In Proceedings of the Sixteenth International Conference on Software Engineering and Knowledge Engineering (SEKE'04),
pp. 117-122, Banff Centre, Banff, Alberta, Canada, June 20 to June 24, 2004.

Case Study Methodology Designed Research in Software Engineering Methodology Validation

Seok Won Lee[1] and David C. Rine[2,3]
[1]*Dept. of Software and Information Systems, University of North Carolina at Charlotte*
9201 University City Blvd, Charlotte, NC 28223, USA.
[2]*Dept. of Computer Science, George Mason University, Fairfax, VA 22030, USA.*
[3]*NASA NET Research Group, George Mason University, USA.*
seoklee@uncc.edu, drine@gmu.edu

Abstract. One of the challenging research problems in validating a software engineering methodology (SEM), and a part of its validation process, is to answer "How to fairly collect, present and analyze the data?". This problem adds complexity, in general, when the SEM involves the use of human knowledge in its methods (phases). How should such created knowledge be captured in the methodology during a SEM process? How can such knowledge be made available for continued SEM process improvement? How can such knowledge be used in validating the SEM? Measuring such knowledge is hard, but we can benefit from the "Case study research design" which is a valuable and an important empirical research alternative in designing a research plan that establishes a logical link from the data to be collected to the initial questions of study. In this paper, a case study research methodology (CSM) designed is presented with its application to the validation of a software requirements engineering methodology (SREM). The preliminary results show the evidence used to validate the SREM as well as the potential usage of CSM as a goal-oriented research design, practice and teaching methodology.

1. Introduction

A case study design, as a technology empirical evaluation research methodology and a way to generalize from observed case study outcomes, builds a basis for valid inferences from the case study events and evidence collected. An invented SEM may be such a technology. For an effective research case study, as an empirical research methodology applied as a validation exercise – applied to an 'invented software (systems) engineering methodology,' it is necessary for the validation exercise to first have designed a case study methodology specific to the characteristics of this invented SEM. More specifically, those characteristics are: 1) the characteristics of this invented SEM that required interventions from the domain Subject Matter Expert (SMEs) or the software engineer in order to perform on demand each appropriate step in the SEM; and 2) the characteristics of the invented SEM validation that cannot favor the invented SEM over alternatives because of the uniqueness of the invented methodology or the relative different level of understanding of the domain and analytical skill of SMEs in the actual case study 'experimental' conditions. Therefore, the consideration of these two above characteristics motivated development of the case study design in this paper, based on the theories and guidelines from [11]. In addition, the theoretical framework of the case study and the application of this case study to the invented Proxy Viewpoints Model-based Requirements Discovery (PVRD) methodology [5, 6] will provide better understanding of the PVRD methodology, and guidance to researchers from academia and real practitioners from industry.

In this paper, the invented SEM (technology) to be validated by a case study empirical approach is the PVRD methodology. In the following sections, a brief introduction of the PVRD methodology and the components in the research case study design will be outlined.

2. The PVRD Methodology

The Proxy Viewpoints Model-based Requirements Discovery (PVRD) is a methodology [5, 6] that provides an integrated framework to reason about "missing natural language system requirements" problems. The PVRD methodology consists of four models: viewpoints model, enterprise model, missing requirements types categorization model, and requirements discovery and analysis model. The viewpoints model [7, 4, 9] represents different perspectives or views for a coverage of direct and indirect stakeholders that need to be identified and incorporated into the legacy status software system requirements. The enterprise model [1] provides a way of categorizing requirements based on systems engineering design process models. The missing requirements types categorization model provides a method to project a requirements space that may contain specific types of missing requirements. The requirements discovery and analysis model provides a method to retrieve requirements of interest by using the requirements term

expansion method [8, 2] that automatically generates a list of "potential query terms" [10] which could assist analysts in acquiring more knowledge about the domain of interest by performing a "complete search" of available requirements resources.

Based on this integrated framework, the PVRD methodology is able to create a proxy viewpoints model and provide a new way of discovering missing natural language system requirements while improving the legacy natural language requirements representation space through the modeling of a new indexing structure that supports multiple viewpoints from many stakeholders in a large-scale complex software system.

The PVRD methodology is applicable in the use of existing natural language software requirements specifications (SRS) in further improved development of legacy systems by 1) discovering missing requirements, especially, when it is necessary to reconstruct the original legacy SRS, and 2) eliciting new requirements for system changes that will take place or creating a new system from a similar legacy system. Figure 1 shows the overview of the steps in the PVRD methodology. More detailed descriptions of the PVRD methodology are in [5, 6].

