In the Proceedings of the 17th ACM Symposium on Applied Computing (SAC 2002) March 10-14, 2002, Madrid, Spain

Special Track on Software Engineering: Theory, Applications and Practice

Proxy Viewpoints Model-based Requirements Engineering

Seok Won Lee
School of Information Technology and Engineering
George Mason University
Fairfax, VA 22030-4444, USA
Science Applications International Corporation (SAIC)
1953 Gallows Rd., M/S V-1-4, Vienna, VA 22182, USA
01-1-703-610-3352
swlee@cs.gmu.edu

ABSTRACT

This paper addresses the problem of the "missing requirements" in software requirement specification (SRS) expressed in natural language. Due to rapid changes in technology and business frequently witnessed over time, the original SRS documents often experience the problems of missing, not available, and hard-to-locate requirements. One of the flaws in earlier solutions to this problem has no consideration for missing requirements from multiple viewpoints. Furthermore, since such SRS documents represent an incomplete domain model, manual discovery (identification and incorporation) of missing requirements is highly labor intensive and error-prone. Consequently, deriving and improving an efficient adaptation of SRS changes remain a complex problem. In this paper, we present a new methodology entitled "Proxy Viewpoints Model-based Requirements Discovery (PVRD)". Through the requirements discovery and analysis process, PVRD methodology provides ways to construct proxy viewpoints model from legacy status requirements. Requirements term expansion technique facilitates the retrieval process of requirements of interest based on the improved requirements representation space in proxy viewpoints model. The PVRD methodology provides an integrated environment that supports requirements discovery process as well as efficient management.

Categories and Subject Descriptors

D.2.1 [**Software Engineering**]: Requirements/Specifications – Methodologies. D. 2.9 [**Software Engineering**]: Management. H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Query formulation, Relevance feedback.

General Terms

Algorithms, Documentation, Management, Measurement, Verification

Keywords

Proxy Viewpoints, Requirements Discovery, Term Expansion

1. INTRODUCTION

This paper addresses the problem of the "missing requirements" in software requirement specification (SRS) expressed in natural language. Due to rapid changes in technology and business frequently witnessed over time, the original SRS documents often experience the problems of missing, not available, and hard-to-locate requirements. This problem can be further decomposed into following subproblems: 1) Earlier solutions do not consider missing requirements from multiple viewpoints; 2) SRS documents with many missing requirements typically tend to be poorly structured and maintained as well as hard-to-trace; 3) SRS documents with missing requirements represent an incomplete domain model; 4) Manual discovery (identification and incorporation) of missing requirements is highly labor intensive and error-prone; and finally 5) These sub problems do not allow efficient adaptation of SRS changes and improvements. Such problems are especially prevalent while dealing with legacy status [1] SRS. Thus, a new methodology that can provide improved solutions to these problems and lengthen the life span of SRS needs to be developed.

In this paper, we present a new methodology entitled "Proxy Viewpoints Model-based Requirements Discovery (PVRD)". Through the requirements discovery and analysis process, PVRD methodology provides ways to construct proxy viewpoints model from legacy status natural language SRS documents. *Proxy viewpoints* is a good enough approximation of viewpoints that would have been constructed if the requirements of domain were well engineered from the beginning of software development life cycle by using one of viewpoints oriented requirements engineering methods such as VORD [6].

The PVRD methodology consists of four models: viewpoints model, enterprise model, missing requirements categorization, and requirements discovery and analysis model. Based on the integrated framework, PVRD methodology is able to create a proxy viewpoints model and provides a new way of discovering missing requirements while improving the requirements representation space through the new indexing structure that supports multiple viewpoints from stakeholders.

2. RELATED WORK

Viewpoints approach to requirements engineering [5][8][13] provides many ways of requirements organization and management. For instance, PREView [12] and VORD provide environments for requirements elicitation, structuring, and management. While these approaches need to be applied at the very beginning stage of software development life cycle, the PVRD methodology provides a way of constructing proxy viewpoints model from the poorly structured requirements of legacy status software systems through several methods in the requirements discovery and analysis model.

