Analysis of Information System Operation Cost Based on Working Time and Unit Cost

Masateru Tsunoda
Nara Institute of Science and Technology
Nara, Japan
Department of Informatics, Kindai University
Osaka, Japan
tsunoda@info.kindai.ac.jp

Akito Monden
Graduate School of Natural Science and Technology
Okayama University
Okayama, Japan
monden@okayama-u.ac.jp

Kenichi Matsumoto
Graduate School of Information Science
Nara Institute of Science and Technology
Nara, Japan
matumoto@is.naist.jp

Sawako Ohiwa, Tomoki Oshino
Economic Research Institute
Economic Research Association
Tokyo, Japan
{er421, er352}@zai-keicho.or.jp

***Abstract*—Recently, information system operation becomes more important because of increasing size of information system and outsourcing the system operation. However, it is not easy for customers to judge the validity of the system operation cost. To provide information which helps the judgment, we analyzed factors which affect system operation cost. Working time of system operation service provider has the strong relationship to the cost. So, if customers know the working time, they estimate the cost properly. However, it is difficult for customers to know the working time generally. So, we assumed that customers estimate unit cost and working time, to speculate total operation cost roughly. To help the estimation, we analyzed factors affected working time and unit cost. The analysis results show that working time is settled based on the software size and the number of users, and the unit cost of the engineers increases when network range of the system is wide.**

Keywords—benchmarking; it operations; work efficiency; regression analysis

I. INTRODUCTION

Recently, size of information system gets large, and the information system operation is often outsourced. Given the situation, the information system operation becomes more important. The information system consists of computers, network and software. In the information system operation, operators manage the computers and the network, e.g., they replace updated software and recover the system when a system failure occurs. To operate the system, standardization of system operation process such as ITIL (Information Technology Infrastructure Library) [1] and ISO20000 get more attention.

Cost is one of the important factors when the system operation is outsourced. However, it is not easy for the customers (system users) to judge whether the operation cost is valid or not. This is because generally, customers do not know the operation cost in other customers' companies, and therefore they cannot compare the cost with others to judge the validity of the cost. The goal of our study is for customers to support the judgement of the validity of the cost.

As explained in the section 2, the relationship between the operation cost and working time of operation service supplier (outsourced operators) was very strong. So, if the working time of the supplier is known, operation cost can be estimated, and the estimated cost can be used to judge the validity of the actual cost. However, it is not easy for the customers to know the working time of the supplier. So, we focus on information which the customers can know (e.g., size of software), to estimate the working time. In the analysis, first, we clarified the factors which affect the working time. Then, data was stratified based on the factors, to show the benchmark of working time. Using the benchmark, customers estimate the working time roughly.

To estimate the operation cost, the estimated working time only is insufficient. The unit cost of outsourced operators is also needed. That is, operation cost can be estimated, by estimated working time multiplied by the estimated unit cost. To estimate the unit cost roughly, we clarified factors which affect the unit cost, and showed the benchmark of the unit cost, stratifying the dataset based on the factors.

Major contribution of the study is not benchmarking work efficiency of system operation, but benchmarking the system operation cost. Although the former is useful for the service suppliers, it is not utilized for the customers to judge the validity of the cost. In contrast, the latter is useful for the judgement. As long as we know, there is no study which analyzed system operation cost for the customers. In our previous study [5], we analyzed factors which affects the system operation cost. In the study, we regarded that the cost is derived by the number of operators, the number of computers, and the unit cost of outsourced operators. This is because in the dataset used in [5], the working time was not recorded in most data points. Instead of the time, we used the number of operators and the number of computers. The analysis of this paper is more proper than the previous study, since we used the working time directly.

The dataset used in the study was collected in 2011 and 2013 by Economic Research Association. It includes 179 data points collected from both customers and operation service suppliers. The size of the systems recorded in the dataset is various (i.e., both small systems' data and large systems' data are included). Therefore, the analysis results derived from the dataset would have generality to some extent.

II. Relationship between Operation Cost and Working Time

In the analysis, we clarified factors which decide the system operation cost. The system operation cost means the total cost required to the operation in a year. Note that the cost means the contract price, and not the real cost in the service supplier. Most of the system operation cost is considered to be based on the labor cost of the operators. In other words, the cost is assumed to relate strongly to the working time of the operators of the service supplier. Note that the system is operated by not only the service supplier but also the customer. So, working time of the service supplier should be distinguished from the working time of the customer, when analyzing the operation cost.

