International Journal of Innovative Research in Information Security (IJIRIS)
Issue 2, Volume 5 (May 2015)
ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

AN ONTOLOGY-BASED TEXT MINING METHOD TO CONSTRUCT D-MATRIX FOR FAULT DETECTION AND DIAGNOSIS USING GRAPH COMPARISON ALGORITHM

Ms. Madhuri M. Varma.
Department of Computer Engineering
Smt. Kashibai Navale College Of Engg.,Vadgaon
Pune-411041, India

Prof. Jyoti Nandimath.
Department of Computer Engineering
Smt. Kashibai Navale College Of Engg.,Vadgaon
Pune-411041, India

Abstract: ***Fault dependency (D)-matrix is a systematic analytic model to catch the different system-level fault diagnostic information consisting of conditions between recognizable symptoms and failure modes connected with a system. Constructing a D-matrix from earlier principles and redesigning it utilizing the space learning is a difficult task and time intensive assignment. Further, in-time augmentation of D-matrix through the revelation of new symptoms and failure modes watched for the first run through is a testing assignment. An ontology based D-matrix[1] depict an ontology based text mining strategy for consequently constructing and redesigning a D-matrix by mining countless repair verbatim (normally written in unstructured text) gathered during the diagnosis scenes. The system constructs the fault diagnosis ontology consisting of concepts and relationships commonly saw in the fault diagnosis space. Next, utilize the content mining calculations that make utilization of this ontology to distinguish the fundamental ancient rarities, for example, parts, symptoms, failure modes, and their conditions from the unstructured repair verbatim content. In our methodology, we first construct the D-Matrices from [1] for different datasets. Next, we generate graph model for each generated d-matrix and used only common patterns from generated graph and develop new graph. And, after this procedure, we create D-matrix depending upon data given by the similarity graph.***

Keyword: Fault diagnosis ontology, D-matrix, Text Mining, Unstructured text, Graph comparison algorithm.

I. INTRODUCTION

Typically, created system lives up to expectations in its pre recognized working conditions and done with its given assignment as expressed. There is nothing to stress over the system until it is working accurately and gives worthy results. In the event that the consequence of the system is not as every the expectation we can say this as introduction of fault in system. Identification of faults and its correction is a subfield of control designing which relate itself with managing a system, recognizing when a fault has happened, what are the reasons behind it and discover the sort of fault and its location [2]. It is important to discover the underlying driver of a fault on the grounds that there may be possibility that other interconnected subsystems might likewise give fault indications that may conceivably shroud the underlying driver.

A complex system interfaces with its enveloping to execute a set of endeavors by keeping up its execution inside a satisfactory extent of resilience. Any deviation of a system from its satisfactory execution is managed as a deficiency. The fault detection and diagnosis (FDD) is performed to find the fault and diagnose the hidden drivers to minimize the downtime of a system. On account of relentlessly creating mechanical advancement that is embedded in the vehicle systems, for illustration propelled programming embedded systems, demonstrative sensors, web, along these lines forward the procedure of FDD becomes a testing development in the occasion of part or system glitch. Systems like On-board diagnostics (OBD) alluding to a vehicle's demonstrative toward oneself and reporting ability. These systems give mechanics or manager of the vehicle access to the status of the different vehicle sub-systems [3]. In the wake of determining the different issues it is essential to note down its causes, impact of the reason on the system in organized way so we can utilize this information recent while creating the system to make it perfect. But when information which we have assembled is in colossal sum there is need to store this unstructured information pertinently with the goal that we can discover faults by gathering its partner symptoms. The reason for Text Mining is to process unstructured data, get out compelling files from the literary data, and make the information available in the data receptive to the different data mining calculations. We are utilizing D-matrix in light of the fact that it is one of the standard indicative models tagged in IEEE Standard 1232[4]. Be that as it may, the development of a D-matrix by utilizing text mining is a testing assignment part of the way because of the commotions saw in the repair verbatim content information.

Regularly the procedure of FD begins by extracting the mistake codes from a target framework and focused around the watched lapse codes the professionals take after particular finding technique alongside their experience to diagnose the shortcomings. At the time of fault determination, a few information sorts are gathered, for example, error codes, examined benefits of working parameters connected with broken part/framework, repair verbatim, and many more. The gathered information is then exchanged to the OEM database and especially the repair verbatim information gathered over a time of time can be mined to create the D-matrix analytic models. Such models can be utilized by the field professionals and other stakeholders to perform precise FDD.

