CONCEPTUAL FRAMEWORK OF BUSINESS INTELLIGENCE ANALYSIS IN ACADEMIC ENVIRONMENT USING BIRT

Julaily Aida Jusoh, Norhakimah Endot, Nazirah Abd. Hamid, Raja Hasyifah Raja Bongsu, Roslinda Muda
Faculty of Informatics,
Universiti Sultan Zainal Abidin,
21300 Kuala Terengganu,
Terengganu
{*julaily, nazirah, rajahasyifah, roslindamuda@unisza.edu.my*}
hakimah1987@yahoo.com

ABSTRACT

The growth of technological development in information technology has caused challenges for many institutions. The enhancements of capabilities to gather and store data in multiple source system make it more difficult for the management in academic institutions to analyze the important data. The objective of this paper is to analyze students' academic achievement based on accommodation factor, in other words, to analyze how accommodation can affects the students' academic achievement. This paper discusses the conceptual framework for business intelligence analysis in academic environment using BIRT tool. BIRT tool will generate a report which illustrate and analyze the comparison of students' academic achievement between the students that stay either in-campus and out-campus.

KEYWORDS

Student Relationship Management (SRM), Data Warehouse, Business Intelligence and Reporting Tools (BIRT)

1 INTRODUCTION

Business Intelligence (BI) comprises of skills, processes, technologies, applications and practices that are functioning as leverage to institutions from internal and external assets and to support and improve decision making [7]. An effective BI solution can be used to integrate data sources that frequently used to complement marketing initiatives and compliance analysis reporting. BI applications that apply the concept of aggregated information are Customer Relationship Management (CRM) and Student Relationship Management (SRM).

CRM is a strategy for managing company interactions with customers, clients and sales prospects. It involves of using technology to organize, automate, and synchronize business processes. The goals of CRM are to find, attract, and win new clients and reduce the costs of marketing and client service in different services such as marketing, customer service, and technical support. Meanwhile, SRM is a concept that has been derived from CRM and can be applied to academic institutions. The goals of SRM are to improve the student experience, reduce drop-out rates, and improve institutions efficiency. The SRM involves automating and synchronizing a number of different processes such as academic advising, counseling, and registration. In SRM, the components of data warehouses have not been fully exploited by users. Data warehouse (DW) can be defined as a relational database that is designed for query and analysis. In this research, DW stores information about assessment, personal and allocation. It is usually contains historical data that has been derived from various sources such as Oracle, Informix, Access and MySQL database.

In addition of DW is functioning as a relational database, DW contains environment that includes extraction, transformation, and loading (ETL) and Online Analytical Processing (OLAP). Designing and maintaining ETL steps are often considered as one of the most difficult and resource-intensive portions of DW project. ETL meant extract, transform and load data across various data sources to create an accurate and quality information. ETL steps are responsible for the operations to take place in the DW architecture. There are three steps involve in ETL as shown in Table 1. Online Analytical Processing (OLAP) can be used to analyze data from multiple databases.

OLAP is an approach to answer multi-dimensional queries that encompasses relational reporting and data mining. OLAP tools enable users to analyze multi-dimensional data from different sources. The databases that have been configured for OLAP can be used for complex analytical and ad-hoc queries with a rapid execution time.

Table 1: Process by ETL tools

STEP	ACTIVITIES
Extracting	The data are extracted from the source data stores that can be On-Line Analysis Processing (OLAP), files, web pages, various kinds of documents (spreadsheets and text documents)
Transforming	Transported to the target system or an intermediate system for further processing
Loading	The data are loaded to the central data warehouse (DW)

In this research, we concentrate more on producing reports by using Business Intelligence and Reporting Tools (BIRT). BIRT can be recognized as a part of BI. BIRT tool can be used to produce reports after data has gone through ETL processes. BIRT reports can be divided into four main parts:

a) Data: involves databases, web services and java object that can supply data to produce report. BIRT provide JDBC, XML, Web Services and Flat File that can support the code from different source.

b) Data transform: data is sorted, summarized, filtered and grouped.

c) Business logic: structure that can be used to create the reports.

d) Presentation: a range of options to present the report. It involves tables, charts, texts and others.

