

# An ontology based project selection using text mining

Dhiraj Kumar, Prajakta Patil, Pratiksha Lokhande, Rupali Khatale

**Abstract**— Selection of research project is very important task for government & other agencies also, because numbers of proposals are received every year, so it's difficult to group them according to their similarities. Previously grouping of proposals are assigned to experts then manual matching is done on the basis of keyword. But this is not an accurate method for grouping of proposals, because it is not compulsory that every expert has adequate knowledge to grouping proposal in correct group so ontology is developed. In this system Text mining is used for clustering proposals according to similarities or keywords by using SOM algorithm. So that it improves efficiency & effectiveness of research project selection. This project will be helpful for Selection of recent research project proposals for government and private agencies. This project will be helpful for grouping of final year projects in educational institute.

**Index Terms**— Clustering analysis, Ontology, R&D, text mining, SOM&GA Algorithm, Knowledge based agent

## I. INTRODUCTION

Selection of research projects is an important and recurring activity in many organizations such as government research funding agencies. It is a challenging multi process task that begins with a call for proposals (CFP) by a funding agency. The CFP is distributed to relevant communities such as universities or research institutions. The research proposals are submitted to the funding agency and then are assigned to experts for peer review. The review results are collected and the proposals are then ranked based on the Aggregations of the experts' review results. In the NSFC, after proposals are submitted, the next important task is to group proposals and assign them to reviewers. The proposals in each group should have similar research characteristics. For instance, if the proposals in a group fall into the same primary research discipline (e.g., supply chain management) and the number of proposals is small, manual grouping based on keywords listed in proposals can be used. However, if the number of proposals is large, it is very difficult to group proposals manually. Although there are several text-mining approaches that can be used to cluster and classify documents. TMMs (Text Mining Method) which deal with English are not effective in processing Chinese text. To solve the aforementioned problems, an ontology-based TMM (OTMM) is proposed. Ontology is a knowledge repository in which concepts and

terms are defined as well as relationships between these concepts. It consists of a set of concepts, axioms, and relationships that describe a domain of interests and represents an agreed-upon conceptualization of the domain's "real-world" setting. Implicit knowledge for humans is made explicit for computers by ontology. Thus, ontology can automate information processing and can facilitate text mining in a specific domain (such as research project selection).

## II. PROBLEM DEFINITION

Research and development (R&D) project selection is an decision making task commonly found in government funding agencies, universities, research institutes, and technology intensive companies. So, here we are introducing the method for grouping proposals for research project selection is proposed using an ontology based text mining approach to cluster research proposals based on their similarities in research area. The method also includes an optimization model that considers applicants' characteristics for balancing proposals.

## III. OBJECTIVE

Proposed system will use to prevent extra time which required manipulating data. Current methods for grouping proposals are based on manual matching of similar research discipline areas and/or keywords. However, the exact research discipline areas of the proposals cannot often be accurately designated by the applicants due to their subjective views and possible misinterpretations. Rich information in the proposals' full text can be used effectively. Text-mining methods have been proposed to solve the problem by automatically classifying text documents, mainly in English.

## IV. EXISTING METHOD

It is given that previously in the selection of research projects, task for grouping that project according to their similarities was assigned to the division managers and program directors. They were grouped the proposals and assigned to the external reviewers. However, it is not necessary that they both have appropriate knowledge of grouping the proposals and due to this there might be possibility of wrong grouping. Hence, to avoid this there will be the urgent need of effective and feasible approach to group the proposals. Hence here we have introduced the ontology based on text mining for grouping the proposals. And the reason to introduce this system is finally that as we know that every system is not perfect it has certain limitations. We came up certain limitations of existing methodology. With above references we are trying to implement the system in which the grouping of proposals is done very accurately and within a limited time.

