

Relational Database SQL Test Auto-scoring System

Tai-Sung Hur*

*Professor, Dept. of Computer Science, Inha Technical College, Incheon, Korea

[Abstract]

SQL is the most common language in data processing. Therefore, most of the colleges offer SQL in their curriculum. In this research, an auto scoring SQL test is proposed for the efficient results of SQL education. The system was treated with algorithms instead of using expensive DBMS(Data Base Management System) for automatic scoring, and satisfactory results were produced. For this system, the test question bank was established out of 'personnel management' and 'academic management'. It provides users with different sets of test each time. Scoring was done by dividing tables into two sections. The one that does not change the table(select) and the other that actually changes the table(update, insert, delete). In the case of a search, the answer and response were executed at first and then the results were compared and processed, the user's answers are evaluated by comparing the table with the correct answer. Modification, insertion, and deletion of table actually changes the data table, so data was restored by using ROLLBACK command. This system was implemented and tested 772 times on the 88 students in Computer Information Division of our college. The results of the implementation show that the average scoring time for a test consisting of 10 questions is 0.052 seconds, and the performance of this system is distinguished considering that multiple responses cannot be processed at the same time by a human grader, we want to develop a problem system that takes into account the difficulty of the problem into account near future.

▶ **Key words:** Relational Database, CBT, SQL(Structured Query Language), auto scoring

[요 약]

오늘날 데이터 처리에 있어 가장 보편적인 언어가 SQL이다. 이를 위해 SQL 교육이 대학에서 진행되고 있다. 따라서 이번 연구에서는 SQL교육의 학습효과를 극대화하기 위한 SQL의 퀴즈 자동 채점 시스템을 제안한다. 본 시스템은 SQL 퀴즈의 자동 채점을 위해 데이터베이스관리시스템을 활용한 알고리즘을 이용하였으며, 만족할 만한 결과를 도출하였다. 본 시스템을 위해 학사관리, 인사관리 데이터베이스에 대해 학사관리의 문제로 문제 은행을 구축하고, 사용자에게 매번 다른 문제를 제공할 수 있도록 하였다. 채점은 테이블에 변화가 없는 검색과 테이블이 변화하는 수정, 삽입, 삭제로 나누어 처리하였다. 검색의 경우 정답과 응답을 실행한 후 실행 결과를 비교하여 처리하였으며, 수정, 삽입, 삭제는 정답과 오답을 실행한 후 테이블을 검색하여 비교함으로써 정답을 확인하도록 하였다. 수정, 삽입, 삭제는 테이블이 변화하였으므로 트랜잭션(transaction) 제어어인 ROLLBACK 명령어를 이용하여 데이터를 원래대로 복원하였다. 본 시스템을 구현하고, 우리대학 컴퓨터정보과 2학년 88명을 대상으로 772회 시행하였다. 시행결과 1회 10문항으로 구성된 시험에 대한 평균 채점 소요시간은 0.052초로 매우 효과적인 것으로 나타났으며, 채점관의 경우 동시에 여러개의 응답을 동시에 처리할 수 없음을 고려한다면 본 시스템의 성능이 월등함을 확인하였다. 향후 정답율을 기초로 문제 난이도를 고려한 문제 시스템으로 발전시키고자 한다.

▶ **주제어:** 관계형 데이터베이스(Relational Database), CBT, SQL(Structured Query Language), 자동채점

-
- First Author: Tai-Sung Hur, Corresponding Author: Tai-Sung Hur
 - *Tai-Sung Hur (tshur@inhatc.ac.kr), Dept. of Computer Science, Inha Technical College
 - Received: 2019. 08. 30, Revised: 2019. 11. 14, Accepted: 2019. 11. 15.

I. Introduction

오늘날 4차 산업혁명의 키워드 중에 하나가 빅데이터이며, 빅데이터란 다양한 종류의 데이터에 대한 생성, 수집, 분석, 표현을 중심으로 데이터로부터 가치를 추출하고 결과를 분석하는 기술이다[1].