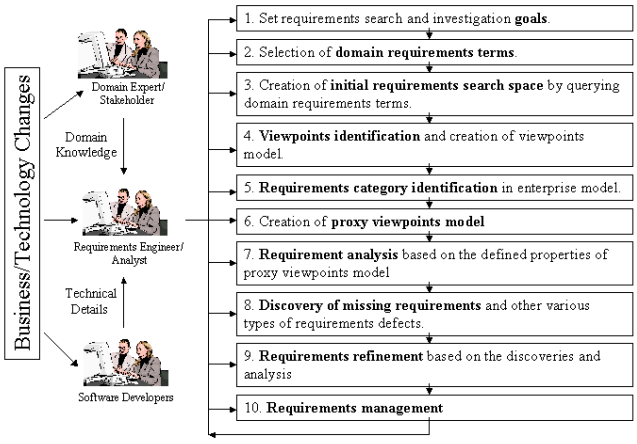

Figure 1. Overview of the PVRD Methodology Steps

3. Case Study Research Design Components

The goal of the PVRD research is "the development of a new methodology that can discover missing natural language requirements and reduce the number of incomplete requirements while reorganizing the requirements representation space that incorporates and supports multiple viewpoints in legacy status, large-scale, information system requirements specifications".

In order to show *how* and *why* ("explanatory" type of case study in [11]) the PVRD methodology can achieve this research goal, the following five important components will be defined during the case study design process [11]:

- A study's questions,
- Study propositions,
- Unit(s) of analysis,
- The logic linking of the data to the propositions, and
- The criteria for interpreting the findings.

Figure 2 shows the inter-relationships between these five components in the explanatory case study design. The study questions can be mapped and further decomposed into a set of more detailed study propositions. These propositions contain metric terms and are used to develop measure data capture questionnaires. The application by the SMEs of the developed PVRD methodology to the units of analysis can then generate results observed and reported by the SMEs in the questionnaires, i.e. measure data collection instruments. The results are then linked back to the study propositions as evidence through the criteria (using metrics) for interpreting these findings.

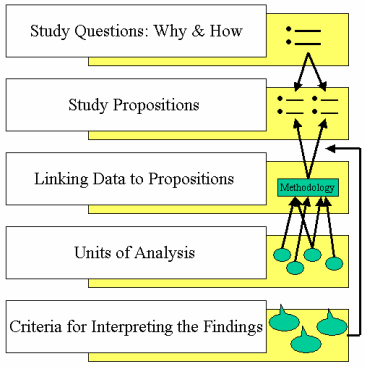

Figure 2. Case Study Design Components

3.1. Study Question

Study questions need to be clarified precisely [11]. For instance, in this case study design, the study question "How and why can the PVRD methodology discover missing requirements from natural language software requirements specification documents that involve multiple viewpoints while reorganizing and improving the quality of a requirements representation space?" needs to be clarified by further decomposition of that research question into sub questions or propositions. In this case study, these natural language software requirements specification documents identified in this study question are the case study units of analysis.

3.2. Case Study Propositions

Case study propositions, derived from the case study questions, become assertions that should be examined, through measure-valued questionnaire items, to answer a

study question within the scope of the case study. For instance, in this case study design, the *general study proposition* is "the PVRD methodology can achieve its research goals because of its integrated framework that benefits by the synergism between the embedded models and methods". And four more *specific case study propositions* derived from this proposition are as follows:

The PVRD methodology and its integrated framework can 1) Reduce the number of incomplete requirements by discovering missing requirements; 2) Discover requirements defects of various types; 3) Discover requirements relationship and workflow process relationship chains in the requirements space; and 4) Create new requirements indexing structures based on the embedded models.

3.3. Units of Analysis

Units of analysis are the selected resources to be examined through the application of the invented technology (e.g. PVRD methodology) by the SMEs in the case study. For instance, in the case study methodology designed for the research reported here, the PVRD methodology will be applied to the (units) "set of software requirements specification documents" that is expressed in natural language. These requirements (units of analysis) are from the legacy status information-based software system that includes many stakeholders (e.g. interactive systems). Also the SMEs and other software development resources (other than SRSs, such as business process descriptions, operational concepts etc.) need to be involved due to their influences in the requirements classification (e.g. assigning requirements to viewpoints) of each model (i.e. Viewpoints Model, Enterprise Model, Missing Requirements Types Categorization Model, Requirements Discovery and Analysis Model) in the PVRD methodology. From the PVRD methodology point of view, the models and methods in the PVRD methodology, and the properties associated with the PVRD methodology, are the subjects to be examined during a case study using this case study design.