We analyzed the relationship between the operation cost and the working time of the service supplier, using the regression analysis. In the regression analysis, the dependent variable was the operation cost, and the independent variable was the working time of the service supplier. Before building the model, log-transformation was applied to the variables. As a result, the built model of R^2 was 0.91. This means the operation cost mostly depends on the labor cost. In other words, if the customers know the working time of the service supplier, the operation cost can be estimated.

As explained above, the cost is the contract price, and the price includes the profit and expenses of the service supplier. They are one of the factors which affect the operation cost. Also, the unit cost of the operators of the service supplier is one of the factors. However, their influence is not very strong, since the model of R^2 was very large (0.91).

Fig. 1 shows the relationship between the operation cost and the working time of the service supplier. To avoid abusing the result, we do not show the operation cost on the vertical axis. Although there is two outliers at the upper left of the figure (their cost is higher than others, considering the working time), the relationship between the cost and the time was strong in other data points. The result is not always natural in software engineering field. For example, in the software maintenance activity, the relationship between the maintenance cost and the working time was relatively weak [6].

III. Factors Affecting Working Time

In the followings, the sum of working time of the customers and the service supplier in a year is called as total working time. As explained in section II, the information system is operated by the customers, not only the service

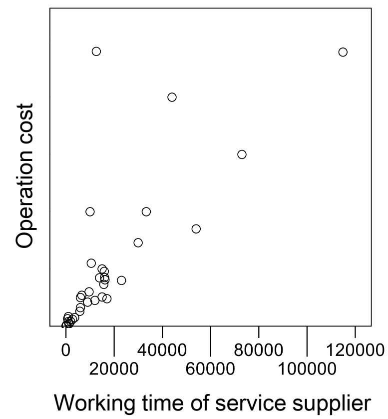

Fig. 1 Relationship between the operation cost and service supplier's working time per year.

suppliers. The ratio of the operators of the customers to the service suppliers was not same in each data point. So, using working time of the service supplier is not appropriate in the analysis. Instead of it, we used the total working time in the following analysis.

To benchmark the total working time, we defined work efficiency as the number of programs divided by the total working time. In the next subsection, we explain why the number of programs is used in the definition. In the following analysis, we show relationship between the work efficiency and other factors using box plots. To estimate the working time roughly, customers multiply the work efficiency shown in the boxplot by the number of programs.

We used multiple regression analysis, to analyze factors affecting the total working time. Before building the model, all variables of the ratio scale were log-transformed. In the model, the dependent variable was the total working time.

A. Relationship to System Size

As explained in section I, main activities of the system operation is to manage computers and network on the system. So, the system size such as software size and the number of computers are considered to be main factor of the total working time. As software size, we used the number of programs, instead of the functional size of software on the system. Since in the number of programs, the number of missing values was relatively small. The dataset has various variables such as the number of Windows servers and the number of mainframes. Based on the preliminarily analysis, we used the number of Windows servers as hardware size of the system.

When the number of the users increases, answering time to the users may also increases. So, we also analyzed the maximum number of users on the system. Using the multiple linear regression analysis, we analyzed the influence of the factors to the total working time. In the model, independent variables were the number of programs, the number of Windows servers, and the maximum number of the users. We applied variable selection to build the regression model.

Table I shows standardized partial regression coefficients in the built model. The number of Windows servers was removed

TABLE I. REGRESSION MODEL USING SYSTEM SIZE

Variable	Std. partial reg. coefficient.	p-value
Maximum number of users	0.41	0.0%
Number of programs	0.49	0.0%

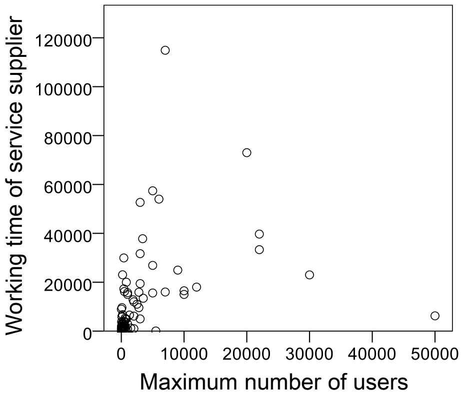

Fig. 2 Relationship between the maximum number of users and service supplier's working time.

