ONTOLOGY BASED CLUSTERING IN RESEARCH PROJECT SELECTION & ASSIGN PROPOSALS TO EXPERTS BY ONTOLOGY MATCHING

M.Lavanya [1], N. Rajkumar[2]

[1]*PG Scholar, Department of Software Engineering (M.E.), Sri Ramakrishna Engineering College, Coimbatore, India*
[2]*Head of the department, Department of Software Engineering (M.E.), Sri Ramakrishna Engineering College, Coimbatore, India*

Abstract

Research Project Selection is a remarkable bustle in all Research funding and Government agencies. Selection and Assignment processes are unproblematic when number of proposal submission is less. But with mounting in numbers, both processes will become complicated. Through traditional methodologies, Proposal Grouping, Experts grouping and Proposal assignment to experts are done manually or via keywords which yield inefficient results. In recent methodologies, Ontology based Text Mining Method (OTMM) has been used to cluster Proposals based on their likeness in research area by means of Research Ontology. Whereas, Reviewers grouping and proposal assignment to experts is still not precise since it is done manually. This paper proposes Ontology based Text Mining Method to cluster reviewers based on their specialism and overcome the Ontology Mapping Problem to allocate proposals to appropriate reviewers analytically. This methodology affords an effective way for research project selection with escalating figure of proposals and reviewers.

Keywords— *Ontology, Research Project Selection, Genetic Algorithm, Grouping process,Ontology Matching*

---***---

1. INTRODUCTION

Research Project Selection is most important and challenging task for awarding the praiseworthy papers. Fig 1 shows steps in Research Project Selection following from early days. The Process begins with call for proposals. Research Funding and Government agencies give an announcement about their research requirements to public. Proposal submission will do by people on or before last date that is announced by agencies. The submitted papers will be grouped according to the Research area. These grouped papers will be assigned to experts for review who already grouped according to their domain specialization and the reviewed papers will be then ranked and awarded.

In current methodology, Proposal grouping, Experts grouping and Proposal Assignment to experts are done manually or using keywords. In manual grouping, reviewers may not have applicants subject view so by misinterpreting the concept of proposals, reviewer may group proposals wrongly.   In keywords based grouping, the proposals are grouped according to keywords similarities but not semantically so this methodology result wrong groups. Poorly grouped proposals will be assigned to poorly grouped experts.

To overcome these problems, Ontology based Text Mining Method (OTMM) has been introduced.   Ontology is a technique that symbolizes knowledge by describing the set of concepts and the relationship between these concepts for specialized domain. Domain experts help has been used to construct ontology by providing all domain concepts and their relationship. By means of this Ontology based grouping, grouping process has been done semantically as well as systematically.

2. RELATED WORKS

Research Project Selection task is mandatory as well as complicated process in Research funding agencies if number of proposal submission increases.   Before using OTMM technique, several techniques are used for Research Project Selection.

Q. Tian, J. Ma and O.Liu [15] proposed a decision models and Knowledge rules system. Decision Models use mathematical formulae which gives the optimized results.   But decision models can used only for structured documents and knowledge rule will give only feasible result.

T.H. Cheng and C.P. Wei [2] gave clustering based category hierarchy integration techniques to overcome the technique clustering based category integration which is used to category documents hieratically. But there is no intermediate category

while clustering and it does not consider the semantic. Also this method is not efficient in heterogeneous scenario.

Next Ontology based systems are introduced. A. J. C. Trappey et al. [16] suggested a novel hierarchal clustering approach for knowledge document self organizing using fuzzy ontology. This method has been used for automatic interpretation and knowledge documents clustering using ontology schema. But this method is pertinent only for patent documents.
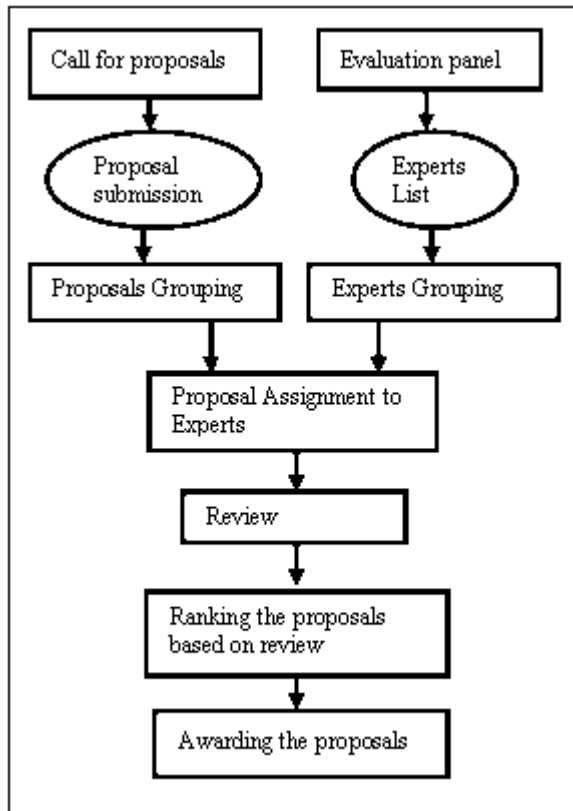