**Manuscript received April 07, 2015.**

Dhiraj Kumar, Computer Engineering, MCERC, Nashik, Maharashtra, India

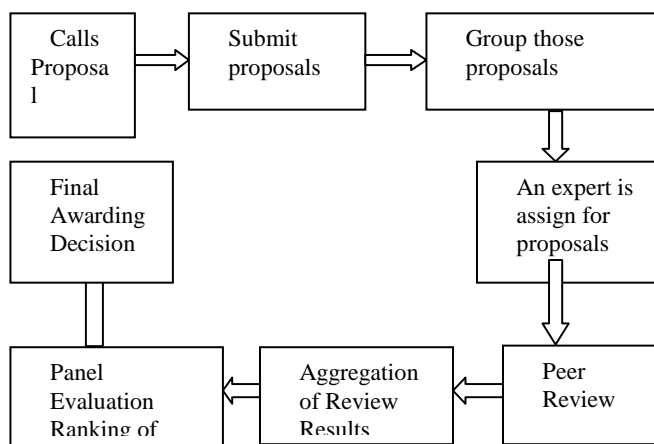
Prajakta Patil, Computer Engineering, MCERC, Nashik, Maharashtra, India

Pratiksha Lokhande, Computer Engineering, MCERC, Nashik, Maharashtra, India

Rupali Khatale, Computer Engineering, MCERC, Nashik, Maharashtra, India

V. RELEVANT THEORY

Fig.1 shows Selection process of project in research i.e. Call for proposals (CFP), proposal submission, proposal grouping, and proposal assign to Experts, Peer review, Aggregation of review results, Panel evaluation & final awarding decision. In past few years many proposals are receive & that proposals are assign to reviewers so as to assure accurate & reliable opinions on proposals. To deal with large number of proposals it's necessary to group them according to their similarities in research discipline. The reviewers or experts then group the proposals & assign then to external reviewers for evaluation. However they may not have adequate knowledge in all research contents of many proposals were not understood when the proposals where group. Hence sit's necessary to design an effective and feasible approach group the submitted research proposals with computer support. So an Ontology based text-mining approach is proposed to solve the problem.



1) Fig.1 Research Project Selection Process

VI. SYSTEM OVERVIEW

The system will be designed as client server based Architecture as shown in Fig 2. Here; the client submits the proposal to the server. All the processing activities will take place at server only and then proposal will be get assigned to particular expert. Expert will get notification about assignment.

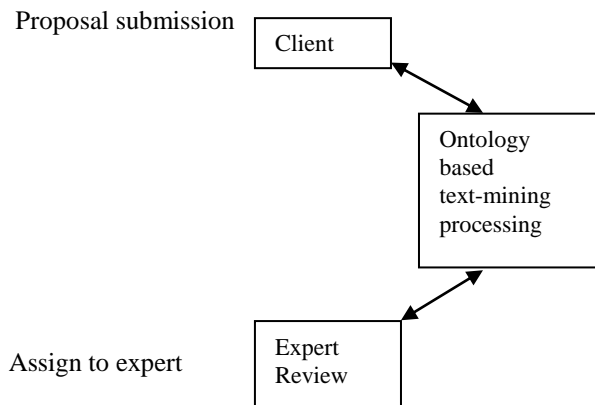


Fig.2 Client-Server Model

VII. FLOW OF SYSTEM

The proposed OTMM is used together with statistical method and optimization models and consists of four phases, as below.

Phase 1: In this phase system store the previous proposals into ontology in data base and it would be to updated continuously. It will help to match the new proposals which will be coming from new users.

Phase 2: In this phase users are going to submit their proposals. After submission of new proposals that proposals are stores into ontology to match with stored proposals. To do this process here Sorting algorithm will work to short new proposals and counting the frequencies of words. After shorting of new proposals those are classified into disciplines.

Phase 3: In this phase actual work of proposed system will work that grouping of proposals according to their similarities. To group proposals according to their similarities text mining will be done by using Self Organize Mapping (SOM) algorithm.

Phase 4: Here clustering of the proposals will be done according to their size and finally Clustered or balanced proposals will be assigned to an expert for peer review.

VIII. CLUSTERING ALGORITHMS

A]. GENETIC ALGORITHM (GA):

- Step 1: Initialize population and parameters; set k=0.
- Step 2: Check stopping condition. If false, continue; If true, stop.
- Step 3: For one generation, perform step 4 to 6.
- Step 4: Breed new offspring through crossover and mutation (genetic operation).
- Step 5: Evaluate the fitness value of parents and offspring.
- Step 6: Select best-ranking offspring to populate and replace worst-ranking parents to form a new generation.
- Step 7: Set k -> k+1; then go to step 2.

