Study on Reliability Improvement for the Acquisition of Web Application among Government Agencies

Kwoung Hee Choi
R&D Evaluation Center
Korea Institute of S&T Evaluation and Planning
Seoul City, ROK

Sung Yul Rhew
School of Computing
Soongsil University
Seoul City, ROK

Abstract - Software reliability prediction is a statistical method to put in place a timely software development practice useful for objective assessment of bidders. The current study suggests one research method that enables reliability assessment on such previous projects by studying user satisfaction and the amount of delay in delivery time. If incorporated into the existing acquisition process, the reliability assessment method will further enhance objectivity and accuracy in bidder selection process. GQM(Goal Quality Method) paradigm was used to identify assessment metrics for bidder evaluation and questionnaires were collected from users to create user satisfaction index. In addition, ‘weight of evidence’, the most appropriate categorical method, was used to isolate attributes of each variable that may contribute to reliability assessment. Also, regression analysis was used to come up with reliability metrics model.

Keywords: Software Reliability, Assessment of Bidder, GQM, Weight of Evidence

1 Introduction

The Software Development Plan 2006 for Government Agencies has earmarked government budget for 250 projects. Out of the 250 projects, 96 projects or 38.4% is worth less than 100 million KRW, and 74 or 29.6% is worth between 100 million to 300 million KRW. Also, 68% of the total is small or mid-sized projects worth less than 300 million KRW while more than 96% of the project developments depend on complete outsourcing or partial outsourcing [1].

Outsourcing partners are screened based on technical proposals submitted which will be assessed against an Evaluation Standard for Software Technology[2]. It is appropriate to employ an evaluation process for large projects only with written proposals as most of bidders are market-proven players. However, bidders for small and mid-sized, short-term projects are mostly small and mid-sized companies and thus, it is required to assess the reliability of the previous projects they have conducted in a practical manner.

The current study suggests one research method that enables reliability assessment on such previous projects by studying user satisfaction and the amount of delay in delivery time. If incorporated into the existing acquisition process, the reliability assessment method will further enhance objectivity and accuracy in bidder selection process.

GQM(Goal Quality Method) paradigm was used to identify assessment metrics for bidder evaluation and questionnaires were collected from users to create user satisfaction index. In addition, ‘weight of evidence’, the most appropriate categorical method, was used to isolate attributes of each variable that may contribute to reliability assessment. Also, regression analysis was used to come up with reliability metrics model.

2 Related Research

2.1 GQM Paradigm

The GQM Paradigm is based on the idea that measurement should be goal-oriented; i.e., all data collection should be based on a rationale that is explicitly documented[7].

![GQM paradigm](image_url)

GQM-based measurement programs should be planned and performed according to the following principles:
1. The analysis task to be performed must be specified precisely and explicitly using a detailed measurement goal;

2. Metrics must be derived in a top-down fashion based on goals and questions;

3. Each metric must have an underlying rationale that is explicitly documented;

4. The data that are gathered for the metrics must be interpreted in a bottom-up fashion using the GQM goal and questions; and

5. The people from whose viewpoint the measurement goal is formulated must be deeply involved in the definition and interpretation of the measurement goal.

Two techniques are useful for developing GQM plans. First, goal templates assist in generating a GQM goal. The template identifies five major aspects, namely the object, purpose, quality focus, viewpoint, and environment of a measurement program.

Second, abstraction sheets are defined to assist in collecting the information necessary to build a detailed GQM plan.

A GQM abstraction sheet is a document, often a single sheet of paper that helps elicit and structure information during an interview and assists in constructing, refining, and reviewing a single GQM plan.

2.2 Metrics on Software Reliability

Software reliability represents a user-oriented view of software quality. Analysis on software reliability is conducted to put in place a timely software development practice via statistical metrics. Reliability analysis consists of prediction and estimation depending on analysis time horizon. Prediction refers to activities taking place from project planning to the phase right before tests and it utilizes empirical data set to find the number of defects during these phases before testing.