Fig 1 Research Project Selection Process

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 and usage preferences. But only three languages are taken for implementation. There is no context sensitive information used.

M. Nagy and M. Vargas-Vera [12] presented multi-agent ontology mapping skeleton which is prerequisite for attaining heterogeneous data integration on semantic web. Using this mature Semantic web applications were developed and is used to interpret and align heterogeneous and distributed ontology on semantic web and it promised "Machine intelligence".

Jian Ma, Wei Xu and et al. [5] proposed OTMM to cluster and to balance research project proposals. Then these grouped proposals have been assigned to reviewers by hand not systematically.

Pavel Shvaiko and Je´ roˆme Euzenat [18] reported Ontology Matching: State of the Art and Future Challenges. In this paper two ontology has been used, so Ontology Matching has to be done to assign Proposals to Experts according to their specialization.

This Paper has been extended the work of Jian Ma, Wei Xu and et al. [5] and Pavel Shvaiko and Je´ roˆme Euzenat [18]. Because of the extention grouped proposals were assigned to grouped reviewers systematically. This methodology affords an efficient and effective means for the selection of research project proposals with the escalating quantity of research proposals and reviewers.

3. RESEARCH METHODOLOHY

Research Project Selection process initiates with Call for Proposals then Proposal submission, Proposal grouping, Experts grouping, Proposal assignment to experts, Review and finally Ranking and Awarding proposals. The paper presents the methodology to overcome the manual or keywords based clustering problems. The project has two phases: 1) Clustering the Research project proposals 2) Clustering the Reviewers and Assigning Clustered Proposals to appropriate Reviewers.

Pre requisites have been done as per the paper "An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection" [5] given by Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang, and Ou Liu.

3.1 Developing a Research Ontology for Proposals Grouping

Research ontology describes various research areas that announced by authority.
Step 1: Collecting the research area topics and its sub-area topics. Ontology is built by adding semantics to topics of the research area that is collected. These topics are collected from past five year proposals. Research area topics it means main area e.g. Data Mining whereas, sub-area topic it means sub-area e.g. Text mining, Image mining, Web mining, Video mining and so on.
Step 2: Research ontology creation. Ontology construction can be done either manually or using tools. Jain Ma et al. has been created ontology manually. Whereas, Here **PROTÉGÉ** tool is used for ontology construction. [17].
Step 3: Revising the research ontology. Revising should be done per annum via topics collected from proposals that are submitted each year.
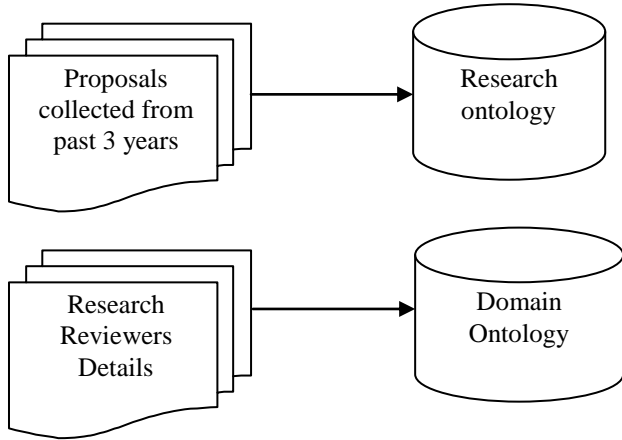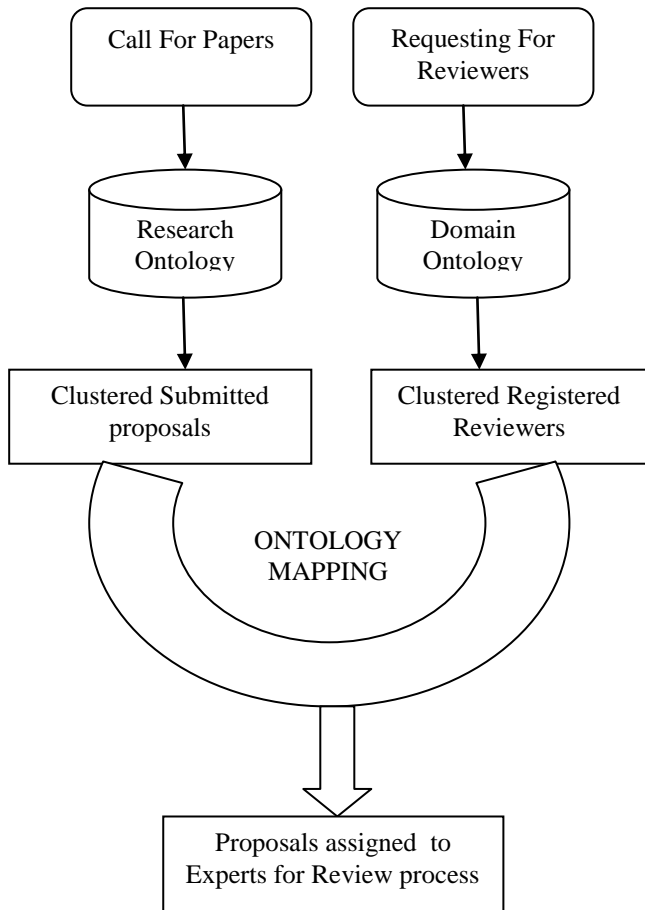
Fig 2 Ontology creation

