

Software Development of Automatic Data Collector for Bus Route Planning System

Adam Hendra Brata***, Deron Liang*, Sholeh Hadi Pramono**

* Department of Computer Science and Information Engineering, National Central University, Taiwan

** Departement of Electrical Engineering, University of Brawijaya, Indonesia

Article Info

Article history:

Received Oct 3, 2014

Revised Nov 23, 2014

Accepted Dec 12, 2014

Keyword:

Automated system

Batch processing

Bus route planning

Data collector

Software engineering

ABSTRACT

Public transportation is important issue in Taiwan. Recently, mobile application named Bus Route Planning was developed to help the user to get information about public transportation using bus. But, this application often gave the user inaccurate bus information and this application has less attractive GUI. To overcome those 2 problems, it needed 2 kinds of solutions. First, a more accurate time prediction algorithm is needed to predict the arrival time of bus. Second, augmented reality technology can be used to make a GUI improvement. In this research, Automatic Data Collector system was proposed to give support for those 2 solutions at once. This proposed system has 3 main functionalities. First, data collector function to provide some data sets that can be further analyzed as an base of time prediction algorithm. Second, data updater functions to provide the most updated bus information for used in augmented reality system. Third, data management function to gave the system better functionality to supported those 2 related systems. This proposed Automatic Data Collector system was developed using batch data processing scenario and SQL native query in Java programming language. The result of testing shown this data processing scenario was very effective to made database manipulation especially for large-sized data.

Copyright © 2015 Institute of Advanced Engineering and Science.
All rights reserved.

Corresponding Author:

Adam Hendra Brata,
Department of Computer Science and Information Engineering,
National Central University,
No. 300, Jhongda Rd., Jhongli City, Taoyuan County 32001, Taiwan, ROC.
Department of Electrical Engineering,
University of Brawijaya,
Jl. Veteran, Malang 65145, East Java, Indonesia.
Email: mizuno.tatsuya.mx@gmail.com

1. INTRODUCTION

Bus rapid transit is one of important public transport modes in Taiwan as well as in capital city of Taiwan, Taipei [1]. In Taipei, many applications that have been developed to assist the passengers who want to travel with public transport in Taipei city [2], [3], but these current applications typically still have some drawbacks like inaccurate informations and have less attractive GUI. These 2 mentioned problems can be solve with better algorithm to make sure the information always accurate [5] and better GUI improvement techniques, like using augmented reality technology [6].

These 2 proposed solutions basically need 1 system to provide them some useful data for further processing. This mentioned application basically has a main function as an server that able to provide data which will be used by them. This system must be able to worked as a single server with multi functionalities to provide them a facility that can be used for those 2 mentioned proposed solutions at once.

Previously, there was a system that have been developed by Software Methodology Laboratory (SML), National Central University, Taiwan which has similar function with the proposed system. This system had a function as a single data updater to provided the updated information to Bus Route Planner mobile application which was also developed by SML. This system used real-time processing scenario to processing the data, but this processing scenario had a drawback for processing the large-sized data [4].

In this paper, automatic data collector system was proposed to gave all-in-one solution from those mentioned problems with effective strategies to processing the data and to manipulates data in database, especially for large-sized data.

2. RESEARCH METHOD

2.1. Automatic Data Collector

The software that will be developed in this research is Automatic Data Collector for bus route planning system. This system has 3 main functionalities, namely Data collector, Data updater and Data management.

Data collector has a function to store the data that have been obtained from various data sources into database frequently. Use case diagram for the Data Collector functionality is shown in Figure 1. Data collector has a function to store the data that have been obtained from various data sources into database frequently. Use case diagram for the Data Collector functionality is shown in Figure 1.

Data updater has a function to make data in database stay updated with the newest data that provided by Taipei Bus API. Use case diagram for the Data Updater functionality is shown in Figure 2.

