A Hybrid Model for Knowledge Acquisition in Expert Systems

Mohamed A. Shouman⁽¹⁾, Mohamed G. Abou-Ali⁽²⁾ and Ayman M. Mostafa⁽³⁾

- (1) Department of Operations Research. University of Zagazig (Egypt) E-mail: profshouman@hotmail.com
- (2) Department of Industrial Engineering. University of Alexandria (Egypt)
 E-mail: m_abouali59@yahoo.com
 - (3) Department of Information Systems. University of Zagazig (Egypt)

 E-mail: eng ayman m@hotmail.com

ABSTRACT

Information Systems (IS) are based upon data collected by means of questionnaires, interviews, and observation. Inexperienced researchers find questionnaires and interviews attractive as a data gathering methodology. Many researchers have discovered that it is not simple to draft a good questionnaire because their answers are very superficial and impact negatively on the research quality. This paper explores a Repertory Grid technique as an alternative method for gathering meaningful data. Also, a hybrid model between questionnaire technique and Repertory Grid technique is presented. The model uses questionnaire as a primary data gathering technique and then the acquired data are automatically transferred to the Repertory Grid. The proposed model is considered an improvement technique of Repertory Grid because it solves many of its problems such as inability to name all the scales in the grid, the size limitation of repertory grid which is opened in the current model, and the expert have to use all elements and constructs in the grid without the ability to leave some of them.

Keywords: Information Systems, Knowledge Acquisition, Repertory Grid

1- INTRODUCTION

Knowledge acquisition is the process of extracting, structuring and organizing knowledge from one or more sources and transferring it to the knowledge base. Acquiring knowledge from different sources of expertise is considered a bottleneck in building expert systems. So, there are different knowledge acquisition techniques for acquiring knowledge from sources of expertise. Questionnaire is a manual knowledge acquisition technique into which users write short answers to open questions or leave the questions blank. Observation technique is a time consuming technique, needs skill and is difficult to observe and analyze what was observed. Interview technique avoid many problems of questionnaire technique but it is hard to set up specially in structured interview where there must be a pre-defined structure and there is no flexibility in acquiring knowledge from knowledge engineer. Repertory Grid technique is considered as a suitable technique because of its bias towards natural lan-

guage, its domain independence, and its built-in methodology for filling in a single repertory grid. By interacting with the domain experts using natural language, some of the difficulty of the knowledge acquisition process is eased. In addition to the cited features, analysis of the elicited requirements from interviewing with repertory grids would provide better understanding about the domain elements. The repertory grid technique faces problems during its processing like inability to name all the scales used in the acquisition process, experts must use all the elements and constructs in the grid without the ability to leave or add some of them for future using, and there is a difficulty in understanding the cluster analysis technique that is used in analyzing repertory grid.

2 - REPERTORY GRID TECHNIQUE

The Repertory Grid is a matrix based technique although it is more complex than simply filling in a matrix of elements. The repertory grid acquires a set of elements and compares elements against their constructs. This technique is used in many fields for eliciting and analyzing knowledge. The researchers used the Repertory Grid technique as a possible solution to their quest to find a more suitable means of collecting data and subsequently analyzing it both for use by themselves and inexperience researchers. Repertory Grid is considered one of the most popular indirect knowledge acquisition techniques which is a knowledge analysis technique derived from Kelly's personal construct theory (PCT) [1]. This theory has been validated and it has been concluded that [1]: -

- 1. Individuals represent their environment using constructs.
- 2. Constructs are organized in interrelated structures which change from time to time.
- 3. The repertory grid technique elicits these constructs accurately and reflects the changes in an individual's construct system over time.
- 4. The grid technique elicits the true structure and organization of the individual's construct system.

2-1REPRETORY GRID DEFINITION AND OBJECTIVES

The Repertory Grid technique is 'a means of surfacing people's perceptions, attitudes or concepts in an uncontaminated way'. The researchers look more thoroughly at the Repertory Grid as a possible solution to their quest to find a more suitable means of collecting data and analyzing it for the creation of knowledge base that is considered the primary component of building expert systems [2]. The repertory grid is considered as an interviewing technique which uses factor analysis to determine an idiographic measure of personality. Factor analysis is a collection of methods used to examine how underlying constructs influence the responses on a number of measured variables [3]. It

can be used to explain variability among observed variables in terms of fewer unobserved variables called factors. The observed variables are modeled as linear combinations of the factors, plus "error" terms. Factor analysis assumes that all the rating data on different attributes can be reduced down to a few important dimensions. The primary use of the Repertory Grid is to investigate or reveal attitudes, beliefs, concepts, assumptions, and perceptions. These personal views are known as constructs. The Repertory Grid is used for evaluation, to compare constructs and hence reveal bias. The comparison of the constructs that participants have proposed gives the technique useful diagnostic qualities and allows assessment of the perceived contribution of that construct to the overall goal [4]. Each construct has two values, one value in the left hand side of the repertory grid and its opposite is put on the right hand side of the repertory grid. The objects that are compared with each construct are called elements.