Fig 3 overall Architecture

3.2 Classifying New-Fangled Research Proposals

New Proposals are given as input to the research ontology that is created. As a result, new proposals are   classified under the accurate Research area.

3.3 Clustering Research Proposals under its Sub-Area Using Text Mining

After classifying, new proposals are clustered under its sub-research area. This is done based on their similarities using text mining. The steps are given below.

Step 1: Collecting the clustered documents. Research Proposals were collected, after classifying each proposal under the discipline areas. These proposals were collected for preprocessing purpose.

Step 2: Preprocessing. Collected Documents contents were generally nonstructured. So, the full documents have been analyzed, extracted and keywords were identified by using Research Ontology that have been constructed earlier. Finally, vocabulary size has been abridged by eliminating words that came into sight only a small amount of times in the document.

Step 3: Encoding. Term Frequency – Inverse Document Frequency (TF-IDF) has been used for encoding of keywords Wi process.   After preprocessing, all documents are segmented and converted to trait vector representation. TF-IDF encoding used to produce a trait by using weighted technique that based on IDF pooled with TF. Trait vectors were put up as a consequence of Encoding process.

The trait was generated by

$$Trait(i) = tfi * \log(N / di) \qquad (1)$$

Where

tfi is the term frequency of the keyword Wi

N is the number of proposals

dfi is the number of proposals comprises the

keyword Wi

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

Step 5: Text Vector clustering. Self Organized Mapping (SOM) algorithm has been used to group the trait vector based on their similarities. As a result of Text Vector Clustering, documents were grouped in apt research area.

3.4 Balancing and Regrouping Research Proposals

If in case, any particular research area may have more papers than other research areas. Before giving it to experts, balancing the clusters should be done. As a result, if there is any possibility for re-grouping it would be done using Genetic algorithm. Fig 4 shown the Pseudo code of genetic algorithm

3.5 Grouping Reviewers by using Domain Ontology

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

Step 1:Collecting information about reviewers. Reviewers those who were eligible can register. By using this registering details reviewers were clustered

Step 2: Constructing the domain ontology. Ontology construction can be done either manually or using tools. Jain Ma et al. has been created research ontology manually. Whereas, Here **PROTÉGÉ** tool is used for ontology construction both research and domain ontology [17]

Step 3: Revising the domain ontology. Revising should be done annually through topics collected from reviewers' details that are registered each year.

3.6 Assign Proposals to Experts by Ontology Mapping

Both Research and Domain Ontology were given as input to this module. Ontology mapping will do between these two ontology. As a result of this process the balanced proposals were assigned to grouped reviewers for further Review process.

Ontology matching is done by using the tools like SAMBO (Linko¨ pings U.), Falcon (Southeast U.), DSSim (Open U., Poznan U. of Economics), RiMOM (Tsinghua U., Hong Kong U. of Science and Technology), Anchor-Flood (Toyohashi U. of Technology), ASMOV (INFOTECH Soft, Inc., U. of Miami), AgreementMaker (U. of Illinois at Chicago). In this paper, SAMBO tool has been used for ontology matching.