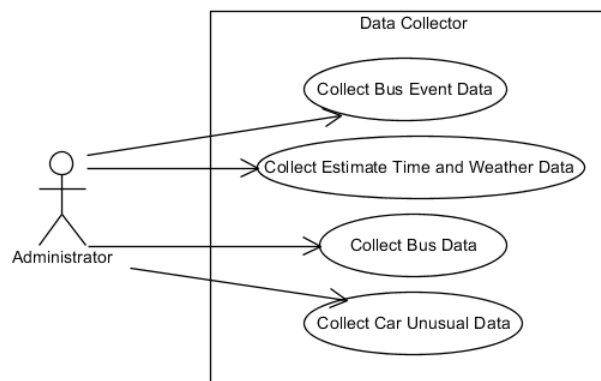


Figure 1. Use case diagram for data collector functionality

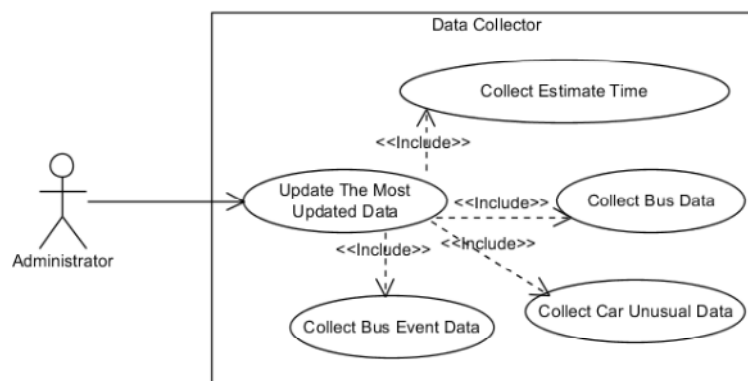


Figure 2. Use case diagram for data updater functionality

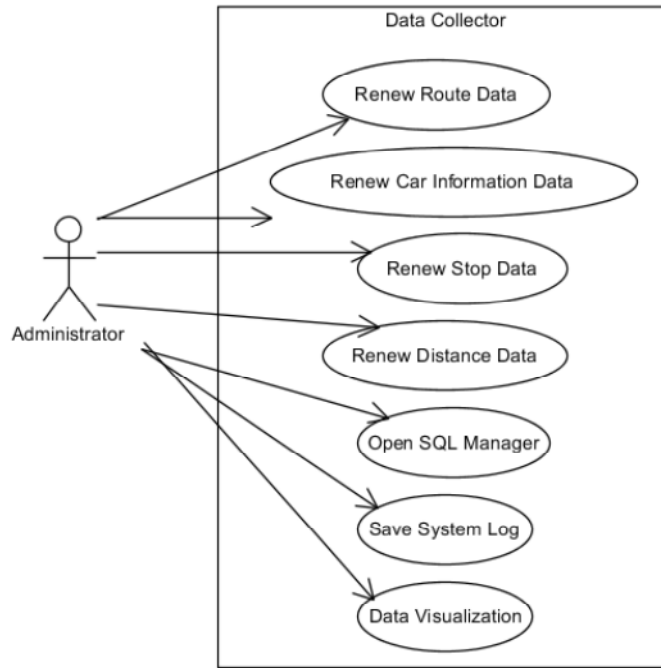


Figure 3. Use case diagram for data management functionality

Data management has a function to perform data management activities to make Administrator easier to manage a whole data that used in this system. Use case diagram for the Data Management functionality is shown in Figure 3.

2.2. Software Design

Software architecture that used in this application actually a simplification of MVC design pattern. This system architecture had an main idea to merge Controller-typed class with Model-typed class into 1 single Core-typed class. This several core classes were designed based on their controlling function based on use case analysis. This application adopt batch processing to do database manipulation. This scenario have an objective to accelerate the processing time and to reduce insertion time. Automatic Data Collector system has a data processing scheme shown in Figure 4. Data processing scheme is divided by 3 main patterns based on Automatic Data Collector system functionalities. Figure 5 shows class diagram of proposed system.