Estimation refers to activities from integration to acceptance test and it utilizes operation profile – the number of defects during observation- to calculate estimated errors as well as associated cost at the point of software delivery and afterwards.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Prediction versus Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Purpose</strong></td>
<td>To determine in advance what practice must be employed (How)</td>
</tr>
<tr>
<td><strong>When they are used</strong></td>
<td>Anytime but usually before system testing</td>
</tr>
<tr>
<td><strong>Data Collected</strong></td>
<td>Empirical or historical</td>
</tr>
<tr>
<td><strong>Entity measured</strong></td>
<td>Defect density</td>
</tr>
</tbody>
</table>

3 Reliability Metrics Design

GQM paradigm was used to identify metrics to assess bidders’ reliability. By use of Object Template, we can derive detail goal.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Reliability Object Template</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object</strong></td>
<td>Major aspects</td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>Project Results</td>
</tr>
<tr>
<td><strong>Quality Focus</strong></td>
<td>Prediction</td>
</tr>
<tr>
<td><strong>View Point</strong></td>
<td>Reliability</td>
</tr>
<tr>
<td><strong>Environment</strong></td>
<td>Acquisitor</td>
</tr>
</tbody>
</table>

Analysis targets include projects that last under 6 months with less than 300 million KRW. For project reliability variation factors, user satisfaction and the amount of delay before project stabilization were selected.

Also, abstraction sheet was used in order to identify factors in detail that are associated with the quality of objective templates found. Abstraction sheet is mostly a one page document and is useful in clearly delivering information on analysis target during interviews or researches. [Table 3] shows an example of abstraction sheet to articulate research objectives and to review the process.
Lastly, it is required to identify a set of questions that can completely define variation factors of reliability abstraction sheet and basic assumptions. Once metrics is designed in line with questions, consistency is ensured for the entire analysis process as each variable will have higher correlation with the research objectives. [Table 4] illustrates a set of questions and metrics based on variation factors of the abstraction sheet and basic assumptions.

<table>
<thead>
<tr>
<th>Objectives</th>
<th>Metrics on the Reliability of Outsourcing Partners from Acquisitor’s Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question</td>
<td>M1 User satisfaction</td>
</tr>
<tr>
<td></td>
<td>M2 Project delay rate</td>
</tr>
<tr>
<td>Question</td>
<td>M3 Reduction in Delay as a result of Improvements (%)?</td>
</tr>
<tr>
<td></td>
<td>M4 How to calculate satisfaction level?</td>
</tr>
<tr>
<td></td>
<td>1 Overall, I am satisfied with how easy it is to use system</td>
</tr>
<tr>
<td></td>
<td>2 I am able to complete my work quickly using system</td>
</tr>
<tr>
<td></td>
<td>3 It is easy to find the information I needed</td>
</tr>
<tr>
<td></td>
<td>4 I like using the interface of system</td>
</tr>
<tr>
<td></td>
<td>5 system has all the functions and capabilities I expect it to have</td>
</tr>
<tr>
<td></td>
<td>Count the number of agreements among 5 questions.</td>
</tr>
</tbody>
</table>

To measure user satisfaction which is highly subjective, such methodologies as user experience research and feasibility test on software quality are available. Among these methodologies, user experience research is used in general and the current study has selected 5 items that suit web application from CSUQ(Customer System Usability Questionnaire) developed by IBM[5].

4 Reliability Prediction

Reliability prediction is based on information gathered from questionnaires that are incorporated with identified metrics and distributed to previous and current clients of the bidder.

4.1 Sampling and Variable Selection

37 samples were gathered out of interviews and questionnaires sent to 60 government agencies in order to do research on the results of outsourced projects. Government agencies here refer to a special corporate body that is mandated by relevant law and that is authorized and commissioned by the government to do tasks. Among the 37 samples, 21 were successful while 16 were not successful in terms of user satisfaction and the period required for project stabilization.