[9][4] discuss several inspection techniques for detecting, diagnosing, and correcting errors in natural language requirements documents. Among them, checklist technique contains a list of questions that the analyst may use to assess each requirement. [6] suggests the list should not normally include more than 10 items and the questions on the checklist should usually be general than restrictive. Improvement of requirements quality through defects discovery involves many issues such as the types of requirements and their representation, types of defects, and the efficiency of the methods and applicability.

SAC 2002, Madrid, Spain
© 2002 ACM 1-58113-445-2/02/03…$5.00.

Therefore, there is no single best universal approach for the requirements defects discovery and should depend upon many factors.

In REVERE [10], a probabilistic natural language processing (NLP) tool is applied to free-text documents to retrieve requirements information. It uses statistical likelihood of the words for the classification purpose. The probabilities are derived from a large corpora of free-text that have already been analyzed and manually tagged for each term's certain set of categories (i.e. syntactic, semantic, or lexical categorizations). The generated log-likelihood figure heavily relies on the correctness and compatibility of the pre-tagged corpora to the application domain. However, the requirements term expansion technique in the PVRD methodology focus on the complete search of requirements of interest through the improvement of quality of query terms.

3. THE METHODOLOGY: PROXY VIEWPOINTS MODEL-BASED REQUIREMENTS DISCOVERY (PVRD)

3.1 Viewpoints Model

In PVRD methodology, *Viewpoints model* represents different perspectives or views that need to be identified and incorporated into the legacy status software system requirements. We assume that legacy software system requirements were initially developed without considering viewpoints concept. The identification and conceptualization of initial viewpoints model takes place at the very first stage of PVRD. However, these initial set of perspectives or views are partial and incomplete descriptions. As shown in Figure 1, the Viewpoints model should adapt to necessary changes and evolve towards an optimal number of viewpoints and descriptions at the end of PVRD.

Although the need for developing viewpoints model in requirements engineering is well studied in [6], constructing viewpoints model from the legacy SRS, that was originally elicited and maintained without viewpoints consideration, remain a hard problem. The Viewpoints model in PVRD methodology attempts to build a good approximation of viewpoints and also facilitates the requirements discovery and analysis process while providing a good requirements management environment.

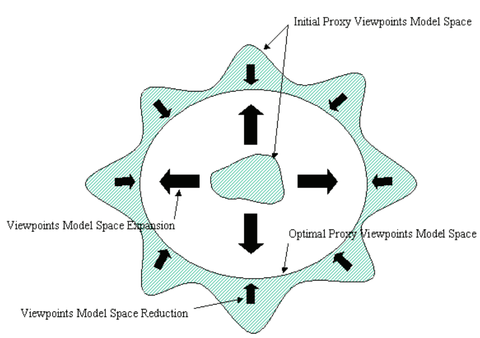

Figure 1. Evolutionary Viewpoints Model Space

A good approximation of the viewpoints' characteristics can be achieved through some analysis techniques such as an interviewing process with subject matter expert/witness, a statistical sampling technique (i.e. stratified sampling technique [3]), or a documentary study of viewpoints. Interviewing process and documentary assessment obviously require a great deal of labor-intensive tasks.

We take the approach of "evolutionary viewpoints model" shown in Figure 1. It starts with an initial viewpoints model that includes a minimum set of viewpoints such as "direct viewpoint" and "indirect viewpoint". Alternatively, the model allows for multiple viewpoints that are partially built from any other available resources from the specific domain. If any "viewpoints template" of a particular domain exists, such template can also be used as an initial viewpoints model as well.

3.2 Enterprise Model

Enterprise model (EM) is a categorization of requirements that are used to define the design problem at various levels of detail in systems engineering [2]. As shown in Figure 2, it usually consists of six sets of requirements categories.