Algorithm: $GA(n, \chi, \mu)$

//**Initialise generation 0:**

$k := 0;$

$P(k) := a$ population of n randomly-generated individuals;

//**Evaluate** $P(k)$:

Compute $fitness(i)$ for each $i \in P(k);$

do
{

//**Create generation** $k+1$:

//**1.Copy:**

Select $(1 - \chi) \times n$ members of $P(k)$ and insert into $P(k+1)$;

//**2.Crossover:**

Select $\chi \times n$ members of $P(k)$;

Pair them up; produce offspring;

insert the offspring into $P(k+1)$

//**3.Mutate:**

Select $\mu \times n$ members of $P(k+1)$;

Invert a randomly-selected bit in each;

//**Evaluate** $P(k+1)$:

Compute $fitness(i)$ for each $i \in P(k);$

//**Increment:**

$k := k+1;$

}

while fitness of fittest individual in $P(k)$ is not high enough;

return the fittest individual from $P(k)$;

where

n is the number of individuals in the population;

χ is the fraction of the population to be replaced by crossover in each iteration

μ is the mutation rate

Fig 4 Genetic Algorithm

3.7 Performance Evaluation

F-Measure value will give the quality of the Cluster and is used to compare how similar two clustering results are. Higher the F-Measure value results higher quality of grouping.

Fig 5 represents the F-measure value against number of proposals. F- Measure value is calculated using Precision and Recall. F-Measure is calculated as follows,

$$F = (2 * precision * recall) / (precision + recall) \quad (2)$$

where:

Precision= (cells correctly put into a cluster) / (total cells put into the cluster)   (3)

Recall = (cells correctly put into a cluster) / (All the cells that should have been in the cluster)   (4)

The experimental result was shown that the quality of the grouping and assigning process has been improved drastically when compared to the TMM. F-Measure value was high for OTMM when compared to TMM. Also, Genetic algorithm and similarity measures (Euclidean distance) improved the Precision and Recall values.

4. CONCLUSIONS AND FUTURE WORK

OTMM has been proposed for grouping Proposals and Reviewers. Genetic Algorithm has been used for proposal balancing and Similarity measure (Euclidean distance) has been calculated for assigning appropriate reviewers for given proposals. And F-Measure values illustrate that OTMM grouping was better than existing technologies that used for grouping. The experimental results showed that OTMM clusters proposals and reviewers in efficient manner and yield better result than previous method TMM.

In Future, this work can be extended to lots of languages by creating multilingual ontology [8]. Finally, with this effort improved end result has been achieved while balancing proposals.
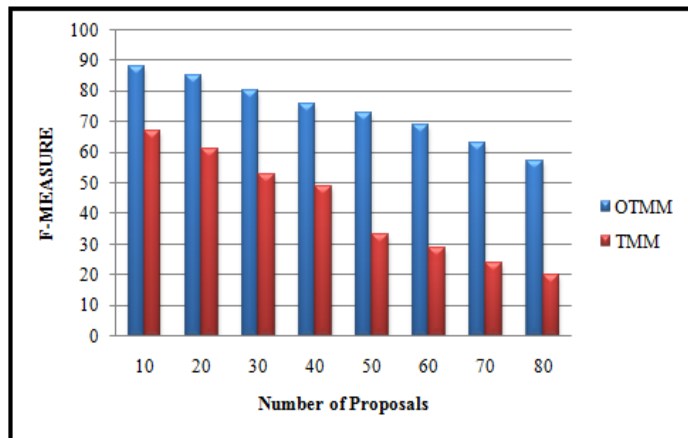

Fig 5 F-Measure in TMM Vs OTMM

ACKNOWLEDGEMENTS

REFERENCES

[1] Cai .M, Zhang .Y .W, and Zhang .K, "ManuHub: A semantic web system for ontology- based service management in distributed manufacturing environments," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 41, no. 3, pp. 574–582, May 2011.

[2] Cheng .T .H and Wei .C .P, "A clustering-based approach for integrating document-category hierarchies," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 38, no. 2, pp. 410–424, Mar. 2008.

[3] Chiang .D .A, Keh .H .C, Huang .H .H, and Chyr .D , "The Chinese text categorization system with association rule and category priority," Expert Syst. Appl., vol. 35, no. 1/2, pp. 102–110, Jul./Aug. 2008.

