A Framework on Ontology Based Classification and Clustering for Grouping Research Proposals

KODAM ANUSHA*1

PG Scholar, Dept of CSE, Kakatiya Institute of Technology and Science, Warangal.

Abstract: The Research project proposals selection is an challenging task in organizations and companies. When large numbers of research proposals are collected then research Project Selection is a remarkable in all Research funding and Government agencies. The process of Assignment and Selection are unproblematic. The number of proposal submissions are very less. But with mounting in numbers both assignment and Selection processes will become complicated. The traditional methodologies, Proposal Grouping and Experts grouping experts are done keywords or manually, which yield inefficient experiment results. In present days methodologies are Ontology based Text Mining Method has been used to cluster Proposals based on their likeness in research area by means of Research Ontology method. Where as Reviewers grouping and proposal assignment to experts is still not precise and it is done manually. In this proposed paper proposes Ontology based Text Mining Method to cluster reviewers based on their specialism and overcome the Ontology Mapping Problem to allocate proposals to relevant reviewers analytically. This methodology an effective way for research project selection with escalating proposals and reviewers.

Keywords—Ontology, Clustering, Concept-based analysis, Research Project Selection, Genetic Algorithm, Grouping process, Ontology Matching.

1. INTRODUCTION:
The Research project proposals selection is an important and challenging task by the private funding agencies and government. When large numbers of research proposals are collected. Optimal allocation of research project proposals is a challenging multi process starts with research proposals by a funding agencies. The proposal invoking is
distributed to relevant communities such as research institutions. The Submission of the research proposals by many institutes organizations and companies are assigned to experts for peer view based on their similarity. The review results are examined and the review proposals are then based on ranking and the aggregation of the experts’ review results. Present methodology, Experts grouping, Proposal grouping and Proposal Assignments to experts are done keywords or manually. In the process of manual grouping the reviewers may not have applicants subject view so by misinterpreting the concept of proposals. The reviewers may group proposals wrongly. In the process of keyword based grouping the proposals are grouped together according to keywords similarities. The grouped proposals are very poorly will be assigned to poorly grouped experts. To overcome all these problems Ontology based Text Mining Method has been introduced. The Ontology is a technique that describing knowledge. The set of concepts and the relationship between these concepts for specialized domains. The Domain experts are helping us to construct ontology by providing all domain concepts and their relationships. This Ontology based grouping process has been done semantically.

2. RELATED WORK:
The Research Project Selection task is mandatory as well as complicated process in Research funding agencies. If number of proposal submissions are increases before using OTMM technique more techniques are used for Research Project Selection. Q. Tian, J. Ma [8] and O.Liu [9] proposed Knowledge rules system decision models. The Decision Models use mathematical formulae which gives the optimized results. But the decision models can used only for structured documents and knowledge rules will give only feasible result. T.H. Cheng and C.P. Wei [5] gave clustering based category hierarchy integration techniques to overcome the technique clustering. which is used to category documents hieratically. But there is no intermediate category while clustering and it does not consider the semantic method. this method is also not efficient in heterogeneous scenarios next Ontology based systems are introduced. A. J. C. Trappey et al. Suggested a novel hierachal clustering approaches for knowledge document self organizing using fuzzy ontology method. This method has been used for automatic interpretation and knowledge documents clustering using ontology schema method. But this method is pertinent only for patent documents.
O. Liu and J. Ma[8] established multilingual ontology framework for R&D project management systems which supports languages and facilitates R&D information sharing among users with different culture backgrounds. In this only three languages are taken for Implementation there is no using of context sensitive information. M. Nagy and M. Vargas Vera presented multi agent ontology mapping skeleton which is prerequisite for attaining heterogeneous data integration on web. Using this mature web applications were developed and is used to interpret and align heterogeneous and distributed ontology on semantic web. Jian Ma, Wei Xu[5] and et al. proposed OTMM to cluster and to balance research project proposals. Then these proposals have been assigned to reviewers by hand. This Paper has been extended the work of Jian Ma, Wei Xu and et al. and Pave Shako because of these extension proposals were assigned to reviewers systematically. This methodology is an efficient and effective means for the selection of research project proposals with the escalating quantity of research proposals.

3. METHODOLOGY-
The Research Project Selection process initiates with Call for Proposals then Proposal grouping, Proposal submission, Proposal assignment to experts, review and ranking proposals. This paper presents the methodology to overcome clustering problems. This project has two phases:
1) The Clustering research project proposals.
2) Clustering the Reviewers and Assigning Proposals to appropriate Reviewers.

3.1 Research Ontology for Proposals:
The Research ontology describes various research areas that announced by authority.

**Step 1:** Collecting the research area topics and its related area topics. It is built by adding semantics to topics of the research areas that is grouping. These topics are collected from past proposals the research area topics. e.g. Data Mining where as sub area topic. It means sub area example Web
mining, Text mining, Image mining, video mining etc.

**Step 2:** In Research ontology creation the Ontology constructions can be done either manually or using some tools. Jain Ma has been created ontology manually.

**Step 3:** the revising the research ontology should be done per annum from proposals that are submitted each year.

**Step 1:** Clustered documents Collection. Present research Proposals were collected after classifying each proposal under the discipline areas. All these proposals were collected for preprocessing.

**Step 2:** Preprocessing Collected Documents contents were generally non structural. The total documents are analyzed and extracted. The keywords were identified by using Research Ontology that has been constructed previously. Finally vocabulary size has been abridged by eliminating or deleting words that came into sight only a small amount of times in the final document.

