Effectual Motive Innovation of Text in the Data set using Text mining

N.Bhuvaneswari[#1], Dr.N.Chandrakala[#2]

MPhil Scholar, Department of computer science, SSM College of Arts and Science, Periyar University, Komarapalayam, Tamilnadu, India.
Professor, Department of computer science, SSM College of Arts and Science, Periyar University, Komarapalayam, Tamilnadu, India.

Abstract—In the effectual motive innovation of the text is using to retrieve the data from the largedata set using text mining. The data mustto characterisedthe text using text mining. In the large scale industries and educational institutes are identify the domain of the area is difficult to search and order the area. In this project used to order the area for easy retrieving. The existing characterised method is identification of the word and characterised the word inaccurate.So the proposed system is based on retrieving text using text mining.In this projects using four steps in the selection work. The proposed system easily identify the particular area comes under the project.

Keywords—*innovation of characterised text, text mining*

I. INTRODUCTION

In the large set of database the terms identify and retrieving is very hard to characterised and monitoring the area of the project. The identification of the domain and the area of the projects todone are not easy in the characterised method.

The proposed innovation of characterised is easy to characterised the region of the task. In searchingthe domain area and to evaluate project and have the outcome of the project selection. A usual procedure that integrates both objective and subjective information on detection is also presented.

The selection process is classified into six groups according to their project-making tasks. These identification groups help with each other to accomplish the overall goal of selecting the best project and domain. They evaluate decision tasks in a certain order. The evaluating department is responsible for the selection tasks, and it divides the tasks to programs. Evaluate managers or decision directors then group the projects and assign them to reviewers for estimate and explanation.

However, they may not have the knowledge in all research order, and contents of projects were not fully understood. Therefore, there was a useful to group the submitted research projects. Effectual motive innovation of text in the data set using text mining is proposed to solve the problem.

II. LITERATURE REVIEW

Classification and plagiarism detection in research proposal selection process.In many organizations like funding agencies or universities research proposal selection is an important activity [1].Ontology based mining techniques for systematic allocation of project proposals to external reviewers. Research project proposals selection is an important and challenging task in organization, when large numbers of research proposals are collected [2].

Survey on an ontology-based text-mining method to cluster proposals for research project selection. The method also includes an optimization model that considers applicants characteristics for balancing proposals by geographical regions. The (OTMM) can also be used to improve the efficiency and effectiveness of research project selection processes in other government and private research funding agencies [3].

Ontox – a method for ontology-driven information extraction (IE). In this paper, we propose an extraction method that utilises the content and pre-defined semantics of ontology formulated in the Web Ontology Language (OWL) to perform the extraction task [4].The ontology extraction maintenance frame work text-to-onto. Ontology play an increasingly important role in Knowledge Management. One of the main problems associated with ontology is that they need to be constructed and maintained [5].

A text mining approach for automatic construction of hypertexts. The hyperlinks were constructed by the creators of the web pages with or without the help of some authoring tools. However, the gigantic amount of documents produced each day prevent from such manual construction [6].

III. PROPOSED METHODOLOGY

In the effectual motive innovation of the text is using to retrieve the data from the data set using text mining. The data retrieving and manipulating the

domain ofthe data mustto characterised the text using text mining. In the large scale industries identify difficult to search and order the area of the work. In this project used to order the area.

To deal with the large volume, it is necessary to group projects according to their same characteristic in research domain and then assigns the projects groups to evaluate. Four to five reviewers are assigned to review each projects so as to assure accurate and reliable opinions on searching the domain of the projects. The area selection and listing the projects are retrieving easily. Evaluating and checking all these proposals according to quantitative and qualitative based, and order the projects.

IV.STEPS IN CHARACTERIZED TEXT

Step 1)Creating the search topics:
The relevant keywords are collected the based on domain. The keywordsof the supportedsearchprojects and their frequencies are counted. The keyword occurrences are the sum of the same keywords that appeared in the ordered store the domain.

Step2)Constructing the search word:
First, the search the database to categorized according to project areas and introduced the domain. It is then developed on the basis of several specific search domain areas.Next it is further divided into some discipline areas; there are some crosschecking project areas. Eg."Data mining "can beplacedunder "Computer Network" domain. Next there are some synonyms used by different projects applicants,which have different names in different projects. At last it leads to project topicsin terms of the feature to set of orders. But represent the equal concepts.

Step 3) Text Document Verification:

After text documents are gathered, they are to be changed into a feature vector format: $VF = (vf_1, vf_2, ..., vf_X)$,where X is the number of features selected and $vf_i (i = 1, 2, ..., X)$ is the keyword v_i.

The searching word describes a weighted method based on inverse document frequency (IDF) combined with the term frequency (TF) to produce the feature v, such thatvi $= tf_i * log(N/df_i)$, where N is the total number of papers in the order, tf_i is the term frequency of the feature word v_i, anddtfi is the number of papers containing the word vi. Thus, research papers can be represented by corresponding feature vectors.

$$v_i = tf_i * log(N/dtf_i)$$

Select the final topic:

When we determine the domain topic and its associated based on weight for each sentence, the selection is most important one and consider it as the main topic for the whole projects.

Step 4) Updating the search database:

Once the database from the project checkingis completed, the researchis updated according to the organization.

V.PROJECT CLASSIFICATION

Projects are classified by the domain areas using to the keyword stored indataset and the topic identified using IdentificationAlgorithm.

VI.CLUSTERING

After the research projects are classified by the domain areas, the projects in eachdomain are clustered using the concept of text-mining technique. The concept-basedmining model fordetecting a raw text document is given as the projects.