따라서 빅데이터 중 데이터 수집 후 저장하고 필요한 정보를 추출하는 방법은 매우 중요하며, 이에 대한 여러 가지 방법이 논의 연구되고 있다. 현재 데이터 저장 관리에 널리 이용되고 있는 것 중 하나가 관계형 데이터베이스 시스템이다. 이러한 관계형 데이터베이스를 이용하여 데이터를 저장 및 추출에 이용하려면 데이터 처리를 위한 SQL(Structured Query Language)의 습득은 필수적이다[2].

SQL은 1970 E.F.codd에 의해 제안된 관계형 데이터 모델에 의한 관계형 데이터베이스관리시스템(Data base Management System)에서 데이터를 관리하기 위해 설계된 특수목적의 프로그램 언어이다[3]. 1970년 제안된 관계형 데이터베이스는 이 후 시장을 확대하면서 SQL은 데이터 처리의 중심에 서게 되었다.

따라서 관계형 데이터베이스 및 빅데이터의 데이터 처리에서 데이터 저장 및 추출하기 위해서는 SQL의 활용은 기본이다.

SQL은 데이터 처리를 위한 특수목적의 프로그램 언어이다. 따라서 반복적인 학습을 통해서 활용기술을 습득할 수 있으며, 반복학습 효과를 극대화할 수 있는 방법 중의 하나가 주기적인 시험이다. 그러나 시험에 따른 채점은 큰 부담이며, 이를 자동 처리하기 위한 연구가 많이 진행되고 있다. 일반적으로 자동 채점에 있어 시험은 객관식시험과 주관식 시험으로 구별되며, 객관식 시험의 경우 OMR 방식 등을 이용해 컴퓨터에 입력되고, 입력된 답안의 채점 프로그램을 이용하여 부담 없이 채점이 이루어지고 있다. 주관식 시험의 경우 자연어 처리를 포함한 여러 알고리즘이 연구되고 있다[4]. 특히 SQL과 같은 특수목적 프로그램 언어의 경우 데이터베이스의 데이터를 이용한 프로그램 실행 결과를 검토해야 함으로 DBMS를 필요로 하며, 채점자에게 큰 부담으로 작용한다. 본 논문에서는 이런 채점과 관련된 문제점을 해결하기 위해 채점 알고리즘을 적용한 시험 관리 시스템을 설계 구현함으로써 관계형 데이터베이스를 공부하는 많은 학생들에게 관계형 데이터베이스의 관리 운영을 위해 필요한 SQL의 습득을 위한 기회를 제공하고자 한다.

본 논문에서는 SQL의 효율적인 습득을 위해 반복학습

을 할 수 있는 시험의 자동 채점 시스템을 구현하였다. SQL의 습득을 위한 시험 자동 채점 시스템을 이용한 SQL 시험 사이트(Allday SQL; 삼시 SQL)를 개설하여 SQL 학습 습을 하고자 하는 사용자에게 시험을 통해 SQL 학습의 기회를 부여함으로써 효과적인 SQL 교육에 기여하고자 한다.

이를 위해 본 연구에서는 SQL 문제 은행을 준비하고, 사용자 각각에 교육 사이트에 접속하여 시험을 실시하고, 자동 채점을 통해 SQL 활용 능력을 향상할 수 있도록 하였다.

본 시스템은 문제의 중복을 피하기 위해 문제를 검색(SELECT), 수정(UPDATE), 삽입(INSERT), 삭제(DELETE)로 구분하고, 각각을 기능별로 다시 소분류 하여 문제 은행을 구성하였으며[5], 사용자가 시스템에 접속할 경우 시험문제의 중복 출제를 방지하기 위한 알고리즘을 적용하였다. 또한 검색의 경우 정답과 응답의 실행 결과를 비교하는 방법으로 채점하였으며, 다만 order by' 절이 포함된 문제의 경우에는 결과의 순서를 확인하기 위해 반복문을 이용하여 처리하였다. 수정, 삽입, 삭제의 경우에는 원본 테이블의 데이터 변경을 방지하기 위해 TCL(Transaction Control Language)인 'ROLLBACK' 명령을 이용하였으며, PL/SQL 프로시저를 이용하였다[5].

