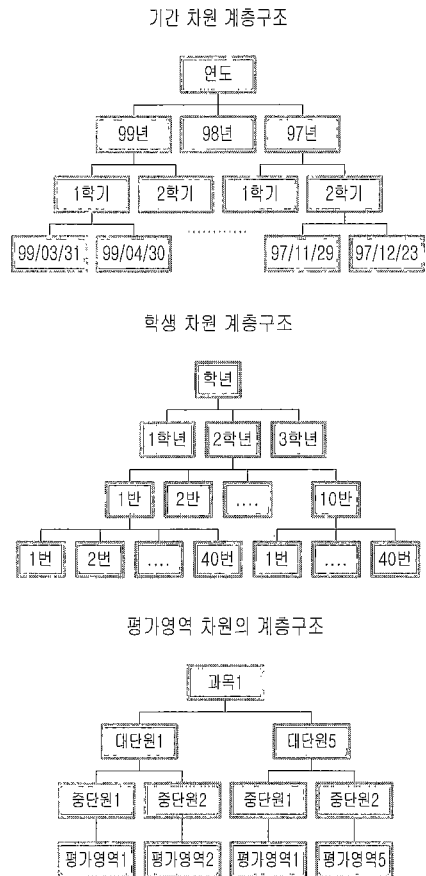


그림 5 SAAS 큐브의 각 차원 계층 구조

3.3 SAAS 시스템 성능 향상 방법

본 연구에서는 SAAS의 성능 향상을 위해 2가지 방법을 사용한다. 첫째, OLAP에서 흔히 사용하는 방법으로써, 분석결과와 일부를 계산해 둔다. 둘째, 대형 배열의 효율적인 저장방법으로써, 배열을 작은 조각 배열인 청크(chunk) 단위로 나누어 저장 [9, 10]해 둔다. 사전 연산 분량과 청크의 크기와 모양은 설계하고자 하는 시스템의 용도에 따라 적절하게 선택되어야 하는데, SAAS에서는 이들을 어떻게 선택하였는지 설명하기로 한다.

3.3.1 사전 연산

사전 연산의 분량은 ① 사전연산에 걸리는 시간, ② 사전연산 후 생성된 데이터의 분량, ③ 질의처리 시간 등을 고려하여 선택해야 한다[7]. 사전 연산을 많이 해 둘수록 질의처리시간이 급속도로 줄지만, 적정점을 지나게 되면 사전 연산 결과를 디스크로부터 읽어오는 것보다 차라리 메모리에 있는 데이터로부터 분석결과를 직접 계산하는 시간이 더 빠르게 된다. 사전 연산의 분량은 또한 사용 가능한 디스크의 분량과 사전 연산에 허용되는 시간을 고려해서 결정되어야 한다. 큐브의 일부만을 계산해 두는 경우에는, 실행시간에 계산하기 힘든 데이터, 질의에 자주 사용되는 데이터, 다른 계산에 자주 쓰이는 데이터를 우선적으로 계산해 두는 것이 효율적이다[7].

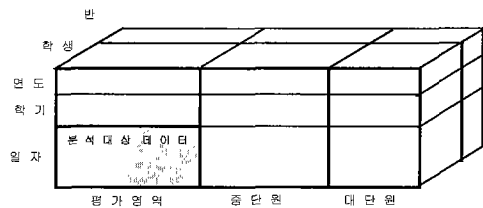


그림 6 계층구조를 갖는 다차원 큐브

사전 연산 비율에 따른 SAAS 큐브의 크기를 계산해 보자. 사전 연산을 전혀 하지 않고 베이스 큐브만 저장해 두는 경우와, 중단원 영역에 대한 사전 연산을 100% 한 경우, 그리고 큐브 전체를 생성한 경우가 표 1에 보인다. 사전 연산 처리시간은 베이스 큐브의 생성시간을 1로 하여 몇 배가 걸리는가를 계산한 것이다.

표 1의 분석결과를 보면, 사전 연산을 100% 하는 경우, 큐브 생성 시간은 베이스 큐브의 생성 시간의 6배가 되고, 사전 연산을 통해 생성된 데이터의 분량은 베이스 큐브의 2배가 된다는 것을 알 수 있다. 반면, 과목영역

에서 조회가 잦은 중단원 부분만을 기준으로 100% 사전 연산을 하는 경우는 큐브의 생성시간이 훨씬 줄어들고, 생성되는 데이터의 분량 또한 베이스 큐브의 1/3 정도에 불과하다. 또한 중단원 영역의 분석결과로부터 그 상위 레벨의 분석결과를 얻고자 할 때는 계산의 양이 많지 않기 때문에 온라인으로 분석결과를 제공하는 경우에도 충분히 그 속도를 보장할 수 있다. 따라서, SAAS는 중단원 영역까지만 사전연산을 해 두는 방법을 쓴다.