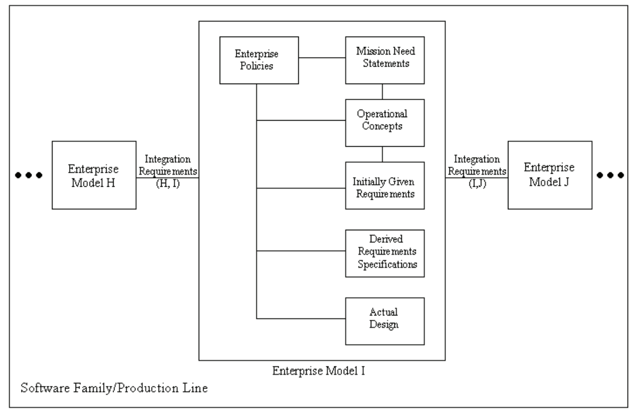

Figure 2. Requirements Categories in Enterprise Model

In PVRD methodology, EM provides a way of organizing and managing requirements from the systems engineering design process point of view. Each requirement is indexed based on the defined roles/scopes of each category of EM as well as the identification of the corresponding viewpoint. The level of abstractness of individual requirement in each category of EM is also determined based on the granularity of viewpoints model. Because each requirements category in EM inherits the characteristics of each viewpoint in viewpoints model, the new indexing structure facilitates the requirements discovery and analysis process as well as the characterization of the requirements. Customization of the categories and their specific roles/scopes depends on the specific domain application.

3.3 Missing Requirements Categorization

[7] describes explorative study of possible missing requirements types in SRS. For instance (from various levels of abstractness and points of view), they are: non-inclusion of all significant requirements; non-conformity to SRS standard; undesired event handling; omitted non-functional requirements; missing requirements due to a single point of failure for a critical system; non-reachable states or operational modes etc. In PVRD methodology, this classification scheme of missing requirements types is used to the projection of requirements space that is associated with corresponding viewpoint and category of EM.

Figure 3 shows the projected requirements space from three different classifications which are: viewpoint, EM category, and missing requirements categorization. RS(i,j,k) represents all requirements space that belongs to viewpoint VP_j, *i*th category of EM: EM_i, and *k*th missing requirements category: MRC_k. Theoretically RS(i,j,k) should satisfy all constraints from all three dimensions. For instance, let's assume that MRC_k is the "non-conformity to SRS standard". More specifically, those standards are "Z39.50 document interface standard" and "ISO 10160-1 document ordering standard". VP_j is the "document standards" and EM_i is the "actual design requirements".

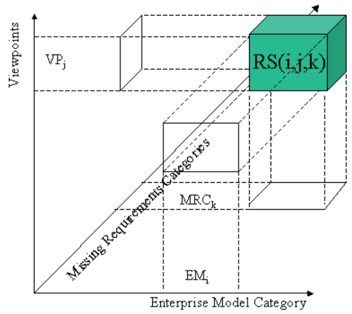

Figure 3. Requirements Space Projection through Missing Requirements Categories

Therefore, all requirements in RS(i,j,k) should conform to those standards. Thus, RS(i,j,k) is assessed for the possibility of missing requirements type MRC_k, and this projection provides a narrow, accurate and effective search space for discovering various types of missing requirements.

3.4 Requirements Discovery and Analysis Model

3.4.1 Requirements Term Expansion

One of the fundamental problems to the complete search in information retrieval is "word mismatch" [14]. Similarly, SRS often contains different terms and descriptions that carry the same contextual information of the domain. Therefore, lack of requirements query terms or non-availability of domain knowledge can result in an incomplete search for specific requirements of interest. Requirements discovery and analysis model uses "requirements term expansion technique" in order to perform complete requirements search in the PVRD methodology.