Table 1 presents the 'Requirements Discovery Summary Sheet' (RDSS) that is used by a team of SMEs during a case study using this case study design. Each step in the RDSS sheet corresponds to the steps in Figure 1. This RDSS sheet is utilized for SMEs to serve the following roles during a case study:

- Provides a template of how to apply the PVRD methodology to the units of analysis in a case study. Combined with the given instructions during a case study, SMEs can identify the units that need to be examined (analyzed) and results to be recorded.
- Captures evidence and findings using the criteria defined by metrics with measures (combination of qualitative and quantitative analysis in section 3.5) that will be used to support/reject the case study propositions (in section 3.2). Without having such specific propositions with metric criteria word attributes in them, an investigator might be tempted to collect "everything" which is impossible to collect [11] and meaningless. For instance, as shown in Table 1, the results from the 'Discovery types' [SID-8-DID-1] and 'Level of significance of discovery' [SID-8-DID-2] from the RDSS sheet will be used to support/reject the case study propositions 1, 2, and 3. Also the results from the 'Viewpoints identification' [SID-4-1], 'Requirements category identification in enterprise model' [SID-5-1], 'Proxy viewpoints model creation' [SID-6-1], and 'Newly indexed requirements' [SID-8-DID-7] will be used to support/reject the case study proposition 4.
- Guides SMEs in a step-by-step approach while conducting the PVRD methodology case study. SMEs follow each step of the PVRD methodology and record their findings and observations (unit by unit) under each step in the RDSS sheet.

In Table 1, steps 1 – 6 focus on the process of the proxy viewpoints model creation, and steps 7 – 8 focus on the requirements analysis and discovery process based on the created proxy viewpoints model. Therefore, the questions under each step from 1 to 6 and the evidence collected by SMEs would capture the idea of "how SMEs created the proxy viewpoints model from the given requirements set". In step 7, SMEs collect their units PVRD methodology analysis results from the created proxy viewpoints model. The fine-grained questions (units) in step 8 would capture the specific evidence to support/reject the propositions related to the requirements discovery process. In the questionnaire, some questions (i.e. SID-8-DID-2, SID-8-DID-7, SID-8-DID-8) will be interpreted based on the qualitative measures and some questions (i.e. SID-8-DID-1) will be interpreted based on the quantitative measures. The following Table 1 summarizes important aspects of the RDSS sheet.

In addition to collecting specific evidence to support/reject specific propositions, it is also important to collect evidence of the 'entire process' wherein the methodology is applied to the given requirements set (units of analysis). This is because the specific evidence is captured in the middle or after the application of the PVRD methodology, and it did not come from an independent evidence collection process. Also the collection of evidence from the 'entire process' must be used, wherein the PVRD methodology is used to support/reject the general proposition in section 3.2 (also the study question in section 3.1).

Having the RDSS with clearly identified steps and interpretation criteria (metric and measures) is important and related to the general 'repeatability' of the case study methodology.