© 2014-15, IJIRIS- All Rights Reserved

ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

II. LITERATURE SURVEY

Ontology-based fault diagnosis for power transformers by Prof. D. Wang proposed an methodology to transformer fault diagnosis is presented focused around the thought of trading information with express, formal and machine open descriptions of importance by utilizing the Semantic Web[5]. An ontology model is created for precise and effective fault diagnosis for power transformers. Through the utilization of this model, different transformer fault symptomatic strategies can be incorporated to depict and derivation among fault phenomena, fault sources and reasons for faults. . Previously, the new concepts can be added into the ontology based on domain [6],but it requires some methods for evaluating and updating the result. So, it is very lengthy process.

Prof. Singh had implemented a paper based on Dynamic multiple fault diagnosis used for performing diagnostic inference for multiple failure modes used in aircraft and automobiles require multi-state component models with multiple test outcomes to reduce complexity in detecting multiple failures and also need to improve convergence[7]. In [8], the matrix is generated for the item sets based on tags, but failed to generate accurate result. Also Saeed Hassanpour [9], due to the informative nature of ontological hierarchies, concept of inter-relationship, it has main focus on identifying only the relevant parts of the text. To concentrate applicable ontology concepts and their relationships from a learning base of heterogeneous content records are proposed by Prof. S. Strasser , a maturation approach which utilizes the diagram theoretic representations of Timed Failure Propagation Graph models and indicative sessions focused around as of late institutionalized demonstrative ontologies to focus measurable inconsistencies between that which is normal by the models and that which has been experienced practically speaking.

In the existing the specialty of fault displaying [10], [11], [12], [13], the restricted endeavors are seen to construct a D-matrix by dissecting unstructured repair verbatim information. Only as of late [16], the device is suggested that finds the information by producing pertinent arrangements from the on-board diagnosis and support information by utilizing the ontology-based information mining. In any case, the basic models proposed in are thought to be finished and static, however in true because of configuration and designing changes and new vehicle structural planning dispatches the new manifestations and disappointment modes are watched making such models outdated. In existing framework construction of D-matrix is done by physically or utilizing first standard. Traditionally, the D-frameworks are constructed by utilizing the history information, building information, and tangible information [10], [11], [12], [13], [14], [15] ,for instance, yet an almost no understanding is given about the revelation of new side effects and disappointment modes watched surprisingly and their inclusion in the D-matrix models.

While developing the D-matrix models, the condensing are disambiguated by consolidating the conflicting disappointment modes into a solitary, predictable term, which gives homogeneous, predictable fault models.

III. PROPOSED SYSTEM

This paper proposes An Ontology-Based Comprehensive D-Matrix Using Graph Comparison Algorithm . This system comprises the developments of D-matrix from the repair verbatim data. After the creation of the D-Matrices from the different datasets, the graph is generated for each D-Matrix. Then, the graphs are combined such that only common patterns are merged from the generated heterogeneous D-Matrices to construct a single, generic D-Matrix.
To construct the D-matrix , following steps has to be created

1. The fault diagnosis Ontology by using dataset.
2. Ontology-based Text Mining.

Ontology-based Text Mining describes three steps, such as Document Annotation, Term Extraction and Phrase Merging.

The proposed system creates two D-Matrix for two dataset respectively. Then, the undirected graph is generated depending on the D-Matrix. At first, the repair verbatim data focuses are collected by recovering them from the OEM's database, which are recorded during field FD. In the first step, the terms, for example, part, symptom, and failure mode, applicable for the D-Matrix are annotated from each one repair verbatim by developing the document annotation calculation. A repair verbatim comprises of a few parts, symptoms, failure modes and actions and the right affiliations must be established between the important terms based on their vicinity with one another. Here, a repair verbatim is first part in different sentences by utilizing the sentence boundary detection principles and the terms showing up in the same sentence are co-related with one another. At last, Naive Bayes likelihood model is developed to disambiguate the abbreviated terms by considering the connection in which they are specified.

From each one annotated repair verbatim the tuples, for example, parts - Pa $\in$ { $P_1, P_2, \ldots . P_i$}, symptoms - Sb $\in$ $\{S_1, S_2, \ldots . , S_j\}$, failure modes - fc $\in$ $\{f_1, f_2, \ldots . f_k\}$, and$(S_b P_a - f_c) \in$ $\{S_1 , P_1 - f_1, S_2, P_2 - f_2, \ldots .., S_j, P_i f_i, S_j, P_j - f_j\}$are constructed by utilizing the term extraction calculation to populate a D-matrix. Toward the end of this step, a few tuples are constructed yet all of them are not discriminating to diagnose the faults observed in relationship with a particular framework. The aggregated normalized recurrence of the tuples is calculated and the tuples with their recurrence over a particular threshold are kept as the valid tuples.

