Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 9, September 2015, pg.154 – 165

RESEARCH ARTICLE

OTMM for Search Proposal Classification

Sunil Datir[1], Arpit Solanki[2]

[1] M. Tech Student

RKDF SOE, Indore Computer Science and Engineering Department & RGPV University, India

[1] sunil.datir@gmail.com ; [2] Er.arpitsolanki@gmail.com

***Abstract*— Research Paper Selection is important decision preparing task for the Government funding Agency, research Institutes. Ontology is Knowledge Repository in which concepts and terms defined as well as relationship between these concepts. In this paper Ontology is old research papers repository of keywords and frequencies of that keywords of the research papers of funding agencies. Ontology makes the tasks of searching similar patterns of text that is to be more effective, efficient and interactive. The current system of grouping of papers for research paper selection based on similarities of Keywords and Frequencies of research papers of ontology. Text mining is the extraction of useful, often previously unknown information from large document. The Research Papers in each domain are clustered using Text mining Technique. Grouped Research papers are allocated to appropriate reviewer or domain experts for peer review systematically. The Reviewer results are collected and papers are getting graded based on experts review results.**

***Keywords*— Clustering analysis, decision support systems, ontology, research project selection, text mining**

I. INTRODUCTION

In many organizations such as funding agencies, institution research project selection is an important and recurring activity. It is many process tasks that starts with a call for proposals (CFP) by a funding agency.it is very challenging process tasks. The CFP is passing out to relevant communities such as universities or research institutions. The research proposals are submitted to the institution, funding agency and then are assigned to experts for peer review. The review results are gathered and the proposals are then ranked based on the aggregation of the experts review results.

In the National Natural Science Foundation of China (NSFC), first the proposals are submitted. The next important activity is to group proposals and allocate them to reviewers. The properties of proposal in each group should have similar. For occurrence, if the proposals in a group fall into the same primary research discipline (e.g., supply chain management) and if 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 [18][19]. Thus, ontology can automate information processing and can facilitate text mining in a specific domain (such as research project selection).

II. METHODOLOGY

The processes of research project selection at the National Natural Science Foundation of China (NSFC) i.e. CFP, as show in Fig 1, proposal submission, proposal grouping, proposal assignment to experts, peer review, aggregation of review

results, panel evaluation, and final awarding decision. These processes are very similar in other funding agencies, except that there are a very large number of proposals that need to be grouped for peer review in the NSFC. In the NSFC, the number of research proposals received has more than doubled in the past four years, with over 110,000 proposals submitted in one deadline in March 2010.

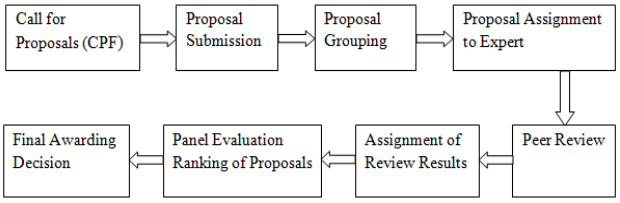


Fig 1: Research Project Selection Process in NSFC[1]

Founded in 1986, the NSFC is the largest government funding agency in China, with the primary aim to fund and manage basic research. The agency is made up of seven scientific departments, four bureaus, one general office, and three associated units. The scientific departments are the decision-making units responsible for funding recommendations and management of funded projects. Departments are classified according to scientific research areas, including mathematical and physical sciences, chemical sciences, life sciences, earth sciences, engineering and material sciences, information sciences, and management sciences. These departments are further divided into 40divisions with a focus on more specific research areas. For example, the Department of Management Science is further divided into three divisions: Management Science and Engineering, Macro Management and Policy, and Business Administration. There was an urgent need for an effective and feasible approach to group the submitted research proposals with computer supports. An ontology-based text-mining approach is proposed to solve the problem.

1. First, a research ontology containing the projects funded in latest five years is constructed according to keywords, and it is updated annually.
2. Then, new research proposals are classified according to discipline areas using a sorting algorithm.
3. Next, with reference to the ontology, the new proposals in each discipline are clustered using a self-organized mapping (SOM) algorithm.
4. If the number of proposals in each cluster is still very large, they will be further decomposed into subgroups where the applicants' characteristics are taken into consideration (e.g. Applicants' affiliations in each proposal group should be diverse). Here we may use of GA (Genetic Algorithm).
5. Finally, the Research project will be assign to expert review.

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.