B]. SELF ORGANIZED MAPPING ALGORITHM (SOM):

- Step 1: Initialize network weight vectors  $w_i$ , initialize learning rate parameter  $\mu$ , define topologies neighborhood function and initialize parameter  $N_q$ , set k=0.
- Step 2: Check stopping condition, if false, continue; If true, stop.
- Step 3: For each training vector  $x$ , perform steps 4 to 7.
- Step 4: Compute the best match of a weight vector with the input

$$q(x) = \max \text{sim}(x, w_i) \cup$$

Where sim can be calculated as cosine value of the angle between vectors.

Step 5: For all the units the specified neighborhood  $i \in N_q(k)$ , where q is the winning neuron update the weight vectors according to:

$$W_i(k+1) = \left\{ \begin{array}{ll} w_i(k) + \mu(k) [x(k) - w_i(k)] & i \in N_q(k) \\ w_i(k) & i \notin N_q(k) \end{array} \right.$$

Where  $0 < \mu(k) < 1$  (the learning parameter).  
Step 6: Adjust the learning rate parameter.  
Step 7: Appropriately reduce the topological neighborhood  
 $N_q(k)$   
Step 8: Set  $k \rightarrow k+1$ ; then go to step 2.

### IX. PROTOTYPES



Fig. 3 Login page

Fig. 3 Shows the First GUI of the Proposed System In which User are going login in with their User Name and Password which will be provided by Email or Contact Number.



Fig. 4 Registration Page

Fig. 4 Shows the GUI of Login Page Here the details of the users will be Submitted.

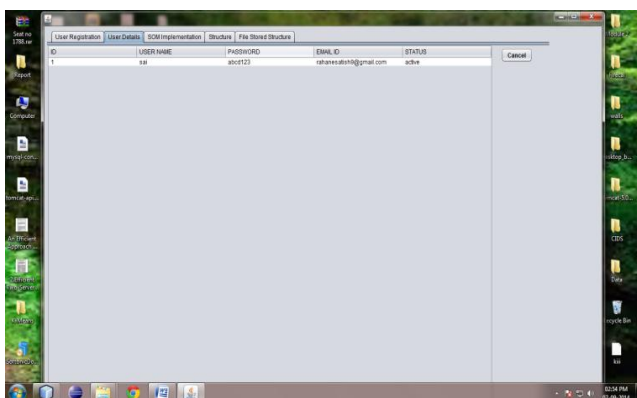


Fig. 5 Log page

Fig. 5 Here the information of users and uploaded file will be visible which would be stored in Data base.

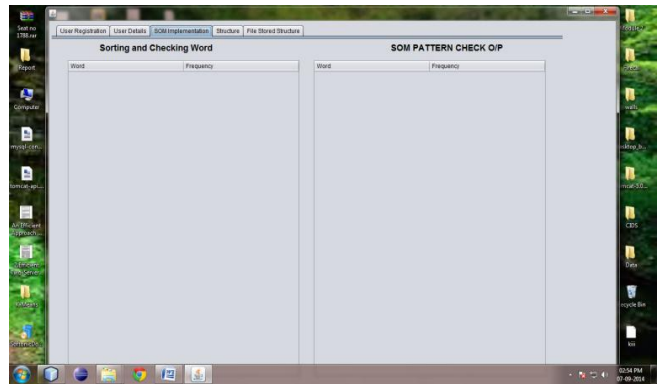


Fig. 6 work page

Fig.6 Here the purpose of implementing Self organized mapping and Genetic algorithm will be working.