© 2014-15, IJIRIS- All Rights Reserved

ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

Next, the Phrase Merging is used to avoid uncertain references of the failure mode phrases, where the failure mode expresses that are composed by utilizing a conflicting vocabulary. The logical data co-happening with the expressions, i.e., parts, symptoms, failure mode, and actions is used to gauge the conditional probabilities and the expressions with their likelihood score over the particular threshold are merged. At long last, the recently constructed D matrix is audited by topic specialists (SMEs) to identify the discovery of new symptoms and failure modes.

A. ALGORITHM

In our system, the graph merging algorithm is used which takes the generated D-Matrix from [1] as input. The same procedure [1] for constructing the D-matrix is done in our proposed system i.e. Ontology-Based Comprehensive D-Matrix Using Graph Comparison Algorithm, for two times for two repair verbatim data. Now, the graph is deployed by using the d-matrix .The columns and rows from the d-matrix are treated as the vertex for the graphs.

As the d-matrix shows the dependencies in the binary format, the edges for the graphs is decided by using this binary information i.e. if 1 is the output in D-Matrix for specific column and row, then, the edge is formed between that column and row. In this manner, the graph is formed from the developed D-Matrix. Next, our system compare two graphs and merged the common details appears in both graph. For this, the system used the graph merging algorithm.

The working of graph merging algorithm includes: First, the graph comparison is done. To compare two graphs it is fundamental with identify corresponding vertices. Then a rundown of correspondence between the vertices are formed as a situated of virtual edges that join the vertices over the two different charts under consideration .In our project, the list of vertex is the list of rows and columns from the two D-Matrix. A schematic representation of the graph comparison is to detect the correlated neighborhood likenesses in two graphs, given a rundown of correspondence between vertices from the two graphs. Initially, each one sets of corresponding vertices is a different group. Secondly, similar gatherings are merged logically by single linkage with a given measure of similitude. After this merging, the similar data from two matrix are gathered. By using this resultant data, we generate newly graph(i.e. similarity between two graphs).This newly graph is used to construct the final D-Matrix. This D-Matrix is the final output for our project which is emerged from the heterogeneous D-Matrices.

B. SYSTEM ARCHITECTURE

The existing system [1] creates the D-Matrix for one dataset. It provides accurate d-matrix but graph is not generated from the matrix. So that, every time, the new d-matrix is created for the dataset. Even if the different datasets contains some similar data, the new D-Matrix is developed for each dataset.

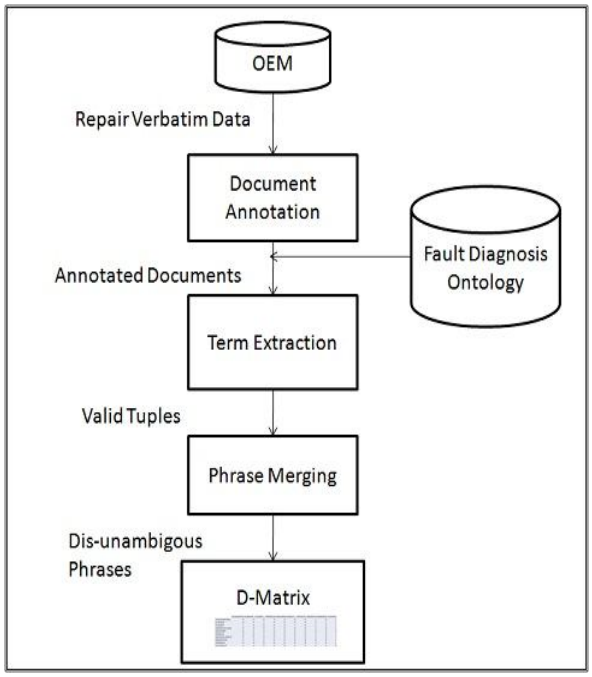

Fig. 1.Text-driven D-matrix development methodology from unstructured text.

© 2014-15, IJIRIS- All Rights Reserved

ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

Our proposed system provides contribution to the existing system. This can be illustrated in fig.2.In our proposed system, first the graph is created for each dataset and then, the similarity between two graphs is found by using graph comparison algorithm. Then, the single D-Matrix is generated by combining different graphs.