Table 1. Evidence Collection through Requirements Discovery Summary Sheet

Questions (Units)	Evidence captured by SMEs (Requirements Discovery Summary Sheet - RDSS)	Step/Model/Method in the PVRD	Related Propositions that support/reject
SID-1-1	SMEs record the *goal(s)* of requirements search/investigation.	Step 1	General Proposition (GP)
SID-2-1 SID-2-2 SID-2-3	SMEs record the selected '*key domain terms*'. SMEs record the *reason* for the selected terms. SMEs record the specific type of '*missing requirements types*' if it is used in the selection of domain requirements terms SID-2-1 (with explanation).	Step 2, Missing Requirements Types Categorization Model	GP
SID-3-1	SMEs record the number of requirements in the *initial requirements search space* created (with requirement ID).	Step 3	GP
SID-4-1	SMEs record identified *viewpoints (VP)* for each requirement with its ID.	Step 4, Viewpoints Model	GP, Proposition 4
SID-5-1	SMEs record identified *category of enterprise model (EM)* for each requirement with its ID.	Step 5, Enterprise Model	GP, Proposition 4
SID-6-1	SMEs check each requirement index based on the VP and EM and create a *proxy viewpoints model* and the layout.	Step 6, Proxy Viewpoints Model	GP, Proposition 4
SID-7-1	SMEs analyze the PVRD layout and record any *discovery patterns* found in the created proxy viewpoints model.	Step 7, Requirements Discovery and Analysis Model	GP
SID-8-DID-1	SMEs record the *types of discovery patterns* found in the created proxy viewpoints model.	Step 8, Requirements Discovery and Analysis Model	GP, Propositions 1, 2 and 3
SID-8-DID-2	SMEs record the *level of significance* of the discovery patterns found.	Step 8, Requirements Discovery and Analysis Model	GP, Propositions 1, 2 and 3
SID-8-DID-3	If the *term expansion method* is used in the discovery process, SMEs record specific steps taken (with detailed explanation of how this method is used and contributed to this discovery process).	Step 8, Term Expansion Method, Requirements Discovery and Analysis Model	GP
SID-8-DID-4	SMEs record the '*requirements distance*' from this discovery.	Step 8, Requirements Discovery and Analysis Model	GP
SID-8-DID-5	If '*missing requirements types*' are used in this discovery, SMEs record the specific type and explanation of how it is used in this discovery.	Step 8, Missing Requirements Types Categorization Model, Requirements Discovery and Analysis Model	GP
SID-8-DID-6	If *new types of missing requirements* are discovered, SMEs record this new type and specific explanation of what they are.	Step 8, Missing Requirements Types Categorization Model, Requirements Discovery and Analysis Model	GP
SID-8-DID-7	SMEs record their observation about the *newly indexed requirements representation* (through the VP and EM), compared to the original requirements structure.	Step 8, Proxy Viewpoints Model	GP, Proposition 4
SID-8-DID-8	SMEs record their observation about the *PVRD methodology contribution for this discovery*.	Step 8, Proxy Viewpoints Model, Requirements Discovery and Analysis Model	GP
SID-8-DID-9	SMEs record their comments about their *experience with the PVRD methodology for this discovery*.	Step 8, Proxy Viewpoints Model, Requirements Discovery and Analysis Model	GP

3.4. Linking Data to Study Propositions

Linking data to propositions represents the data analysis step in the case study design research [11]. The PVRD methodology is applied to the units of analysis and plays a role in connecting the generated measure data results back to the study propositions.

In a case study using this case study design, the generated measure data results can be based on any mix of *qualitative* and *quantitative* evidence. In this case study, the generated results will be collected by SMEs in the form of RDSS sheet in Table 1 and also the notes from the lessons learned meeting. In addition, a case study need not always include a direct, detailed observation as a source of evidence [11]. Therefore, in this case study, a set of evidence, its analytical interpretations, and lessons learned are sources of evidence as summarized in Table 1.

From the findings through the RDSS sheet, Table 1 serves as qualitative and quantitative evidence, such as 1) whether or not the discovered missing requirements are defining, mandatory or optional requirements (as qualitative measures defined in section 3.5); and 2) numbers and types of discoveries found (as quantitative measures).

Figure 3 shows how the findings are linked in the support of corresponding study propositions (an explosion of the link between the 'study propositions' and 'linking data to propositions' in Figure 2). The findings will be collected through the RDSS sheet in Table 1 by SMEs during a case study.