2-2 REPERTORY GRID COMPONENTS

Every grid consists of four components. These components are topic, constructs, elements, and ratings. A grid is always conducted about a particular topic, with the intention of eliciting just those constructs which the person uses in making sense of his domain. The main basic unit of description and analysis is called a construct. Constructs are qualities that describe the elements. Constructs can be generated in several ways; they can be elicited from triads and supplied by the researcher or combine elicited and supplied constructs [5]. Elements are the objects of an individual's thinking to which they relate their concepts or values. Finally, each element is rated on each construct, to provide an exact picture of what the person wishes to say about each element within the topic.

2-3 REPERTORY GRID MECHANISM

In order to create a Repertory Grid, three stages must be satisfied [6]: - Stage 1

The elements are selected for the grid. It is found the number of elements used by researchers generally ranged from eight to 30. In destination studies the number of elements has ranged from 6 to 40. For the technique to be successful and not to take too much time to operate, the number of elements should be not more than 15 objects [7]. A set of about the same number of constructs is also required.

Stage 2

This stage involves the rating of each element against each construct. A numerical scale is used for the rating process. The best recommended scales are (2, 3, 5, 7, and 9). Figure 1 exhibits the resulting grid created after stage 1 and stage 2.



Figure 1 A Sample Repertory Grid

Stage 3

The ratings are applied to a statistical calculation called Hierarchical Cluster Analysis (HCA) to create the analyzed grid. These calculations used for finding homogeneous clusters of cases based on measured characteristics. It stars with each element or construct in a separate cluster and then combines the clusters sequentially, reducing the number of clusters at each step until only one cluster is left. When there are N elements or constructs, this involves N-1 clustering steps [8]. Figure 2 exhibits the resulting analyzed grid.

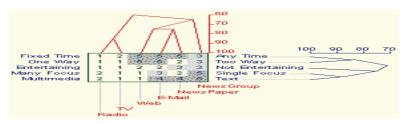


Figure 2 Cluster Analysis of the Repertory Grid

2-4 REPERTORY GRID STRENGTHS AND WEAKNESSES

The Repertory Grid Technique has the following strengths [2]: -

- 1. Repertory grid is flexible because it elicits both qualitative and quantitative data.
- **2.** This technique emphasizes relevance as it focuses on the perceptions and personal constructs.
- **3.** Repertory grid is considered as an improvement on questionnaires where the researcher formulates the questions and decides unilaterally which issues are important.
- **4.** The participant suggests both elements and constructs based on his own words experience.
- **5.** Repertory grid technique is less resource hungry than participant observation.
- **6.** It provides a structured approach which is easy to explain and relatively quick to complete.
- **7.** The common constructs and the scores assigned by the individuals allowed the group to compare their assessments and then to analyze differences in interpretation of the constructs.

The following weaknesses could be noted as [2, 9]: -

- 1. Participants may have misconceptions leading to invalid conclusions.
- 2. The interpretation of the data might not be significantly easier when the

repertory grid is used in an ideographic or qualitative way.

- **3.** The larger and more reliable the grid, the longer it takes to complete.
- **4.** If the number of interviewees involved, more time is necessary for the analysis process.
- **5.** The information which had been acquired flexibly is applied inflexibly.

3- THE PROPOSED HYBRID MODEL

When the domain expert deals with the repertory grid technique, he or she must enter the scaling rate, the number of elements (objects), the number of constructs (attributes), and finally the scale number for each element against its constructs. The resulting grid is analyzed without any intermediate explanation for the analysis process. It is a solid and a boring process for acquiring data. In this paper, a hybrid model between Questionnaire technique and Repertory Grid technique is presented. The proposed model uses the Questionnaire technique as a primary data gathering technique for the purpose of acquiring data in an efficient and flexible way, then the acquired data are automatically transferred to the Repertory Grid. In the proposed model the data are filled automatically without an intermediate way for asking questions about the data acquired. The number of objects (elements) and attributes (constructs) are not limited. The expert enters elements and constructs of the repertory grid with no unused elements and constructs. In such structure of the proposed model, the analysis processes of Hierarchical Cluster Analysis (HCA) of repertory grid is not vague and of explanatory and the intermediate scales are of meaningful. This can overcome the different interpretation of the meaning of the scale [6]. The major steps of knowledge acquisition process using the proposed hybrid model is explained in figure 3.