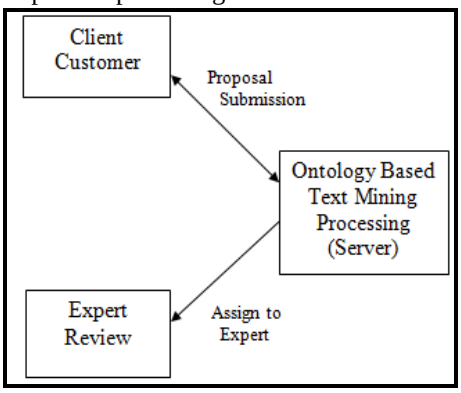


Fig 2. Client - Server Model

The OTMM is used together with statistical method and optimization models and consists of four phases, as shown in Fig.3 First, a research ontology containing the projects sponsored in latest five years is established according to keywords, and it is updated year after year (phase 1). Then, using sorting algorithm the new research proposals are classified according to discipline areas (phase2). Next, with address to the ontology, using self-organized mapping (SOM) algorithm the new proposals in each discipline are clustered (phase 3). Finally, (phase 4) if the number of proposals in each cluster is still very large, they will be further decomposed into subgroups where the applicants' characteristics are taken into consideration (e.g. Applicants' affiliations in each proposal group should be diverse).

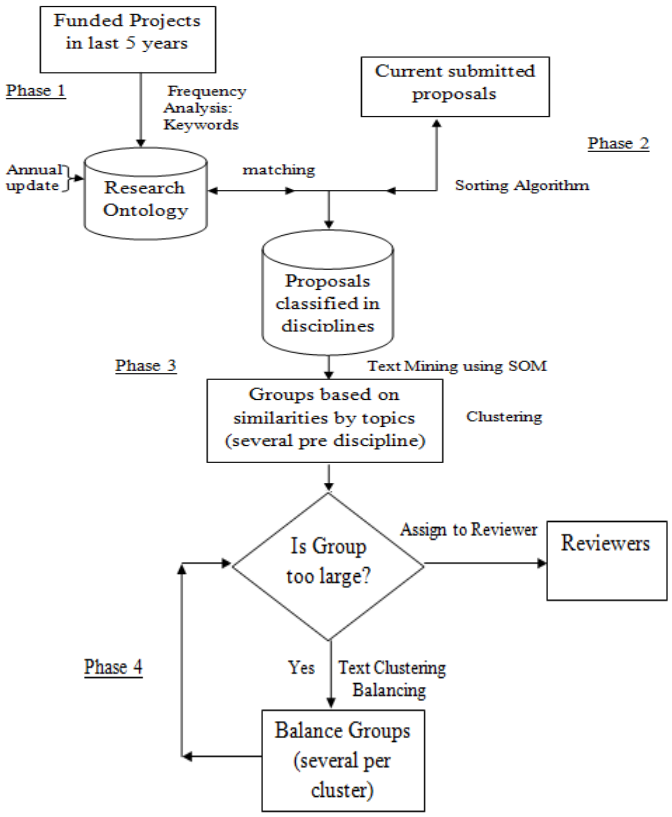


Fig 3. Process of OTMM[1]

III. PROPOSED WORK

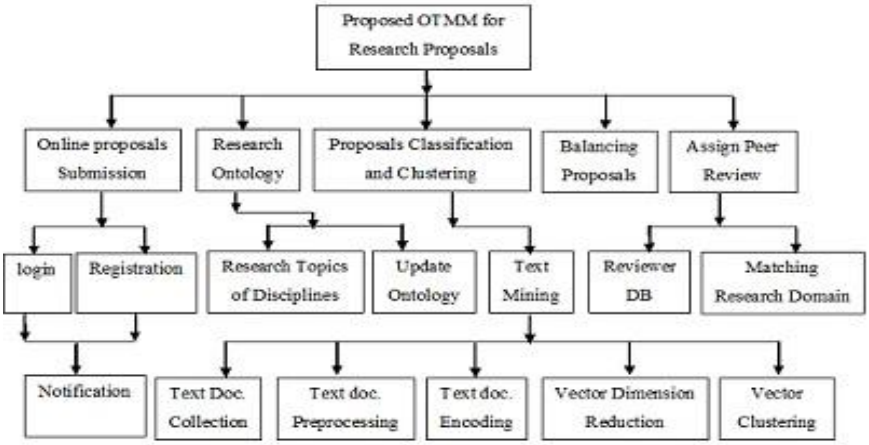


Fig. 4. Project Work Breakdown Structure (Implementation)

A. ONLINE PROPOSAL SUBMISSION

As mentioned in client-server model, role of the client is to submit research proposal. It comes under first module. Client will be going to establish the http connection with server. All the further processing will take place at server side. Implementation of server is discussed in further modules. Client will receive notification about their proposal submission.