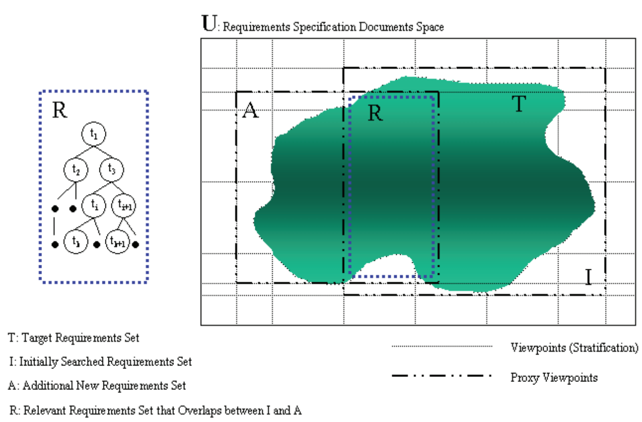

Figure 4. Requirements Search Space

Figure 4 shows how this technique can be applied to the requirements search space. U represents the entire requirements space available for searching. T (shaded area) represents a target requirements space that is most relevant to the user's interests. I represents a set of requirements retrieved through the initial query. Our goal is to find and retrieve the set of requirements A that spans the remaining part of the target requirements space T, which was not retrieved in the space I, using the query expansion technique. In order to do that, query expansion technique automatically generates a list of terms in space R that can potentially characterize the set of requirements A. These terms are the key words that are actually relevant to the requirements that only belong to A. Depending on the granularity of rectangular boxes in the requirements space, it is possible to approximate proxy viewpoints based on the actual viewpoints that should have been constructed. Therefore, it is also possible to define relationships and roles of query expansion in PVRD to the construction of proxy viewpoint model.

As a query expansion technique, term weight function, which is a term frequency-based information retrieval technique [11], is used.

Definition: Requirements Term $\mathbf{T_{ij}}$ is the *i*th unique single word or concept (words or phrases) that are used in the given requirement D_j.

Definition: Requirements Term Weight $\mathbf{WT_{ij}}$ is a measure associated with each requirements term T_{ij} that represents the total topical importance in the given set of requirements documents.

$$WT_{ij} = TF_{ij} * TR_i$$

where, TF_{ij}: term frequency, the average number of times a particular term T_i occurs in a given document D_j . TR_i: term relevance factor, where

$$TR_i = \frac{r_i / (R - r_i)}{s_i / (I - s_i)}$$

R is the total number of relevant documents.
r_i is the number of relevant documents that contain term T_i.
I is the total number of irrelevant documents.
s_i is the number of irrelevant documents that contain term T_i.

Figure 5 shows a procedural description of the requirements term expansion. Requirements terms generated from R and I are the "potential query terms" that can be used to retrieve additional requirements for the complete search. Therefore, these terms comprise a conceptual space that can characterize requirements space R. It is a highly labor-intensive task for users to generate such a long list of elaborative (similar/alternative) queries with an expectation of not missing any important requirements terms. Therefore, automatic query expansion techniques attempt to minimize such user's effort by improving the quality of "queries" by "expanding" them with additional terms (unknown in advance) either automatically or by choosing in cooperation with the user.

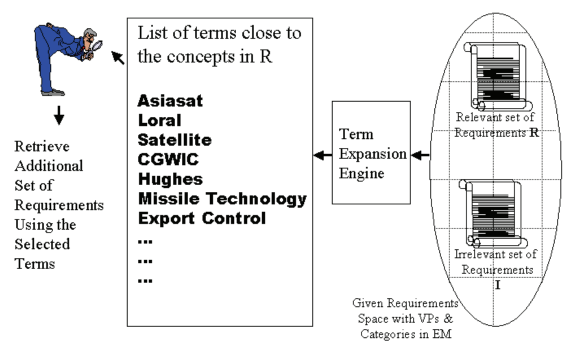

Figure 5. Complete Search through Requirements Term Expansion

3.4.2 Requirements Relation Chain Discovery

Some of requirements often establish certain relationships between them and create relation chains across the categories in EM. For example, these relations include causal, is-a membership, general, special, any feature relationships etc. with many-to-one, one-to-one, or one-to-many correspondences as shown in Figure 6.

r_a ([$R_{ij\text{-}1k}$, $R_{i+1j\text{-}1k}$], $R_{i\text{-}1jk}$) represents many-to-one relationships. r_b ($R_{i\text{-}1j\text{-}1m}$, $R_{i\text{-}1jk}$) represents one-to-one relationship. r_c ($R_{i\text{-}1jk}$, [R_{ij+1k}, $R_{i+1j+1k}$]) represents one-to-many relationships. Relation chains such as r_ar_c and r_br_c can also be discovered. These relationships represent not only relationships between requirements but also relationships of viewpoints and categories of EM.