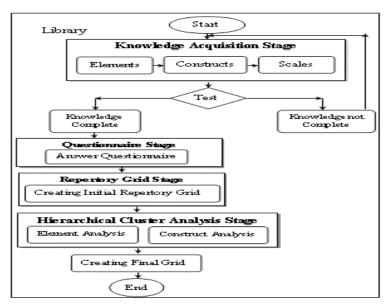


Figure 3 The Hybrid Model Flowchart

4- CASE STUDY IMPLEMENTATION

In the proposed hybrid model, data in the field of computer science are acquired. A size of 8 elements * 7 constructs is used. The data are acquired using Questionnaire technique as an intermediate layer for acquiring data then these data are transferred automatically to the Repertory Grid technique. The curriculums in the field of computer science are presented as elements in the Repertory Grid technique with their main constructs. Constructs are entered into the repertory grid technique in the form of an attribute in the left hand side (LHS) and its opposite in the right hand side (RHS). The hybrid model was created using Microsoft Visual Studio 2005 Windows Application because it adds innovative language constructs, new compiler features, enhanced developer productivity, and an improved debugging experience.

5- THE HYBRID MODEL ANALYSIS

The proposed model is illustrated in two stages: the knowledge acquisition stage and the analysis stage using hierarchical cluster analysis.

5-1 KNOWLEDGE ACQUISITION STAGE

In this stage the element library, construct library and scale library are acquired. This stage is considered as an improvement over the traditional repertory grid because the expert has the ability to insert large number of elements, constructs and scales into the hybrid model and acquires some or all of them according to his or her requirements.

STEP 1 CREATING ELEMENT LIBRARY

The proposed model starts by entering all the elements available and saves it into an element Library until the expert uses some or all of them according to his or her requirements.

STEP 2 CREATING CONSTRUCT LIBRARY

The proposed model enters all the constructs available and saves it into a construct Library until the expert uses some or all of them according to his or her requirements. This is considered an improvement over traditional repertory grid.

STEP 3 CREATING SCALE LIBRARY

Figure 4 explains the entry operation of the number of scales (2, 3, 5, 7, 9, or 11) with the name of each scale. This operation solves the problem of different interpretation of the intermediate scales in traditional repertory grid because each scale number has its unique name which is never used by any other scale number.

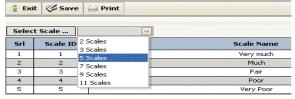


Figure 4 A screen for entering the number of scales with their names

STEP 4 QUESTIONNAIRE TECHNIQUE

Figure 5 explores the questionnaire technique that serves as an intermediate layer between domain experts and the other components of the model. It can be used for entering information about the element **Organizational Behavior** (**OB**). The expert must select only one choice for each construct. The process continues for the next element until all required elements are satisfied. This is explained in figure 6.

	Very much-1	Much-2	Fair3	Poor4	Very Poor-5	
Other People in preparation					~	I can prepare this by my self
Succeed with no preparation			~			Require some preparation
May bore the students					~	Interests the students
Less injoyable for me					~	More injoyable for me
Need Pre-course			~			Doesn't need Pre-course
Expensive				~		Economical
Require equipement					~	No equipement

Figure 5 A Questionnaire step for entering each element against its constructs

STEP 5 CREATING INITIAL REPERTORY GRID

Figure 6 explains the initial grid resulted from the questionnaire operation. The creation of this grid is more flexible than creating it using traditional repertory grid because the user answers the Questionnaire in a flexible way and the answers are converted automatically to the Repertory Grid. The grid is analyzed to determine how each element and construct are close to each other.

