FAULT DETECTION AND DIAGNOSIS TO CONSTRUCT D-MATRIX USING ONTOLOGY BASED TEXT MINING

Sushant Patil,(you'r designation here) , Computer Eng. Department, SJCET Palghar, Mumbai University, India

Shreeya S. Desai, Student, Computer Eng. Department, SJCET Palghar, Mumbai University, India

Priya D'silva, Student, Computer Eng. Department, SJCET Palghar, Mumbai University, India

Bhargava G. Dubal, Student, Computer Eng. Department, SJCET Palghar, Mumbai University, India

Abstract
Fault dependency (D-matrix) is a diagnostic model which is used to catch the fault system data and its causal relationship at the hierarchical system-levels. It consists of dependencies and relationship between observable failure modes and symptoms associated with a system. Here, we describe an ontology based text mining approach for fault detection and diagnosis for automatic construction of D-matrix by mining hundreds of repair verbatim gathered during diagnosis episodes. In our approach, we first perform the document annotation phase consisting of search module, which includes the term extraction phase as well and retrieves the queries relevant to the problem statement using data mining algorithms. Next, a D-matrix granting possible solutions is constructed using the processed information. Proposed method is implemented as a prototype model and validated by real-life data collected from automobile domain.
Index Terms- Fault diagnosis, fault detection, information retrieval, D-matrix, text mining. ***(key words)***

I. INTRODUCTION

Fault detection and diagnosis is termed as the process of identifying the cause of malfunction by observing its effects at various monitoring points in a system. With the rapid advancement in technology there is a significant increase in the complexity and sophistication of system. Recently the number of possible causes have increased while reduction in test points have resulted in reduced fault observability making it very difficult to troubleshoot these systems. In the previous years testability has been an adhoc, manual effort, in which the engineers attempt to identify an efficient method of troubleshooting for the given system with no idea of the system design. However, poor fault observability in complex systems have driven up the life-cycle maintenance cost of the systems over 10 times the maintenance cost.. Fault is defined as any abnormal behaviour of the system[1].

Ontology is the study of nature of being, becoming, reality, what is existing as well as the basic categories of being and their relationship. Ontology deals with questions concerning what entities exist or can be said to exist, and how such entities can be grouped, related within a hierarchy, and subdivided according to similarities and differences. Ontological engineering have become an efficient methodology for knowledge representation and management in many domains and tasks. Design, methodologies and approaches of ontology are very important factors to build ontologies for specific task[2]. The ontological engineering methodology is widely used in many aspects of cooperative information systems, information retrieval and extraction, knowledge representation, DBMS, computer and information science. In this modern culture, text is the most common vehicle for the formal exchange of information. Although extracting useful information from texts is not an easy task, it is a need of this modern life to have a business intelligent mechanism which is able to extract useful information as quick as possible and at a low cost. Text mining is a new and exciting research area that tries to take the challenge and produce the intelligence mechanism. The tool is a text mining system which has the capability to analyse large quantities of natural language text and detects lexical and linguistic usage patterns in an attempt to extract meaningful and useful information. The aim of text mining is to be able to answer sophisticated questions and perform text searches with an element of intelligence. Among the features of text mining include:

- a user centric process which leverages analysis technologies and computing power to access valuable information within unstructured text data sources ;
- text mining processes are driven by natural language processing and linguistic based algorithm
- eliminate the need to manually read unstructured data sources.

The major challenging issues in text mining arise from the complexity of a natural language itself. The natural language is not free from the ambiguity problem. Ambiguity is the capability of being understood in two or more possible senses or ways. Ambiguity gives a natural language its flexibility and usability, and consequently, therefore it cannot be entirely eliminated from the natural

language. One word may have multiple meanings. One phrase or sentence can be interpreted in various ways, thus various meanings can be obtained. Thus the input is disambiguated discarding inconsistent terms, merging common terms into a single term.