B. EMAIL NOTIFICATION

Once the proposal is submitted to the system the students will get the email notification. The email notification will include the domain of the paper as per SOM algorithm, the name of guide to which paper is forwarded. This Email notification will work through SMTP protocol which will automatically direct the email from user to guide and vice versa. With the help of this students can keep the track of their research paper.

- **What is SMTP?**

SMTP is one of the parts of the application layer of the TCP/IP protocol. It works like "store and forward," SMTP send and receive your email on and over the networks. It works similar something called the Mail Transfer Agent (MTA) to send your communication to the specific computer and inbox of email. SMTP explain step-by-step and directs how your email transferred between more than one computer MTA. Using that "store and forward" feature described before, the message travel in number steps from your computer to its destination computer. Simple Mail Transfer Protocol is doing its work at each step.

C. CONSTRUCTING A RESEARCH ONTOLOGY

Funding agencies such as the NSFC maintain a directory of discipline areas that form a tree structure. As a domain ontology, a research ontology is a public concept set of the research project management domain. Let K is discipline areas, and Ak represents discipline area k(k = 1, 2, . . . , K).

Research ontology can be constructed in the following three steps to represent the topics of the disciplines. The example model of structure of research ontology:

Creating the research topics of the discipline Ak, (k = 1, 2, . . . , K)

The keywords of the supported research projects each year are collected, and their frequencies are counted. The keywords as well as their frequencies are represented by the feature set (Nok, IDk,year, {(keyword1, frequency1), (keyword2,frequency2),. . . , (keyword, frequency)}), where Nokis the sequence number of the kth record and IDkis the corresponding discipline code. For instance, if discipline Ak has two keywords in 2007 (i.e., "data mining" and "business intelligence") and the total number of counts for them are 30 and 50, respectively, the discipline can be denoted by (Nok, IDk, 2007, {(data mining, 30), (business intelligence, 50)}).

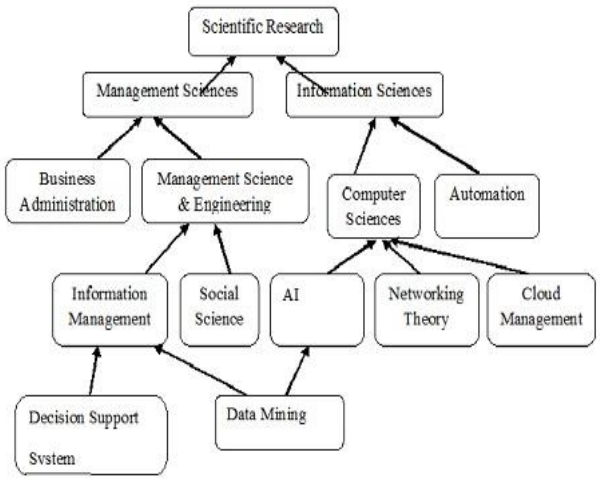


Fig.5 Structure Of Research Ontology

In this way, a feature set of each discipline can be generated. The keyword frequency in the feature set is the addition of the same keywords that occurred in this discipline during the most recent five years, and then, the feature set of Ak is denoted by (Nok, IDk, {(keyword1, frequency1)(keyword2, frequency2), . . . ,(keyword, frequency)}).

First, the research ontology is categorized according to scientific research areas introduced in the background. It is then developed on the basis of several specific research areas. Next, it is further divide into some arrowed discipline areas. Finally, it leads to research topics in terms of the feature set of disciplines created in. It is more complex than just a tree-like structure. There are some cross-discipline research areas (e.g., "data mining" can be placed under "Information management" in "Management Sciences and Engineering" or under "Artificial Intelligence" in "Information Sciences"). Therefore, the research ontology allows more complex relationship between concepts besides the basic tree-like structure. Also, to deal with proposals in English text, there are some synonyms used by different project applicants, which have different names in different proposals but represent the same concepts. Therefore, the research ontology allows more complex relationship between concepts besides the basic tree-like structure.

- **Updating the research ontology**

Once the project funding is completed each year, the research ontology is updated according to agency's policy and the change of the feature set.

