

## A text Ontology Method based on mining Develop D –MATRIX

Tanaji L Dhikale  
Dept. of Computer Science and Engineering,  
Vathsalya Institute of Science & Technology,  
Anantharam, Bhongir, Nalgonda-508116  
Email: [tanaji.dhikale88@gmail.com](mailto:tanaji.dhikale88@gmail.com)

Asso. Prof. Mr. Ajay Kumar Kurra (HOD)  
Dept. of Computer Science and Engineering,  
Vathsalya Institute of Science & Technology,  
Anantharam, Bhongir, Nalgonda-508116  
Email: [ajaykumarkurra@gmail.com](mailto:ajaykumarkurra@gmail.com)

**Abstract:** In this issue, we demonstrate a text mining method of ontology based on the development and updating of a D-matrix naturally extraction of a large number of verbatim repairs (written in unstructured text) collected during the analysis stages. dependence (D) Fault - Matrix is a systematic demonstrative model is used to capture data symptomatic level progressive elimination system including dependencies between observable symptoms and failure modes associated with a frame. Matrix is a time D-long process. The development of D-matrix from first standards and update using the domain information is a concentrated work. In addition, increased D-die time for the disclosure of new symptoms and failure modes observed for the first race is a difficult task. In this methodology, we first develop the fault diagnosis ontology includes concepts and relationships regularly seen in fault diagnosis field. Then we use text mining algorithm that make use of this ontology to distinguish basic items, such as coins, symptoms, failure modes, and conditions of the unstructured text verbatim repair. The proposed technique is tools like a prototype tool and accepted using real - life information collected from cars space.

**Keywords:** Data Mining, fault analysis, fault diagnosis, information retrieval, word processing.

\*\*\*\*\*

### 1. Introduction

A complex system works with around perform a set of tasks keeping performance within an acceptable range of tolerances. Any deviation from a part of his worthy performance is treated as a default. Fault detection and diagnosis (FDD) is carried out distinguish defects and diagnose root causes minimize the downtime of a system. Again because of becoming technological complexity which is inserted into the vehicle system for the complex programming case installed [2] system, diagnostic sensors, web, and thus the FDD methodology gets to be a difficult activity in the During a party or a system malfunction. As everyone could wait, after each stage of the trial lessons learned are kept in some databases (eg, error codes are stored on board aircraft PC) to detect and diagnose faults. Diagnostic information generally held book comes as repair unstructured verbatim (also referred as medication records of patients in the medical industry), which gives an rich source of diagnostic data. She understands indications relating to defective parts, watched the disappointment modes and movements performed repairs to fix defaults. More verbatim this repair is collected. However, we submit that there is an urgent need to use the information to , improve the fault diagnostics ( FD) . However, the timing size of the repair information verbatim limit capacity his powerful use during the time spent FD . Text mining [3] pick up a real review because of its ability to naturally find unstructured knowledge assets covered by text. In this article , we propose a text mining strategy guide the separate analytical information of unstructured repair verbatim in a D matrix [4] . D matrix is a standard defined by the IEEE standard analytical models 1232 standard [ 5] . Anyway , the development of a D matrix using text extraction is a difficult task partly because of the noise observed in the repair verbatim text information. Short text entries: it is used to record the terms and it is vital to disambiguate their meanings, for example, loose door en chkd repair performed. Incomplete text entries: Incomplete repair

information it is difficult to determine the exact learning from data. disambiguation term: the same term is composed of using an incoherent vocabulary, for example, FTPS-Inop and Short-internal FTPS. A methodology is proposed in principle to build a D-matrix analytical model by dissecting the repair of unstructured information related to the verbatim various systems in parallel through the promotion of text mining algorithms based on ontology. It beats constraint facing the real business of having to build the D-matrix diagnostic models using first principle. In addition, in our methodology, we have the ability to catch inter-system dependencies that have made a difference essentially improving the implementation of FDD. Typically, Dmatrix constructed using the history information, engineering information and sensory information to the sample [6]. However, an understanding almost nothing is given disclosure of new symptoms also failure modes looked first time and their incorporation into the D-matrix models. In our occasional growth methodology cosmology deficiency find help content extraction calculation to develop the line D-matrix.