표 1 사전 연산 분량에 따른 성능 분석

사전 연산 분량	저장공간 (바이트)	사전연산 처리시간
(1) 큐브 100% 계산	744,150	6
(2) 중단원 영역 100% 계산	139,740	2
(3) 0% 사전 연산	384,000	1

3.3.2 큐브의 청크 단위 저장

SAAS의 질의는 전형적인 OLAP 연산들이다. 따라서 큐브의 데이터는 특정 차원의 값을 기준으로 슬라이스(slice) 또는 다이스(dice) 형태로 잘라져서 메모리에 로드되는 경우가 대부분이다. 대형 다차원 배열을 그대로 디스크에 저장한다면, 이러한 연산을 처리하기 위해서 필요 이상으로 많은 양의 데이터가 메모리로 수시로 로드되어야 한다. 이 문제를 해결하기 위해 대형 다차원 배열을 모든 차원에 공평한 기회를 주도록 작은 다차원 배열형태로 디스크에 저장해 두는 방안이 [9, 10]의 연구에서 제안되었다.

연산들이 필요로 하는 데이터가 어떤 것인지 파악되지 않는 상태에서는 모든 차원에 공평한 기회를 주도록 청크 모양을 만드는 것이 바람직하지만, SAAS의 연산의 특징을 고려해 볼 때, 청크의 모양을 달리하는 것이 시스템의 효율을 높이게 된다. SAAS의 경우 100% 사전 연산을 하는 경우, 한 학년, 한 과목의 큐브가 차지하는 공간은 778,569 바이트 수준이다. 만약 전체 학년, 전 과목의 정보를 필요로 하는 경우는 대략 20M 바이트 이상의 공간을 필요로 하게 된다. 그림 7은 SAAS의 베이스큐브에 대한 청크 단위 저장구조를 보인다. SAAS는 평가영역을 분리하지 않기 때문에, 시스템 특성상 평가영역차원을 하나로 묶어 큐브들을 자른다. 청크는 디스크 블록 크기인 4K 바이트씩 나누었다. SAAS의 경우는 한 학생의 전체 평가 자료가 960 바이트이므로, 4

명 학생의 평가 자료를 하나의 청크에 저장한다. 이렇게 저장함으로써, 한 반에 대한 분석을 할 경우, 10개의 청크만 메모리에 로드시키면 된다.

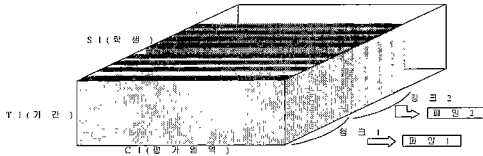


그림 7 S₁T₁C₁ 서브큐브의 청크단위 저장

4. SAAS의 설계 및 구현

SAAS 시스템은 그림 8과 같이 4개의 기본 모듈과 2개의 추가 서비스 모듈로 구성된다. 베이스 큐브 생성 모듈과 서브큐브 사전연산 모듈은 필요한 분석자료를 생성하고, 완성된 큐브 상에서 큐브 연산 모듈이 슬라이스, 다이스 등의 연산을 실행한다. 성취도 분석 인터페이스는 분석결과를 편리한 인터페이스를 통해 사용자에게 제공한다. 2개의 추가 서비스 모듈 중의 하나인 채점 모듈은 성취도 분석을 위한 사전 준비에 쓰이고, 학년말 평어 산출 모듈은 분석 후 처리로 SAAS의 사용 효율성을 높인다. 각 모듈별 기능은 다음과 같다.