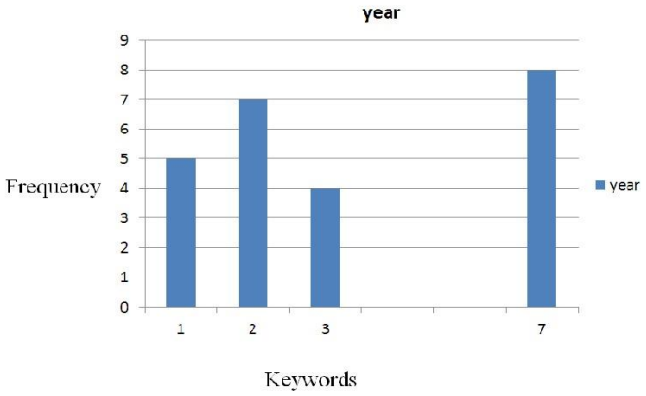


Fig.6 Keywords in the year

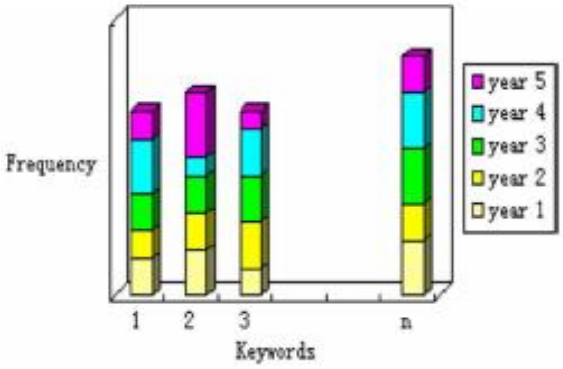


Fig.7 Feature set of keywords

D. CLASSIFYING NEW RESEARCH PROPOSALS INTO DISCIPLINES AREAS

Classifications of Proposals are based on discipline areas. A sorting algorithm is used for classification of proposals. This is done by reference of the research ontology as follows. Let K is discipline areas, and Ak represents area k(k = 1, 2, . . . , K). Pi represents proposals i (i = 1, 2, . . . , I), and Sk represents the set of proposals which based on area k. A sorting algorithm can be performing to classify proposals to their discipline areas.

E. CLUSTERING RESEARCH PROPOSALS THROUGH SIMILARITIES USING TEXT MINING

After the research programmed are categories by the discipline areas, the proposals in each discipline are clustered using the TMM technique. The main clustering process following five steps: Text document collection, Text document pre-processing, Text document encoding, vector dimension reduction, and Text vector clustering. The descriptions of each step are as follows.

- **TEXT DOCUMENT COLLECTION**

After the research programmed are categories to states as the discipline areas, the proposal documents in each discipline Ak(k = 1, 2, . . . , K) are gathered for text document pre-processing.

- **TEXT DOCUMENT PRE-PROCESSING**

The contents of proposals are usually unstructured. Because the texts of the proposals consist of Chinese characters which are difficult to segment, the research ontology is used to analyze, extract, and identify the keywords in the full text of the proposals. For example, "Research on behavior modeling and detection methods in financial fraud using ensemble learning" can be divided into word sets {"behavior modeling," "detection method," "financial fraud," "ensemble learning"}. Finally, a further next reduction in the vocabulary size can be reached through the removal of all words that occurred only a less times (say less than 5 times) in the proposal documents.

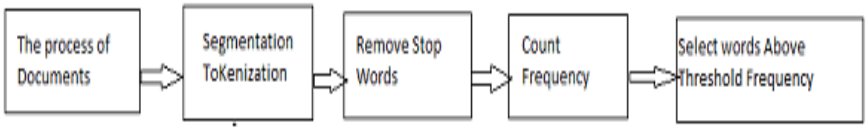


Fig. 8 Pre-processing steps of text document for text clustering**.**

- **AFTER TEXT DOCUMENTS ARE TEXT DOCUMENT ENCODING**

Segmented, they are transferred into a feature vector denoted by: V = (v1, v2, . . . , vM), where M is the number of features selected and vi(i = 1, 2, . . . , M) is the encoding of the keyword wi. The feature v, to the extent that vi= tfi∗ log(N/dfi), where N is the total number of proposals in the discipline, tfi is the term frequency of the feature word wi, and dfi is the number of proposals carry the word wi. Thus, research proposals can be denoted by corresponding feature vectors.

- **VECTOR DIMENSION REDUCTION**

Feature vectors dimension is often too large; thus, its required reduction in the vectors' size by automatically selecting a subset carry the most important keywords in frequency terms. To solve the problem using Latent Semantic Indexing (LSI). It reduces the dimensions of the feature vectors effectively as well as creates the semantic relations among the keywords. For the deduction of the dimensions of the document vectors with correct place the needful information in a proposal, a term-by-document matrix is produced, where there is one column that correlates to the term frequency of a document. Additionally, the term-by document matrix is break into a set of eigen-vectors using angular-value decomposition. The eigenvectors that has the minimum effect on the matrix are then discarded.