II. LITERATURE SURVEY

In [3] Prof. Felke had implemented an application using D- matrix technology to find out dependencies between symptoms, failure modes and repairs by analysing the structured service manual data. But the model that is too general will not provide the required accuracy. Text mining also searches for patterns in unstructured text. In addition, Information Retrieval (IR) techniques are widely used for tasks such as matching document and grouping [4].
Fault detection and diagnosis is a main part of many operations which manage automation systems. Fault diagnosis includes fault detection, fault isolation & fault identification. Fault detection: Indicating if there is a fault. Fault isolation: Determining where the fault is. In Fault isolation, when an unexpected outcome is found, the negative outcomes are limited. By limiting the scope of problems minimize the degree for damage and makes systems easier to maintain [5]. Detection + Isolation = Diagnosis.
In [6] A FDI approach used which is based on analytical redundancy to achieve maximum robustness by decoupling the effects of faults and errors. But analytical redundancy- FDI method is applied on non-linear discrete time systems only.
In [7] Prof. Singh had implemented a paper 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.

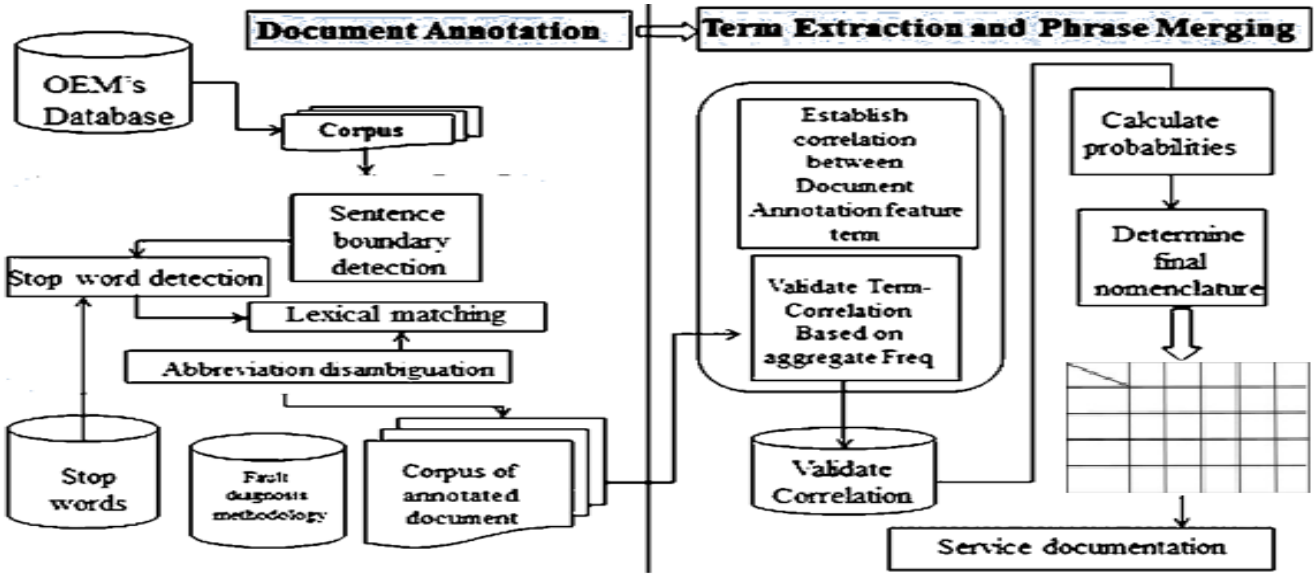

Fig.1. Ontology based FDD framework

Existing system [8] is a text mining method to map the diagnostic information extracted from the unstructured repair verbatim in a D-matrix. However, the construction of a D-matrix by using text mining is a challenging task partly due to the noises observed in the repair verbatim text data abbreviated text entries the abbreviation are used to record the terms and it is crucial to disambiguate their meaning, incomplete text entries the incomplete repair information makes it difficult to derive the precise knowledge from the data; term disambiguation the same term is written by using inconsistent vocabulary.

Existing system consists of ontology based text mining in which the three phases used are document annotation, term extraction and phrase merging.

Document annotation: The document annotation helps to filter out the information that is not related for analysis and it provides a specific background for the reliable understanding of the data.

Term extraction: Using this phrase, the vital terms desirable for the development of a D-matrix, i.e., symptoms and failure modes are extracted by using the term extractor algorithm.