	Subject Curriculum								
	OB	Mathematics	Physics	Data Structure	Image Processing	Network	Artificial Intelligence	GIS	
Other People in preparation	5	1	3	1	3	3	2	2	I can prepare this by my self
Succeed with no preparation	3	2	4	2	2	4	3	3	Require some preparation
May bore the students	5	4	2	4	2	2	5	5	Interests the students
Less injoyable for me	5	3	1	4	3	4	3	2	More injoyable for me
Need Pre-course	3	4	5	2	4	5	2	3	Doesn't need Pre-course
Expensive	4	2	4	3	2	2	4	1	Economical
Require equipement	5	1	3	2	1	1	2	2	No equipement

Figure 6 Initial Repertory Grid Resulted from Questionnaire Step STEP 6 ELEMENT ANALYSIS OPERATIONS

In this step, each element is analyzed against all its constructs. As explained in figure 6, the element named by **Mathematics** has a scale (1) with the two constructs named by **other people in preparation** and **require equipment**. The element **Mathematics** has no scale (5), so it takes (Null)...etc

STEP 7 CONSTRUCT ANALYSIS OPERATIONS

In this step, each construct is analyzed against all its elements. As explained in figure 6, the construct named by other people in preparation – i can prepare this by myself takes scale (1) with the two elements named by Mathematics – Data Structure and this construct has no scale (4), so it takes (Null)...etc

STEP 8 SCALE ANALYSIS OPERATIONS

In this step, each scale is analyzed against its elements and constructs at the same time. With scale (1), the element named by **OB** takes **null** because it doesn't take scale 1 in figure 6 while element **Mathematics** takes scale (1) with the two constructs **other people in preparation** and **require equipment**.

5-2 HIERARCHICAL CLUSTER ANALYSIS STAGE

5-2-1 FLEMENT ANALYSIS

Using Hierarchical Cluster Analysis, the degree of similarity and dissimilarity between elements is calculated using the following equation: -

Similarity Score between Elements =
$$100 - (\frac{SD}{(LR-1)*C})*100$$
 (1)

Where **SD** is the sum of absolute values between each two elements, **(LR-1)** is largest rating scale (5 on a 5-point scale; 7 on a 7-point scale) minus 1, and **C** is the number of constructs. The similarity score between element **OB** and element **Mathematics** is required to be calculated, the **absolute value** between them from figure 6 should be evaluated first: -

OB-Mathematics =
$$4 + 1 + 1 + 2 + 1 + 2 + 4$$

Sum of Difference (SD) = 15

On a 5-point scale, (LR-1) = 5 - 1 = 4

Similarity Score =
$$100 - (\frac{15}{4*7})*100 = 46.4\%$$

This process will be continued until the entire upper triangular matrix in figure 7 is calculated. It is clear that values under the diagonal are the same over the diagonal. From figure 7, the similarity score between elements **Mathematics**—**Image Processing** and elements **Image Processing—Networks** = <u>85.7%</u>. This means that these elements are closely related to themselves and this is highly relevance in practice.

Subject Curriculum								
	ов	Mathematics	Physics	Data Structure	Image Processing	Network	Artificial Intelligence	GIS
ОВ		46.4	50	57.1	46.4	46.4	67.9	57.1
Mathematics			53.6	82.1	85.7	71.4	71.4	75
Physics				50	67.9	75	60.7	57.1
Data Structure					67.9	60.7	82.1	71.4
Image Processing						85.7	64.3	67.9
Network							57.1	60.7
Artificial Intelligence								82.1
GIS								

Figure 7 Similarity Score between elements in percentage (%)

5-2-2 CONSTRUCT ANALYSIS

Using Hierarchical Cluster Analysis, the degree of similarity and dissimilarity between constructs is calculated. It is clear that constructs have to sides; left hand side and right hand side. The reverse relationship is involved. The conventional way to do this is to spread the range of possible percentages over a 200-point scale. This means that you multiply the proportion by 200, not 100 [5]. The equation for construct analysis is: -

Similarity Score between Constructs = $100 - (\frac{SD}{(LR-1)*E})*200$ (2) Where SD is the sum of absolute values between each two constructs, (LR-1) is largest rating scale (5 on a 5-point scale; 7 on a 7-point scale) minus 1, and E is the number of elements. The similarity score between construct named by **other people in preparation** and construct named by **succeeds with no preparation** is required to be evaluated, the **absolute value** between them is evaluated first from figure 6: -

Other People in preparation - Succeeds with no preparation=2+1+1+1+1+1+1+1 Sum of Difference (SD) = **9**

A reverse relationship is calculated. The second construct named by **succeeds with no preparation** is converted and its opposite named by **requires some preparation** is taken in the following rule [5]: -

5- Point scale (1 5, 2 4, 3 3) because 3 is the intermediate scale so it remains as it is.

Requires some preparation = 3 4 2 4 4 2 3 3

Other People in preparation - Requires some preparation = 2+3+1+3+1+1+1+1

Sum of Difference (SD) = 13

The smallest number from this operation = (9) is selected.

