THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

An Ontology Based Text-Mining to Clustering the Research Projects Based on Fuzzy Technique

D. Saravana Priya
Assistant Professor, Department of Information Technology,
P. A. College of Engineering and Technology, Affiliated to Anna University, Chennai, India
Dr. M. Karthikeyan
Professor, Department of Electronics and Communication Engineering,
Tamil Nadu College of Engineering, Affiliated to Anna University, Chennai, India

Abstract:
Both the internet and the intranets contain more resources and they are called as text documents. Research and development (R&D) scheme selection is a type of decision-making normally present in government support agencies, research institutes, technology intensive companies and universities. Text Mining has come out as an authoritative technique for extracting the unknown information from the large text document. Ontology is defined as a knowledge storehouse in which concepts and conditions are defined in addition to relationships between these concepts. Ontology's build the task of searching alike pattern of text that to be more effectual, efficient and interactive. The present method for combine proposals for selection of research project is proposed by ontology based text mining technique to the data mining approach of cluster research proposals support on their likeness in research area. Though the research proposal regarding particular research area is cannot always be accurate. This paper proposes an ontology based text mining to group required data such as research proposals and external reviewers based on their research area. The proposed method like Fuzzy SOM and the NRGA algorithm is used to cluster here. The proposed method is efficient and effective for clustering research proposals. The experimental result is evaluated based on F-measure which proves that the proposed approach gives improved results.

Keywords: *Clustering analysis, Fuzzy SOM, knowledge based agent, NGRA Algorithm, Ontology, R&D and Text mining.*

1. Introduction

Owing to fast development of digital data that are present in current years, the knowledge discovery and data mining are the approaches plays a significant role for revolving such data into useful information and knowledge. Speaking is a wide task is familiar with the efforts for the purpose of two major things. The first thing is the researchers and practitioners functioning in the research areas to get back the information and text mining to find the textual resources by automatic methods. This method is to define a metric on a text or a document to group the neighbouring documents into resourceful document groups is called text clustering [i]. The next thing accumulated, is a metric on a document allocated to list of targeted lists. By this the new documents are consigned to labels and it is called as text classification [ii]. In the areas of thesauri [iii] and ontologies [iv] the researchers are working and a certain documents are assigned to form the theoretical structures.

In government research agencies the research projects are selected and it is the frequent activity that to be happens over there. The selected research proposals are made to review. Utmost five members are allocated to each proposal to estimate the correctness of it. If the proposals are there in excess, it is required to cluster or group the proposals according to their unique nature. The grouped proposals are taker to the reviewer belongs to their respective disciplines. Generally the project selection is done once in a year by scheduling all the projects to calculate the quantitative and qualitative criteria of the projects.

In Research and Development during selection process the decision makers are separated into six groups, they unite together in linked manner to select the best proposals. They select the best proposal in definite manner and the output of one group is given as an input to another group. The R&D would be in charge for the selection of the proposal and it offers a task to program. The program directors send the selected proposal to the external reviewers for the evaluation. Though, the external reviewers do not contain sufficient knowledge in all the research areas. So an effective approach called an ontology-based text-mining approach is proposed to cluster the submitted research proposals to the external reviewers with the computer supports.

2. Literature Survey

In [v] Text-To-Onto ontology based approach is introduced by means of supervised learning. The group decision support approach is proposed in [vi]. The author Choi and Park introduced a text mining approach for R&D to select proposal [vii]. Roussinov and Chen introduced Document clustering for an electronic meeting is explained in [viii]. A preference and content based approaches are introduced to increase the efficiency of the document clustering is studied in [ix]. Runkler and Bezdek introduced a Web mining with relational clustering [x]. Minimum entropy clustering method is proposed to gene expression [xi]. Gauch, Chaffee, and Pretschner[2003] proposed a Ontology-based personalized search and browsing[xii]. Meade and Presley [2002] established an R&D project range using the analytic network process [xiii]. An Automatic Topic Identification Algorithm is developed in [xiv]. Hettich and Pazzani [2006] proposed a text-mining approach to cluster the proposals, identify reviewers, and assign reviewers to proposals [xv].