Variables for the research are 28 in total: 17 HR-related information and 11 methodologies. To come down with finalized variables for modeling, processes illustrated in [Figure 2] were followed.

Variables to be used for reliability assessment were drawn from a set of 148 variables developed by SofeRel, a specialized reliability research agency, based on its 15 year study regarding software reliability[3]. The variables can be classified into 5 categories as in [Table 5].
As the research targets web application at the stage of acquisition, 28 variables were selected for the first round of study in terms of methodologies and HR.

Among the selected variables for the first round, Information Value was used to see if the attributes of the variables are appropriate to show its impact on reliability.

\[(p_{ij} - q_{ij})w_{ij}\]  \(1\)

Note, \(p_{ij}\) = the number of good quality with attribute \(j\) from a variable with \(i\) attribute /total number of good quality

\(q_{ij}\) = the number of defective quality with attribute \(j\) from a variable with \(i\) attribute /total number of good quality

\(w_{ij}\) = \(\ln\left(\frac{p_{ij}}{q_{ij}}\right)\)

In general, the relevant variable is selected for the model if information value that exceeds 0.1. In information theory, evidence weight is an index that represents the importance of specific information and it also shows how much information is provided in determining “reliable” or “non-reliable” for reliability assessment. The evidence weight that is greater than 0, means that reliability rate is higher than non-reliability rate.

The selected variables for the first round are sorted out depending on their attributes and on evidence weight. [Table 6] show selected variables that have information value exceeding 0.1.

### 4.2 Reliability Model

To test the relationship between finalized variables and project outcomes, a regression analysis was conducted with the success of the project as dependent variable. [Table 7] shows result of regression analysis.

\[
\begin{array}{cccccc}
\text{Variable} & \text{B} & \text{Std. Dev.} & \text{Beta} & \text{t} & \text{p} \\
\hline
\text{Constant} & -0.094 & 0.144 & -0.651 & 0.519 \\
\text{Risk Management} & 0.422 & 0.159 & 0.365 & 2.650 & 0.01 \\
\text{Timely HR Input} & 0.313 & 0.135 & 0.316 & 2.319 & 0.02 \\
\text{Analyzed Scope} & 0.305 & 0.131 & 0.302 & 2.317 & 0.02 \\
\end{array}
\]

Model \(R\) \(R^2\) Modified \(R^2\) SE \(p\)

Reliability 0.674 0.454 0.405 0.387 0.0001

As the impact of finalized variables on project deliverables was tested, the following regression model can be derived to explain project reliability.

\[Y = -0.094 + 0.422 V1(\text{bidder’s risk management}) + 0.313 V2(\text{timely input of HR}) + 0.305 V3(\text{analysis on development scope})\]  \(2\)

The model shows that the most important variable in assessing reliability of small/mid-sized, short term projects is the timely input of human resources if other conditions are the same. It was also found that development scope analysis at the beginning of the project, should be accurate and that bidders should cope with risk situations in an appropriate manner.

The score of bidder’s reliability can be classified into 3 levels as in [Table 8].

### 4.3 Model Evaluation

To measure the fitness of model, following [table 9] can be used.
Sensitivity: The percentage of classifying the actual Good Group as good.
\[ \left( \frac{19}{21} \right) \times 100 = 90\% \]

Specification: The percentage of classifying the actual Bad Group as bad.
\[ \left( \frac{12}{16} \right) \times 100 = 75\% \]

Accuracy: Probability of classifying the Good Group and Bad Group correctly.
\[ \left( \frac{12 + 19}{37} \right) \times 100 = 84\% \]

5 Conclusions

Statistical assessment can serve as an objective tool to assess bidders whose reliability has yet to be objectively measured for acquisition projects of government agencies. Based on the most optimal categorical method of 'evidence weight', variables were selected to draw a line between 'reliable' and 'non-reliable' and furthermore, regression analysis was employed to construct a simple model.

In future, it is required to find additional variables that have higher applicability by doing more detailed case studies for diverse projects.

6 References