Phrase merging:
Here the same failure mode phrases that are generally written with the help of a conflicting vocabulary are merged into a single, consistent failure mode phrase using phrase merging algorithm.
The existing system uses the same Mathematical programming model but with optimized by combined the two phases Term Extraction and Phrase Merging to optimize more time for processing.
• Creating a vocabulary for ontology is to extract important terms from text documents related to a particular domain.
• The corpus is then parsed into tokens or terms.
• Unstructured text in the corpus becomes a structured data object via the creation of a term- by-document frequency matrix.
• Frequency weights of those concepts can be adjusted to account for the distribution of terms across documents.
• Natural language processing (NLP) and text mining techniques are effective for information extraction from text documents.

III. PROPOSED SYSTEM

The proposed system consists of ontological based text mining method to detect and diagnose faults for a system and provide accurate results in a D-matrix. This method comprises of three building blocks which are document annotation, term extraction and phrase merging. The proposed system is based on a combination of document annotation and term extraction blocks for better performance and optimization of time along with phrase merging block.
Normally the process of fault detection and diagnosis starts by extracting the error codes from the target system and then on the basis of their experience and observed error codes the technicians/engineers try to follow a specific diagnosis procedure to diagnose the faults. During this procedure, several data types such as repair verbatim, error codes, scanned values of faulty component/system, and so on are taken into consideration.

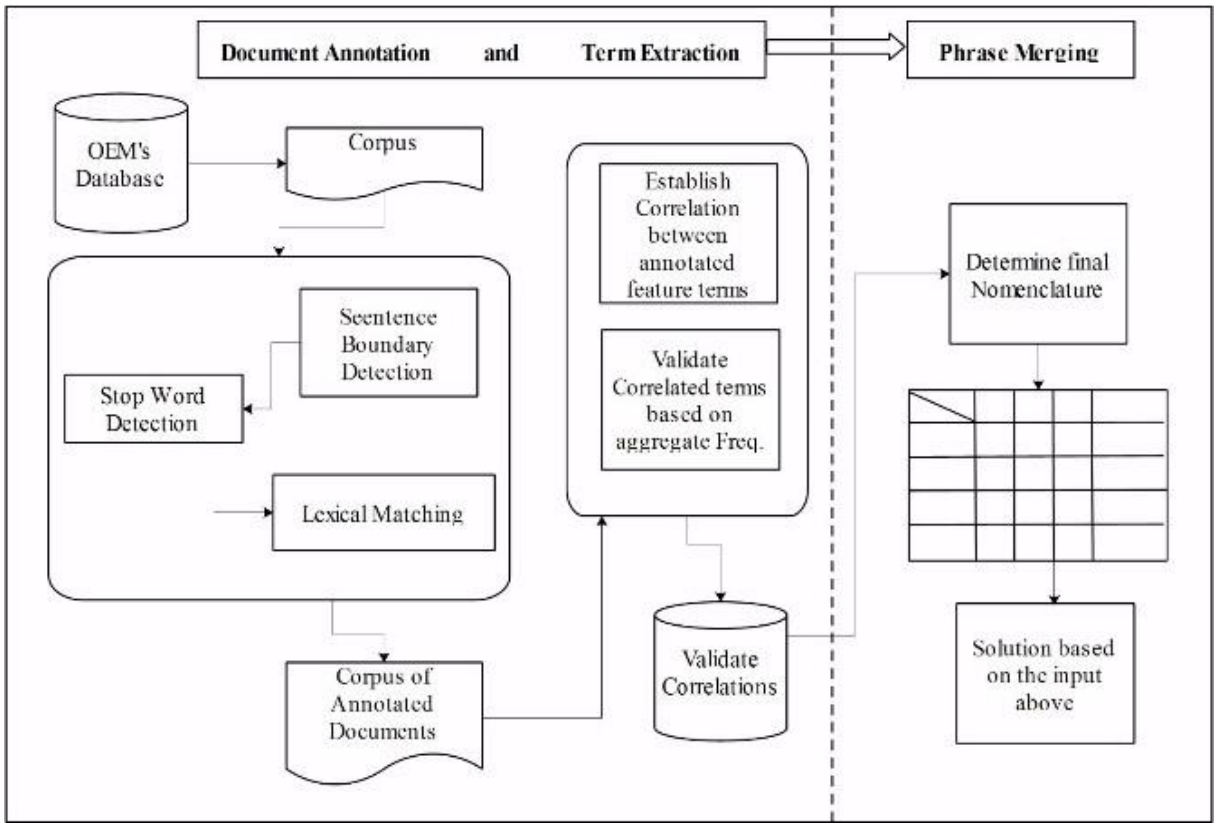

Fig.2. Framework for constructing D-matrix

This collected data is then transferred to the OEM database and the repaired verbatim data collected over a period is mined to develop dependency matrix models which are further used to perform accurate fault detection and diagnosis.