In [xvi] the author proposed an optimal allocation method to the reviewers to select a best proposal. A rotation program method is proposed in [xvii]. Choi and Park proposed a text-mining approach for R&D proposal selection in [xviii]. Girotra et al. [xix] presented an experiential study to value projects in a selection. Sun et al. proposed a decision support system to evaluate reviewers for research project selection in [xx]. Hybrid knowledge-based and modeling approach is proposed in [xxi].

An OWL Web Ontology Language is established in [xxii] for selecting the best research proposals. An ontoX—A method is designed for ontology-driven information extraction [xxiii]. Jian Ma, Wei Xu, Efraim Turban, Shouyang Wang, Yong-hong Sunpresents an method to Cluster Proposals for Research Project. The author develops an Ontology-Based Text-Mining Method based on their analogous discipline areas. In this method the research proposal is cluster effectively both English and Chinese texts [xxiv].

3. Ontology Based Text Mining

The important work of the R&D is to cluster the proposals after they are submitted. The proposals are separated into certain groups and the groups which are belong to same research area. For example, if the group of proposals is belong to identical research area and it contains less number of proposals, manual grouping is done based on keywords presented in the proposals and give them to reviewer.

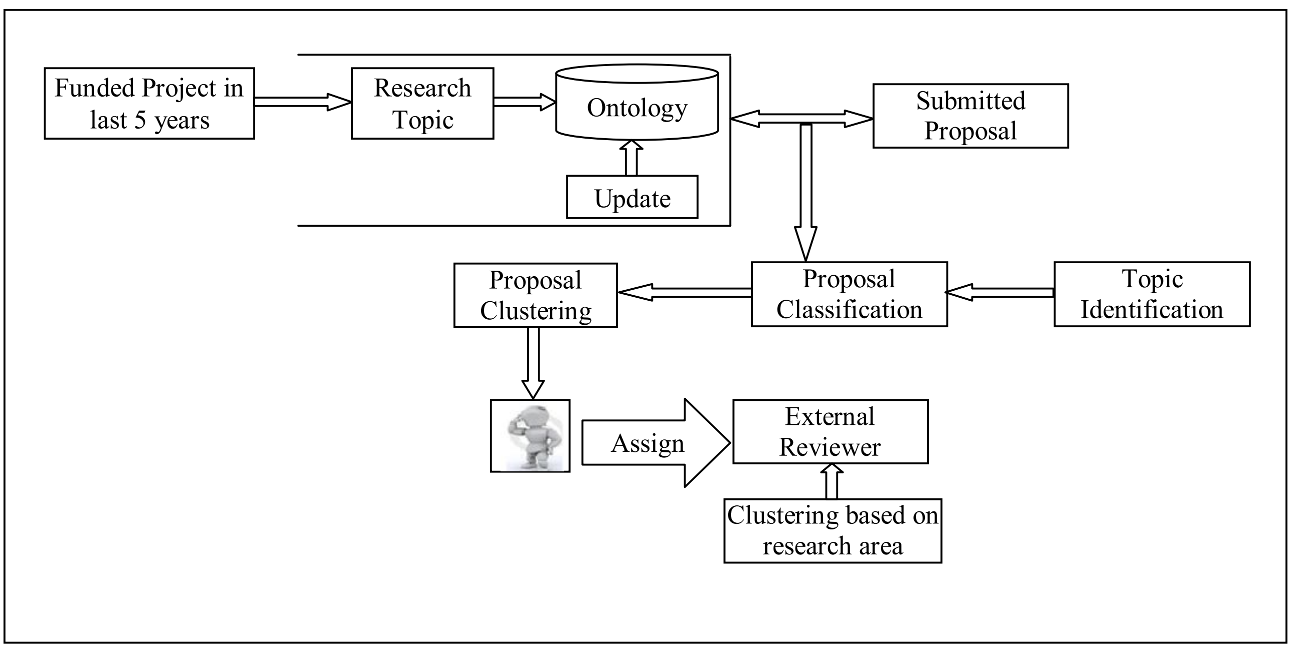

Figure 1: Proposed framework

But if the required proposal volume is large it is complex to group them manually, so to classify the large volume of the proposals an ontology based approach is used and to cluster the proposal using text mining approach at last submit to the reviewer systematically.