Each document has mention sentence compulsory. Each sentence in the document is marked repeat and must have one or more noted verb argument format. The total amount of marked information is totally correct information present in the sentence. The document contained many marked verb format joins many verbs connected with their areas.

VII. EXPERIMENTAL ANALYSIS

The experimental and result deals the algorithm. The result shows better results than the analysis shown in below.
The Table 1 shows the performance of the existing and proposed analysis. The table is described the speed of the speed of the two methods. In this speed testing the searching speed is fast and save the time.

Table 1: Time based Analysis

No	Terms	Existing speed	Proposed speed
	Time based Analysis (ms)		
1	100	2.5	1.39
2	200	2.8	2.01
3	300	3.5	2.8
4	400	4.5	4.25

Proposed algorithm gives as better result than existing algorithms. The algorithm gives an increase in speed to searching the word and finds the domain easily and quickly.
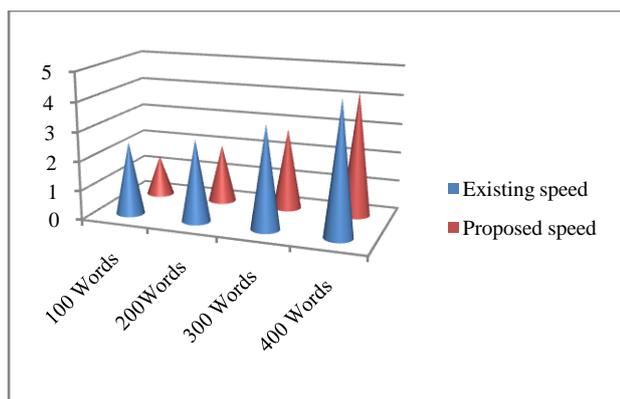
Figure 1: COMPARISION FOR SPEED AND TIME BASED ANALYSIS

The Figure 1 shows the comparison of the speed with existing and with proposed algorithm. The result shows the gives more speed than other existing algorithm.

The speed is calculated based on the execution and response time of the project. The speed improvement gives more efficient and result. The speed is increasing the searching process using this algorithm. The existing algorithms give less performance. However the time is reduced to detect the domain of the proposal. The existing algorithm gives less solution. The proposed methodology gives more efficient result than the existing algorithm. Therefore an attempt is made to increase the speed and decrease the time of detecting domain area.

VIII. CONCLUSION

This paper has presented a framework on innovation of text in the dataset based text mining for grouping search projects and assigning the grouped projects to experts systematic. Research is constructed to categorize the concept terms in different domain areas and to order relationships among them. It using text-mining and cluster research proposals based on their relationship and then to assign them to experts according to their clear domain area. The projects are order toexperts with the help of knowledge based. Future work is needed to replace the work of reviewer by system.

REFERENCES

1. Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, ShouyangWang"An Ontology-Based Text- Mining Method to Cluster Proposals for Research Project Selection",IEEE Trans an systems and humans vol.42,no.3 May2012
2. Xin Xia; Lo, D.; WeiweiQiu; Xingen Wang; BoZhou "Automated Configuration Bug Report Prediction Using Text Mining" Computer Software and Applications Conference (COMPSAC), 2014 IEEE 38th AnnualYear: 2014
3. Wu Maowen; Zhang Cai Dong; LanWeiyao; Wu Qing Qiang "Text topic mining based on LDA and co-occurrence theory", Computer Science & Education (ICCSE), 2012 7th International Conference onYear: 2012
4. Preethi, T.; Lakshmi, R. "An implementation of clustering project proposals on ontology based textmining approach",Information Communication and Embedded Systems (ICICES), 2013 International Conference on Year: 2013
5. 5.Zurada, J.M.; Ensari, T.; Asl, E.H.; Chorowski, J."Nonnegative Matrix Factorization and its application to pattern analysis and text mining"Computer Science and Information Systems (FedCSIS), 2013 Federated Conference onYear: 2013
6. 6.Hoai Nam Vu; Tuan Anh Tran; In Seop Na; SooHyung Kim "Automatic extraction of text regions from document images by multilevel thresholding and k-means clustering" Computer and Information Science (ICIS), 2015 IEEE/ACIS 14th International Conference on Year: 2015Pages: 329 - 334, DOI: 10.1109/ICIS.2015.7166615IEEE Conference Publications
7. 7. Djellali, C. "Enhancing text clustering model based on Truncated Singular Value Decomposition, fuzzy ART and Cross Validation" Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on Year: 2013 Pages: 1078 – 1083IEEE Conference Publications
8. Agnihotri, D.; Verma, K.; Tripathi, P. "Pattern and Cluster Mining on Text Data ",Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference onYear: 2014Pages: 428 - 432, DOI: 10.1109/CSNT.2014.92IEEE Conference Publications
9. Fonda, W.; Purwarianti, A. "Experiments on keyword list generation by term distribution clustering for text classification",Advanced Computer Science and Information Systems (ICACSIS), 2014 International Conference on Year: 2014Pages: 297 - 301, DOI: 10.1109/ICACSIS.2014.7065879IEEE Conference Publications
10. Skabar, A.; Abdalgader, K. "Clustering Sentence-Level Text Usinga Novel Fuzzy Relational Clustering Algorithm "Knowledge and Data Engineering, IEEE Transactions on Year: 2013, Volume: 25, Issue: 1 Pages: 62 - 75, DOI: 10.1109/TKDE.2011.205 Cited by: Papers (3)IEEE Journals & Magazines.
11. V.Vijayadeepa,N.Gomathi, "Data Sharing in the Cloud Computing Security Using JAR" IJPTT-V3i6P103.