C. MATHEMATICAL MODEL

System S is represented as S = {D, A, T, M, F,O}

A]. Database

D= {d1, d2, d3, ...,dn}

Where, D is the set of documents and d1, d2, d3, ...,dn are the number of documents.

B]. Fault diagnosis ontology

F={G, I}

G= {g1, g2, g3, ...gn }

Where G is represent as a set of concepts and attributes and g1, g2, g3, ^agn is a number

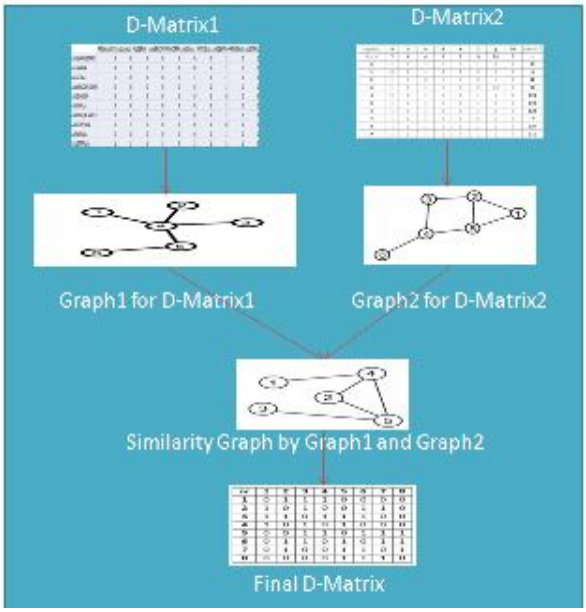

Fig. 2. Graph Comparison to Develop Single D-Matrix

of concepts and attributes.

I= {i1, i2, i3, ..., in }

Where I is the set of instances and i1, i2, i3,...,in represent as a number of instances.

C]. Document Annotation

A= {P, C}

Where P= {W, E}

Where, P is represent as a set of preprocessing and

W= {w1, w2, w3, ...,wn}

Where W is represent as a set of Stop Words and w1, w2, w3, ...wn number of stop words and

E= {e1, e2, e3, ...,en}

Where E is represent as a set of steaming words and e1, e2, e3, ...en is a number of steaming word.

C= {c1,c2,..,cn}

Where C is the set of corpus annotated document and c1, c2^acn represent as a number of corpus annotated document.

D]. Term Extraction

T= {U, V}

Here T is a set term extraction and

© 2014-15, IJIRIS- All Rights Reserved

ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

U= {u1, u2, u3, ...un }
Where U is represent as a set of feature terms and u1, u2, u3, ...,un is a number of feature terms.
V= {v1, v2, v3, ...vn }
Where V is represent as a set of valid correlations and v1, v2, v3,....vn is a number of valid correlations.

E]. Phrase Merging

M = {X, R}
Where M is represent as a set of Phrase Merging
X={x1, x2, x3, ..,xn }
Where X is represent as a set of context information and x1, x2, x3, ...xn is a number of context information.
R={r1, r2, r3, ...rn }
Where R is represent as a set of merged phrases and r1, r2, r3, ...rn is a number of merged phrases.

F]. Graph Comparison Algorithm

O= {o1, o2, o3, ...on }
Where O is the set of D-Matrix graphs and o1, o2, o3, ...,on represent as a number of D-Matrix graphs.
Q= {q1,q2,q3….qn}
Where Q is the set of similarities in the D-matrix graphs and q1,q2,q3,….,qn is the number of similarities in D-matrix graphs.

IV. EXPERIMENTAL WORK AND RESULTS

The proposed system compares two graphs generated by separate d-matrix and merging the similar patterns to develop the single d-matrix. So, it has to be tested likewise to check the performance of the text-driven d-matrix against the historical data-driven d-matrix. For this, two testability and diagnosability measurements have been utilized to focus the execution of the content driven D-matrix—fault detection, fault isolation.

A. Fault detection (P_D):

Fault detection characterized as the percent of deficiencies discovered by the symptoms by watching the failure modes connected with a system. It was utilized to focus the deficiency scope and evaluate undetected shortcomings to focus worthiness. Short of what 100% shortcoming detection demonstrated that there are shortcomings, which can't be recognized by the accessible symptoms. In our proposed system, the final output i.e. D-matrix is also tested against historical data- driven D-matrix for fault detection.