The ontology based text mining consist of following four phases

- Constructing a Research Ontology
- Classifying Disciplines based on New Research Proposals
- Similarity based Research proposal clustering using text mining
- Research Proposals Balancing and Regrouping Research Proposal by Considering Applicants' Characteristics

4. Research Ontology Construction

4.1. Step1: Creating the research topics:
While making a research projects, many keywords are used over there and also in supporting research documents and their corresponding frequencies are considered for every year. The keyword frequency is the sum of the similar keywords that shown in the regulation for the period of the five years.

4.2. Step2: Constructing the Research Ontology
According to scientific research area the research ontology is classified and it is improved on specific research area.
Then it is divided into their respective discipline. Lastly the research topic is selected in terms of feature set. The researchers may construct their research proposal in different names with similar concepts.

4.3. Step 3: An Automatic Approach for Topic Identification

4.3.1. Split the Text into Sentences
The algorithm in which the first step is to split the sentence, a smallest text part contains a topic. Therefore, split the document into equivalent sentences. In this research use a Proxem Antelope (Proxem, 2009) which provides an open-source abundance of NLP tool. This tool contains a text splitter to split the sentence, by which would have a set of sentences.

4.3.2. Pars the Sentences
This step is to pars the sentences to find the candidate name initially to stay away from unwanted calculation. Syntactic parts like Noun Phrase (NP) and Verb Phrase (VP) are the most significant roles to present the meaning of the sentence and consider them instead of grammatical roles like noun and verb to spot the candidate topic for each sentence. These syntactic parts are easy to get through a dependence syntactic parser.

4.3.3. Select the Candidate Parts
Select noun phrase (NP) and the head of a Verb Phrase (VP) as an alternative of just pairs of nouns and noun-verb. Assume that the most significant parts from a sentence are the NP's that function as subject and the head of the VP. The combination of three topics is considered as a candidate topic.

4.3.4. Weight Calculation for Candidate Topic
In this step calculate the IDF and SNV only for essential syntactic parts. The formula to calculate the weight is

$$SNV\ (NP, head\ (VP)) = IDF\ (NP).IDF \frac{(head\ (VP))}{D\ (NP, head\ (VP))}$$

4.3.5. Select the Final Topic
While verifying the candidate topic and its connected weight for each sentence, select the most weighted one and think it as the main topic for the whole document. Suppose there are more than one candidate topics with greatest weight, consider all of them as the main topic.

4.4. Step 4: research ontology updation
The research ontology is updated once the project funding is completed each year, according to agency's policy.

4.4.1. Classifying Disciplines based on New Research Proposals
Proposals are classified by the discipline areas according to the keyword stored in ontology and the topic identified using Topic Identification Algorithm. In this paper a modified SOM that hybridizes with fuzzy c-means rules to cluster the proposals.

4.4.2. Standard SOM
SOM is like artificial neural networks with respect to the unsupervised learning. The SOM is an easy implementation tool and it is used to reduce the dimensionality reduction. To get accurate result instead of SOM, fuzzy SOM is proposed.

