An Ontology Based Text Mining Method For Result Analysis

Supriya Sonawane ,Yogesh Kumavat, Nachiket Ukande,Pallavi Dube

Abstract— **An Ontology Based Text Mining Method For Result Analysis is used for performing analysis and operations on result pdf. Text mining is gaining a serious attention due to its ability to automatically discover the knowledge assets buried in unstructured text. In this system we take the result pdf as input.System is used to take a sorted result format.The sorting will be done as per Department,Subject and Student. This system is intended to be used in various universities.All institutes have some basic needs to analyze results of Department ,Student and Subject in particular format.In this system an attempt is made to design a system for students results that make it in analytical format.The main advantage of this system is to get structured and analytical data.Document Annotation Algorithm,Term Extraction Algorithm and Text Mining algorithms are used.**

Index Terms— **Document Annotation,Term Extraction,Text mining**

I. INTRODUCTION

We describe an ontology based text mining method for analyzing result by mining hundreds of student results from the given Result PDF. In our approach, we first use itextsharp.dll to read PDF and convert it into Text format. After converting classification will be performed as per student, department and subject, then analysis will be perform on classified data. The output of the system is given according to department, student and subject(like how many student Have first class,distinction,Second class, pass class,ATKT). Proposed system is used in various colleges and universities for student result analysis.It can be done by mining data from result pdf using ontology technique. The purpose of our system is to generate student result analysis report.The system will acquire marks of student from result pdf and analyses the obtained data then declare the results based on the grade criteria's.

II. PROBLEM STATEMENT

This system will be useful for university and college staff to analyze student result. Presently, Pdf analyze are done manually. One has to calculate total number of students marks, subject marks. It has following limitations:
1. Manual recording consumes excess time.
2. Chances of error.
3. Repeated recording of some subject.

Focusing on drawbacks and inadequacies of existing process, Definitely there is need of an efficient system. The proposed system rectifies the demerits and defects of existing process to a greater extend.

III. LITERATURE SURVEY

In existing system Ontology technique used to Develop D-Matrix from Unstructured Text. Fault dependency (D)-matrix is a systematic diagnostic model to capture the hierarchical system level fault diagnostic information consisting of dependencies between observable symptoms and failure modes associated with a system. Constructing a D-matrix from first principles and updating it using the domain knowledge is a labor intensive and time consuming task.Further,in-time augmentation of D-matrix through the discovery of new symptoms and failure modes observed for the first time is a challenging task. Here, is an ontology based text mining method for automatically constructing and updating a D-matrix by mining hundreds of thousands of repair verbatim (typically written in unstructured text) collected during the diagnosis episodes. In this approach, they first construct the fault diagnosis ontology consisting of concepts and relationships commonly observed in the fault diagnosis domain. Next, they employ the text mining algorithms that make use of this ontology to identify the necessary artifacts, such as parts, symptoms, failure modes, and their dependencies from the unstructured repair verbatim text. The method is implemented as a prototype tool and validated by using real-life data collected from the auto-mobile domain.

IV. PROPOSED SYSTEM

In the proposed system first we take result pdf as input, We cannot perform analysis on pdf directly thus we use itextsharp.dll to read pdf and convert it to the text format. Next, we will perform classification on that text format. We use splitters to classify this

data student, department, and subject wise. Then identification operation perform on classified data. After identification data is stored in sql tables. We retrieve stored data and analysis will be perform on it. In analysis system automatically sort out the data and Generate report
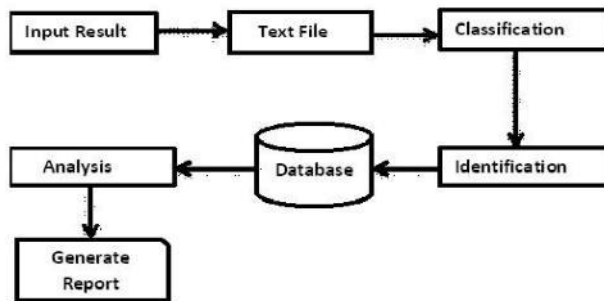

Fig1 :Proposed System

V.ALGORITHM USED

DOCUMENT ANNOTATION:

Document Annotation is used to highlight or mark sentence in pdf which are needed. Sentence boundary detection (SBD) is used to detect boundary in a given pdf. It separate sentences, stop words are deleted to remove the non-descriptive terms, and the lexical matching identifies the correct meaning of abbreviations.

TERM EXTRACTION:

By using reference of terms annoted, only the correct terms are extracted ignoring unwanted terms. These terms are entered into a particular classified section.

TEXT MINING:

The phrase merging is used to avoid ambiguous references of the failure mode phrases, where the failure mode phrases that are written by using an inconsistent vocabulary, are merged into a single, consistent failure mode phrase, maintain the homogeneity. The contextual information co-occurring with the phrases, i.e., parts, symptoms, failure mode, and actions is used to estimate the conditional probabilities and the phrases with their probability score above the specific threshold are merged. A priori our system does not have the knowledge about which two failure modes can be merged; hence the similarity of each failure mode is calculated with every other failure mode to derive the similarity score.

VI. CONCLUSION

An ontology-based text mining methodology has been proposed to university result pdf analysis by automatically mining the unstructured result data collected during result analysis. Ontology basically used for text mining and it provide more accuracy of result, and low time complexity. In our approach natural language processing algorithms were proposed to automatically analyze result from the unstructured result pdf.

REFERENCES

[1] An Ontology-Based Text Mining Method to Develop D-Matrix from Unstructured Text Dnyanesh G. Rajpathak, Member, IEEE and Satnam Singh, Senior Member, IEEE,VOL. 44, NO. 7, JULY 2014.

[2] T. Hearst, Untangling text data mining, in Proc. 37th Annu.Meeting Assoc. Comput. Linguist., 1999, pp. 310.

[3] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri, A review of process fault detection and diagnosis Part I: Quantitative modelbased methods, Comput. Chem. Eng., vol. 27, no. 3, pp.293311, 2003.

[4] V. Venkatasubramanian, R. Rengaswamy, and S. Kavuri, A review of process fault detection and diagnosis Part II: Qualitative models and search strategies, Comput. Chem. Eng., vol. 27, no. 3, pp. 313326,2003.

[5] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. Kavuri, A review of process fault detection and diagnosis Part III: Process history based methods, Comput. Chem. Eng., vol. 27, no. 3, pp. 327346, 2003.

[6] IEEE standard for artificial intelligence exchange and service tie to all test environments (AI-ESTATE), IEEE Std. 12322002, 2002.

[7] J. Sheppard, M. Kaufman, and T. Wilmering, Model based standards for diagnostic and maintenance information integration, in Proc. Autotestcon IEEE, 2007, pp. 304310.

[8] T. Felke, Application of model-based diagnostic technology on the Boeing 777 airplane, inProc.13th AIAA/IEEE DASC,1994, pp.15

[9] G. Ramohalli, The Honeywell on-board diagnostic and maintenance system for the Boeing 777, in Proc. IEEE/AIAA DASC, 1992,pp