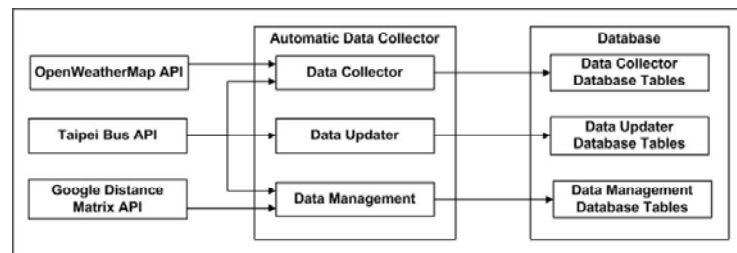


Figure 4. Data processing scheme

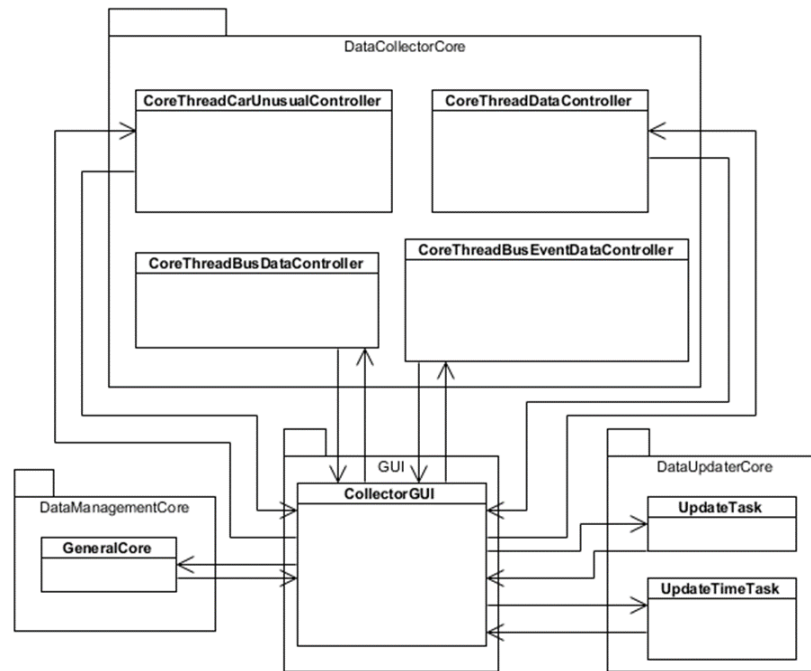


Figure 5. Class diagram

2.3. Software Implementation

2.3.1 Java XML Parser

There are 3 basic methods to parse XML in Java, SAX, StAX and DOM. DOM is good for back-and-forth data access. Zao and Bhuyan said, DOM is better for complex and frequent XML parsing. SAX and StAX are appropriate for applications with extremely restrictive memory but not for back and forth access or modification [7]. Nicola and John said that DOM is the most suitable for database applications, while SAX and StAX are more appropriate for streaming applications [8]. This proposed system used DOM model as XML parser, because this type is suitable for database application.

2.3.2 Java Batch Processing and JDBC Native Query

Batch processing is have a goal to process a large set of data in a specific way, automatically, without needing any user intervention. The data is first collected, during a work day, for example, and then batch-processed, so all the collected data is processed in one go [9]. In Java, batch processing function can be used with calling the `addBatch()` and `executeBatch()` methods for `Statement` or `PreparedStatement` objects using JDBC [10]. Dwyer said batch processing is improving the efficiency of transaction processing systems especially if dealing with big data [11]. JDBC is designed to allow Java user manipulates SQL database from Java program with simple, fast and efficient way with using standard native SQL query.

2.3.3 Graphical User Interface Implementation

The graphical user interface for this application consists of 2 main pages, Data Collector and Data Management. These 2 pages represent 3 main functionalities of Automatic Data Collector system. Figure 6 shows the graphical user interface for Data Collector page. Figure 7 shows the graphical user interface for Data Management page.

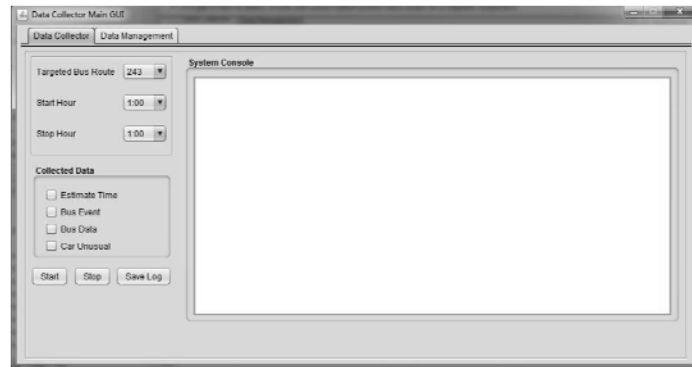


Figure 6. GUI for data collector page

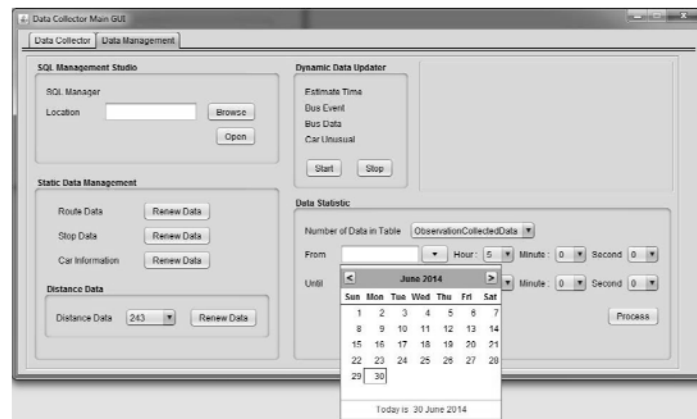


Figure 7. GUI for data management page

2.4. Software Testing

2.4.1 Data Collector Testing

This testing has objective to know the performance of the proposed system's data collector capability.