4.4.3. Fuzzy SOM
Let X_{PN}denote the input space (matrix), M_{LN} the weight vectors (codebook), $d_{PL}$the Euclidean distance measure, R_{PL} the fuzzy membership function, and $h_j$the neighborhood function, where P, N and L are positive integers.
Step 1: Initialize the weight vectors (codebook).
Step 2: Compute d, the Euclidean distance measure:

$$d_{lj}(t) = \sqrt{\sum_{i=1}^{N}\left(X_{li} - M_{ji}(t)\right)} \begin{matrix} l = 1,2,3, \ldots .. P \\ j = 1,2,3, \quad L \end{matrix} \quad (1)$$

Step 3: membership computation of every map unit:

$$R_{lj} = \frac{\left(\frac{1}{d_{lj}^2(t)}\right)^p}{\sum_{m-1}^{L}\left(\frac{1}{d_{lm}^2(t)}\right)^p} \quad \begin{matrix} l = 1,2,3,\ldots..P \\ j = 1,2,3, \quad L \end{matrix} \qquad (2)$$

where p is a constant integer whose value is arbitrary and dependant on the size of the dataset.
Step 4: Update the weights vector:

$$M_{ji}(t+1) = M_{ji}(t) + h_j(t).\frac{\sum_{l=1}^{M} R_{lj}(t).\left(x_{li} - M_{ji}(t)\right)}{\sum_{l=1}^{M} R_{lj}(t)} \quad \begin{matrix} l = 1,2,3,\ldots..P \\ j = 1,2,3, \quad L \end{matrix} \qquad (3)$$

where, t is an iteration step, instead of a variable. This algorithm runs iteratively for a set number of iterations. The stopping criteria used in this work is a fixed training length, where the number of training period is based on the dimensions of the dataset and number of map units used. The original SOM toolbox uses this stopping criterion, and it remains unchanged for the implementation of the MFSOM algorithm. At the every iteration one random input vector is chosen. Subsequently, the smallest Euclidean distance of the selected input vector to the codebook is chosen as the best matching unit (BMU). This is used for the calculation of the simple Gaussian neighbourhood function used in the neighbourhood function, h, where it is given as:

$$h_{ci}(t) = a(t).e^{\frac{d_{ci}^2}{2\sigma^2(t)}} \qquad (4)$$

Where d_{ci} is the distance between the BMU and the current map unit, σ (t) is the current neighbourhood radius and α (t) is the training rate (non-increasing function).

The training rate α (t) used is defined as:

$$a(t) = a_0\left(\frac{a_T}{a_0}\right)^{\frac{1}{T}} \qquad (5)$$

where $\propto_0$ is the initial learning rate, $\propto_T$ the final learning rate, t is the iteration step and T is the learning length. The learning rate is determined based on the size of the dataset.

5. Similarity Based Research Proposal Clustering Using Text Mining

Once the classification of research proposals in discipline areas, the proposals in each discipline are clustered using the text-mining technique [xviii], [xix]. The main clustering process consists of five steps, as shown in Fig. 3: text document collection, text document pre-processing, text document encoding, vector dimension reduction, and text vector clustering.

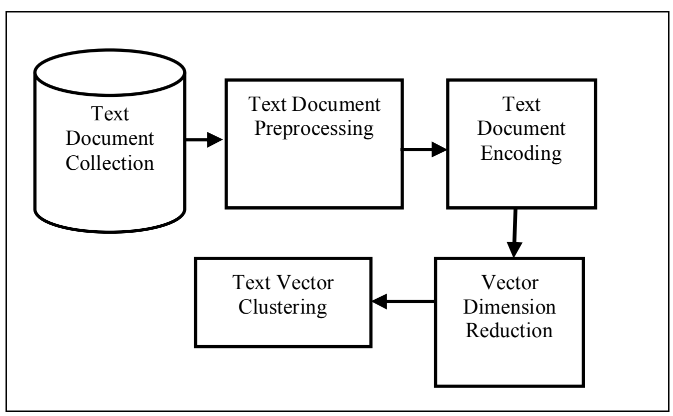

Figure 2: Main process of text mining

5.1. Step 1: Text document collection

The proposal documents in each discipline $A_k(k = 1, 2, \ldots, K)$ are collected after the research proposal is classified.

5.2. Step 2: Text Document Pre-Processing

The contents of proposals are generally nonstructural. Because in the proposal consist of Chinese characters which are hard to segment, the research ontology is used to study, extract, and identify the keywords in the full text of the proposals.