The d-matrix captures component and system level dependencies between a single or multiple failure modes, parts with a single or multiple symptoms in a structured way. These dependencies help the technicians or stakeholders to detect a set of failures modes and parts causing symptoms along with the probability of detection.
The proposed d-matrix is a text driven binary d-matrix which has probability value of either 0 or 1, where 0 indicates no fault detection and 1 indicates complete fault detection of a particular failure mode using a particular symptom. The values intermediate between 0 and 1 depict the level of detecting a failure mode causing symptom. Therefore the d-matrix development methodology would capture the dependencies between symptoms and failure modes by mining the unstructured repair verbatim data. This process would result into a text-driven d-matrix as a diagnostic model. The following described are main steps of combined document annotation and term extractor, along with phrase merging phase to construct a d-matrix:
Document Annotation and Term Extraction: Due to different types of noises observed in the unstructured repaired verbatim data the process of construction of d-matrix becomes difficult.
Therefore we make use of document annotation phase which helps us to filter out the data which is irrelevant for further analysis. The main blocks of document annotation phase include SBD, Stop Words Removal and Lexical matching. The following are the pre-processing steps for each block: SBD: SBD known as Sentence Boundary Detection is used to split the repair verbatim data into separate sentences by identifying its beginning and its end.

Stop Words Removal: Stop Words are the non-descriptive words like a, an, the, then there etc. which just add extra noise to the repair verbatim data. So it becomes necessary to remove these words for accurate analysis of data. Stop Words Removal help to remove such non-descriptive terms and only keep descriptive terms like parts, symptoms and failure mode phrases which are needed to further analyse the data.

Lexical Matching: Repair verbatim data comprises of lot of abbreviations having multiple meanings. for ex: TPS could be Tire Pressure Sensor or Tank Pressure Sensor so it becomes necessary to find out their correct meaning before constructing a d-matrix. So lexical matching helps to find the exact meaning of abbreviations and match those using instances.

Term Extraction: Term Extraction phase helps to extract the critical terms like parts, failure modes and symptoms which are further needed to establish correlations and validate them based on aggregate frequency. The main building blocks of term extractor are establishing correlations between featured terms and validating term-correlations based on aggregate frequency.

Phrase Merging: Phrase Merging is used to merge two different failure modes if they are the variations of the same failure mode. Phrase Merging consists of the following building blocks: collecting contextual information, calculating probabilities, determining final nomenclature, merging phrases and then populating a d-matrix.

IV. RESULTS

The D-Matrix framework model is created using proposed methodology. The real-life data is collected from the automobile car system. A text driven D-Matrix is created using the symptoms shown in column and error code in rows. Also, in contrast to existing results, which showed D-matrices with 0 and 1 that represented the probability of detection of faults, proposed system gives only one D-matrix with the choice of giving an input for getting the solution as well which is not in existing system.