So, the construct other People in preparation - succeeds with no preparation is selected.

On a 5-point scale, (LR-1) is = 5-1 = 4
Similarity Score =
$$100 - \frac{9}{4*8} * 200 = \underline{43.8\%}$$

The similarity score between construct **other people in preparation** and construct **may bore the students** is evaluated. The **absolute value** between them is evaluated first from figure 6: -

Other People in preparation – May bore the students = 0+3+1+3+1+1+3+3Sum of Difference (SD) = **15**

A reverse relationship is calculated. The second construct **may bore the students** is converted and its opposite **Interests the students** is taken in the same rule as mentioned.

Interests the students = 1 2 4 2 4 4 1 1

Other People in preparation – Interests the students= 4 + 1 + 1 + 1 + 1 + 1 + 1 + 1 + 1

Sum of Difference (SD) = 11

The smallest number from this operation = $\underbrace{ 11}$ is selected.

So, the construct other people in preparation - interests the students is selected.

On a 5-point scale, (LR-1) = 5-1 = 4

Similarity Score =
$$100 - \frac{11}{4*8} * 200 = 31.25\%$$

This process is continued until the entire upper triangular matrix in figure 8 is determined. It is clear that values under the diagonal are the same over the diagonal.

	Subject Curriculum								
	Other People in preparation	Succeed with no preparation	Interests the students	Less injoyable for me	Need Pre-course	Expensive	Require equipement		
Other People in preparation		43.8	31.2	43.8	25	37.5	68.8		
Succeed with no preparation			50	37.5	56.2	56.2	37.5		
Interests the students				12.5	43.8	31.2	12.5		
Less injoyable for me					18.8	31.2	25		
Need Pre-course						12.5	6.2		
Expensive							43.8		
Require equipement									

Figure 8 Similarity Score between constructs in percentage (%)

6 - RESULTS

The final hybrid model grid resulted from the process of (Hierarchical Clustering Analysis) of elements and constructs is exhibited in figure 9. The proposed hybrid model has the ability to be used for unlimited number of elements and constructs. To make sure that the final hybrid model grid is working properly, the final grid is examined against similarity scores of both elements and constructs.

	Mathematics	lmage Processing	Network	Physics	Data Structure	Artificial Intelligence	OB	GIS	
Less injoyable for me	3	3	4	1	4	1	5	2	More injoyable for me
Require equipement	1	1	1	3	2	2	5	2	No equipement
Other People in preparation	1	3	3	3	1	2	5	2	I can prepare this by my self
Need Pre-course	4	4	5	5	2	2	3	3	Doesn't need Pre-course
Succeed with no preparation	2	2	4	4	2	3	3	3	Require some preparation
Expensive	2	2	2	4	3	4	4	1	Economical
Interests the students	2	4	4	4	2	1	1	1	May bore the students

Figure 9 Final Hybrid Model Grid

From figure 7, the highest similarity score is (85.7%) between elements Image Processing – Mathematics and elements Networks – Image Processing. This means that they are closely related to themselves as resulted in figure 9.

From figure 8, the highest similarity score is (68.8%) between the constructs other people in preparation – require equipment. The second highest similarity score is (56.2%) between both expensive – succeed with no preparation and need precourse - succeed with no preparation. This means that they are closely related to themselves as resulted in figure 9.

From figure 9, each rating score (1, 2, 3, 4, or 5) is very near to his similar score. This can explain that the resulted hybrid model is in a high dregree of accuracy and effeciency.

In order to ensure the effeciency of the resulted grid and be certain that the hierarchical cluster analysis has been executed more accurately, it is better to compare the resulted hybrid model grid with the output of the traditional grid. Figure 10 exhibits the output of the traditional repertory grid.

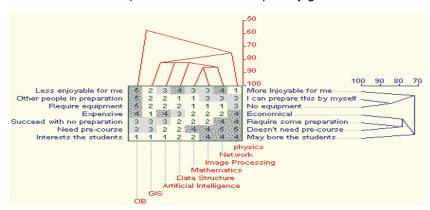


Figure 10 Traditional Repertory Grid Output

For more improvement in the proposed hybrid model, a graphical representation from the final hybrid model grid is presented in figure 11. This graphical representation takes each element and connects it to its constructs through the scale name.