- Run proposed system in 3 days, 17 hours collecting time (6 a.m – 11 p.m).
- The targeted bus route is bus route number 243, with 53 bus stops (25 bus stops for go bus and 28 bus stops for back bus).
- Measure the number of data that collected by the proposed system.
- The compared collected data is taken from one of database tables which has the biggest amount of data among the others.
- This testing use this simple calculation to give a performance standart :
 - System will grab the new data from API in every 25 seconds.
 - Testing assumptions :
 - Total data in 1 day = $((17 \times 60 \times 60) / 25) \times 53 = 129744$ data
 - Make a comparison between the data that collected by the proposed system and the total data standart which have been determined before.
 - Testing done by ignoring all the other processes running on the computer and other factors relating to the performance of the system.

2.4.2 Processing Time Testing

This testing has objective to measure how much the time that needed for the proposed system to process some data and make a database manipulation to database server. This test will compare the proposed system with the current server system that previously developed by former public transportation team of

Software Methodology Laboratory [4]. This testing will measure the total processing time, since the database manipulation method was called in the first time until this method produce the output.

- Running the programs to execute the query multiple times with different number of queries.
 - 10 queries
 - 100 queries
 - 1000 queries
 - 10000 queries
 - 100000 queries
- Measure the processing time of the proposed system and the current system to know the processing time and compare the result.
- Testing done by ignoring all the other processes running on the computer and other factors relating to the performance of the system.

3. RESULTS AND ANALYSIS

3.1. Result of Data Collector Testing

This testing stage has intention to measure the performance of the proposed system’s data collecting capability. The testing result will be shown in Table 1 and Figure 8. The usage of simpler architecture with little loops and branches made the complexity of system architecture was very low. The usage of SQL native query to make an database operation and the usage of batch processing scenario made the the system ran faster. Although overall system performance was good, but the proposed system still had little data losses. This data losses occured because sometimes connection from system to data provider has timed out since number of request to API has a limit. Beside of that, this data losses also occured because system connection to database server has timed out after several hours and made the system failed in 1 or more iterations. In the future work, connection pooling can be used to made better database connection, so the risk of connection time out can be reduced. URL encoding technique also can be used to solve the limitation of API request problem. The percentage of data that successfully collected by the proposed sytem shown in Figure 9.

Table 1. Data collector testing result

Day	Colleted data (record)	Data standard (record)
1	126790	129744
2	127041	129744
3	123636	129744

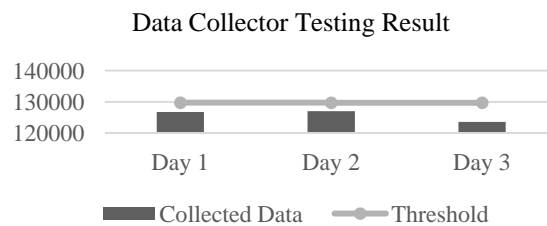


Figure 8. Data collector testing result

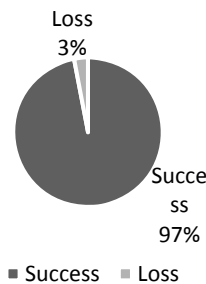


Figure 9. Percentage of collected data

3.2. Result of Processing Time Testing

Testing result for processing time testing shown in Table 2 and Figure 10.

Number of queries	Current system (milisecond)	Proposes system (milisecond)
10	640	421
100	813	577
1000	2031	950
10000	31784	21116
100000	274340	211585

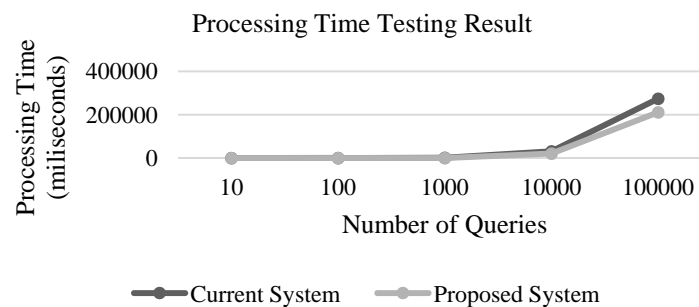


Figure 10. Processing time testing result

From this testing known that the proposed system had a better performance to do database operations than the current system. The proposed system using batch processing scenario and the current system using real-time processing scenario. This batch processing scenario is effective to dealing with the data that have huge quantity, because with batch processing the system can done database processing in a bulk and execute it in 1 time, so only need a little time to execute this database processing.