### 2. Literature review

In the paper [16] , this document almost optimal algorithm for dynamic multiple fault diagnosis problems ( DMFD ) close to home results perfect testing. The DMFD question is the focus no doubt the progress of component states, special case that better clarifies the results of observed tests. Here we look at four DMFD details of the problem. These incorporating the deterministic circumstances relevant to impeccably observed coupled Markov decision process some part of the path covered factorial looked Markov models ranging from the situation where the impure test the test results are all elements in the situation where the test results are elements of gaps and testing, and further if false alarms are connected with the ostensible (flawless) cases. All these definitions are unmanageable combinatorial optimization NP-hard problems. In [17] paper, principal

component analysis (PCA) wide application found in the monitoring process, slow and normal process occurs frequently in the actual methods. In this paper, we propose two recursive algorithms for PCA Monitoring of the adaptation process. The paper begins with a effective way to deal with the redesign of the relationship the recursive matrix. Calculations, using a rank modification and Lanczos tridiagonalization are then proposed. The amounts of the main components and the extent as possible for process monitoring are decided recursive. A complete algorithm for adaptive monitoring addresses issues of missing values and is described displayed. Finally, the proposed calculations are connected to a rapid thermal annealing method of the semiconductor treatment for adaptive monitoring. In [18] paper, in this paper, we conduct a study on the content grouping regularly using elements games. The principle the commitment of this article is three collectors. First, we present a review on existing methods for grouping files using common reason. Second, another strategy called Maximum capture is proposed for clustering document. Third, the tests are completed to evaluate the proposed technique in comparison with the methods CFWS, CMS, FTC and HIFC. In addition, subjects produced by Capturing Maximum clusters distinguished from each other and can be used as groups of documents labels. In the paper [19], we depict Latent Dirichlet Allocation (LDA), a generative probabilistic model for discrete accumulations data, e.g., text corpora. LDA is a three level Bayesian hierarchical model, in which each set of accumulation is demonstrated as an infinite mixture on a hidden arrangement of subjects. Each subject was, in turn, shown as an endless mix of hidden arrangement of probabilities points. In connection showing the content, subject probabilities give unequivocal representation of an archive. We exhibit effective rough inference techniques for the variational methods and an inference technique to the correct paper Bayes estimate. We report brings the text classification, collaborative filtering, also oriented community screening, contrasting with a unigrams model mixture LSI probabilistic model. In [ 20] , here we display a model for trainable limits of distinctive phrases into plain text . Given a corpus clarified with sentence boundaries , our model on the figures how to group all occurrences . , ? And / or as a valid or sentence invalid limit. The preparation system does not require handcrafted rules , lexicons, parts of speech labels, or area including information . The model can be formed thusly effortlessly on any kind of English, and should be trainable any other language of the Roman alphabet. Performance is substantially the same or greater than that of the execution of the same systems, we emphasize ease of recycling for the new domain. In this article [21], we show a language independent, unsupervised manner to face the sentence boundary detection. It is given the hypothesis that a significant number of ambiguities in the sentencing border can be removed once abbreviations were identified. Rather than relying spelling rooms information, the proposed framework has the ability identify with precision accuracy using three criteria just require data on the type of candidate himself and are regardless of the context: The abbreviations can be characterized as a tight collocation including a

truncated word and last period, abbreviations forms are generally short, and condensation sometimes contain internal periods. US too demonstrate the ability to demonstrate collocation for two vital subtasks of sentence boundary disambiguation, such that the detection of initial and ordinal numbers. Proposal framework was judged largely on eleven different languages and various classifications of text. he accomplished great results without any other changes to the other resources specific to the language. We find his execution three against baselines and contrast various separate sentence boundary detection frames proposed in Literature. In this article, [22] with the use of natural development Treatment of technical languages (NLP) for data extraction and indexing of ideas in the biomedical field, a strategy quickly and effectively relegates the feeling of a right Biomedical ambiguous term in a given context is needed at the same time. The current state of the word meaning disambiguation (WSD) in biomedicine is that craft rules are used taking into account the relevant documents. Bad service this methodology are (i) generate WSD rules is a long and tedious task manually, (ii) maintenance of rule sets is becoming increasingly difficult over time, and (iii) craft rules are often incomplete and poor results in new areas including specialized vocabularies and different kinds of text .. This paper displays a strategy unattended two-step the manufacture of a WSD classifier for biomedical equivocal W. The main phase term makes naturally marked sense- W corpus, and the second phase infers a W classifier using strong sense of corpus determined set preparation. A development review was made, who the hell that began graders prepared on the corpora marked determined direction reached a general accuracy of nearly 97%, with greater than 90% precision for each individual term uncertain.