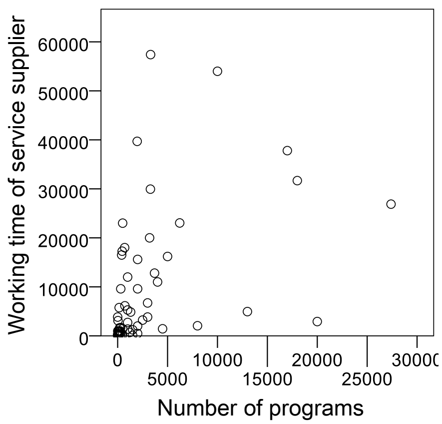

Fig. 3 Relationship between the number of programs and service supplier's working time.

from the model by the variable selection. This means the number of Windows servers is not indispensable to estimate the total working time. The partial regression coefficient of the number of programs was larger than the maximum number of the users. So, we used the number of programs to define the work efficiency. Note that the latter was not small, and hence the influence of the maximum number of the users was not ignorable. Hence, we also used the maximum number of the users to build regression models in the followings.

Adjusted R^2 was 0.53, and this means that the total working time is explained by the number of programs and the maximum number of the users, since Adjusted R^2 was larger than 0.5. Fig. 2 and Fig. 3 show the relationships between total working time and the variables. In the figures, data points are scattered, and these suggest the estimation is difficult, using the number of programs or the maximum number of the users only. So, in the following analysis, we focused on other factors which were considered to affect the total working time.

TABLE II. REGRESSION MODEL USING PROCESS STANDARDIZATION

Variable	Std. partial reg. coefficient.	p-value
Maximum number of users	0.47	0.0%
Number of programs	0.47	0.0%
Process Standardization	-0.16	9.3%

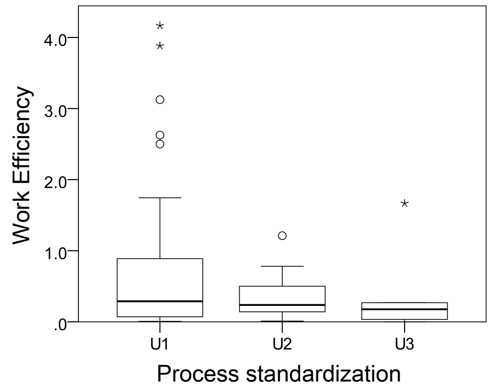

Fig. 4 Relationship between process standardization and work efficiency.

B. *Relationship to Process Standardization*

In the system operation, the standard procedure of the operation is sometimes defined, to enhance the efficiency of the operation and to suppress system trouble. We assumed that when the operation process is standardized, the work efficiency is improved, and it affects the total working time.

We applied multiple regression analysis, to clarify the relationship between process standardization and the total working time. In the analysis, independent variables were the number of programs, the maximum number of the users, and the extent of the process standardization. The extent is signified by three levels. Level U1 means process is standardized, level U2 means process standardization is work in progress, and level U3 means process is not standardized. We made a dummy variable which denotes the process standardization is level U1 or not, because the distribution of level U1 was different from other levels (see Fig. 4).

Standardized partial regression coefficients of the model are shown in Table II. Adjusted R^2 of the model was 0.54, and it was almost same as the model without the process standardization built in section III.A, although the regression coefficients of the process standardization was not very small. P-value of the partial regression coefficient of the process standardization was smaller than 10%. Therefore, the process standardization has the relationship to the total working time. The regression coefficient of the process standardization was negative. So, when the process is standardized (i.e., level U1), the total working time decreases, and work efficiency becomes high.

Fig. 4 shows the relationship between the work efficiency and the process standardization. In the figure, there were many data points whose work efficiency were high in level U1, although the variance of level U1 was large. Based on the boxplot and the regression analysis, the process standardization is regarded to have the relationships to the work efficiency.

C. *Relationship to Social Impact*

Some information systems are used as the infrastructure of the society. If some trouble occurs on the system, many social activities are affected by the trouble. We call the influence as the social impact of the system. When an information system has high social impact, the system should be operated carefully. It may need more working time than the system which has small social impact.

To analyze the influence of the social impact to the total working time, we applied multiple regression analysis. In the analysis, independent variables were the number of programs, the maximum number of the users, and the extent of the social impact of the system. The social impact is denoted by three levels. Level C1 denotes the social impact of the system is almost nothing, level C2 denotes the system has the social impact, but it is limited, and level C3 denotes the social impact is very large. We made a dummy variable which denotes the social impact is level C1 or not, because level C3, the number of data points was few.