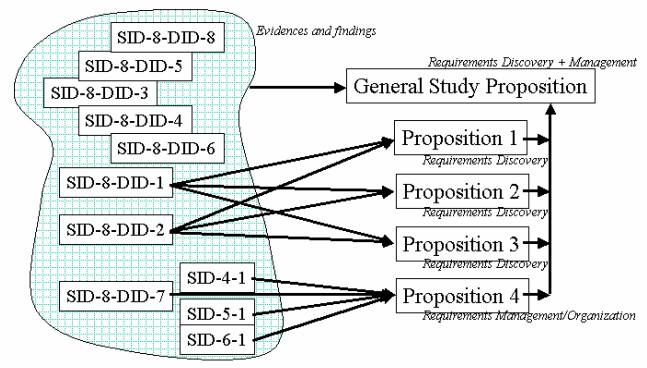

Figure 3. Linking Data to Study Propositions through Discovery Summary

3.5. Criteria for Interpreting a Case Study's Findings

Criteria for interpreting a case study's findings correspond to the metric and measures used in evaluating the results from the properties of requirements defects types defined in the PVRD methodology (such as incomplete, inconsistent, redundant, and ambiguous, as well as requirements relationship chain and workflow process relationship chain). In other words, the results from the requirements discovery and analysis model [5, 6] focus on the findings of significant defects of requirements and improvement of requirements quality as well as the quantitative analysis of how many requirements defects discovered. One example of such findings of significant defects that will be focused on is whether or not the discovered missing requirements or defects are one of defining, mandatory or optional requirements. The discovery of defining or mandatory requirements is much more critical than discovery of optional requirements. Therefore, the "importance/significance of the discovered requirements defects" serves as a *metric* in the analysis of the findings and three different requirements types "defining, mandatory or optional requirements" serve as *qualitative measures* in deciding the significance of the requirements. Also the number of requirements defects of various types serves as *quantitative measures*. Therefore, the metrics and measures will be all interpreted from a combined qualitative and quantitative analysis perspectives based on the summary of RDSS sheet in Table 1. All five components in the case study design described will guide a case study and become the fundamental basis in validating the PVRD methodology.

4. Multiple Case Studies

One of the most important points made in Yin's case study design approach [11] is the design of a theoretical case study framework which is presented in section 3. Also, it is important to understand the importance of '*analytical generalization*' – case studies (as with experiments), compared to 'statistical generalization' – survey research, in case study design. In statistical generalization, an inference is made about a population (or universe) on the basis of empirical data collected about a sample (i.e. surveys). In analytical generalization, the investigator is striving to generalize a particular set of results to some broader theory [11].

The evidence from multiple cases is often considered more compelling, and the overall study is therefore regarded as being more robust [3, 11]. A theory must be tested through replications of the findings in a second or more case that will lead to an analytical generalization. Under the development of a theoretical framework in [11], a *literal replication* (each case predicts the similar results) can explain the conditions under which a particular phenomenon is likely to be found, a *theoretical replication* (each case produces contrasting results but for predictable reasons) can explain the conditions when it is not likely to be found. Multiple case studies were carried out for the PVRD methodology validation and established a literal replication. For each individual case, collected evidence indicated how and why a particular proposition was demonstrated (or not demonstrated).

5. A Case Study in Educational Information Management System (EMS)

This particular case study is performed by a team of domain independent SMEs from industry (more than 10 years software requirements engineering & software development experiences) based on the CSM in order to formally confirm and validate the case study propositions of the PVRD methodology. For this case study, SMEs are trained to understand and apply the PVRD methodology through an orientation/workshop to practice performing a case study independently. The purposes of the workshop/orientation are to educate SMEs about the PVRD methodology case study, the methodology, embedded models, and methods, and going through step-by-step approach. It is also important that the researcher conducting the case study must have no interaction with the SMEs once the case study exercise is underway so as not to bias or prejudice SMEs' judgement. The EMS requirements documents size was over 300 pages. The findings and evidence are recorded and collected through the RDSS sheet using a given set of instructions.

Figure 4 shows the PVRD model that SMEs constructed during this case study in which became the basis of the discovery process. Each requirement represented in this PVRD model has been applied to each model and method embedded in the methodology, and corresponding relationships are established across the models.

By using the designed case study methodology as presented in this paper, evidence that can support/reject the case study propositions is collected as well as all the detailed observational descriptions by SMEs. Three investigations' results are summarized in Table 2. Other case studies that were carried out (but are not reported in this paper), more detailed description of the case and experimental set up are in [5].