[4] Fan .Z .P, Chen .Y, Ma .J, and Zhu .Y, "Decision support for proposal grouping: A hybrid approach using knowledge rule and genetic algorithm," Expert Syst. Appl., vol. 36, no. 2, pp. 1004–1013, Mar. 2009.

[5] Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang, and Ou Liu, " An Ontology-Based Text-Mining Method to Cluster Proposals for Research Project Selection," IEEE Trans. Syst., Man, Cybernetics. A, Syst., Humans, vol. 42, no. 3, pp. 784–790, May. 2012.

[6] Liang .Q, Wu .X, Park .E .K, Khoshgoftaar .T .M, and Chi .C .H, "Ontology based business process customization for composite web services," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 41, no. 4, pp. 717–729, Jul. 2011.

[7] Lim .G .H, Suh .I .H, and Suh .H, "Ontology-based unified robot knowledge for service robots in indoor environments," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 41, no. 3, pp. 492–509, May 2011.

[8] Liu. O and Ma .J, "A multilingual ontology framework for R&D project management systems," Expert Syst. Appl., vol. 37, no. 6, pp. 4626–4631, Jun. 2010.

[9] Liu .Y, Jiang .Y, and Huang .L, "Modeling complex architectures based on granular computing on ontology," IEEE Trans. Fuzzy Syst., vol. 18, no. 3, pp. 585–598, Jun. 2010.

[10] Liu .Y, Wang .X, and Wu .C, "ConSOM: A conceptional self-organizing map model for text clustering," Neurocomputing, vol. 71, no. 4–6, pp. 857–862, Jan. 2008.

[11] Lu .C, Hu .X, and Park .J .R, "Exploiting the social tagging network for web clustering," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 41, no. 5, pp. 840–852, Sep. 2011.

[12] Nagy .M and Vargas-Vera .M, "Multiagent ontology mapping framework for the semantic web," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 41, no. 4, pp. 693–704, Jul. 2011.

[13] Razmerita .L, "An ontology-based framework for modeling user behavior—A case study in knowledge

management," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 41, no. 4, pp. 772–783, Jul. 2011.

[14]   Sun .Y .H, Ma .J, Fan .Z .P, and Wang .J, "A hybrid knowledge and model approach for reviewer assignment," Expert Syst. Appl., vol. 34, no. 2, pp. 817–824, Feb. 2008.

[15]   Tian .Q, Ma .J, Liu .O, Liang .J, Kowk .R, and Zhang .Q, "An organizational decision support system for effective R&D project selection," Decis. Support Syst., vol. 39, no. 3, pp. 403–413, May 2005.

[16]   Trappey .A .J .C, Trappey .C.V, Hsu .F .C, and Hsiao .D .W, "A fuzzy ontological knowledge document clustering methodology," IEEE Trans.Syst., Man, Cybern. B, Cybern., vol. 39, no. 3, pp. 806–814, Jun. 2009.

[17]   Website: http://www.Protégé.Standford.edu

[18]   Pavel Shvaiko and Je´roˆme Euzenat, " Ontology Matching: State of the Art and Future Challenges," IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 1, Jan 2013

BIOGRAPHIES

Dr.N. Rajkumar obtained his Bachelor's degree in Computer Science and Engineering from Madurai Kamaraj University in 1991 and His Masters in Engg. the same stream in the 1995 from Jadavpur university, Kolkata. He has completed his Masters in Business Administration from IGNOU in the year 2003. His doctorate is in the field of Data Mining, which he completed in 2005 from PSG College of Technology, Coimbatore. He is currently the Head of the department of Computer Science and Engineering at Sri Ramakrishna Engineering College, Coimbatore, Tamilnadu. He has served in the field of education for over 20 years at various Technical Institutions. He has been instrumental in the conduct of 30 short-term courses and has also attended 20 courses conducted by other institution and organizations. He has authored for 2 books for the benefit of the student community in Networking and Computer Servicing. He has published as many as 50 papers in International Journals, Conferences and at the National level in his area of expertise namely Data Mining, Networking and Parallel computing respectively. He has guided 100- Project Scholar's to-date. His E-mail id is nrk29@rediffmail.com

M.Lavanya received her B.E CSE from Sri Ramakrishna Engineering college, Coimbatore affiliated to Anna University, Chennai in 2012. At present, She is pursuing her M.E Software Engineering in the same institute affiliated to Anna University Chennai. Her E-mail id is iamlavs38@gmail.com