- (1) 베이스큐브 생성 모듈: 채점 모듈을 통해 입력되고 저장되어 있는 학생 성적을 베이스 큐브로 로딩시킨다. 각 학생의 성적을 베이스 큐브의 특정 위치로 로딩하기 위해서는 각 차원의 멤버가 베이스 큐브의 해당 축의 어떤 위치에 해당하는가를 신속하게 찾아야 하는데, 이를 위해 각 차원별로 해쉬 테이블이 생성된다. 성적은 해쉬 테이블이 참조되면서 1개씩 3차원 베이스큐브에 로드된다.
- (2) 서브큐브 사전연산 모듈: 베이스큐브로부터 계층 구조에 따른 서브 큐브를 생성한다. 각 차원의 계층 구조 정보는 각 차원 테이블로부터 얻는다. 계층 구조에 따라 bottom-up 방식으로 각 성적을 집계 계산해 가면서 서브 큐브들을 생성한다.
- (3) 큐브 연산 모듈: 과목, 학년을 입력 파라미터로 하여 해당 서브큐브를 찾아 메모리에 로드시킨다. 성취도 분석 인터페이스를 통해 입력되는 연산 (슬라이스, 다이스, 등)을 실행한다.
- (4) 성취도 분석 인터페이스: 과목, 기간, 학생 차원의 값을 입력받아 해당 슬라이스를 화면에 표시한다. 각 차원 계층 정보를 입력받아 roll-up, drill-down 연산을 가능하게 한다. 분석 결과는 테이블 형식이나 차트

(막대, 꺾은선) 형식으로 표현 가능하다.

(5) 채점 모듈: 시험성적을 입력받아 저장한다. 학생들의 답안과 모범 답안을 비교하여 답안의 정확도를 “상/중/하”로 입력할 수 있도록 한다. 채점 화면에는 상/중/하의 기준도 정확하게 명시되어 있다.

(6) 학년말 평어산출 모듈: “상/중/하”로 되어있는 시험 성적들을 종합하여 학년말 “수/우/미/양/가” 평어를 산출한다. 여러 회차를 통해 치러진 시험 성취도 결과에 0 - 2까지의 가중치를 주어 합산/평균하여 과목별 “수/우/미/양/가” 평어으로써 학생들의 학생생활기록부에 기재한다.

SAAS는 다음과 같이 구현되었다. 이 시스템은 비주얼 베이직과 C 언어로 구현되었고, Windows 95 이상의 운영체제의 PC 환경에서 실행된다. 시험 채점은 그림 9와 같은 화면을 통해 이루어진다. 시험문제가 문젠행으로부터 추출된 CAT형 평가라하면, 그림과 같이 화면에 보이는 문항 문제를 보면서 채점하게 되고, 종이에 출력된 시험이었다면, 교사는 평가결과만을 입력한다. 평가결과는 ‘상/중/하’로 입력한다. 한 문항의 채점이 끝나면 채점결과는 ‘답안 데이터베이스’에 기록된다. 입력된 채점 결과는 문젠행 정보와 함께 평가결과 자료로 쓰이게 된다. 그림 10은 학기말 평어 산출 화면을 보여준다

큐브 생성 컴포넌트는 분석할 ‘학년’과 ‘과목’ 데이터를 입력으로 받아서, 큐브를 생성한다. 이 단계에서, 시험 성적 테이블 자료가 베이스 큐브에 로드되고, 해당 서브큐브와 각 서브큐브의 인덱스가 생성된다. 그림 11-16는 학업성취도 분석 화면들이다. 학업 성취도 분석 화면을 보면 과목, 학생, 기간의 3차원별로, 그리고