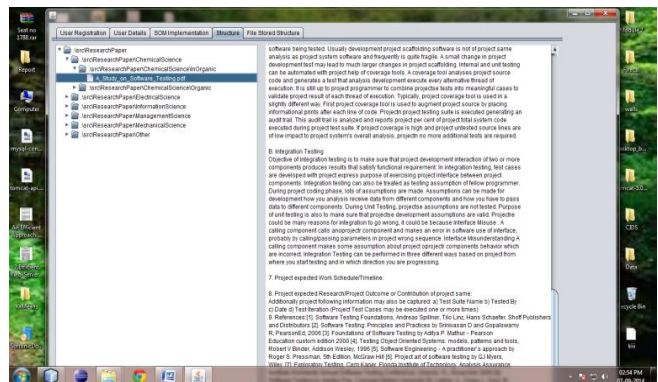


Fig. 7 structure file

### X. APPLICATIONS

1. Useful in any institute private funding agencies and many others.
2. Used in proposal grouping process in the National Natural Science Foundation of India (NSFC).

### XI. ACKNOWLEDGEMENT

We take immense pleasure in thanking our college management for having permitted us to carry out this project work. Words are inadequate in offering our thanks to our respected co-operation in carrying out project work in the starting phase of the project like requirement analysis and literature survey. We also express our at most gratitude to the department of Computer Engineering, for this valuable co-operation in selecting the project topic and guiding us stepwise ahead. We also express our sincere gratitude towards Savitribai Phule Pune University for giving an opportunity to publish our ideas. Inspiration and guidance are invaluable in every aspect of life, especially in field of education, which we have received from our respected project guide who has helped a lot In first two phase of project like topic selection then information gathering and guide us to throughout the project work and gave earnest co-operation whenever required. We would like to express sincere gratitude towards him. At last, we would like to take this opportunity to convey thanks to all my staff members, who directly or indirectly encouraged and helped us project work. Finally yet

importantly, we would like to express heartfelt thanks to our beloved parents for their blessings, our friends/classmates for their and helps and wishes.

## XII. CONCLUSION

From given References and doing literature survey on ontology “Based Project selection method” Using Text mining is to categories concept terms in educational area. This system is mainly useful for grouping of the proposal according to their similarities of Projects in disciplined manner. It also cluster external reviewers based on their research areas and to assign grouped research proposals to review systematically.

This system will help to avoid the process of manual matching and there is no chance to make any misconception during selection of proposals by an expert.

## REFERENCES

- [1] Jian Ma, Wei Xu, Yong-hong Sun, EfraimTurban, Shouyang Wang, and Ou Liu, “An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection,” IEEE
- [2] Transaction on System, man, and cybernetics May, 2012.
- [3] N.Arunachalam, E.Sathya, S.Hismath Begum and M.Uma Makeswari, “An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection,” International Journal of Computer Science & Information Technology (IJCSIT) Vol 5, No 1, February 2013
- [4] MS. K.Mugunthadevi, MRS. S.C. Punitha, Dr..M. Punithavalli, “Survey on Feature Selection in Document Clustering,” International Journal on Computer Science and Engineering (IJCSE), Feb. 2000.
- [5] J. Butler, D. J. Morrice, and P. W. Mullarkey, “A multiple attribute utility theory approach to ranking and selection,” Manage. Sci., vol. 47, 6, Jun.2001.
- [6] C. H. Loch and S. Kavadias, “Dynamic portfolio selection of NPD programs using marginal returns,” Manage. Sci., vol. 48, no. 10, pp. 1227– 1241, Oct. 2002.
- [7] L. M. Meade and A. Presley, “R&D project selection using the analytic network process,” IEEE Trans. Eng. Manag., vol. 49, no. 1, pp. 59–66, Feb. 2002.
- [8] M. A. Greiner, J. W. Fowler, D. L. Shunk, W. M. Carlyle, and R. T. Mcnett, “A hybrid approach using the analytic hierarchy process and integer programming to screen weapon systems projects,” IEEE Trans.Eng. Manag, vol. 50, no. 2, pp. 192–203, May 2003.
- [9] Q. Tian, J. Ma, J. Liang, R. Kowk, O. Liu, and Q. Zhang, “An organizational decision support system for effective R&D project selection,” Decis. Support Syst., vol. 39, no. 3, pp. 403–413, May 2005.
- W. D. Cook, B. Golany, M. Kress, M. Penn, and T. Raviv, “Optimal allocation of proposals to reviewers to facilitate effective ranking,” Manage.Sci., vol. 51, no. 4, pp. 655–661, Apr. 2005.