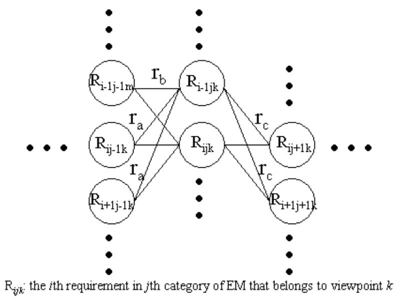

Figure 6. Relation Chain Discovery between Requirements

The relation chains can be also applied to Software Product Line (PL) or Software Family (SF) concepts. A set of verified requirements relation chains can be the "requirements feature template". This template plays a key role in requirements verification of similar module/component in the PL or SF.

3.5 Requirements Verification

Verification process checks whether the component, element, segment, or the system is built correctly based on the requirements. Figure 7 shows the top-down verification approach in the PVRD methodology.

Verification process starts from the most general requirements to the most detail requirements as categorized in EM. For instance, the general requirement R_{i2k} (i.e. mission need statements category) should have more detailed corresponding descriptions $\{R_{i-13k}, R_{i3k}, R_{i+13k}\}$ in the next level of requirements category (i.e. operational concepts category). Also this can be done from a specific viewpoint (viewpoint k in Figure 7) or take other viewpoints into consideration.

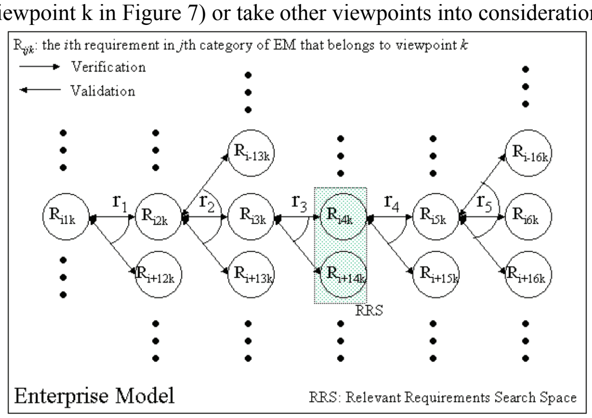

Figure 7. Requirements Verification & Validation

3.6 Requirements Validation

Validation process, which is often combined with acceptance testing, demonstrates that the system satisfies the users' needs. Figure 7 shows the bottom-up validation approach in the PVRD methodology. Validation process starts from the most detail requirements to the most general requirements as categorized in EM. For instance, a set of detailed requirements $\{R_{i4k}, R_{i+14k}\}$ (i.e. initial requirements) should be checked whether it satisfies the corresponding general requirement R_{i3k} (i.e. operational concept). Similar to verification, validation process takes not only individual requirement quality but also associated viewpoints into consideration as well. Validation process is different than the verification in terms of satisfaction of subjectivity of users needs. Therefore, complete enumeration of set of detailed requirements that satisfies general requirements is important. In this phase, requirements term expansion can also be used to retrieve additional requirements from available resources. In Figure 7, set RRS is used for this purpose.

3.7 The PVRD Methodology

Previous sections described the models and methods that are embedded into the PVRD methodology. Figure 8 shows the architecture of the PVRD methodology and Table 1 describes its algorithm in detail.

Users need to understand the requirements domain and identify an initial set of viewpoints of stakeholders. This initial viewpoints model expands whenever new viewpoints are identified and reduces whenever multiple viewpoints need to merge to a single viewpoint. As a result, nearly optimal set of viewpoints for the given requirements will be created in the final stage of the PVRD methodology.