**Step 3:** The Encoding term Frequency Inverse Document Frequency has been used for encoding of keywords. After preprocessing all documents are Segmented and converted to trait vector representation. Encoding of TF-IDF used to produce a trait by using weighted techniques. These techniques based on IDF. The Trait vectors were put up as a consequence of process of Encoding.

\[
\text{Trait } (i) = t_f(i) \times \log \left( \frac{N}{d_i} \right)
\]

**Step 4:** Vector dimension reduction is obligatory to diminish the Dimension of the generated trait vectors. it is too outsized. Latent Semantic Indexing (LSI) has been used to trim down the dimension of the trait vector to generate the semantic relations
among keywords.

**Step 5:** Clustering of Text Vector. The Self Organized Mapping algorithm has been used to group the trait vector based on their similarities. This result of Text Vector Clustering, documents was grouped in research area.

### 3.3 Grouping Reviewers by using Domain Ontology:

The Domain ontology describes various domains. Research and Domain Ontology are almost referring the same but Research ontology has been created for grouping of proposal and Domain ontology has been created for grouping of reviewers.

**Step 1:** The Collecting information about reviewers. The Reviewers who were eligible can register by using this registering details reviewers were clustered

**Step 2:** The Constructing the domain ontology construction can be done either manually or using some tools. Jain Ma et al. has been created research ontology’s manually. Where as here PROTEGE tool is used for ontology construction of research and domain ontology.

**Step 3:** The Revising the domain ontology. Domain ontology Revising should be done annually through topics collected from reviewers details that are registered every year.

### 3.4 Assign Proposals to Experts by Ontology Mapping:

The Research and Domain Ontology were given as input to this method. Ontology mapping will do between them. As a result of this process the balanced proposals were assigned to grouped reviewers for future Review process. The process Ontology matching is done by using the tools like SAMBO Falcon Poznan, Ri MOM U. of Science and Anchor Flood ASMOV Agreement Maker In this SAMBO tool has been used for ontology matching

```plaintext
Algorithm: GA(\eta,\gamma,\mu)
//Initialise generation 0:
\text{\hspace{1cm}}k := 0;
\text{\hspace{1cm}}P(k) := \text{a population of \mu randomly-generated individuals};
//Evaluate \text{ } P(k);
\text{\hspace{1cm}}\text{Compute } \text{fitness}(i) \text{ for each } i \in P(k);
\text{\hspace{1cm}}\text{do}
\text{\hspace{2cm}}//Create generation \text{ } k + 1:
\text{\hspace{3cm}}//1. \text{Copy:}
\text{\hspace{4cm}}\text{Select } (1 - \gamma) \times \mu \text{ members of } P(k) \text{ and insert into } P(k + 1);
\text{\hspace{3cm}}//2. \text{Crossover:}
\text{\hspace{4cm}}\text{Select } \gamma \times \mu \text{ members of } P(k);
\text{\hspace{4cm}}\text{Pair them up, produce offspring;}
\text{\hspace{4cm}}\text{insert the offspring into } P(k + 1)
\text{\hspace{3cm}}//3. \text{Mutate:}
\text{\hspace{4cm}}\text{Select } \mu \times \mu \text{ members of } P(k + 1);
\text{\hspace{4cm}}\text{Invert a randomly-selected bit in each;}
\text{\hspace{4cm}}//Evaluate \text{ } P(k + 1);
\text{\hspace{4cm}}\text{Compute } \text{fitness}(i) \text{ for each } i \in P(k);
\text{\hspace{1cm}}\text{Increment:}
\text{\hspace{2cm}}k := k + 1;
\text{\hspace{1cm}}\text{while fitness of fittest individual in } P(k) \text{ is not high enough;}
\text{\hspace{2cm}}\text{return the fittest individual from } P(k);
\text{\hspace{1cm}}\text{where}
\text{\hspace{2cm}}\eta \text{ is the number of individuals in the population;}
\text{\hspace{2cm}}\gamma \text{ is the fraction of the population to be replaced by crossover in each iteration}
\text{\hspace{2cm}}\mu \text{ is the mutation rate}
```

Fig 4: Genetic Algorithm
3.5. Information Retrieval:
For information retrieval knowledge based agent I used. Which comprises a knowledge base and inference system? Systematic allocation of retrieved grouped proposals to grouped external reviewers is done by knowledge base agent E. the Proposals Assign to Reviewers The Final step of this approach is to assign the Research Proposals group to the External Research Reviewers group systematically. The Proposals of the particular Discipline area is assign to the Reviewers having the same research area. they can examine the proposals efficiently for the peer review.

4. CONCLUSIONS AND FUTURE WORK:
This paper has presented a framework on ontology based classification and clustering for grouping research proposals as well as reviewers. which will be used by research funding Agencies for grouping and assigning the grouped proposal to reviewers group systematically OTMM has been proposed for grouping Proposals and Reviewers. The Genetic Algorithm has been used for balancing and Similarity measure has been calculated for assigning appropriate reviewers for given proposals. And F Measure values, these values illustrate that OTMM grouping was better than existing technologies. The experimental results showed that OTMM clusters proposals efficient manner and yield better results than previous methods. In Future this work can be extended to lots of languages by creating multilingual ontology. Finally with this effort improved. The end results have been achieved while balancing proposals.

Fig 5 : F-Measure in TMM Vs OTMM

REFERENCES:
2008.


ABOUT THE AUTHORS:

KODAM ANUSHA is currently pursuing her M.Tech Software Engineering Department in Kakatiya Institute of Technology and Science, Warangal. She
received her B.Tech in Information Technology from SR institutes of technological sciences, WARANGAL. Her area of interests includes Data mining and Network security.