5.3. Step 3: Text Document Encoding

After text documents are segmented, they are converted into a feature vector illustration: $V = (v_1, v_2. .. v_M)$, where M is the number of features selected and $v_i (i = 1, 2, \ldots, M)$ is the TFIDF encoding [xviii] of the keywordw_i. TF-IDF encoding describes a weighted method based on inverse document frequency (IDF) combined with the term frequency (TF) to produce the feature v, such that $v_i = tf_i * \log(N/df_i)$, where N is the total number of proposals in the discipline, tf_i is the term frequency of the feature word w_i, and df_i is the number of proposals containing the word w_i. Thus, research proposals can be represented by corresponding feature vectors.

5.4. Step 4: Vector Dimension Reduction

The feature vector dimension is large, so that the vector size is reduced automatically by selecting a subset which consists of more number of keywords. Latent semantic indexing (LSI) is used to solve the problem [xxvi]. It not only reduces the dimensions of the feature vectors also generate the semantic relations between the keywords. LSI is a technique for replacement of the original data vectors with shorter vectors in which the semantic information is conserved. Without losing the information in a proposal, a term-by-document matrix is created, where there is one column that corresponds to the term frequency of a document. Also, the matrix is decayed into a set of eigenvectors by means of singular-value decomposition. Thus, the document vector formed from the term of the enduring eigenvectors has a very small dimension and retains approximately all of the related original features.

5.5. Step 5: Text Vector Clustering

This step uses a Fuzzy SOM algorithm to cluster the feature vectors based on similarities of research areas.

4.3 Research Proposals Balancing and Regrouping Them by Considering Applicant's Characteristics

Even though the number of proposals in particular cluster is large the candidate's characteristics like affiliated universities are considered. As in Fan et al. [xvii], the proposal group composition should be varied. Previous years the reviewers sometimes knob the proposals improperly, having poor group work like the same affiliation in a specific proposal group. Reviewers may feel confused and sore when evaluating proposals that may have poor group composition, so it is wise that the applicants' characteristics in each proposal group should be as varied as possible. Besides, the group size in each group should be similar. This is done by Non-dominated Ranking Genetic Algorithm (NRGA) algorithm as follows.

At first, a random parent population P is formed. The sorting of the population is based on the non-domination. Each solution is allocated as fitness equivalent to its non-domination level.

The normal reproduction operator is the comparable reproduction operator. For its fitness a string with probability proportional is chosen for mating pool. In the population the i^th string which is probability proportional chosen to Fitness value 〖F〗_v . in a simple GA the size of the population is always kept fixed. The string chosen for the mating pool should satisfy the following condition

sum(probabilty of each string))=1 (6)

Therefore the possible way for selecting the i^th string is

$$\rho(i)=F_v/\left(\sum_{i=1}^{n} F_v\right) \quad \text{Where n population size} \qquad (7)$$

In the proposed work, this selection procedure is implemented by imagine a Roulette- Wheel (RW) in which its circumference marked for each string proportionate to the string's fitness. The RW is spun number of times. Every time the RW pointer chose an instance of a string. The RW makes $F_v/\bar{A}$ copies of the strings. The formula for the average fitness is

$$\bar{A}=\sum_{i=1}^{n} F_v \qquad (8)$$

The algorithm represents that Non-Dominated Ranked Genetic algorithm is simple and straightforward. A combined population P∪Q is created. The mixed population is of size 2N then obtained; according to the non-domination the mixed population is sorted. Each and every population members are incorporated in the mixed population elitism is guaranteed. Out of 2N the solution N will be chosen by this process.

The new population of size 'N' is utilized for choosing. next two tiers, ranked dependent on roulette wheel selection [xxviii] and [xxix]; one tier to choose the front and the other to choose solution from the front, the highest probability results obtained for the finest Non-Dominated set F1 to be chosen. Therefore, results from the set F2 are selected with small probability than results from the set F1 and so on. The next procedure is generation of population RP of size N. this will be done by crossover and mutation. The diversity between non-dominated results is established by the second tier of ranked dependent Roulette wheel selection that ranks the results according to their crowding distance. The results with lesser crowding distance will have the higher probabilities.