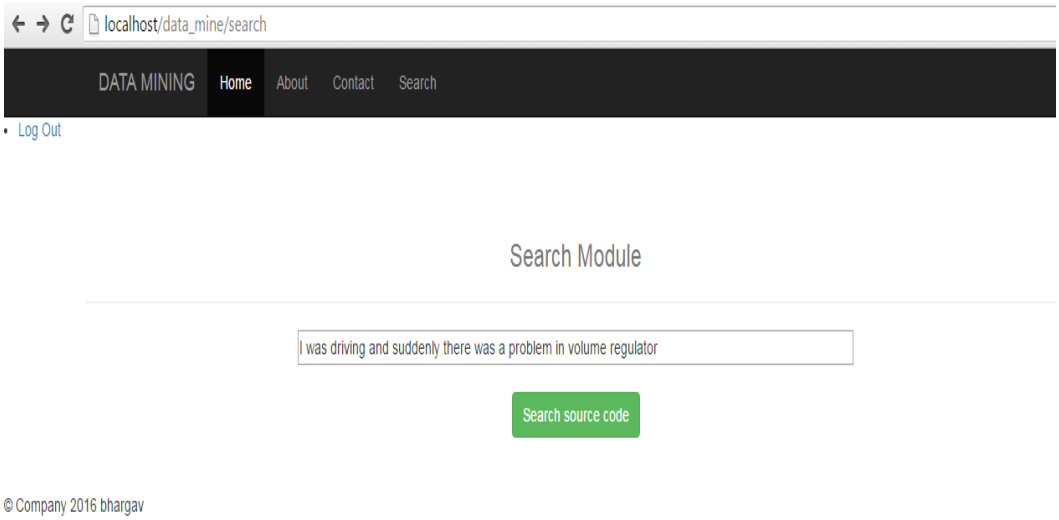

Fig.3. Search module-1

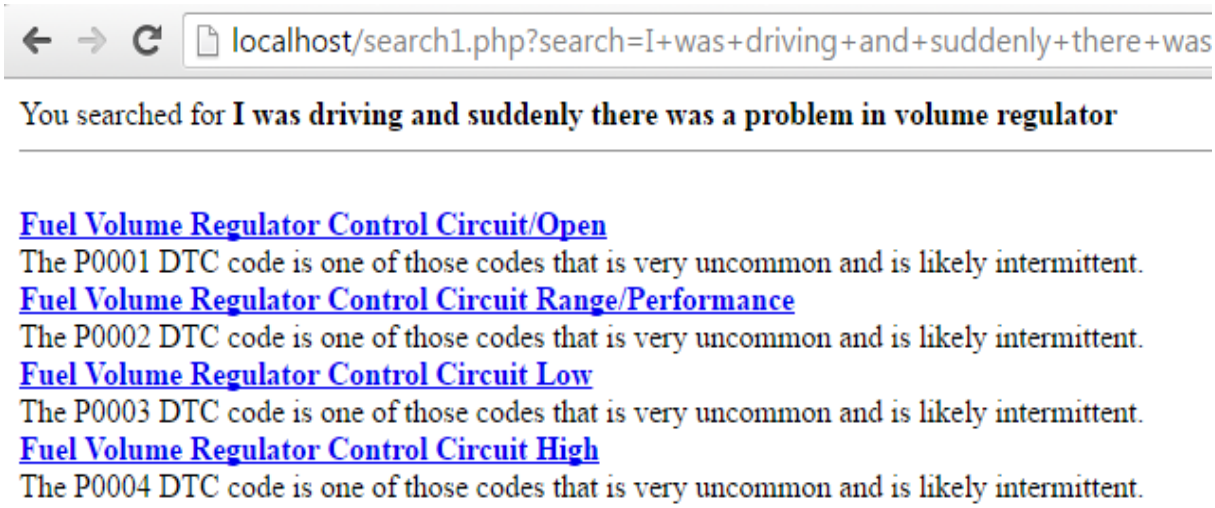

Fig.4. Search module-2

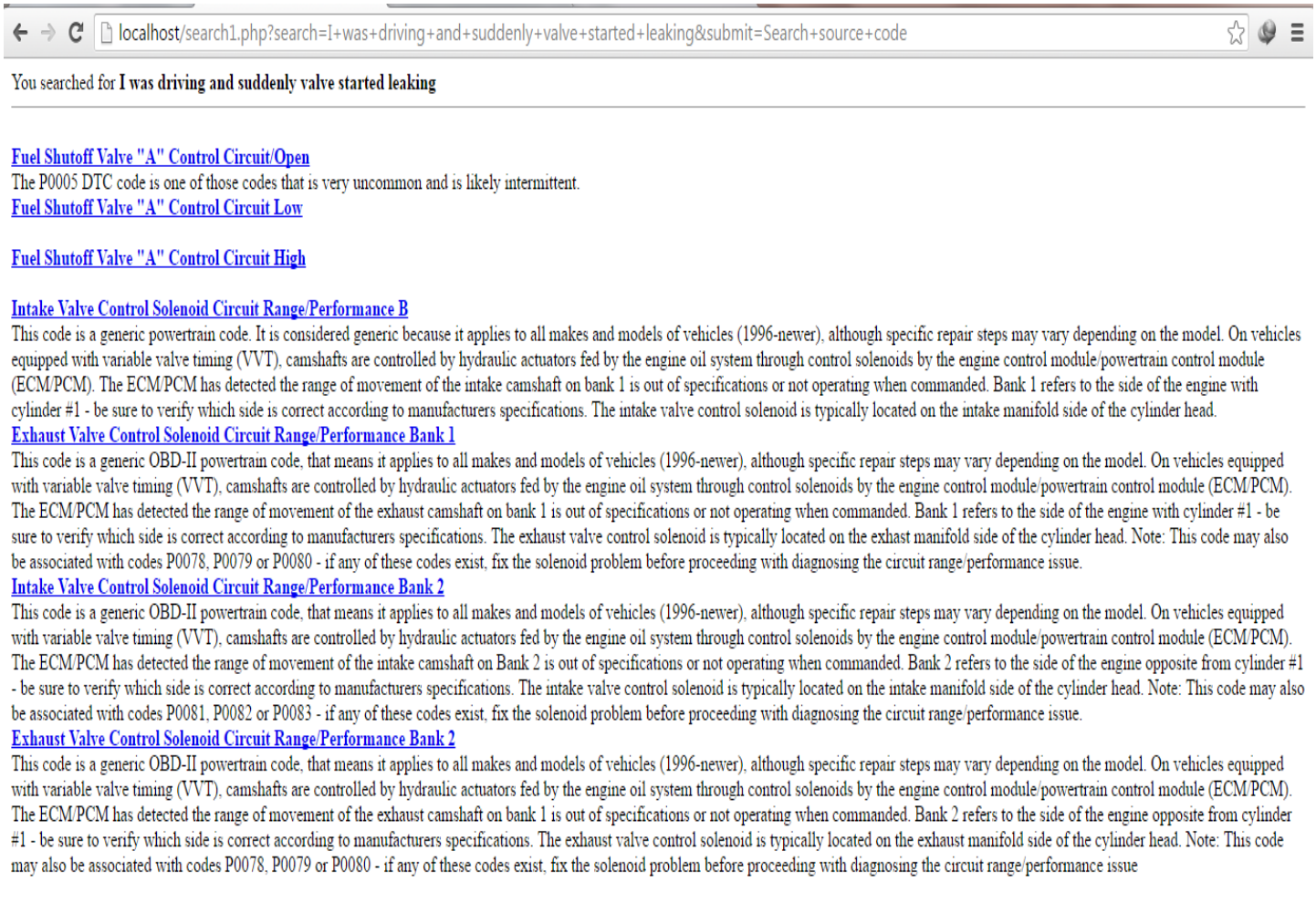

Fig.5. Search module-3

V. CONCLUSION

It is concluded to construct D-matrices by the automatic mining of the structured repair verbatim which was collected during the fault diagnosis phase.

VI. REFERENCES

[1] S. Singh, S. W. Holland, and P. Bandyopadhyay, "Trends in the development of system-level fault dependency matrices," in *Proc. IEEE Aerosp.Conf* ., 2010, pp. 1–9.

[2] M. M. VarmaJyoti and Nandimath, " An Ontology-based Comprehensive D-matrix Construction for Accurate FDD," in *Proc.* International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

[3] IEEE standard for artificial intelligence exchange and service tie to all test environments (AI-ESTATE), IEEE Std. 1232–2002, 2002.

[4] T. Felke, "Application of model-based diagnostic technology on the Boeing 777 airplane," in Proc. 13th AIAA/IEEE DASC, 1994, pp. 1–5.

[5] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACM Press, New York, 1999.

[6] Izadi, H.A. Dept. of Mech. & Ind. Eng., Concordia Univ., Montreal, QC, Canada Gordon, B.W. ; Youmin Zhang "A data-driven fault tolerant model predictive control with fault identification" Conference on Oct. 2010

[7] 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.

[8] D. 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. 1–8.

[9] R. Deore, "An Epistemological Text Mining Method by Generating D-Matrix for Unstructured Text" in (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (1), 2015, 473-475.