4. CONCLUSION

This research dealt with the problem of developing automatic data collector system to gave all-in-one support solution with effective strategies to processing the data and to manipulates data in database. From software testing result was known that all of the proposed system functionalities work well, and from the data collecting testing was known the performance report of the proposed system's data collecting capability. In the next testing, the proposed system and the current server application were compared to proof the proposed system was better than the current server application. The proposed system was much faster than the current server application, it was because the proposed system was use batch processing scenario and on the contrary the current server application use real-time processing.

ACKNOWLEDGEMENTS

The authors would like to thank Komang Candra Brata, Lutfi Fanani and all members of Software Methodology Laboratory of National Central University Taiwan for insightful discussions during this research. We would also like to thank the Planning and Cooperation of Foreign Affairs of Ministry of Education and Culture of Indonesia for the International Dual Degree scholarship so that this research could be done.

REFERENCES

- [1] Lan, Lawrence W., Ming-Te Wang and April Y. Kuo, "Development and deployment of public transport policy and planning in Taiwan", *Springer Journal on Transportation*, Volume 33, Issue 2, pp 153-170, 2006.
- [2] Taiwan Ministry of Transportation (MOTC), "Fun Travel in Taipei", 2011. Taken from : <http://english.dot.taipei.gov.tw/ct.asp?xItem=51165215&ctNode=65619&mp=117002> , Last Access : May 7, 2014.

- [3] Software Methodology Laboratory (SML), "Bus Route Planner Mobile Application", National Central University, *Private Communication*, Taiwan, 2013.
- [4] Software Methodology Laboratory (SML), "Bus Route Planner Server Application", National Central University, *Private Communication*, Taiwan, 2013.
- [5] Fanani, Lutfi, "Bus Arrival Prediction – to Ensure Users Not to Miss the Bus (Preliminary Study based on Bus Line 243 Taipei)", National Central University, *Master Thesis*, Taiwan, 2014.
- [6] Brata, Komang Candra, "Developing Location-Based Augmented Reality Information Support for Bus Route Planning System", National Central University, *Master Thesis*, Taiwan, 2014.
- [7] L. Zhao and L. Bhuyan, "Performance Evaluation and Acceleration for XML Data Parsing", *Proc. 9th Workshop Computer Architecture Evaluation Using Commercial Workloads (CAECW 06)*, Texas, USA, 2006.
- [8] Nicola, Mathias and Jasmi John, "XML Parsing: A Threat to Database Performance", *Proc. 12th Int'l Conf. Information and Knowledge Management (CIKM 03)*, pp. 175-178, ACM Press, New York, USA, 2003.
- [9] Walker, Michael, "Batch vs. Real Time Data Processing", 2013. Taken from : <http://www.datasciencecentral.com/profiles/blogs/batch-vs-real-time-data-processing> , Last Access : May 7, 2014.
- [10] Oracle, "Oracle Times Ten In-Memory Database Java Developer's Guide", 2014. Taken from : http://download.oracle.com/otn_hosted_doc/timesten/1122/doc/timesten.1122/e21638/tuning.htm#TTJDV251, Last Access : May 7, 2014.
- [11] Dwyer, Barry, "The Automatic Design of Batch Processing Systems", University of Adelaide, *Doctoral Thesis*, Australia, 1999.

BIOGRAPHIES OF AUTHORS



Adam Hendra Brata is an International Dual Degree Master student between University of Brawijaya, Indonesia and National Central University, Taiwan. He completed his Bachelor degree in Department of Informatics Engineering, University of Brawijaya, Indonesia. His research interest area is in the computer science and information technology areas, especially in software engineering, augmented reality and game development.



Deron Liang is an Professor at the Department of Computer Science and Information Engineering, National Central University, Taiwan. He got Ph.D degree from University of Maryland, USA. His research interest are in the distributed systems, fault tolerant, security auditing, and object oriented. He presently work as Head of Software Methodology Laboratory in Department of Computer Science and Information Engineering, National Central University, Taiwan.



Sholeh Hadi Pramono is an senior lecturer in Department of Electrical Engineering, University of Brawijaya, Indonesia. He is expert in Electrical Engineering research area, especially in optical telecommunication and technology of antenna. He got Doctor degree from University of Indonesia, Indonesia. He presently work in Telecommunication Laboratory, Department of Electrical Engineering, University of Brawijaya, Indonesia as an optical telecommunication specialist.