Users need to define the requirements categories in the enterprise model depends on the nature of domain. Each requirements category inherits the characteristics of viewpoints model and creates a new structure, which is a proxy viewpoints model. Based on this model, users can organize and manage individual requirement that is consistent to the roles and scopes of each requirements category in the enterprise model. Also the level of abstractness of requirements descriptions in enterprise model is determined by the granularity of the viewpoints model.

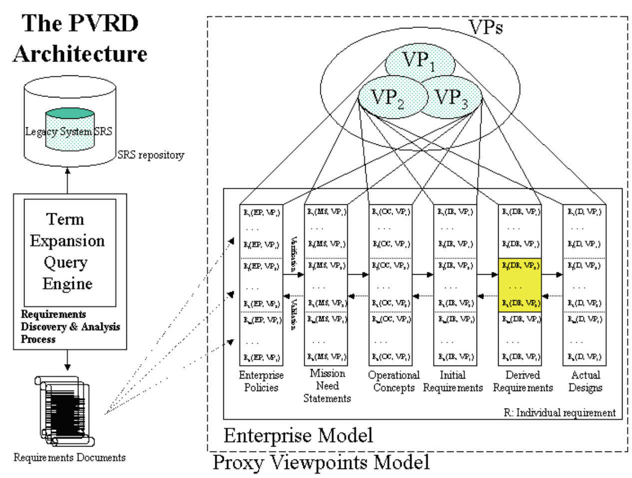

Figure 8. The Architecture of Proxy Viewpoints Model-based Requirements Discovery (PVRD) Methodology

Requirements discovery and analysis process retrieves specific requirements of interest through the requirements term expansion technique. Each requirement is assessed and indexed based on the proxy viewpoints model. This includes the identification of corresponding viewpoint(s) and the category of enterprise model. In this phase, either new requirement can be derived and constructed or viewpoints model can be expanded or reduced. As requirements are being indexed and getting organized, the proxy viewpoints model repeatedly guides the discovery process by applying the term expansion technique.

During the requirements analysis phase, requirements relation chain can also be discovered. Requirements verification and validation are done not only to the individual requirement but also to the embedded models in the proxy viewpoints model. Use of missing requirements categorization also enables users to project a set of requirements space from a specific category of enterprise model and viewpoints model. These set of requirements are further assessed to determine missing requirements that belong to the given missing requirements types.

Table 1. PVRD Methodology

Input to the methodology:

R: A set of legacy status requirements documents expressed in natural language

S: Other external resources that are directly/indirectly related to the system implementation, operation, and organization's structural information etc.

Output from the methodology:

R⁺: A set of requirements documents that have reduced level of incompleteness, inconsistency and are well-organized and maintained in the enterprise model

PVM: A proxy viewpoints model that represents multiple perspectives/views of stakeholders of the system domain within the defined scopes of categories in the enterprise model

Process

Step1:

System domain understanding

Identification of initial set of viewpoints $VP = (VP_1,\ldots,VP_n)$ of stakeholders in Viewpoints Model VP.

Step2:

Identification of roles/scopes of each requirements category $EM = (EM_1,\ldots,EM_m)$ in Enterprise model EM.

Step3: Requirements Discovery Process

While **R** is not empty
  Compose an initial query term T_i for searching target requirements documents.
  Retrieve a set of requirements documents $D(T_i)$
  Categorize $D(T_i)$ into relevant(RV) and less-relevant(LV) sets of documents
  Use term expansion algorithm to generate a list of terms L that can characterize RV to LV
  Compose a set of query term(s) T_{qe} from the list L
  T_{qe} represents the abstract concept of requirements of interest
  Retrieve a set of requirements documents $D(T_{qe})$ that contains requirements $r_1,\ldots,r_k$
  Record T_i and T_{qe} in the requirements term thesaurus
  For all requirements $\{r_1,\ldots,r_k\}$
    Analyze each requirement r_i and identify each corresponding VP_i in VP and EM_i in EM.
    If there is no corresponding VP_i exist in VP, Viewpoints Expansion is necessary.
    If there are more than one corresponding VP_is, Viewpoints Reduction/Resolution is necessary.
    Create a $r_i(EM_i, VP_i)$ index and move to the corresponding EM_i storage place.
    Remove r_i from **R**
  End For
End While