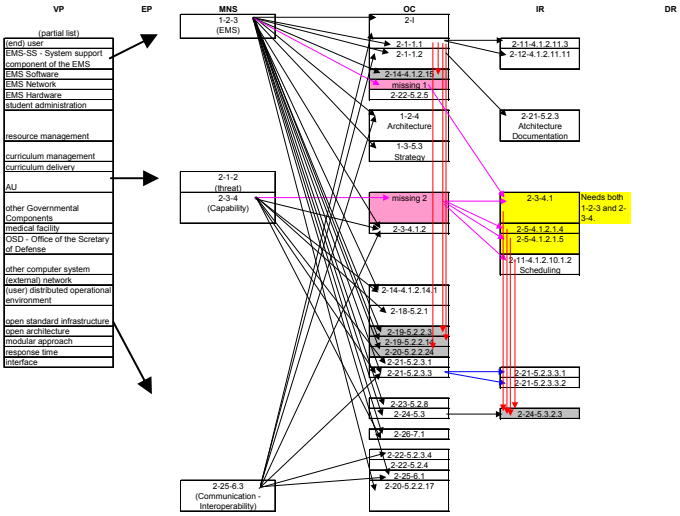

Figure 4. The PVRD Model

Table 2. EMS Case Study Results Summary

Results Category	Investigation 1 (SME #1)	Investigation 2 (SME #1)	Investigation 3 (SME #2)
# Discoveries	4	3	2
Missing Reqts	1	1	1
Other types of defects	Redundant (1) Inconsistent/ Ambiguous (1)	Redundant (1) Inconsistent/ Ambiguous (1)	
Reqts Relationship Chains	1 (Requirements relationship)		1 (Workflow process relationship)
# 'Defining' Reqts Discovery	1		
# 'Mandatory' Reqts Discovery	2	1	2
# 'Optional' Reqts Discovery	1	2	
New missing reqts types category, reqts distance, etc.	New missing reqts type, Requirements distance (>20)	Requirements distance (>20)	Requirements distance (>20)
The PVRD methodology contribution to the discovery process	Strongly (1) – missing requirements, Moderately (3) – others	Strongly (1) – missing requirements, Moderately (2) – others	Strongly (2)
Supporting Propositions	General, Propositions 1,2,3, and 4	General, Propositions 1,2, and 4	General, Propositions 1, 3, and 4
Rejecting Propositions		Proposition 3	Proposition 2
Overall Supporting Propositions	General Proposition, Propositions 1,2,3 and 4		

6. Conclusion and Future Work

Case studies are multi-perspective/dimensional analyses that need to consider many aspects in collecting evidences from various resources during their design and executions. For example, the use of the CSM in validation of the PVRD methodology considers not only the technical aspects but also the interactions with SMEs in capturing data and knowledge acquired.

The CSM can leverage its usage as follows: 1) a "goal-oriented" research design, practice and validation methodology (through effective evidence collection, presentation and analysis); 2) a flexible but theoretically powerful and solid methodology that can cover various interdisciplinary research domains' characteristics; and 3) a teaching methodology in education to cultivate student's integrative analytical and problem-solving skills. As a future work, more in-depth study of the application of the CSM to knowledge-intensive software engineering methodology validation is planned.

References

[1] Buede, D. M. *The Engineering Design of Systems : Models and Methods*, New York: Wiley, December. 1999.

[2] Faloutsos, C. and Oard, D., A Survey of Information Retrieval and Filtering Methods, CS-TR-3514, University of Maryland, 1995.

[3] Herriott, R.E. and Firestone, W.A. Multisite qualitative policy research: Optimizing description and generalizability. *Educational Researcher*, 12, 14-19. 1983.

[4] Kotonya, G. and Sommerville, I. Requirements Engineering with Viewpoints, *BCS/IEE Software Engineering Journal*, Vol. 11, No. 1, pp.5-18. 1996.

[5] Lee, S. W. "Proxy Viewpoints Model-based Requirements Discovery" *PhD Dissertation*, George Mason University, Fairfax, VA. 2003.

[6] Lee, S.W. and Rine, D.C. "Missing Requirements and Relationship Discovery through Proxy Viewpoints Model," In *Proceedings of the 19th annual ACM Symposium on Applied Computing (SAC 2004)*, Software Engineering, March, Cyprus, ACM Press, 2004.

[7] Nuseibeh, B., Kramer, J. and Finkelstein, A. A Framework for Expressing the Relationships Between Multiple Views in Requirements Specification, *IEEE Trans on Software Eng*, 20(10):760-773, IEEE CS Press, October. 1994.

[8] Salton, G. *The SMART Retrieval System – Experiments in Automatic Document Processing*. Prentice-Hall Inc., Englewood Cliffs, New Jersey. 1971.

[9] Sommerville, I. and Sawyer, P. Viewpoints: Principles, Problems and a Practical Approach to Requirements Engineering. *Annual Soft Eng*, Vol. 3. pp. 101-130. 1997.

[10] Xu, J. and Croft, B. Improving the Effectiveness of Information Retrieval with Local Context Analysis. *ACM Trans on Info Systems*, Vol.18, No.1, pp.79-112. Jan. 2000.

[11] Yin, Robert. K. 1994. *Case Study Research: Design and Methods*. 2nd Edition. Thousand Oaks: Sage Publications, Applied Social Research Methods Series Vol. 5.