Table III shows standardized partial regression coefficients of the model. Adjusted R^2 of the model was 0.54, and it was almost same as the model built in section III.A. However, the regression coefficients of the process standardization was not very small. P-value of the coefficient of the social impact was relatively small, although it was larger than 10%. Therefore, it is probable that the social impact has the relationship to the total working time. The regression coefficient of the social impact was negative. Hence, when the social impact is small (i.e., level C1), the total working time decreases, and the work efficiency becomes higher.

The relationship between the work efficiency and the social impact is shown in Fig. 5. In the level C3, the number of data points was two. So, we disregard level C3. Although the median of level C1 and C2 is almost same, variance of C1 was large, and in level C2, there were few data points whose work efficiency were high. That is, when the social impact is not small, work efficiency tends to be lower.

D. *Analysis of Working Time Considering Multiple Factors*

If the process standardization and the social impact has the relationship, one strongly relates to the work efficiency, but the other may not relate very much. To identify the major factors of the total working time, we built a multiple regression model. In the model, independent variables were the number of programs, the maximum number of the users, the process standardization, and the social impact.

Variable selection was applied to build the model. As a result, the social impact was removed from the model. This means the total working time is explained by the number of programs, the maximum number of the users, the process standardization. So, the relationship between the total working time and the social impact was weaker than others. Note that we do not show the built model because it is same as the model built in section III.B.

TABLE III. REGRESSION MODEL USING SOCIAL IMPACT

Variable	Std. partial reg. coefficient.	p-value
Maximum number of users	0.37	0.0%
Number of programs	0.44	0.0%
Social impact	-0.16	10.6%

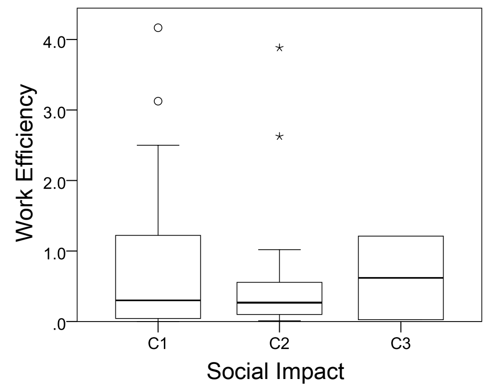

Fig. 5 Relationship between social impact and work efficiency.

IV. FACTORS AFFECTING UNIT COST

To estimate the operation cost, the estimated working time only is insufficient. The unit cost of outsourced operators is also needed, as explained in section I. We defined the unit cost of operators as the total cost divided by the working time of service supplier. If there is a factor which affects the operation cost, in addition to the working time, the factor is considered to affect the unit cost. So, we analyzed factors affecting the operation cost, using multiple regression analysis. In the model, the dependent variable was the operation cost, and all variables of the ratio scale were log-transformed.

A. *Relationship to Network range*

Information system sometimes connects to other systems using the network. When the network is complex, system operators who have high network skills are needed to operate the system, and it may increase the unit cost of operators. So, we analyzed the relationship between the network range and the operation cost, i.e., the influence of the network range to the unit cost. The network range is signified by five levels. Level N1 means the network is between customers and other organizations, level N2 means it is between customers and other branch offices, level N3 means it is within a branch office, level N4 means it is within a department, and level N5 means process the system does not connect to other system.

In the multiple regression analysis, independent variables were the working time of service supplier and the network range. We treated the network range as ordinal scale, considering the relationship between the unit cost and the network range (see Fig. 6). We replaces level N5 as level N3, since the number of the data point was one, and there is no data point whose network range was level N4.

Standardized partial regression coefficients of the model are shown in Table IV. Adjusted R^2 of the model was 0.92. P-value of the partial regression coefficient of the process

TABLE IV. REGRESSION MODEL USING NETWORK RANGE

Variable	Std. partial reg. coefficient.	p-value
Working time of service supplier	0.95	0.0%
Network range	-0.12	2.0%

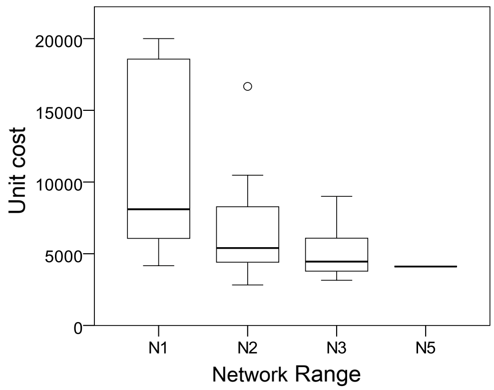

Fig. 6 Relationship between network range and unit cost.

standardization was smaller than 5%. Therefore, the network range has the relationship to the operation cost. The regression coefficient of the process standardization was negative. So, when the network range is narrow (e.g., level N3), the operation cost decreases, and the unit cost becomes low. Fig. 6 shows the relationship between the network range and the unit cost. As explained above, the number of the data point of level N1 was one, and there is no data point whose network range was level N4. Currency unit of the vertical axis is Japanese yen. The figure shows also when the network range is narrower, the unit cost is lower.