본 시스템의 성능을 확인하기 위해 인하공업전문대학 컴퓨터정보과 2학년 학생들의 데이터베이스 강의에 활용하였으며, 향후 사용자의 시험 결과 데이터를 활용하여 난이도를 이용한 시험시스템으로 확장할 계획이다.

II. Preliminaries

1. Related works

학업 성취도 평가에 있어 가장 중요하게 고려할 부분은 출제의 공정성과 난이도, 채점시간을 최소화하는 것이다. 따라서 세계 각국은 국제적 규모의 학업성취도 평가에서 컴퓨터 기반 평가 체제의 도입을 적극적으로 모색하고 있다. 실제로 미국의 GMAT(Graduate Management Admission Test)와 NAEP(National Assessment of Educational Progress), 네덜란드의 MATHCAT(대학의 수학 강의 초기 배치평가), 일본의 CASEC(Computerized Assessment System for English Communication) 등의 컴퓨터 기반 평가가 실시되고 있으며, PISA(Programme for International Student Assessment)와 ICILS(International Computer and

Information Literacy Study)와 같은 국제 학업성취도 평가에서도 컴퓨터 기반 평가를 도입·시행하고 있다[6].

특히 대규모 학업 성취도 평가의 경우 학생들의 답안을 빠르게 수합하고 채점하여 그 결과를 신속히 처리하기 위해 컴퓨터를 이용한 평가를 주로 이용하고 있으며, 이를 극대화하기 위해서는 무엇보다 기계에 의한 자동채점이 필수적으로 실현되어야 한다.

일반적으로 학업 성취도 평가에서 시험은 객관식과 주관식으로 구별한다. 평가와 관련해서는 평가의 효율화는 매우 중요하며, 객관식 문제의 경우 컴퓨터 채점 프로그램을 이용하여 손쉽게 적용할 수 있다. 그러나 답안의 내용을 검토하여 채점하는 서술형 문제의 경우 해결해야 할 문제가 있다. 현재의 학업성취도 평가와 같이 교과 지식 기반으로 답안 내용이 제한적이고 그 언어 단위 길이가 짧을 수록 기계에 의한 자동채점 가능성과 채점 신뢰도는 높일 수 있다. 따라서 현재의 자연언어처리 기술로 해결 가능한 단어·구 수준의 짧은 답안부터 자동으로 채점할 수 있는 프로그램을 이용하고 있으며, 많은 연구가 진행되고 있다. 주관식 문항 자동채점 프로그램은 컴퓨터 기술을 이용해서 서술식으로 기술된 답안을 자동으로 평가하고 채점하는 것으로, 크게는 에세이형 채점 프로그램과 단문형 채점 프로그램으로 나뉜다. 에세이형 채점 프로그램은 보통 답안에 포함된 내용 자체의 정확성보다 글 쓰는 방식에 초점을 두는 반면, 단문형 채점 프로그램은 자연언어처리 기반으로 내용의 정확성을 중심으로 채점한다[7]. 현재 우리나라의 대규모 평가에서 활용되는 서술형 문항은 한 교과 영역에서 주요 개념의 이해 정도를 평가하기 위한 단문 형태의 답안을 요구하므로, 우선적으로 단어·구 수준의 한국어 서답형 문항 자동채점 프로그램을 개발하는 데 참조할 만하다[8].

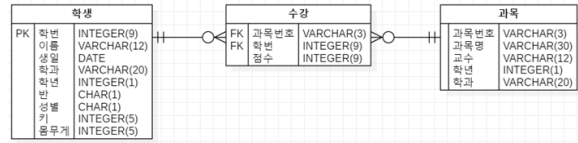
그러나 본 연구에서는 주관식, 객관식 문항 이외의 학생의 답안에 대한 처리결과를 확인해야 함으로 기존의 주관식, 객관식 문제에 대한 자동 채점과는 달리 학생들의 답안을 처리한 결과를 채점에 이용하는 프로그램이 필요하다.