This paper is organized as follows. Section 2 describes related research on what and how far does business intelligence can assist in academic environment. Section 3 explains business intelligence tool called BIRT and conceptual framework of business intelligence analysis in academic environment is described in Section 4. We present experimental testing in Section 5 and lastly we conclude this work in Section 6.

2 RELATED WORKS

Recently, many researches have been conducted on fundamental knowledge of the realities in education. Many institutions have problems in academic results which related to the students' activities**.** In order to overcome these problems, some strategies have been proposed to improve education systems.

A web based software using data mining and Quality Function Deployment (QFD) was proposed by [2] to assist higher institutions in processes such as assessing, predicting and managing issues that related to the student success. [2] highlighted that the biggest challenge in higher education was to predict the suitable academic path for the students. This situation happened because the institutions had lack of information to guide the students in selecting a suitable program and course for them. Consequently, many students who did not interested in the enrolled courses could not concentrate and fail to complete their study.

[1] proposed CRM in education to manage relationship across student lifecycle. CRM strategies had been suggested to manage relationship more effectively between company and their clients, in this case, between an institution and their students. CRM strategies offered solutions with more powerful, different features and functionality, and also could help to determine quality of the student across student lifecycle [1].

SRM were introduced in 2008 to support the BI system in higher education [3]. [3] used DW and data mining to analyse the students' data. System architecture of SRM was developed to identify students' activities while strategies were built to solve problems among the students. The concepts, practices and strategies that support education in acquisition of knowledge to make decision making process also discussed in [3]. Table 2 describes the

processes and activities that involve in SRM architecture.

Table 2: Phase and activities in SRM architecture

PHASE	ACTIVITIES
Concept	Discuss process how student acquired the knowledge
Practice	Discuss about the activities that guarantee a personal contact with student an effective, adequate and closely monitoring of academic performance.
Strategy	Discuss about vision and mission of the institution.

BI concepts could be adopted to support the teaching and learning process and also to manage student and lecturer relationship in higher institutions environment. The main objectives of [3] were to prepare the technological tools that could be used to support the education process and also to support the decision making process. [3, 4] mentioned that one of the contribution of student's success factor was by closely monitoring the student's participation in academic activities. However this factor did not occur in many higher institutions. Students' academic failure often occurred during the first year because the institutions unable to monitor their academic activities [4]. For that reason, a conceptual framework and technological infrastructure were proposed by [4] and then integrated them with the SRM system.

A case study which involved 70 students was done by [4] to show that the SRM system was suitable to be used in the learning process. In data gathering process, student's information was collected from presential component, e-learning system and assessment. The presential components comprised of theoretical, practical and tutorial orientation. Several information such as curriculum contents, exercises and project guidelines were included in the e-learning process. The assessment involved were the normal assessment period and exam assessment period. The authors used OLAP, data mining, DW and ETL tools to describe the case study and to examine relevant data [4]. OLAP was used to analyze data from different sources while data mining was used to identify the model and pattern of the data. Besides that, OLAP cubes were used to analyze the student's results verified the teaching learning process and assessment. The ETL process extracted data from the database. The data was cleaned and transformed to the targeted database or data warehouse.

3 BIRT TOOLS

Business Intelligence and Reporting Tools (BIRT) has grown successfully since 2004. The successful of this growing technology is achieved when the user use the BIRT technology to produce reports in business environment. The BIRT reports allow user to recognize trends in data and present the report in interactive ways and make the reports look professional [6].

BIRT can be categorized into two categories. The first category is report development environment that can be used to design and develop the reports while report development environment which is similar to Macromedia Dreamweaver is functioning to design web pages. This category is connected with the Java Database Connectivity (JDBC). JDBC is used to connect with the MySQL database to produce reports. The Structured Query Language (SQL) in MySQL consist of different statement such as select, delete, update and adding statement. The second category is Java APIs. It provides Java Server Report for integrating between BIRT Report Engine and Apache Tomcat [5].

In this research, the graph generated by BIRT will be used to produce the students' report. This report can be used to analyze the correlation of academic results between the students that stay in the hostel provided by the university (in-campus) and those who are renting outside (out-campus accommodations). In this study, we anticipate that accommodation is one of the factors that might influence students' academic results.