### 3. Methodology

In this development methodology includes D -MATRIX the accompanying building block of the annotation of documents, term extraction and melting phrases. First, repair verbatim axes data are collected by recovering from Database of OEM, which are recorded in field FD . In the first step, terms such as a part , a symptom, and failure Fashion, relevant , applicable for the D matrix are explained each verbatim repair constructing the document algorithm annotation. Servicing includes a verbatim several sections, symptoms, modes and failure actions good affiliations must be built between important which, given their proximity to each other. here is a Service is first part verbatim in different sentences using the detection rules for sentence boundaries and the terms showing in the same sentence are co-related to each other.

**(I). Fault Diagnosis Ontology:** the ontology field building attracted serious attention from the artificial intelligence community in the 90s An ontology can be regarded as a information model Expressly depicts different concepts that exist in a domain speech and their properties. Fault diagnosis Ontology is a specifically lightweight ontology that is formalized by using the ontology development methodology. He grabs the terms and relationships

discussed in the field of vehicle fault diagnosis. Cases are formalized by using the asset representation scheme that facilitate the exchange, storage and readability Machine ontology.

**(II). Ontology-Based Text Mining:** In this model, we describe some few steps to know. extractor terms, the document annotation and sentence merging involved in the ontology the construction of text mining based on a D-matrix

**(a). Document Annotation:** Due to the different types of us seen in our information (Section I) of the company distinguish the main elements of D-matrix, Example. parts, symptoms and failure modes in a nontrifling activity. The markup document serves to channel on data that is irrelevant to our inquiry and gives a special connection to the coherence and shared interpretive information. Initially, the following preprocessing steps, the sentence boundary detection (SBD), are used to separate repair verbatim in partitioned sentence, the stop words are deleted in the non-expressive words, and lexical pairing identifies the right abbreviations. Therefore, under prepared in full are coordinated using the examples the fault diagnosis ontology.

**(B). Extractor term:** In this method, we exposed the the terms, the essential requirements for the development of a Dmatrix namely symptoms and failure modes are fired by using the term extraction algorithms. First, causality connection between the relevant failure mode symptom sets stands to check than just good games extirpated. Existing methods for frequent object sets mining ignores the application in which the term expressions are recorded in the document, in any case, we must maintain this asking how fault diagnosis is performed.

**(C). Fusion sentence:** In this point, because of the man action while catching repair a verbatim data the irregularity was seen in the details Term used to record the failure modes, eg container Tank Pressure Sensor Sensor\_Inop or internal short. So we check disappointment two distinctive modes are varieties which are essentially the same failure mode so that they can be converged before filling a D-Matrix. The phrase fusion algorithms takes like entering a chess arrangement modes freed by the long-term extraction algorithm. A priori Our system does not have the information two failure modes can be combined ; and the proximity of all the failure mode is coincident with each other, for example , Evap leak in the fuel tank and drain tank at the neck and we must treat these phrases more. Some procedures are proposed to study the problem of incompatibility of vocabulary , stalking, translation models , expansion of the application, LSI , between other. The term fusion algorithm is compared in the soul with query expansion procedures , where every failure mode is treated as a potential application and the middle of the stage of development of additional data is collected as attributes .

#### 4. Conclusion

In a future perspective the main objective of each Dmatrix graphic form and develop the graphs of comparison algorithms. With the help of the text based on technology we ontology developed D-Matrix automatically using extraction different structure repair verbatim information is collected during fault diagnosis time. The manual development diagnostic model D-matrix, it would be involved in integrate information of SMEs. In most of the cases, SMEs may not be able to realize all the dependencies between fault diagnosis and failure nodes. The main approach of these the system is to overcome the natural limits where language algorithms are automatically develop D-matrices of unstructured repair verbatim. We have also compared to the historic Diagnosability measures datadriven D-matrix and testability. In section four, we discussed the text of D-driven matrix approach reached more an insulation fault, the less the size of the group of ambiguity due to the DTC symptoms. Finally, we discussed the effectiveness of text driven D-matrix which is compared to LDA. the used main LDA improves fault isolation and default rates detection.