III. System configuration

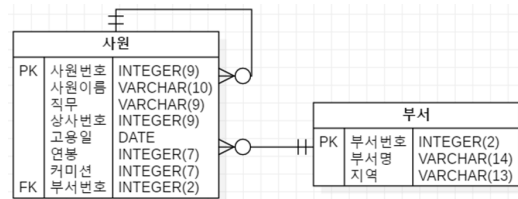
1. Item pool

본 시스템은 문제은행식으로 운영되며, 초기 문제 은행에는 Fig. 1과 같이 현재 데이터베이스 교재 등에서 가장 많이 사용되고 있는 학사관리 데이터베이스와 오라클의

참고용 데이터베이스인 인사관리 데이터베이스를 사용하였다[5]. 문제는 학사관리 528문제, 인사관리 393문제를 준비하였다.



a. The ERD of Academic management DB



b. The ERD of Personal management DB

Fig. 1. DB ERD in test question bank

본 시스템의 문제 은행은 문제가 한 영역에 치중되지 않도록 하기 위해 기본적으로 데이터 처리의 검색, 수정, 삽입, 삭제 4개 영역으로 구분하였다.

Table 1에서 보는바와 같이 검색의 경우에는 일반 SELECT, JOIN, 부질의(subquery)를 포함하는 질의로 나누어 구분하였으며, 검색의 일반 SELECT는 단순 SELECT, order by 포함, group by 포함, 함수 포함, 복합 영역으로 세분류하였으며, JOIN의 경우에는 일반 JOIN, SELF JOIN, OUTER JOIN, 복합영역으로 세분류하여 구성하였다[5].

시험 문제지는 1회에 10문제로 구성하였으며, 문제지는 매회 다른 문제로 구성하기 위한 알고리즘을 적용하였다.

Table 1. SQL instruction problem categories

Category	SQL	Subcategory
Read	SELECT	normal SELECT
		include order by
		include group by
		include function
	JOIN	complex SELECT
		normal JOIN
		SELF JOIN
SUB QUERY	OUTER JOIN	
	complex JOIN	
Modify	UPDATE	-
Insert	INSERT	-
Delete	DELETE	-

2. System Specification

Fig. 2에서 처럼 본 시스템은 25개의 프로세스로 구성되어 있으며, Fig. 3에서 보는바와 같이 10개의 테이블로

구성되어 있다. 또한 채점을 위해 검색과 수정, 삽입, 삭제를 위한 2개의 PL/SQL 프로시저를 이용하였으며, 문항의 통과를 계산, 문제별 응시 횟수 추가, 각 사용자의 문제 분류별 평균 정답률 계산, 시험 시작 시 시험 정보 주가, 시험 종료 시 총 점수, 정답자 수를 변경하기 위해 5개의 트리거 모듈을 이용하였다. 또한 사용자들의 학업 성취를 확인하기 위해 그래프 기능을 이용하여 시각적인 효과를 극대화하였다.