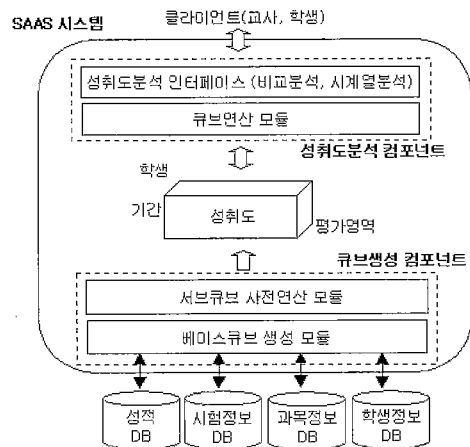


그림 8 SAAS 시스템의 구조

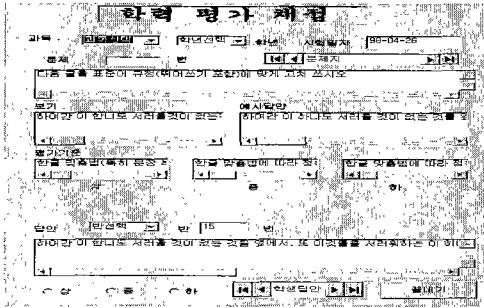


그림 9 평가 결과 채점 화면

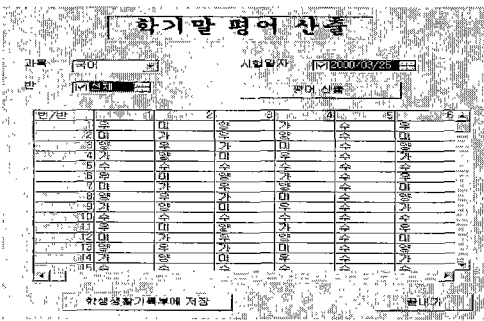


그림 10 학기말 최종 평어 산출

각 차원의 계층별로 분석이 가능하도록 한다는 것을 알 수 있다. 사용자는 원하는 차원과 계층을 편리한 사용자 인터페이스를 통해 선택할 수 있다. 분석 결과 자료는 화면에 단독으로 또는 다른 자료들과의 비교 화면으로 보여진다. 분석결과는 테이블 형태로 볼 수도 있고, 막대 그래프나 꺾은선 그래프 형태로도 볼 수 있다.

여러 번에 걸쳐 시행된 학력 평가 결과는 학기말에 종합되어 '학생생활기록부'에 기재된다. 이 때, '상/중/하'의 평가결과는 '수/우/미/양/가'의 평어로 환산된다. 교사들이 합산하고자 하는 평가들의 시험일자를 다중 선택한 후 평어산출 버튼을 누르면 학생들의 최종 성적결과가 반별로 정렬되어 그림 10과 같이 화면에 나타난다. '학생생활기록부에 저장' 버튼을 누르면 이 자료들은 학생생활기록부 데이터베이스에 해당 과목, 해당 학년별로 기록되어진다.

그림 11-16은 한 학생의 평가결과를 기준으로 여러 가지 분석결과 자료를 산출한 예를 보여준다. 그림 11은 1998년 5월 20일에 시행한 국어과목에 대해 1학년 1반 10번 학생의 학업 성취결과를 그 학생이 속한 1반 전체 평균과 그 학생이 속한 1학년 전체 학생들의 학업 성취 평균을 함께 비교한 것이다. 해당 학년 전체와의 상대 비

교를 통해 그 학생의 우수/열등 정도를 알 수 있다. 중단원 1, 2, 3 영역에 대한 평가가 이루어졌음을 알 수 있고, 해당 학생은 중단원 2에서는 또래 학생들에 비해 우수한 성취도를 보이지만 중단원 1에서는 반 평균에 비해 열등한 결과를 보인다는 것을 알 수 있다. 그림 12는 이 결과를 막대차트로 보인다.

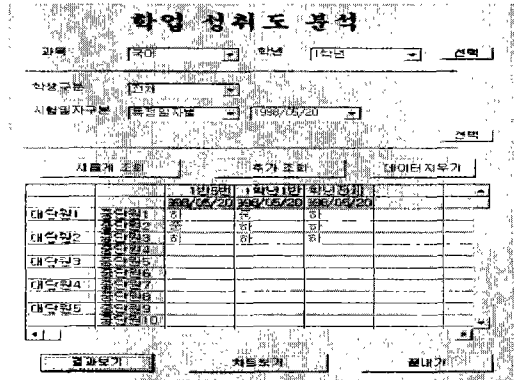


그림 11 반/학년평균과의 학업성취도 비교(테이블)

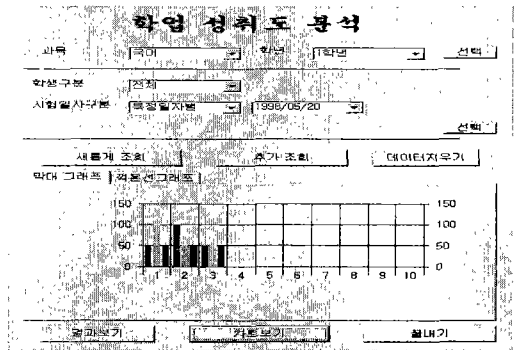


그림 12 반/학년 평균과의 학업 성취도 비교(막대 차트)

그림 13은 시계열에 의한 학업 성취도를 분석 한 경우이다. 1학년 1반 15번 학생의 학업 성취 결과에 대하여 1998년 3월부터 1998년 6월까지 시간의 흐름에 따른 추이를 파악하려는 것이다. 이 학생은 중단원 1에서는 지속적인 향상을 보이나, 중단원 2에서는 별다른 학업 성취상의 진전을 보이지 않는다. 이러한 비교 분석은 학생들의 학업 성취 이해도 및 향상도를 분석하려할 때 유용하다. 이러한 분석의 경우는 그림 14와 같이 꺾은선이 경우 꺾은선 차트 형태의 그림 13은 더욱 학업 성취 향상에 대한 이해를 도울 수 있다.

