

Design and Implementation of Data Integration and Analysis System for Campus Card

Zhao Lei
Shandong University of
Science and Technology,
Qingdao 266590, China

Wang Le
Shandong University of
Science and Technology,
Qingdao 266590, China

Cong Liu*
Shandong University of
Science and Technology,
Qingdao 266590, China

Abstract: Campus card is used to record learning, living and entertainment data of students during school learning period. It contains numerous valuable information of student daily behavior. The effective integration of campus card data is the basis for further mining and analyzing of student behavior. This paper presents the design and implementation of a data integration and analysis system for campus card. This system is analyzed and designed by Unified Modeling Language (UML) and implemented by Java. The main input involves the records of consumption, entrance guard, library and other types of related data. This system can be used to integrate common types of source data and provide several kinds of statistical charts and tables to illustrate the analysis results. Moreover, the system supports to add new functions in terms of plugins.

Keywords: Data integration and analysis; student campus card data; statistical charts; system analysis and design

1. Introduction

The rapid development of information and network technology inspired the widely application of campus cards in lots of schools, especially in universities and colleges. The campus card records almost all the information of students' daily living in school, including borrowing books from library, consumption, dining, shopping, surfing the Internet, and seeking doctors and so on. These data imply many valuable information that can assist the administration managers and teachers to improve the effectiveness and efficiency. Nowadays, the research on the management and mining of campus card data become a hotspot [1]-[2].

Bowman et al. [1] first introduce the development of campus card monitoring system using Cacti Framework and Nagios Plug-in. It can monitor network traffic to evaluate operation of network. *Li et al.* [2] introduce some ways to improve the campus card system's stability and reliability, taking the campus card system of Fudan University as the example. The ways include hierarchical authority, improvement of system automation, and system monitoring etc. *R. Hu* [3] introduces the composition, function, system design, system security and network design of a campus card system. *X. Dong et al.* [4] analyze the effectiveness of some clustering algorithm, including *CobWeb* and *EM*, in mining campus card data, based on consumption data and *Weka* platform. *Xu et al.* [5] use decision-tree-classification algorithm to analyze campus card data about physical exercise. Through this way, students' situation of exercise could be evaluated. *Jiang et al.* [6] analyze students' consumption behavior and learning behavior based on campus card data. This paper uses K-means clustering algorithm to analyze the consumption location and consumption level about boys and girls in school, and it also analyzes the relations among the number of using card, the number of borrowing books, scholarship and academic achievement. *Chen* [7] makes use of the campus card data to excavate and analyze the user's behavior of dining and consumption, and puts forward some suggestions on the operation of dining hall and management of school, taking Liaoning Medical University as an example. *Fei et al.* [8] optimize impoverished students' determination by K-means

clustering algorithm to analyze the consumption data of campus card. *Li* [9] analyzes the students' consumption and behavior according to campus card data of *Beijing Jiaotong University*. *Jin et al.* [10] expound the support of campus card data to schools' decision making based on the analysis of dining records, consumer records, etc.

Most of the above-mentioned references mainly focus on data mining and analysis for the campus card data, and lack of the universal platform of campus card's data integration and analysis. To solve this problem, this paper describes the design and implementation of a data integration and analysis system for campus card, to support the improvement of university information management.

2. Basic Introduction of Data Integration and Analysis Systems for Campus Card

This system mainly realizes the integration, statistics and analysis of campus card log data. The detailed requirements are as follows.

(1) *Data Integration*. The system integrates data from diverse data sources into the system. The supported data sources include Excel, HTML, CSV, structured text, XML and database. During integration, the users choose the file to store the source data in first, and then specify the relationship between target data format and source data format. Finally, the system will check the format and import the source data if the source data is okay.

(2) *Data Query*. It supports basic query, complex query and expert inquiry. The basic query allows users to select one or a few conditions for a query. The complex query enables users to query by multiple conditions and combination of multiple tables. The expert inquiry supports users to write *Structured Query Language (SQL)* statements autonomously or load files containing queries to query.