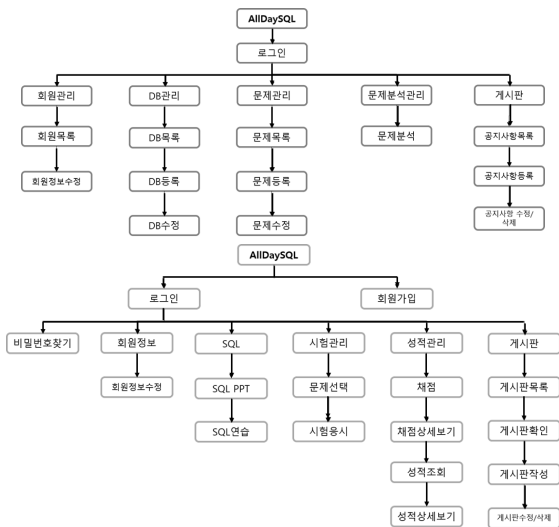


Fig. 2. System configuration

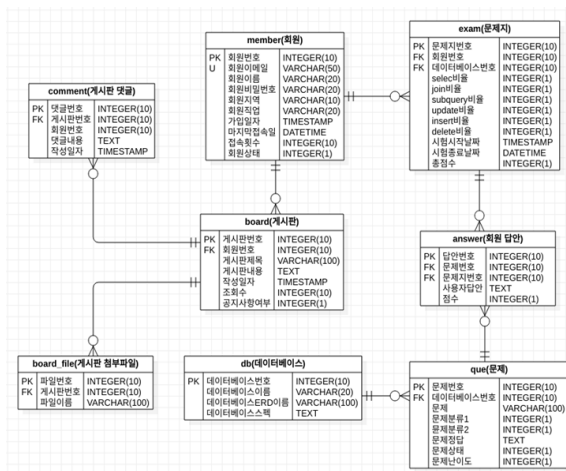


Fig. 3. System ERD

3. Problem Select Algorithm

시험 문제지를 구성하기 위한 문제지는 매번 다른 문제를 선택하기 위해 Fig. 4와 같은 알고리즘으로 한다.

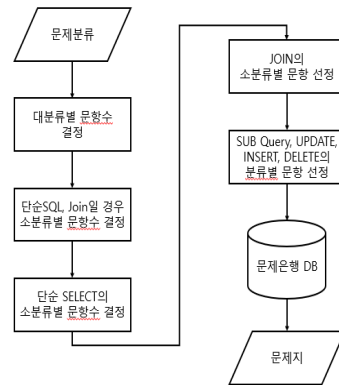


Fig. 4. algorithm of problem selection

본 시스템은 Fig. 4에서와 같이 문제지 작성을 위해 문제의 분류별 문항수를 할당하고, 이를 토대로 각 분류별 문제를 선택한다. 이때 다양한 문제를 선택하기 위해 'order by'절에 RAND 함수를 이용하여 사용자에게 이전 문제지에서 최소 선택된 문제 별 전체 사용자에게 최소 선택된 문제 중 임의의 문제를 선택하도록 하였다.

4. Scoring Algorithm

채점은 Fig. 5에서 보는바와 같이 DBMS의 종류에 따른 문자변환 및 함수 치환 후 테이블의 변화가 없는 검색과 테이블을 변화하는 수정, 삽입, 삭제로 나누어 PL/SQL 프로시저로 처리한다. 또한 채점의 경우 데이터베이스 관리 시스템별로 함수명이 차이가 나는 문제를 해결하기 위해 Table 2에서 처럼 본 시스템에서 사용한 Maria 데이터베이스 관리시스템과 차이가 나는 함수명을 치환하는 절차를 거치게 된다.

Table 2. Function name of DBMSs

Function description	Oracle	MSSQL	MySQL	MariaDB
standard deviation	STDDEV_SAMP STDDEV_POP	STDEV STDEVP	STD STDDEV	STD STDDEV
variance	VAR_SAMP VAR_POP	VAR VARP	VAR VARP	VAR VARP
current date	SYSDATE()	GETDATE()	NOW()	NOW()
cast	TO_CHAR, TO_NUMBER	CAST	CAST	CAST
substring	SUBSTR	SUBSTRING LEFT RIGHT	SUBSTRING LEFT RIGHT	SUBSTRING LEFT RIGHT
concatenate	CONCAT	CONCAT	CONCAT	CONCAT
null value	NVL NVL2	ISNULL	IFNULL	IFNULL