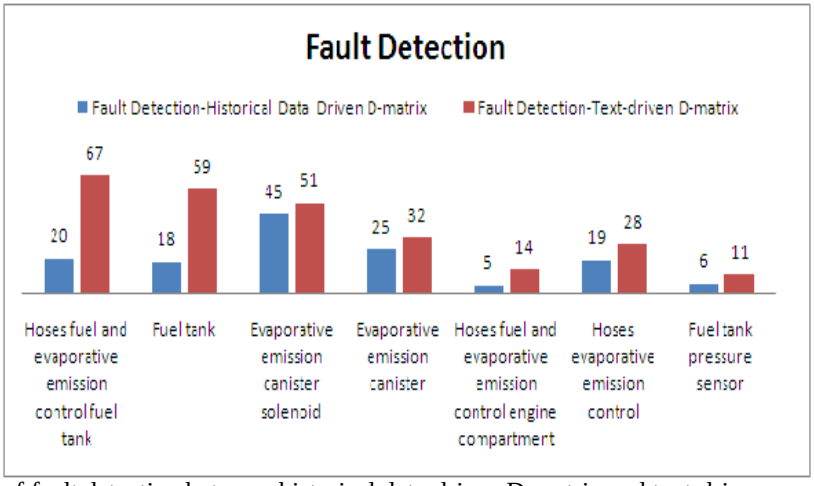

Fig.3. Comparison of fault detection between historical data-driven D-matrix and text-driven graph based D-matrix.

B. Fault Isolation(F_1):

The Fault Isolation is the likelihood that the symptoms isolate the faults of a system for the failure modes connected with a system. The percent issue isolation registered as the percent of aggregate blames that can be exceptionally segregated utilizing for unweighted case. In our proposed system, the final output i.e. D-matrix is also tested against historical data- driven D-matrix for fault isolation.

Depending on these two testability measurements, thr proposed system is compared against the existing as well as with the historical data-driven d-matrix.As stated,it is assumed that the performance, fault detection and fault isolation of the d-matrix is extremely high than the historical data-driven D-Matrix.

© 2014-15, IJIRIS- All Rights Reserved

ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

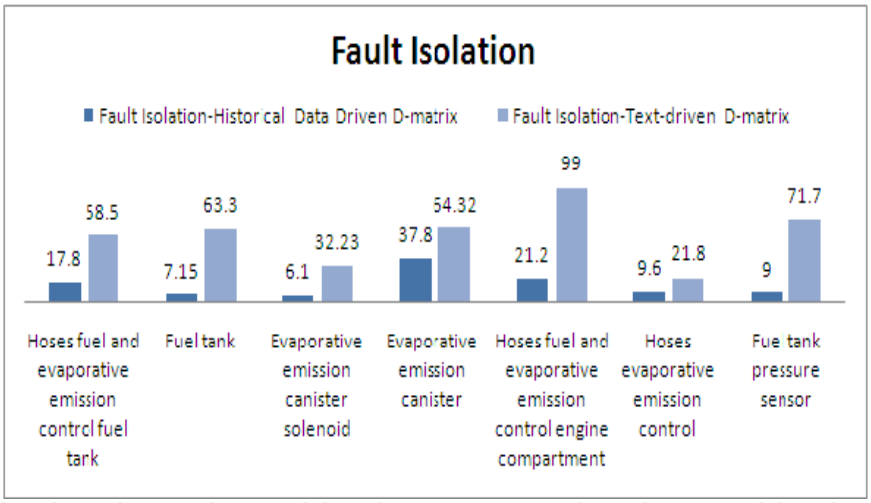

Fig.4. Comparison of fault isolation between historical data-driven D-matrix and text-driven graph based D-matrix

V. CONCLUSION

In this paper, An Comprehensive D-Matrix is developed using text mining method which is based on ontology by which we can store unstructured information obtained during fault recognizing & fault solving practices. Truth be told, the manual development of a D-matrix analytic model relating to the complex frameworks is not reasonable as it would include huge push to coordinate the knowledge from SMEs and speak to it in a D-matrix. By and large the SMEs may not in any case have the capacity to figure it out all the conditions between disappointment modes and symptoms coming about into incomplete backing to perform shortcoming judgment. Ontology based text mining to develop D-Matrix conquered these impediments where regular natural language transforming algorithm were proposed to consequently create the D-matrix from the unstructured repair verbatim. In our approach, we develop d-matrices from different datasets. After that, we represent each d-matrix in graph and by using graph merging algorithm, we combine common patterns of generated graph into new graph. This newly generated graph is then used to construct D-matrix which is comprehensive and combination of two d-matrixes. In future, our aim is to reuse this newly generated d-matrix whenever the relevant dataset will come to generate the D-matrix. So, this will save the time and useful in many manners.