(3) *Data Statistics*. It supports simple and comprehensive data statistics. The former implements simple statistics of the frequency of consumption, amount of consumption, number of access dormitory and so on. Comprehensive statistics realize the statistical analysis of multiple data.

(4) *Data Presentation*. For statistical results, it supports the ordinary ways to display, such as line chart, histogram, scatter plot and other types of charts and tables. As for the students' activities, it can show them by two ways.

a) *Directed Graph*. It includes static and dynamic illustrations. Static display shows students' activity trajectory by weighted directed graph, which uses edge weight and thickness of line to represent the number of times through the corresponding path. Similar to "step into" debug method of a program, the manual dynamic display allows users to follow the student's trajectory step by step by an interactive way. Instead, the automatically dynamic display method illustrates students' trajectory automatically according to the user-defined time interval.

b) *Sequence Diagram*. It shows the changes of students' location according to the time sequence. Each point of a sequence diagram represents a location where the given student used his/her campus card.

(5) *User-defined Data Mining and Analysis Plugin*. The system provides interfaces to import user-defined plug-ins. Through this way, the users can realize special query and statistics functions.

3. System Design

3.1 Overall Design of System

As we can see, the system adopts *Model-View-Controller (MVC)* model. MVC model achieves the separation of logic, data and user interface, and it integrates the business logic into one component. Therefore, it is not required to rewrite the business logic while improving customized interface and user interaction. The basic process is:

- (1) The controller creates the model and view, and the model is associated with the view.
- (2) The user interacts with the system through the controller to change the state of the model.
- (3) When the state of the model changed, the controller automatically refresh the related view to show query results.

3.2 Metadata Definition

In order to support diverse data sources, the metadata should be given in first. Taking consumption record, accessing dormitory records and library records of campus card as examples, the metadata standards are defined as follows.

- (1) Consumption Record. Name, POS (Point of Sale) number, system code, operation type, balance of the card, account balance, occurrence time, recorded time, account number, account of the other side, number of using card, consumption type, serial number, student number.
- (2) Access Dormitory Record. Date, time, student number, name, location, card number, department, event description.
- (3) Library Record. Number, date, time, card number, card version, the type of card, controller number, read head number, door number, event code, illegal card number, controller number, door ID, work number, department ID, work number version, photo tagging. The first twelve fields are not allowed to empty, the rest of the fields are allowed to be empty.

3.3 Database Design

Converting consumption records, entrance guard records and library borrowing records to entities, the entities are consumption entity, dormitory building entity and library entity.

In order to support the above entities, it need the following entities and their attributes:

- (1) POS entity. Number, IP address, system code, area code, Merc code (Merchant code), POS type, location.
- (2) Student entity. employee number, first name, last name, ID department, work number, traffic password, telephone number, room number, whether to display name, whether to allow the dial, display name, employee version number, department effective time.
- (3) Plug-in entity. Plug-in number, plug-in name, plug-in path, plug-in package name, and class name.

4. Key Technologies of Systems

4.1 Data Integration and Query

The source data may be raw or wrong, so the system use the *Extract-Transform-Load (ETL)* method to integrate a given data source into the system. The main techniques include:

- (1) Different procedures are used to import different data sources into this system. Formatted text files can be direct imported into database using SQL statements. Excel files should be read into list, and then are imported to the database. XML files are imported into database after being traversed by the data in each node. The other database exported files should be used character stream to parse, and then, extract SQL statement and use `addBatch` operation method to realize the bulk insert.
- (2) Some views of the data in database should be constructed to improve the speed of query in the cascade query. These views include `Iodetails`, `PaymentAndPos`, `PositionChanged`, etc.
- (3) For analyzing students' behavior trajectory, the semantics of consumption locations are not clear. Hence, the consumption locations are classified into restaurants, hospitals, Internet bars, supermarkets, water room and other places according to their properties. User can view students' track in the some locations where they want to see, and ignore the meaningless sites for them. It is easy for user to analyze students' behavior.