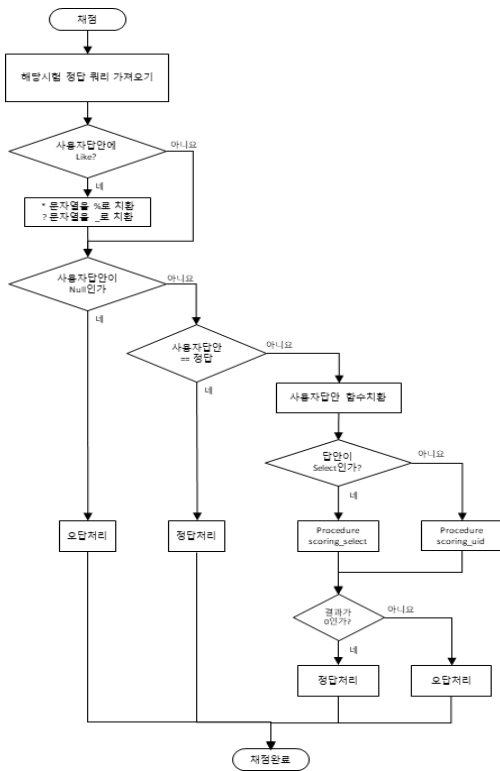


Fig. 5. Scoring algorithm

4.1 SELECT Scoring Algorithm

검색의 경우 테이블의 변화는 없으므로 Fig. 6에서 보는 바와 같이 정답과 응답의 실행결과를 비교하여 사용자의 응답의 정답여부를 확인한다. 다만, 실행 결과의 순서화를 위한 'order by' 절이 포함된 문제의 경우 별도로 결과를 행별로 비교함으로써 정답여부를 확인하도록 하였다.

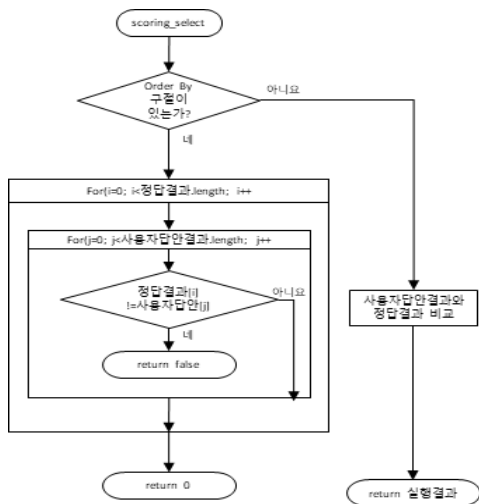


Fig. 6. Scoring algorithm of SELECT question

4.2 UPDATE, INSERT, DELETE Scoring Algorithm

수정, 삽입, 삭제의 경우에는 Fig. 7과 같이 정답과 응답을 실행한 후 테이블을 검색하여 비교함으로써 정답여

부를 확인하도록 하였다. 다만, 수정, 삽입, 삭제의 경우에는 테이블이 변화함으로 채점 후 테이블을 원래대로 환원하기 위해 트랜잭션 제어 언어(TCL)인 'ROLLBACK' 명령어를 이용하였다.

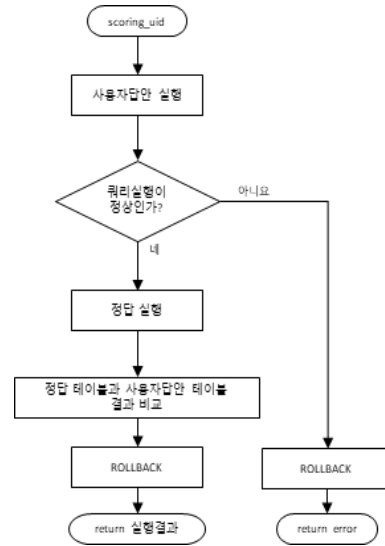


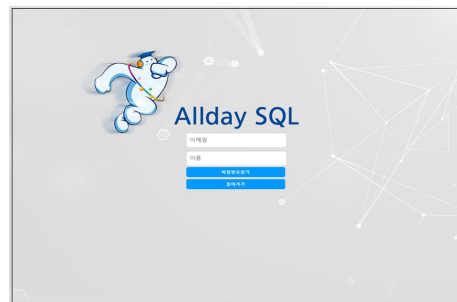
Fig. 7. Scoring algorithm of UPDATE, INSERT, DELETE question

IV. System Implementation

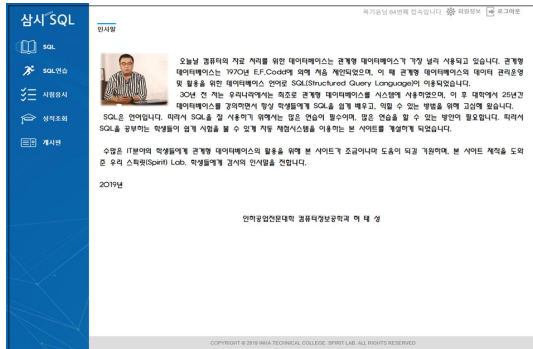
본 연구에서 시스템 구현을 위해 Node.js+ express[9,10]와 Maria 데이터베이스 관리시스템[11]을 이용하였다.