REFERENCES

[1].Dnyanesh G. Rajpathak, "An Ontology-Based Text Mining Method to Develop D-Matrix from Unstructured Text"*Member, IEEE* and Satnam Singh, *Senior Member, IEEE*.
[2].O. Benedittini, T. S. Baines, H. W. Lightfoot, and R. M. Greenough,"State-of-the-art in integrated vehicle health management," *J. Aer. Eng*.,vol. 223, no. 2, pp. 157–170, 2009.
[3].Hassan, N.N. ; Dept. of Electron. Eng., NED Univ. of Eng. & Technol., Karachi, Pakistan ; Arif, A. ; Pervez, U. ; Hassam, M. "Micro-controller Based On-Board Diagnostic (OBD) System for Non-OBD Vehicles" Computer Modeling and Simulation (UK), 2011 UkSim 13th International Conference on March 30 2011-April 1 2011
[4].*IEEE standard for artificial intelligence exchange and service tie to all test environments (AI-ESTATE)*, IEEE Std. 1232–2002, 2002.
[5].Wang, W. H. Tang, and Q. H. Wu, "Ontology-based fault diagnosis for power transformers," in Proc. IEEE Power Energy Soc. Gen. Meeting, 2010, pp.
[6].Rung-Ching Chen, Ya-Ching Lee, Ren-Hao Pan," Adding New Concepts On The Domain Ontology Based On Semantic Similarity," Department of Information Management, Chaoyang University of Technology 168, Jifong East Road, Wufong Township, Taichung County 41349, Taiwan ROC.
[7].S. Singh, A. Kodali, K. Choi, K. R. Pattipati, S. M. Namburu, S. C. Sean, D. V. Prokhorov, and L. Qiao, "Dynamic multiple fault diagnosis: Mathematical formulations and solution techniques," IEEE Trans. Syst., Man Cybern. A, Syst. Humans, vol. 39, no. 1, pp. 160–176, Jan. 2009.
[8].Harpreetsingh and renuDhir "A New efficient Matrix based frequent Item sets Mining Algorithm with Tags" August, 2013.
[9].Saeed Hassanpour, Amar K. Das," Ontology-Based Text Mining of Concept Definitions in Biomedical Literature," Stanford Center for Biomedical Informatics Research,Stanford, CA 94305, U.S.A.
[10]. Miguelanez, K. E. Brown, R. Lewis, C. Roberts, and D. M. Lane,"Fault diagnosis of a train door system based on semantic knowledge representation railway condition monitoring," in *Proc. 4th IET Int. Conf* ., 2008, pp. 1–6.

© 2014-15, IJIRIS- All Rights Reserved

ISSN: 2349-7017(O)
ISSN: 2349-7009(P)
www.ijiris.com

[11]. J. Sheppard, M. Kaufman, and T. Wilmering, "Model based standard for diagnostic and maintenance information integration," in *Proc. IEEE AUTOTESTCON Conf.*, 2012, pp. 304–310.
[12]. M. Schuh, J. Sheppard, S. Strasser, R. Angryk, and C. Izurieta, "Ontology-guided knowledge discovery of event sequences in maintenance data," in *Proc. IEEE AUTOTESTCON Conf* ., 2011, pp. 279–285.
[13]. S. Strasser, J. Sheppard, M. Schuh, R. Angryk, and C. Izurieta, "Graphbased ontology-guided data mining for d-matrix model maturation," in *Proc. IEEE Aerosp. Conf* ., 2011, pp. 1–12.
[14]. S. Deb, S. K. Pattipati, V. Raghavan, M. Shakeri, and R. Shrestha, "Multi-signal flow graphs: A novel approach for system testability analysis and fault diagnosis," *IEEE Aerosp. Electron. Syst.*., vol. 10, no. 5, pp. 14–25, May 1995.
[15]. R. Studer, V. R. Benjamins, and D. Fensel, "Knowledge engineering: Principles and methods," *J. Data Knowl. Eng*.., vol. 25, nos. 1–2, pp. 161–197, 1998.
[16]. M. Schuh, J. W. Sheppard, S. Strasser, R. Angryk, and C. Izurieta,"A Visualization tool for knowledge discovery in maintenance event sequences," *IEEE Aerosp. Electron. Syst. Mag*., vol. 28, no. 7, pp. 30–39, Jul. 2013.

© 2014-15, IJIRIS- All Rights Reserved