4 CONCEPTUAL FRAMEWORK

The purpose of this framework is to identify flow of activities involved in generating related reports. Figure 1 shows the conceptual framework of the research.

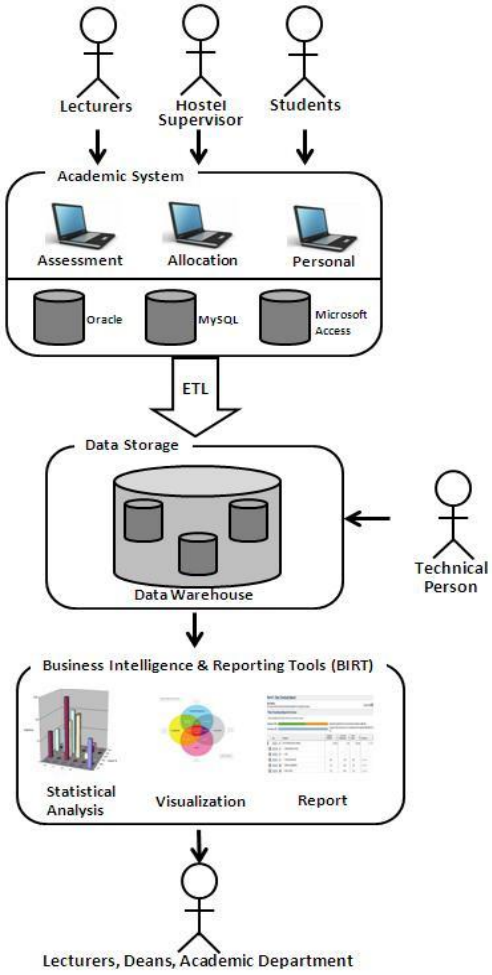

Figure1: Conceptual Framework

The framework can be divided into three phases consist of input, processing and output. The descriptions of the framework are as follow:

Phase 1: Input

In this research, the corresponding users are lecturers, hostel supervisors and students. In Academic System, we consider 3 existing sub-systems that consist of assessment, allocation and personal system. The assessment system contains information regarding students' assessments such as quizzes, assignments, mid-term tests and final examinations. Lecturers are responsible to key-in students' marks in the assessment system. Next is the allocation system that serves the purpose to store students' accommodation data (in-campus or out-campus). This system will be managed by the supervisors who are in-charge of students' accommodation. In addition, personal system is used to record students' personal information such as their addresses and contact numbers.

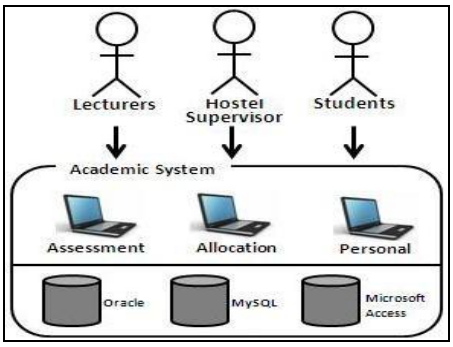

Figure 2: Input Phase

Since the existing systems were usually developed using different database formats (eg: Oracle, MySQL and Microsoft Access), there is the need to process the data from different databases into a single data format prior to further processing. This is where the ETL process comes into play where various data formats can be extracted, transformed and loaded into a unified data format. The three steps for ETL are:

Step 1: Extraction

In this process, data will be extracted from the various sources of the sub-systems and converted into a single database format which is appropriate for the next process.

Step 2: Transformation

Next, a series of rules or functions will be applied to the extracted data from the source sub-systems for loading into the end target database format. This step allows the data from multiple database sources to be joined, filtered and sorted using specific attributes. In this step, certain fields or columns from the source sub-systems with different names but similar attributes (eg : both Matric_No and No_Mat represent the same attribute, ie. students' matric numbers) will be selected and reformat before being loaded into the targeted database format. This process allows data cleansing by removing duplicates and consequently enforce consistency by using a single name of the same attribute.

In addition, only some attributes deemed necessary will be selected from the data sources depending on the type of analysis that is going to be performed. For example, the examination table from assessment system consists of Name (NAME), Matric Number (MATRIC_NO), Course (COURSE), Semester (SEM), Assessment_marks (ASSESS_MARK) and Cumulative Grade Point Average (CGPA). Only important data such as MATRIC_NO, SEM and CGPA will be selected for the analysis. Other data in the table such as NAME, COURSE and ASSESS_MARK will be ignored.