본 개발 시스템은 사용자들의 관계형 데이터베이스 언어인 SQL학습을 위한 시스템으로 시스템의 이름은 Allday SQL(삼시 SQL)로 규정하였다.

Fig. 8은 시스템의 초기 로그인 화면이다. 사용자는 회원가입을 이용하여 회원이 되면 시스템을 이용할 수 있다.



a. Login screen



b. System first screen

Fig. 8. Login, first screen

로그인 후 SQL에 대한 간단한 교육용 강의의 자료가 있어 SQL의 학습을 할 수 있도록 하였다.

SQL시험을 보기 위해서는 Fig. 9에서 보는바와 같이 학사관리, 인사관리 데이터베이스를 선택한 후 시험범위를 선택하게 된다.



Fig. 9. Selection screen of DB and test categories

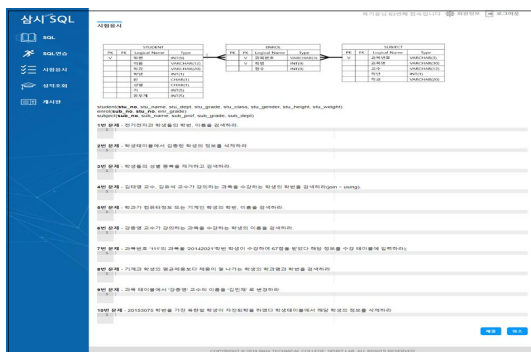


Fig. 10. Testing screen

데이터베이스와 시험 범위를 선택하면 매회 10개의 문제가 선택되어 문제지가 작성되며, Fig. 10에서 보는바와 같이 시험을 실시하게 된다. 시험 시행 후 제출 버튼을 클릭하면 바로 채점이 실시되며, Fig. 11과 같이 채점결과를 확인할 수 있다.

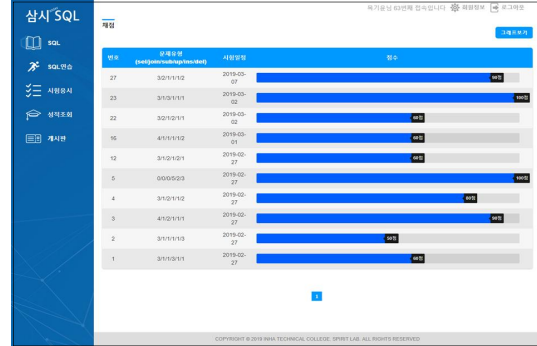


Fig. 11. Scoring history screen

사용자는 SQL을 학습하는 초보자이므로 일정 기간 본 시스템을 이용하여 시험을 통해 SQL을 숙지하게 되며, SQL학습에 대한 결과를 Fig. 12와 같이 그래프로 확인할 수 있다.



Fig. 12. Scoring graph screen

V. Experiments

본 시스템을 구현하고, 2019학년도 1학기 인하공업전문대학 컴퓨터정보과 2학년 학생 88명을 대상으로 772회 시험을 시행하였다. 학사관리 데이터베이스는 354회, 인사관리 데이터베이스는 368회 시행하였다. 시험 1회 시행에 10문항으로 하였다.

Table 3에서 보는바와 같이 10문항 1회분 채점에 소요되는 시간은 평균 0.052초로 나타났으며, 표준편차는 0.021, 최대 소요 시간은 0.265초, 최소 소요시간은 0.023초로, 사람이 채점할 경우 10문항을 채점하는데 걸리는 시간은 실측이 어려우나 2~3분 정도 소요된다고 가정할 경우 채점자에 비해 월등히 높은 효율을 보였다. 또한 채점관에 의한 채점의 경우 동시에 학생들의 답안을 채점하는 것은 불가능하므로 자동 채점의 효율은 매우 효과적으로 평가된다.