As solutions contend with their crowding distance, no extra niching attribute is needed. The crowding distance is obtained in the parameter space Even though it is computed in the objective function space, The objective function space niching is utilized in this proposed approach [xxx].

Pseudo code for NRGA Algorithm:

```
Initialize Population P
 {
  Generate random populations of  P– size n         Evaluate population
objective values J based on 1- NN for P
Assign rank (level) for random Populations of P based on pareto dominance
sort
   }
 {
          Ranked based roulette wheel selection                Recombination
          and mutation
}
                                        Q ∈ F_i
for i=1 to g do

          for each member of the combined population     (PUQ) do
        Assign rank (level) based on Pareto-sort
        Generate sets of non-dominated fronts
          Calculate the crowding distance between members of each front

        end for

(elitist) Select the members of the combined population based on least
dominated n solution t_i make the population of the next generation. Ties are
resolved by taking the less crowding distance
Create next generation
{
     Ranked based Roulette wheel selection
     Recombination Mutation
}
 end for
```

6. Experimental Results

To validate the proposed approach, several experiments are conducted using the previous granted research projects. Research projects from various disciplines are considered for the experimental evaluations. Artificial intelligence, information management, image processing, data mining, networking and software engineering are the domains and disciplines taken into consideration. Moreover, the typical criterion for text clustering F measurement is used to measure the quality of clustering research projects. As mentioned in [31], for generated cluster c and predefined research topic t, the corresponding Recall and Precision can be calculated as follows:

$$Precision(c,t) = \frac{n(c,t)}{nc} \qquad (8)$$

$$Recall(c,t) = \frac{n(c,t)}{nt} \qquad (9)$$

Where n(c, t) is the project number of the intersection between cluster c and topic t. $n_c$is the number of projects in cluster c, and n_t is the number of projects in topic t. F measurement between cluster c and topic t can be calculated as follows:

$$F(c,t) = \frac{(2 * \mathrm{Recall}(c,t) * \mathrm{Precision}(c,t))}{(\mathrm{Recall}(c,t) + \mathrm{Precision}(c,t))} \qquad (10)$$

The F measurement can be given by

$$F = \sum_i \frac{n_i}{n} \max\{F(i,j)\} \qquad (11)$$

where n is the whole number of research projects and i is each predefined research topic.

6.1. Performance Evaluation

In order to compare the clustering quality of the proposed Fuzzy SOM ontology based Text Mining Method (FSOM based TMM), the existing Ontology based TMM and clustering based TMM is taken for consideration.

6.2. F Measurement Evaluation

Figure 4. shows the F-measurement comparison of the proposed approach with the existing approaches. It is clearly observed from the figure that the proposed approach outperforms the other two approaches. The F measurement of the proposed FSOM based TMM is better than the other two approaches.