4.2 Display of Statistical Analysis Using Directed Graph

Based on the campus card data, a student's trajectories during a certain period can be acquired. There are three kinds of display directed graphs: summary display, interactive single step display and automatic single step display.

To realize this function, the observer pattern is used to achieve a one-to-many dependency between objects. The class `BTGraphY` is observed, and observers include `BTView`, `BTNext`, `BTDynamic` and so on. The method "GraphArray" of class `BTGraphY` obtains the results by querying the database, then the method "NotifyGraphBT" notifies its boservers. The observers automatically update their views in next. This design pattern is easy to extend new function to reduce the times of data query. The class diagram is shown in *Fig. 1*.

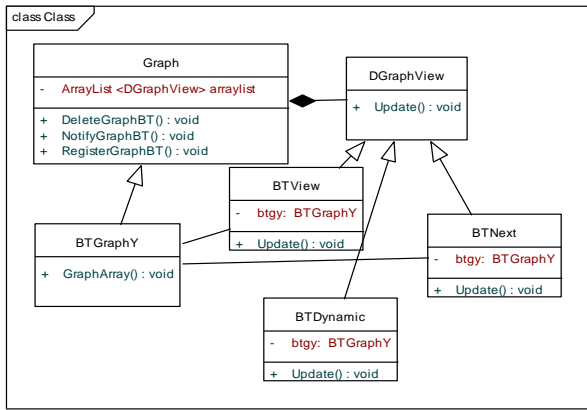


Figure 1. Class Diagram

Besides, we adopt the following methods to show the directed graph. The directed graph is shown Fig. 2.

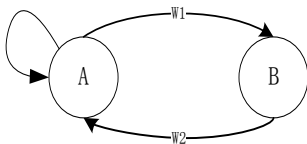


Figure 2. Directed Graph

(1) The point *A* and *B* represent location where student used campus card. The edge from *A* to *B* (or from *B* to *A*) represents student moved from position *A* to position *B* (or from position *B* to position *A*).

(2) Edge weight. The weight w_1 (or w_2) is the number of times of a student's going through the corresponding path.

(3) Edge thickness. We set an upper and lower bounds for the thickness of edges. The Edge thickness is given by equation

$$ET_i = \frac{(W_{max} - W_{min} + 1)}{(W_n - W_0 + 1)} \times (W_i - W_{min}) + W_0 \quad (1).$$

ET_i is the thickness of edge i . W_{max} is the maximum of weight. W_{min} is the minimum of weight. W_n is the upper bounds for the thickness of edges. W_0 is the lower bounds for the thickness of edges. W_i is the weight of edge i .

4.3 Demonstration

The system can automatically demonstrate not only the trajectory of a single person, but also trajectories of multi-persons. The realization of automatic demonstration needs to add a timer to achieve a regular task. For the single step automatic display of directed graph, the differences of demonstrations between single person and multi-persons are as follows:

(1) For multi-persons' results, each person is allocated to a random color based on the person's ID, and the colors are stored in a hash table to prevent duplication of color.

(2) The color of an edge is the color of the corresponding person.

(3) The results are sorted according to two keys: The first is persons' ID, and the second is record's time.

4.4 User Defined Plug-in

The key to achieve the user defined plug-in accessing are as follows:

(1) Configuration plug-in of main program: It uses plug-in ID, plug-in name, plug-in address, package name and class name in database.

(2) Instantiate object of plug-in: Dynamic loading of Java is mainly achieved by ClassLoader. Java has provided a ready-made ClassLoader to implement class URLClassLoader, which can load the jar specified by URL.

This system provides the interface of basic data integration, query and statistics, so that the user can integrate the data mining analysis plug-in into system.