Table 3. Distribution of time required for scoring per 10 questions

평균 소요시간	0.052sec
최대 소요시간	0.265sec
최소 소요시간	0.021sec
표준편차	0.021

VI. Conclusions

SQL은 1970 E.F.codd에 의해 제한된 관계형 데이터 모델 기반의 한 관계형 데이터베이스 관리시스템의 데이터 처리를 위한 특수목적 프로그램 언어로 오늘날 데이터 처리에 있어 중요한 자리를 차지하고 있다. 이에 대한 프로그램 입문과정의 사용자에게는 SQL을 쉽게 배우고 사용할 수 있는 교육방법이 필요하다. 따라서 본 연구는 SQL의 교육을 위한 시험 시스템을 도입하여 많은 데이터 처리 입문자들에게 기회를 제공하고자 한다.

시험에서 고려할 점은 채점의 자동화이며, 이에 대한 많은 연구가 이루어지고 있으나, 객관식, 주관식 문제에 국한되어 있으며 본 연구에서처럼 학생들의 답안을 컴퓨터로 처리한 결과로 채점이 이루어지는 경우 기존 연구와는 달리 또 다른 방안이 모색되어야 한다. 특히 본 연구에서처럼 SQL의 실행은 데이터베이스 관리시스템이 필요로 하며 본 연구에서는 이를 간편하게 처리할 수 있도록 알고리즘으로 구현하였다.

본 시스템을 위해 Node.js-express와 Maria 데이터베이스 관리시스템을 이용하였다. 또한 채점을 위해 PL/SQL 프로시저와 채점에 따른 데이터 변화를 자동 수정하기 위해 5개의 트리거를 이용하였다.

본 시스템 구현 후 우리대학 컴퓨터정보과 2학년 83명 학생들을 대상으로 총 772회 시험을 시행한 결과 1회 10문항으로 처리한 경우 평균 채점 소요시간은 0.052초로 매우 효과적인 것으로 나타났다. 또한 채점관의 경우 대량의 학생답안을 동시에 채점할 수 없다는 점을 감안한다면 본 연구는 매우 효율적으로 이용될 수 있을 것이다. 향후 문항별 통과률을 바탕으로 문제를 출제할 때 난이도를 고려하는 방안을 추가하여 SQL 입문 과정의 학생들에게 시험을 통한 반복학습으로 SQL 습득의 기회를 제공하고자 한다.

ACKNOWLEDGEMENT

This work was supported by INHA TECHNICAL COLLEGE Research Grant.

REFERENCES

- [1] Wikipedia, https://en.wikipedia.org/wiki/Big_data
- [2] Song Tae Min, Song juyoung, "Checking th Big Data Analysis", Hannarae Publishing, 2015, pp. 23-26.
- [3] Wikipedia, <https://en.wikipedia.org/wiki/SQL>
- [4] Dikli, Semire, "An Overview of Automated Scoring of Essays", The Journal of Technology, Learning and Assessment, Vol. 5, No. 1, Aug. 2006, pp.4.
- [5] Tai-Sung Hur, "ORACLE SQL&PL/SQL", BooksHolic publishing, pp.38-175, 2016.
- [6] Kim, Kyunghee and others, "Introduced plans for Computer Based National Assessment of Educationa Achievement", Korea Institute for Curriculum and Evaluation, 2012, pp.1-3
- [7] Butcher, P.G., Jordan, S.E., "Acomprison of Human and computer markong of short free-text student responses", Vol. 55, Issue 2, Sep. 2010, p.489.
- [8] Noh, Eunhee, Sung, Kyunghee, "Developing an Automatic Content Scoring Program for Short Answer Korean Items in Large-Scale Assessments", The Journal of Curriculum and Evaluation, Vol. 17, No. 2, 2014, pp.101-104.
- [9] JavaScript, <https://developer.mozilla.org/ko/docs/Web/JavaScript/Reference/>
- [10] Node.js Express, <https://expressjs.com/ko/>
- [11] Matia DB, <https://mariadb.com/kb/ko/mariadb/>

Authors



Tai-Sung Hur received the B.S degree in Dept. of Computer Science from Inha University in 1984, and M.S degree in Dept. of Computer engineering from Soongsil University in 1987, and Ph. D. degree in

Dept. of Computer engineering from Inha University in 1992. Dr. Hur has over 30 years of computer education. He is currently a Professor in the Dept. of Computer Science, Inha Technical College. He is interested in Big data, Database and Internet of Things.