B. Relationship to Contract Types

There are some contract types of the system operation. Contract type means the working type of the operators. The contract type has four types. Type K1 is dispatched workers, type K2 is quasi-mandate workers, K3 is contract workers, type K4 is mixing quasi-mandate and contract workers, and type K5 is others. When the contract type is different, the unit cost of the operators may also different.

To clarify the relationship between the contract type and the operation cost, we applied multiple regression analysis. In the analysis, independent variables were the working time of service supplier and the contract type. We made dummy variables which denote the contract type. The dummy variables of type K5 was not used, because the number of data points of them was one. Note that we did not make the dummy variables of type K4, since it can be denoted by the variables of type K1, K2 and K3.

Table V shows standardized partial regression coefficients of the model. Adjusted R^2 of the model was 0.91. P-value of the coefficient of the contract type (type K3) was smaller than 5%. Therefore, the contract type of K3 has the relationship to the operation cost. The regression coefficient of the value was negative. Hence, when the contract type is K3, the operation cost decreases, and the unit cost also becomes low. The relationship between the operation cost and the contract type is shown in Fig. 7. The number of data points of type K1 and K5 was few. So, we disregard type K1 and K5. The figure shows that type K3 is the lowest unit cost.

TABLE V. REGRESSION MODEL USING CONTRACT TYPE

Variable	Std. partial reg. coefficient.	p-value
Working time of service supplier	0.97	0.0%
Contract type K1	-0.07	19.1%
Contract type K2	-0.01	87.9%
Contract type K3	-0.12	4.5%

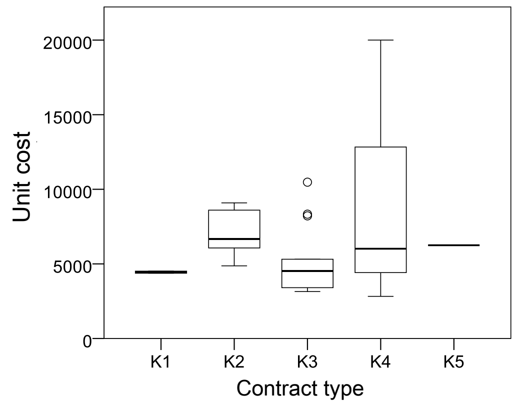

Fig. 7 Relationship between contract type and unit cost.

C. Relationship to SLA

When the system operation is outsourced, SLA (Service Level Agreement) is sometimes made. SLA assures service level of the operation such as system recovery time in advance. While it ensures quality of the system operation, it may also enhance the unit cost of the operators. So, we applied multiple regression analysis to clarify the influence of SLA to the operation cost (i.e., the unit cost). The independent variables were the working time of service supplier and SLA status. SLA status denotes SLA of each data point. Status S1 means SLA is made (goals are guaranteed), status S2 means SLA is made (goals are not guaranteed), and status S2 means SLA is not made. We made dummy variables which denote status S1 and S2.

Standardized partial regression coefficients of the model are shown in Table VI. Adjusted R^2 of the model was 0.91. P-value of the coefficient of the SLA status (status S1) was relatively small, although it was larger than 10%. The regression coefficient was positive. This means when the SLA is made (i.e., status S1), the operation cost increases, and the unit cost becomes high. Fig. 8 shows the relationship between the SLA status and the unit cost. In the figure, although the median of status S3 was lower than others, the distribution was not very different. When SLA is not made, the unit cost may be low. However, the relationship between SLA status and the unit cost was not explicit.