Step 3: Data Loading

In this research, this process involves writing the transformed data into MySQL database format (end target database). Any constraints and indexes are disabled prior to the start of the process and re-enabled after the loading process has completed.

Phase 2: Processing

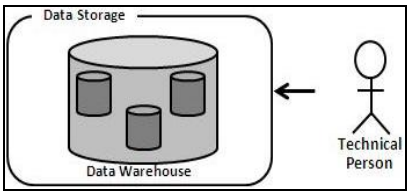

Figure 3: Processing Phase

After the ETL process, the data will be stored in a data warehouse where it can be used for the analysis purposes in order to generate appropriate reports. The data warehouse will be managed by the technical authority.

Phase 3: Output

The Business Intelligence and Reporting Tool (BIRT) will be used to represent the necessary reports. BIRT plays a significant role in the process of statistical analysis and visualization to produce the reports by collecting and analyzing data from the database and displaying students' academic achievement using graphs. The different graphs obtained for in-campus and out-campus students can be compared in order to determine whether the factor of accommodation influences students' academic achievement. Users of the system (lecturers, deans and academic management) can view the graphs analysis which in turn can be used to support the institution in developing a strategy to improve students' achievement.

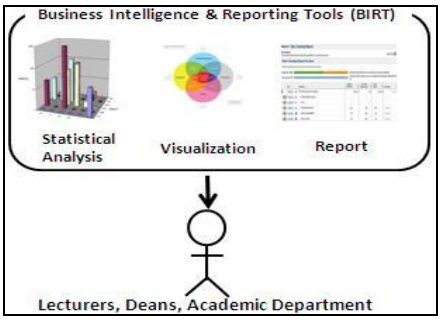

Figure 4: Output Phase

5 EXPERIMENTAL TESTING

To prove the proposed framework, we use dummy data for analysis simulation from various data source format. We assume that the data is collected from two sub-systems consist of assessment and allocation system. In this research, ETL process is use to extract data from those databases and then clean the data and make it uniform to be transferred into one central repository which is MySQL database. Table 3 and 4 illustrates the example of data extracted from academic system.

Table 5 table display the Matric Number (MATRIC_NO), Semester (SEM) and Cumulative Grade Point Average (CGPA) for students of Bachelor of Computer Science (BCS) from semester II 2010/2011.

Table 3: Example of data extracted from the assessment system.

MATRIC_NO	SEM	CGPA
025417	sem 2 2010/2011	2.99
024567	sem 2 2010/2011	3.21
013425	sem 2 2010/2011	3.43
027820	sem 2 2010/2011	3
027615	sem 2 2010/2011	2.77
025419	sem 2 2010/2011	2.8
026781	sem 2 2010/2011	2.7
024178	sem 2 2010/2011	3.1
026710	sem 2 2010/2011	3.15
025695	sem 2 2010/2011	3.33
021569	sem 2 2010/2011	3.5
027899	sem 2 2010/2011	3.55
026008	sem 2 2010/2011	3.6
027611	sem 2 2010/2011	3.58
025410	sem 2 2010/2011	3.62
021755	sem 2 2010/2011	2.5
022371	sem 2 2010/2011	2.48
022581	sem 2 2010/2011	2.47
022890	sem 2 2010/2011	2.6
028711	sem 2 2010/2011	2.82

Table 4: Example of data extracted from the allocation system.

MATRIC_NO	ACCM	ADDRESS
025417	In Campus	Kindi Block
024567	Out Campus	Tok Jembal
013426	In Campus	Khawarizmi Block
027820	In Campus	Kindi Block
025419	In Campus	Wakaf Tengah
026781	In Campus	Fatimah Block
027859	Out Campus	Gong Badak
026008	In Campus	Fatimah Block
021775	In Campus	Kindi Block
022371	In Campus	Khawarizmi Block

Table 4 comprises of Matric Number (MATRIC_NO), Accomodation (ACCM) and Block Name (ADDRESS). The data contains information of students that stay either inside or outside of campus in the same semester.