Step4: Requirements Analysis Process

For each VP_i in VP and EM_i in EM
  Check if any VP_i or EM_i has empty set of corresponding requirements documents.
  If any empty VP_i found then missing requirements exist for that VP_i
  If any empty EM_i found then missing requirements exist for that EM_i
  Derive and construct missing requirements for such VP_i or EM_i if any
End For

For all EM_i in EM // Verification Step
  For all requirements r_i in EM_i
    Find corresponding more detailed requirements r_d in EM_{i+1}
    If such requirements not found then derive new missing requirements in EM_{i+1}
    Check missing requirements category
    Construct Relation Chain by identifying any important relationships between r_i and r_d
  End For
End For

For all EM_i in EM // Validation Step
  For all requirements r_i in EM_i
    Find corresponding more general requirements r_g in EM_{i-1}
    If such requirements not found derive new missing requirements in EM_{i-1}
    Check missing requirements category
    Construct Relation Chain by identifying any important relationships between r_i and r_g
  End For
End For

4. CONCLUSIONS AND FUTURE WORK

The PVRD methodology is unique in its architecture in that it provides an integrated environment for requirements discovery as well as requirements organization and management through the proxy viewpoints model. Term expansion technique and missing requirements categorization play a major role in the requirements discovery process. Combined viewpoints model and enterprise model establish the new requirements classification mechanism for requirements organization and management. These models and various methods facilitate the requirements discovery, analysis, verification and validation processes.

The PVRD methodology is also very useful as a training purpose for the users (i.e. software/system engineer, data analyst) who are not yet familiar with the requirements domain and associated processes. In real business practice, rapid changes of organization and business result in discontinued knowledge and require a lot of efforts to learn and adapt to such changes. Requirements term expansion technique with proxy viewpoints model can be used to reacquire such domain knowledge and changes.

Based on this research, explanatory scenario-based case studies will be developed in the near future as a way of the PVRD methodology validation and exemplary guidance for both researchers from academia and real industry practitioners.

5. REFERENCES

[1] Alderson, A. and Shah, H. Viewpoints on Legacy Systems. *Communications of the ACM*, Vol.**42**, No.3. pp. 115-116, March. 1999.

[2] Buede, D. M. *The Engineering Design of Systems : Models and Methods* (Wiley Series in Systems Engineering), New York: Wiley, December. 1999.

[3] Cochran, W. G. *Sampling Techniques*, Third Edition, John Wiley & Sons. New York. 1977.

[4] Kirner, T.G. and Abib, J.C. Inspection of Software Requirements Specification Documents: A Pilot Study. In *Proceedings of the 15th annual international conference on Computer documentation*. pp. 161-171. Snowbird, UT. 1997.

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

[6] Kotonya, G. and Sommerville, I. *Requirements Engineering: Processes and Techniques* (Worldwide Series in Computer Science), J. Wiley, New York, September. 1998.

[7] Merriman, M. FIERS: Categorization of missing requirements. *Technical Report*, School of Information Technology and Engineering, George Mason University, Fairfax, VA. 2001.

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

[9] Porter, A.A., Votta, L.G., and Basili, V.R. Comparing Detection Methods for Software Requirements Inspection: A Replicated Experiment. *IEEE Transactions on Software Engineering*, Vol. **21**, No. 6. pp.563-575. 1995.

[10] Rayson, P., Garside, R. and Sawyer, P. Assisting Requirements Recovery from Legacy Documents. *Technical Report* CSEG/8/2000, Cooperative Systems Engineering Group. Computing Department, Lancaster University, United Kingdom. 2000.

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

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

[13] Spanoudakis, G., Finkelstein, A. and Emmerich, W. (Editors). *Joint proceedings of the second international software architecture workshop (ISAW-2) and international workshop on multiple perspectives in software development (Viewpoints '96) on SIGSOFT '96 workshops*. ACM Symposium on Foundations of Software Engineering. October 14-15. San Francisco. ACM Press. 1996.

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