D. Analysis of Operation Cost Considering Multiple Factors

There may be relationships between the network range, the contract type, and SLA status. If there is the relationships, a factor strongly affects the operation cost, but the others may

TABLE VI. REGRESSION MODEL USING SLA

Variable	Std. partial reg. coefficient.	p-value
Working time of service supplier	0.93	0%
SLA status	-0.09	10%

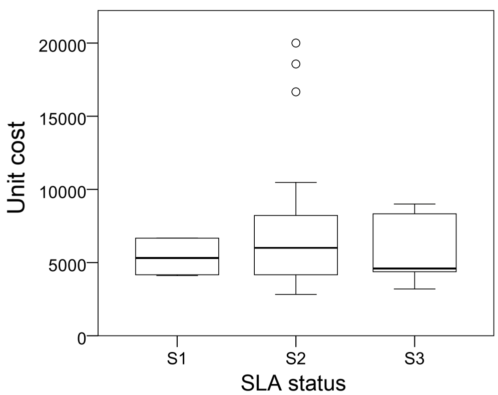

Fig. 8 Relationship between SLA and unit cost.

TABLE VII. REGRESSION MODEL USING NEWWORK RANGE AND CONTRACT TYPE

Variable	Std. partial reg. coefficient.	p-value
Working time of service supplier	0.97	0.0%
Network range	-0.10	4.6%
Contract type K3	-0.09	5.6%

not affect it very much. To clarify the major factors of the operation cost, we applied multiple regression analysis. In the analysis, independent variables were the working time of service supplier, the network range, the contract type, and SLA status.

Table VII shows standardized partial regression coefficients of the model. Variable selection was applied to build the model, and SLA status was removed from the model. So, the relationship between the operation cost and SLA status was weaker than others. On the contrary, the network range and the contract type is not ignorable to estimate the unit cost.

V. RELATED WORK

While there are some studies treating IT operations, they mainly focus on the process of IT operations. For instance, Brooks [2] showed metrics for evaluation of IT operations process, but the metrics are not effective for IT operations cost estimation. Pollard et al. [4] performed case study of the United States and Australian companies' IT operations, and identified some critical success factors (CSF) for successful ITIL implementations. The research also focused on IT operations process, and did not treat cost estimation.

Mannino et al. [3] analyzed efficiency of data warehouse operations by data envelopment analysis (DEA). Although data warehouse operations is one part of IT operations, their results cannot apply to staff or cost estimation.

To our knowledge, there is no study which treats quantitative IT operations staff (cost) estimation. ITIL explains cost estimation based on cost types such as equipment cost or organization (staff) cost [1]. However, it does not tell how to estimate the number of staff quantitatively.

VI. CONCLUSIONS

To support for customers to validate the operation cost of information system. Using a cross-company dataset, we clarified the factors which affect the working time, and the unit cost of the operators. The dataset is stratified based on the factors, to estimate the working time and the unit cost roughly.

- The operation cost is settled based on the working time of operators
- The process standardization relates to the working time of operators.
- The network range and the contract type relates to the unit cost of operators.

Main contribution of the study is to benchmark the system operation cost, and not to benchmark the work efficiency of the system operation. The benchmark of the work efficiency is useful for the service suppliers, but it is not useful for customers in most cases. In contrast, the benchmark of the cost is useful for customers. As long as we know, there is no study which analyzed the system operation cost, considering the customers. One of our future works is to analyze whether the similar tendency is observed or not in the emerging Web-based or cloud-based systems.

ACKNOWLEDGMENT

This research was partially supported by the Japan Ministry of Education, Science, Sports, and Culture [Grant-in-Aid for Scientific Research (C) (No. 25330090)]

REFERENCES

[1] J. Bon, ed.: Foundations of IT Service Management: based on ITIL, Van Haren Publishing, 2005.

[2] P. Brooks, "Metrics for IT Service Management," Van Haren Publishing, 2006.

[3] M. Mannino, S. Hong, and I. Choi, "Efficiency evaluation of data warehouse operations, Decision Support Systems," vol.44, no.4, pp.883-898, 2008.

[4] C. Pollard, and A. Cater-Steel, "Justifications, Strategies, and Critical Success Factors in Successful ITIL Implementations in U.S. and Australian Companies: An Exploratory Study," Information System Management, vol.26, no.2, pp.164-175, 2009.

[5] M. Tsunoda, A. Monden, K. Matsumoto, A. Takahashi, and T. Oshino, "An Empirical Analysis of Information Technology Operations Cost," Proc. of International Workshop on Software Measurement (IWSM/Metrikon/Mensura), pp.571-585, 2010.

[6] M. Tsunoda, A. Monden, K. Matsumoto, S. Ohiwa, and T. Oshino, "Benchmarking Software Maintenance Based on Working Time," In Proc. of Applied Computing and Information Technology (ACIT), pp.20-27, 2015.