#### References:

- [1] T. Felke, —Application of model-based diagnostic technology on the Boeing 777 airplane,| in Proc. 13th AIAA/IEEE DASC, 1994, pp. 1–5.
- [2] G. Ramohalli, —The Honeywell on-board diagnostic and maintenance system for the Boeing 777,| in Proc. IEEE/AIAA DASC, 1992, pp. 485–490.
- [3] P. M. Frank and J. Wunnenberg, —Robust fault diagnosis using unknown input observer schemes,| in Proc. Fault Diagnosis Dynamical Syst.: Theory Appl., 1989, pp. 47–98.
- [4] N. Viswanadham and R. Srichander, —Fault detection using unknown Input observers,| Control-Theory Ad Tech., vol. 3, pp. 91–101, 1987.
- [5] P. M. Frank, —Fault detection in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results,| Automatica, vol. 26, no. 3, pp. 459–474, 1990.
- [6] Y. Dingli, J. B. Gomm, D. N. Shields, D. Williams, and K. Disdell, —Fault diagnosis for a gas-fired furnace using bilinear observer method,| in Proc. Amer. Control Conf., 1995, pp. 1127–1131.
- [7] H. Yang and M. Saif, —Nonlinear adaptive observer design for fault detection,| in Proc. Amer. Control Conf., 1995, pp. 1136–1139.
- [8] V. Venkatasubramanian and S. H. Rich, —An object-oriented two-tier architecture for integrating compiled and deep-level knowledge for process diagnosis,| Comput. Chem. Eng., vol. 12, no. 9–10, pp. 903–921, 1988.
- [9] C. Charniak and D. McDermott, Introduction to Artificial Intelligence. Reading, MA, USA: Addison Wesley, 1985.
- [10] V. R. Benjamins, —Problem-solving methods for diagnosis and their role in knowledge acquisition,| Int. J. Expert Syst.: Res. Appl., vol. 8, no. 2, pp. 93–120, 1995.
- [11] T. Umeda, T. Kuriyama, E. Oshima, and H. Matsuyama, —A graphical approach to cause and effect analysis of chemical processing systems,| Chem. Eng. Sci., vol. 35, no. 12, pp. 2379–2388, 1980.

- [12] M. A. Kramer and B. L. Palowitch, —A rule based approach to fault diagnosis using the signed directed graph,| *AIChE J.*, vol. 33, no. 7, pp. 1067–1078, 1987.[26] R. F. Li and X. Z. Wang, —Qualitative/quantitative simulation of process temporal behavior using clustered fuzzy digraphs,| *AIChE J.*, vol. 47, no. 4, pp. 906–919, 2001.
- [13] J. F. McGregor and T. Kourti, —Statistical process control of multivariate processes,| *Control Eng. Practice*, vol. 3, no. 3, pp. 403–414, 1995.
- [14] W. Li, H. Yue, S. Valle-Cervantes, and S. J. Qin, —Recursive PCA for adaptive process monitoring,| *J. Process Control.*, vol. 10, no. 5, pp. 471–486, 2000.
- [15] S. Singh, H. S. Subramania, and C. Pinion, —Data-driven framework for detecting anomalies in field failure Data,| in *Proc. IEEE Aerosp. Conf.*, 2011, pp. 1–14.
- [16] Satnam Singh, Member, IEEE, Anuradha Kodaly, Kihoon Choi, Krishna R. Pattipati,| *Dynamic Multiple Fault Diagnosis: Mathematical Formulations and Solution Techniques*|
- [17] Weihua Li, H. Henry Yue, Sergio Valle-Cervantes, S. Joe Qin,| *Recursive PCA for adaptive process monitoring*|, Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, USA .
- [18] Wen Zhang, Taketoshi Yoshida , Xijin Tang , Qing Wang,| *Text clustering using frequent itemsets*|, Internet Software Technologies, Institute of Software, Chinese Academy of Sciences, Beijing 100190.
- [19] David M. Blei, Andrew Y. Ng, Michael I. Jordan,| *Latent Dirichlet Allocation*|, Computer Science Division and Department of Statistics University of California Berkeley, CA 94720, USA .
- [20] Jeffrey C. Reynar and Adwait Ratnaparkhi, |*A Maximum Entropy Approach to Identifying Sentence Boundaries*|, Department of Computer and Information Science University of Pennsylvania, USA
- [21] Tibor Kiss, Jan Strunk ,| *Unsupervised Multilingual Sentence Boundary Detection*|
- [22] Hongfang Liu,\*<sup>1</sup>, Yves A. Lussier,<sup>†,2</sup> and Carol Friedman,| *Disambiguating Ambiguous Biomedical Terms in Biomedical Narrative Text: An Unsupervised Method* |