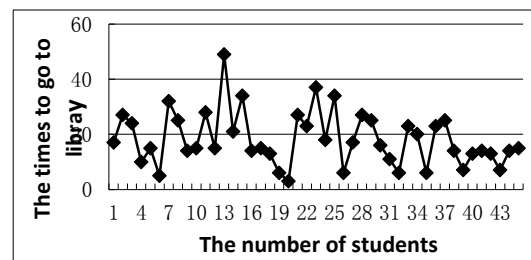
5. Implementation and Application

We can analyze the relationship between student's achievement and the number of times of accessing library, using the records of the second semester in 2012-2013 school year. The students are selected from different classes of different professions, and half of them are the top 30% in the class and the remains are the bottom 30%. Fig.3 shows the query result of the library records of 90 students, and we can draw the following conclusions based on these results:

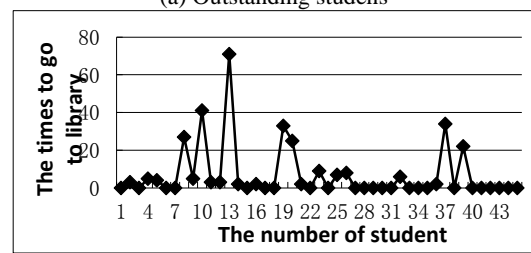
(1) Generally speaking, more than 82% excellent students go to the library more frequently than the others. However, there also a few excellent students rarely go to the library. Numeric result.

(2) More than 84% poor students barely even went to library. However the times of 15.5% poor students to go to library is more than 20 times. There also exists poor students who like to study in the library, but the number of them is smaller.

(3) Overall speaking, There not many students frequently go to the library whether poor students or eugenics.



(a) Outstanding students



(b) poor performance students

Figure 3. Statistical results of library entry and exit of outstanding students

6. Conclusion and Future work

This paper presents a comprehensive introduction of the design and implementation of a Data Integration and Analysis System for Campus Card. The system is designed based on UML (unified modeling language) and is implemented using Java. We adopt MySQL as our database. The main functions of the system include data integration, data statistics, data query, data mining and data analysis. Compared with the other similar software, the proposed system has the following advantages:

(1) Our focus is the campus card data, so it can meet the

requirements of universities better than the other general data integration system; (2) The system uses dynamic graphs to show the results, which are more vividly and easily to acquire model for the behavior of the students; (3) With classified and inductive data, we can get the student's behavior regularity more easily; and (4) It has better scalability with the open interfaces to allow users to import their own plugins.

However, this system can be improved at least in the following aspects: (1) we plan to provide more open interfaces to import plugins based on the other programming languages besides Java; and (2) we are trying to build the behavior model of students with cluster analysis and analyze their typical behavior patterns.

7. Reference

- [1] Bowman, M., Debray, S. K., and Peterson, L. L. 1993. Reasoning about naming systems. [1] W. Zhou, C. Chen, S. Li: Experimental technology and management, 2011, 28(4): 246-249.
- [2] S. Li, W. Zhou, Y. Mi: Experimental technology and management, 2011, 28(6): 7-10.
- [3] R. Hu: Research in laboratory, 2012, 31(7): 404-420
- [4] X. Dong, H. Zhang: Computer systems & applications, 2014, 23(1):158-161.
- [5] H. Xu, B. Sun, J. Gao, et al, in: Proc. Int. Conf. Comp. Sci. Sports Engin., 2010: 583-586.
- [6] N. Jiang, W. Xu: Microcomputer Applications, 2015, 31(2):35-38.
- [7] F. Chen: The Chinese Journal of ICT in Education, 2012, 9: 47-49.
- [8] X. Fei, X. Dong, H. Zhang: Computer Knowledge and Technology, 2014, 10(20): 4934-4936.
- [9] S. LI: Railway computer application, 2010, 19(6) : 55-58.
- [10] P. Jin, X. Wang: Journal of East China Normal University (Natural Science), 2015, S1:525-529.