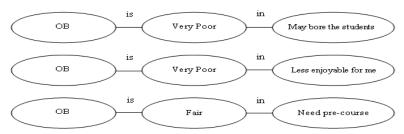


Figure 11 A graphical representation for elements, scales, and Constructs

100

As it is exhibited from the proposed hybrid model, the execution time of the analysis is calculated by determining the time resulted from elements - constructs analysis. Different number of elements and constructs are used to calculate the execution time is presented in table 1. Table 2 presents the execution time of traditional repertory grid which is limited to 15 elements and 15 constructs. It is clear that the hybrid model saves much time in the analysis operation than the traditional one.

	No of Elements									
		5	10	15	20		100			
	5	109ms	125ms	140ms	150ms					
ucts	10	125ms	146ms	156ms	172ms					
onstı	15	141ms	160ms	187ms	207ms					
io of Constructs	20	156ms	187ms	219ms	250ms					

Table 1 Execution time between Elements – Constructs for proposed hybrid model

Table 2 Execution time between Elements – Constructs for traditional repertory grid

	No of Elements								
-25		5	10	15					
structs	5	332 ms	585 ms	639 m <i>s</i>					
No of Constr	10	739 ms	765 ms	908 m <i>s</i>					
2	15	859 ms	1023 ms	1133 ms					

A graph of the relationship between elements- constructs and time is presented in figure 12. From figure 12, it is clear that the execution time resulted from the analysis process of more than 15 elements and 15 constructs is not increasing incredibly.

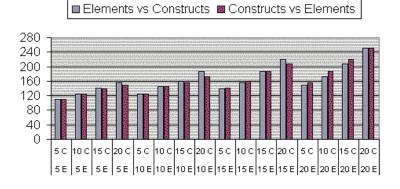


Figure 12 Execution Time for Element and Construct Analysis for proposed hybrid model

7- CONCLUSIONS

Repertory Grid Technique is an interview technique that utilizes individual's ability to compare elements, elicit attitudes, category making, assessing criteria and some personal tacit knowledge. It is a qualitative method in which statistical approaches may be used to enhance analysis. A hybrid model between questionnaire technique and Repertory Grid technique is presented. This model uses questionnaire as a primary data gathering technique then the acquired data are automatically transferred to the Repertory Grid. This model is considered as an improvement of Repertory Grid because it solves many of its problems such as creating a library for elements, constructs, and ratings for storing large number of them, the ability to acquire some or all elements, constructs or ratings for future using, and the ability to name all scales in the grid. In the proposed model, the grid size has become unlimited, and the final grid has become more efficient than the traditional repertory grid. For more enhancements in the proposed hybrid model, an additional category for generating rules from the final grid can be added. This category can classify the hybrid model as an automated tool of knowledge acquisition.

REFERENCES

- [1] P. Crowther, J. Hartnett, and R. Williams, "Teaching Repertory Grid Concepts for Knowledge Acquisition in Expert Systems: An Interactive Approach", proc. Ascilite International Conference, Melbourne, Australia, 2007
- [2] P. Alexander and J.V. Loggerenberg, "The Repertory Grid: "Discovering" a 50-year-old Research Technique", proc. ACM International Conference, South African Institute for Computer Scientists and Information Technologists White River, South Africa, pp.192-199, 2005
- [3] T.W. Schubert, F. Friedmann, and H.T. Regenbrecht, "Decomposing the sense of presence: factor analytic insights", proc. Second International Workshop on Presence, University of Essex, Colchester, 1999
- [4] F.B. Tan and M.G. Hunter, "The repertory grid technique: A method for the study of cognition in Information Systems", MIS quarterly journal, Vol. 26, no. 1, pp. 39-57.7, 2002
- [5] Jankowicz D., *The Easy Guide to Repertory Grids*, John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, England, pp. 101-115, 2004
- [6] N. Niu and S. Easterbrook, "Discovering Aspects in Requirements with Repertory Grid", proc. ACM International Conference on Software Engineering, Shanghai, China, pp.35-42, 2006
- [7] S.D. Pike, Repertory Grid Analysis in group settings to elicit salient destination image attributes, Multilingual Matters & Channel View Publications, Vol. 10, no. 4, pp. 378-392, 2007
- [8] N.L. Zhang, "Hierarchical Latent Class Models for Cluster Analysis", Machine Learning Journal, vol. 5, no. 26, pp. 697-723, 2004
- [9] R. Wolf and H.S. Delugach, "Knowledge Acquisition via Tracked Repertory Grid", CiteSeer Journal. Scientific Literature Digital Library, 1996

Int.J. of Software Engineering, IJSE Vol.2 No.1 January 2009