Figure 1 summarize the entire of analysis process such as where and how the data are obtained, extract, transform and load to make the data accurate in generating reports or graph. This experimental testing highlight four important data to be extracted and analyzed such as matric number (MATRIC_NO), semester (SEM), cumulative grade point average (CGPA) and accomodation (ACCM). Those data will be combined to determine and analyse the correlation of students' academic result and accomodation. In BIRT, the SQL statement was created and used in producing the graph. Below is the example of the SQL statement.

```
Select M.MATRIC_NO, M.CGPA,
C.ACCOMODATION FROM
Examination as M, Allocation as C
WHEREM.MATRIC_NO= C.MATRIC_NO
```

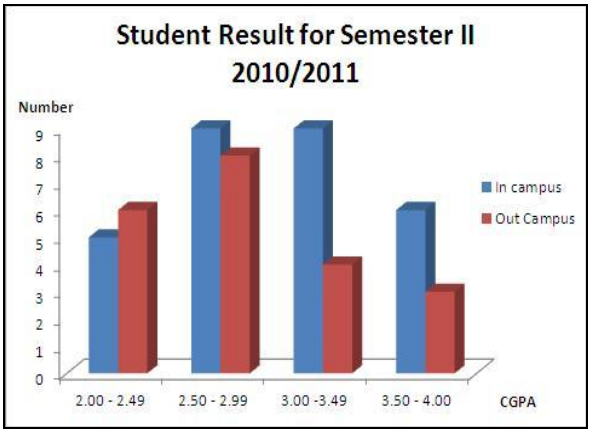

Figure 5: Analysis Graph of Students' Academic Result

Figure 5 depicts the example of the correlation analysis between students' academic result and accommodation for semester II 2010/2011. This figure shows the expected graph of students' results generated by the BIRT tool. From the graph, we can conclude that most in-campus students achieved better results compared to out-campus students. It is apparent that the deployment of the tool provides significant information regarding students' academic achievement which can be used further in the strategizing how to improve the academic result among out-campus students. We anticipate that a better analysis will be achieved if real life data is used in the analysis.

6 CONCLUSIONS AND FUTURE WORK

This research proposed a conceptual framework employed in an academic environment utilizing BIRT tool that assists the process of generating reports for students' academic achievement. The significance of this research lies in the process on how to analyze any factors that could influence students' academic achievement. In future, we are going to implement this framework using real cases in academic environment. Not only the framework can be used to help to predict the category of subjects (eg: programming, database, mathematics) that students can perform well, it can also be used to find the correlation in students' academic achievement for every semester. Eventually, we are hoping to be able to use the proposed framework to predict and identify any patterns associated with students' academic achievement.

7 REFERENCES

1. Engelbert, N., *Best practice with CRM in higher education:Managing relationship across the student lifecycle.* 2007: p. 1-13.

2. Sahay, A., Mehta, K. *Assisting Higher Education In Assessing, Predicting, and Managing Issues Related To Student Success:A Web Based Software Using Data Mining And Quality Function Deployment.* Academic and Business Research Institute Conference, Las Vegas, 2010.

3. Maria Beatriz Piedade, M.Y.S., *Student Relationship Management:Concept, Practice and Technological Support*. In Proceedings of the IEMC-Europe 2008-IEEE International Engineering Management Conference (Estoril, Portugal, June 28-30). ISBN 978-1-4244-2289-0.

4. Maria Beatriz Piedade, M.Y.S., *Business Intelligence Supporting The Teaching Learning Process*. Proceedings of the 5th WSEAS International Symposium on Data Mining and Information Processing (Budapest, Hungary, September 3-5). WSEAS Press, pages 292-297, ISBN:978-960-474-113-7.

5. Ward, J., *Practical Data Analysis And Reporting With BIRT*. 2008.

6. Sheldon Lee-Loy, J.F., *Advance Charting In Birt*. 2008: p. 1-43.

7. Dave Sharman., *An Introduction to Business Intelligence for Higher Education*. Collegiate Project Srvices, 2010.

8. Maria Beatriz Piedade, M.Y.S., *Promoting Students Success with a Business Intelligence System*. Proceedings of CSEDU - 2nd International Conference on Computer Supported Education (Valencia, Spain, April 7-10). INSTICC Press 2010, pages 46-52, ISBN 978-989674-024-5.