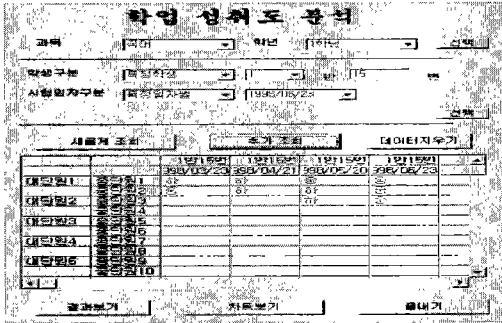


그림 13 시계열에 의한 학업 성취도 비교(테이블)

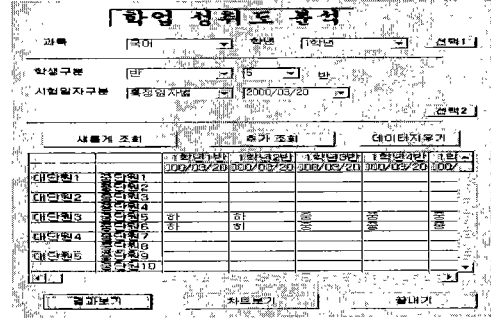


그림 16 반별 학업성취도 비교(테이블)

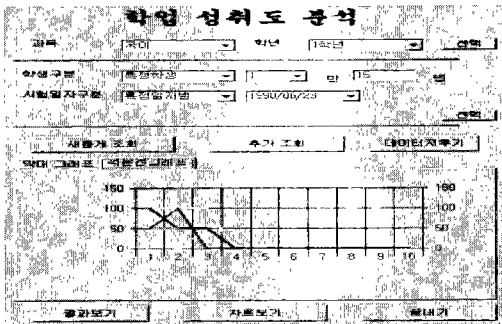


그림 14 시계열에 의한 학업 성취도 비교(꺾은선 차트)

그림 15-16은 교사들이 수업의 진도를 결정할 때나 반별 학생들의 교과 이해도 분석을 하는 것을 보인다. 1반과 2반 학생들은 다른 반 학생들에 비해 중단원 5, 6에서 뒤처지는 점수를 보인다. 이 경우, 1반, 2반 학생들에 대한 보충 강의를 계획할 수 있고, 해당 반의 수업 태도에 대해 검토해 볼 수도 있다. 이와 같이, SAAS는 교사와 학생들에게 정확하고 구체적인 학업 성취 결과를

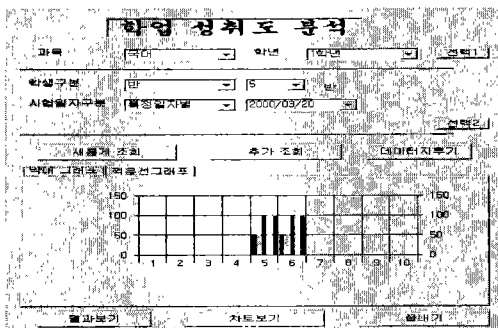


그림 15 반별 학업성취도 비교(막대 차트)

보이는 시스템이다. 이 시스템이 제공하는 정보는 교수-학습에 있어 중요한 의사 결정 시점에서 결정적인 자료로 쓰일 수 있다.

5. 결론

본 연구에서는 고등 학생의 학업 성취도를 다차원적으로 분석하여 그 결과를 온라인으로 제공하는 학업 성취도 분석 시스템인 SAAS를 설계하고 구현하였다. 이 시스템은 시험 채점으로부터 평가 결과의 분석에 이르는 학습 평가 전체 과정을 전산화하여 여러 사용자가 정보를 공유, 검색, 관리, 분석할 수 있도록 한다. SAAS는 최근 비즈니스 테이타의 다차원적 테이타의 분석에 활발히 쓰이는 OLAP 기술을 적절하게 적용하여 테이타의 저장공간을 효율적으로 사용하면서 신속한 분석 결과를 낸다. 평가결과의 분석에 필요한 연산의 특징 상, 흔히 쓰이는 RDB에 자료를 저장해두고 SQL로 사용자의 질의에 응답하는 것보다 훨씬 신속하게 작업이 처리된다.