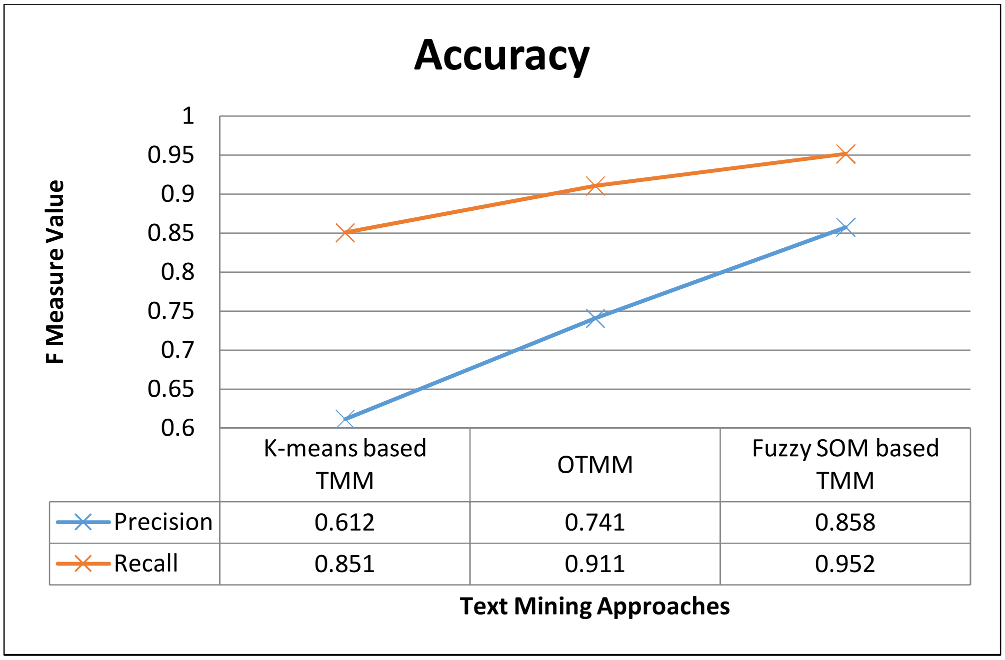

Figure 3: Accuracy of algorithm.
Precision and Recall Comparison

Table 1. shows the comparison of the precision and recall comparison of the proposed and the existing approaches.
It is observed from the table that the precision and recall attained by the proposed Fuzzy SOM based TMM is higher when compared with the existing K-means based TMM and OTMM approaches.
From the above mentioned experimental results, the performance of the proposed approach is evaluated using F measurement. The classification accuracy of the proposed performance is 0.94 than existing approach such as k-means based TMM and OTMM are 0.92 and 0.87. Hence the proposed approach performs better than the existing approaches.

7. Conclusion

This paper has presented a structure on ontology based text mining for grouping research proposals and conveying the grouped proposal to reviewers analytically. Research ontology is designed to separate the concept tasks in various regulations areas and to form a concurrent relationship with them. It assists with text-mining and optimization techniques to cluster research proposals based on their resemblance and then based on the research area it is allocated to reviewer. With the help of knowledge based agent the proposals are assigned to the reviewer. In future this reviewer work is replaced by the system. Also, there is a need to empirically compare the results of manual classification to text-mining classification.

8. Conflict of Interests

Here we declare that there is no conflict of interests regarding the publication of this paper.

9. References

i. G. Salton. Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer. Addison-Wesley, (1989).
ii. Fabrizio Sebastiani. Machine learning in automated text categorization. ACM Computing Surveys, 34(1):1–47, (2002).
iii. D J Foskett. Thesaurus. In P. Willett and K Sparck-Jones, editors, Reproduced in Readings in Information Retrieval, pages 111–134. Morgan Kaufmann, (1997).
iv. S. Staab and R. Studer, editors. Handbook on Ontologies. Springer, (2004).
v. Maedche and S. Staab, “The Text-To-Onto ontology learning environment,”in Proc. 8th Int.Conf. Conceptual Struct., Darmstadt, Germany, pp. 14–18, (2000).
vi. E. Turban, D. Zhou, and J. Ma, “A group decision support approach to evaluating journals,” Inf. Manage., vol. 42, no. 1, pp. 31–44, Dec. (2004).
vii. C. Choi and Y. Park, “R&D proposal screening system based on text mining approach,” Int. J.Technol. Intell. Plan., vol. 2, no. 1, pp. 61–72,(2006).