SAAS는 현재 고등학교 교육에 초점이 맞춰져 있으나, 이는 초/중학교의 교육에도 별다른 구조상의 수정사항 없이 사용할 수 있다. 또한 분석처리가 매우 신속하게 이루어지기 때문에 전국규모의 모의고사의 분석에도 쓰일 수 있는 확장성을 지닌다. 현재는 한 명의 사용자가 쓰도록 구축되어 있으나, 다수의 사용자가 웹환경을 통해 액세스할 수 있도록 하는 시스템을 구축할 때에도, 자료의 저장/관리/분석 엔진으로 쓰일 수 있다. 질의에 고성능으로 응답할 수 있도록 구축되어 있어서, 다수의 사용자가 동시에 사용해도 고성능이 유지될 수 있다.

SAAS가 제공하는 다양한 분석자료는 교사들의 교과 수업 연구에도 큰 도움을 줄 수 있다. 점차 확대 시행되고 있는 교과 교실제에서는 교과 담당교사가 그 교과를 이수하는 학생들에 대한 모든 정보를 관리해야 한다. 그

경우, 교과별로 학생들의 평가결과를 분석하도록 하는 SAAS를 사용하여 교과 담당 교사는 그 과목을 수강하는 학생들의 학업 이해도, 수업 행태, 교과 난이도 등을 학기 중에 신속하게 파악하여 교사의 수업 계획에 적용할 수 있다.

참 고 문 헌

- [1] 백순근 외 3명, "국가교육과정에 근거한 평가 기준 및 도구 개발", 한국교육과정평가원, 1998.
- [2] 최종학, "評價結果의 피드백 方法이 學業成就에 미치는 影響", 고려대학교 교육대학원 석사학위 논문, 1996.
- [3] 한국교육과정평가원, "CT&CAT, "http://kice2.kice.re.kr/ctcat99, 2000.
- [4] Pilot Software Inc., "An Introduction to OLAP Multidimensional Terminology and Techmology," Technical White Paper, <http://www.pilotsw.com/olap/olap.htm>, 1999.
- [5] Thomsen, Erik, "OLAP Solutions: Building Multidimensional Information Systems," John Wiley & Sons, New York, 1997.
- [6] 조재희, 박성진, "OLAP 테크놀로지", 시그마컨설팅 그룹, 2000.
- [7] Business Intelligence Ltd, "The Olap Report : Database Explosion," <http://www.olapreport.com/DatabaseExplosion.htm>, 2000.
- [8] Kim, Won and Myung Kim, "Performance and Scalability in Knowledge Engineering: Issues and Solutions," Journal of Object-Oriented Programming, Vol. 12, No. 7, pp. 39-43, Nov/Dec. 1999.
- [9] Zhao, Yihong, Prasad Deshpande, and Jeffrey Naughton, "An Array-Based Algorithm for Simultaneous Multidimensional Aggregates," Proc. ACM SIGMOD '97, pp. 159-170.
- [10] Sarawagi, Sunita and Michael Stonebraker, "Efficient Organization of Large Multidimensional Arrays," Proc. of 10th Data Engineering Conference, Feb. 1994.



김 명

1981년 이화여자대학교 수학과 학사. 1983년 서울대학교 계산통계학과 석사. 1990년 미네소타대학교 컴퓨터학과 석사. 1993년 캘리포니아 주립대학교(산타바바라 소재) 박사. 1993년 ~ 1994년 캘리포니아 주립대학교(산타바바라 소재) Postdoc, 강사. 1995년 ~ 1999년 이화여자대학교 컴퓨터학과 조교수. 2000년 ~ 현재 이화여자대학교 컴퓨터학과 부교수. 관심분야는 지식공학, OLAP, 인터넷 기술, 고성능 컴퓨팅 등



박 미 현

1995년 한국외국어대학교 컴퓨터 공학과 학사. 2000년 이화여자대학교 교육대학원 컴퓨터교육학 석사. 2000년 ~ 현재 부산대학교 환태평양컴퓨터 강의 및 부경대학교 전자컴퓨터 정보통신학과 시간강사. 관심분야는 컴퓨터 교육, 지식공학 등