viii. D. Roussinov and H. Chen, "Document clustering for electronic meetings:An experimental comparison of two techniques," Decis. Support Syst.,vol. 27, no. 1/2, pp. 67–79, Nov. (1999).
ix. C. Wei, C. S. Yang, H. W. Hsiao, and T. H. Cheng, "Combining preference- and content based approaches for improving document clustering effectiveness," Inf. process. Manage.,vol. 42, no. 2, pp. 350–372,Mar. (2006).
x. T. A. Runkler and J. C. Bezdek, "Web mining with relational clustering,"Int. J. Approx. Reason., vol. 32, no. 2/3, pp. 217–236, Feb. (2003)
xi. H. Li, K. Zhang, and T. Jiang, "Minimum entropy clustering and applications to gene expression analysis," in Proc. 3rd IEEE Comput. Syst. Bioinform. Conf., Stanford, CA, pp. 142–151 (2004).
xii. S. Gauch, J. Chaffee, and A. Pretschner, "Ontology-based personalized search and browsing," Web Intell. Agent Syst., vol. 1, no. 3/4,pp. 219– 234, Dec. (2003).
xiii. 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).
xiv. Hossein Shahsavand Baghdadi and Bali Ranaivo-Malançon ,"An Automatic Topic Identification Algorithm," Journal of Computer Science 7 (9): 1363-1367, ISSN 1549-3636, (2011)
xv. S. Hettich and M. Pazzani, "Mining for proposal reviewers: Lessons learned at the National Science Foundation," in Proc. 12th Int. Conf. Knowl. Discov. Data Mining , pp. 862–871, (2006)
xvi. 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).
xvii. Arya and B. Mittendorf, "Project assignment when budget padding taints resource allocation," Manage. Sci., vol. 52, no. 9, pp. 1345–1358, Sep. (2006).
xviii. Choi and Y. Park, "R&D proposal screening system based on text mining approach," Int. J. Technol. Intell. Plan., vol. 2, no. 1, pp. 61–72, (2006).
xix. K. Girotra, C. Terwiesch, and K. T. Ulrich, "Valuing R&D projects in a portfolio: Evidence from the pharmaceutical industry," Manage. Sci., vol. 53, no. 9, pp. 1452–1466, Sep. (2007).
xx. Y. H. Sun, J. Ma, Z. P. Fan, and J. Wang, "A group decision support approach to evaluate experts for R&D project selection," IEEE Trans. Eng. Manag., vol. 55, no. 1, pp. 158–170, Feb. (2008).
xxi. Y. H. Sun, J. Ma, Z. P. Fan, and J. Wang, "A hybrid knowledge and model approach for reviewer assignment," Expert Syst. Appl., vol. 34, no. 2, pp. 817–824, Feb. (2008).
xxii. S. Bechhofer et al., OWL Web Ontology Language Reference, W3C recommendation, vol.10, p. 2006-01, (2004).
xxiii. B. Yildiz and S.Miksch, "ontoX—A method for ontology-driven information extraction," in Proc.ICCSA (3), vol. 4707, Lecture Notes in Computer Science, O. Gervasi andM. L. Gavrilova, Eds., , pp. 660–673, Berlin,Germany: Springer-Verlag. (2007).
xxiv. Jian Ma, Wei Xu, Yong-hong Sun, Efraim Turban, Shouyang Wang"An Ontology-Based Text- Mining Method to Cluster Proposals for Research Project Selection",IEEE Trans an systems and humans vol.42,no.3 May (2012)
xxv. Q. Tian, J. Ma, and O. Liu, "A hybrid knowledge and model system for R&D project selection," Expert Syst. Appl., vol. 23, no. 3, pp. 265–271, Oct. (2002).
xxvi. R. Feldman and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge Univ.Press, (2007).
xxvii. D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning. Redwood City: Addison-Wesley, (1989).
xxviii. S. Wu and T.W.S. Chow, PRSOM: A new visualization method by hybridizing multidimensional scaling and self-organizing map, IEEE Transactions on Neural Networks, 16(6), 1362-1380, (2005).
xxix. L. Xiao, K. Wang, Y. Teng and J. Zhang, Component plane presentation integrated self-organizing map for microarray data analysis FEBS Letter, 538(1) , 117-124. (2003).
xxx. P. Toronen, M. Kolehmainen, G. Wong and E. Castren, Analysis of gene expression data using self-organizing maps, FEBS Letters, 451(2), 142-146. (1999).