- **TEXT VECTOR CLUSTERING**

This step very important that uses an SOM algorithm to cluster the feature vectors based on correspondence of research areas. This algorithm is a typical unsupervised learning NN Model that clusters given data with similarities. Detail of SOM algorithm [16][17] .

- **FEASIBILITY ASSESSMENT OF SOM**

All problems are coming under NP category. SOM problem comes under NP-Complete area. SOM is the unsupervised learning neural network. Hence, we have to take decision about winning neuron and such kind of decision problems falls under NP-Complete.

In unsupervised learning, training of network is entirely data driven and no target result for input data vectors is provided. Input data vector may fall under any of the output unit which is non-deterministic. SOM provides a topology preserving mapping from high dimensional space to map units. Map units or neurons, usually forms two dimensional (2D) lattice and thus mapping is mapping from high dimensional space onto plane.

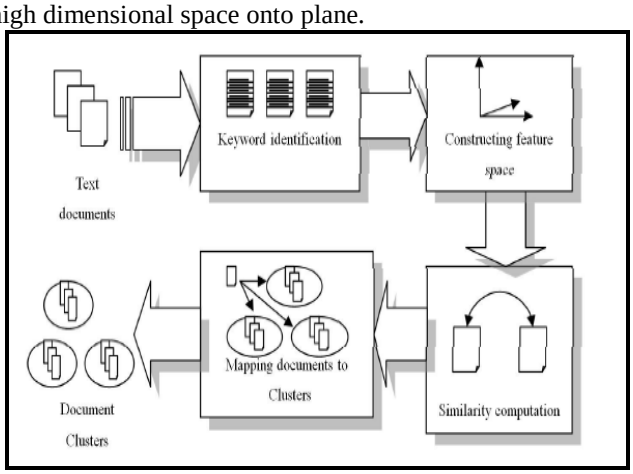


Fig. 9 Text clustering system

- **Input:**
  Training data: vectors, X
  – Vectors of length *n*
  (x1,1,x1,2, ..., x1,i,…, x1,n)
  (x2,1, x2,2, ..., x2,i,…, x2,n)
  …
  (xj,1, xj,2, ..., xj,i,…, xj,n)0
  …
  (xp,1, xp,2, ..., xp,i,…, xp,n)
  – Vector components are real numbers
- **Outputs**
  – A vector, Y, of length *m*: (y1, y2, ...,yi,…, ym)Sometimes m < n, sometimes m > n, sometimes m = n
  – Each of the *p* vectors in the training data is classified as falling inone of *m*clusters or categories
  – That is: Which category does the training vector fall into?
- **Generalization**
  – For a new vector: (xj,1, xj,2, ..., xj,i,…, xj,n)
  – Which of the *m* categories (clusters) does it fall into?

- **Network Architecture**
  Two layers of units:
  – Input: *n* units (length of training vectors)
  – Output: *m* units (number of categories)
  Input units fully connected with weights to output units.

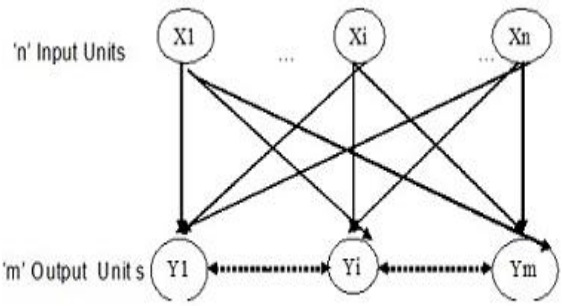


Fig. 10 Network Architecture

- **SOM ALGORITHM USING FOR OTMM.**

1. Select output layer network topology.
   1.1 Initialize current neighbourhood distance, D(0), to a positive value
2. Initialize weights from inputs to outputs to small random values
3. Let t = 1
4. While computational bounds are not exceeded do
   4.1 Select an input sample
   4.2 Compute the square of the Euclidean distance offrom weight vectors (wj) associated with each output node.
   $\sum^{n}_{k=1}(i_{l,k}- w_{j,k}(t))^2$
   4.3 Select output node *j** that has weight vector with minimum value from step
   4.4 Update weights to all nodes within a topological distance given by D(t)from *j**, using the weight update rule:
   $w_j(t+1) = w_j(t)+\eta\ (t)(i_l- w_j(t))$
   4.5 Increment t
5. End while.

Learning rate generally decreases with time:

$0 < \eta\ (t) \leq \eta\ (t-1) \leq 1$

F. BALANCING RESEARCH PROPOSALS AND REGROUPING THEM BY CONSIDERING APPLICANT'S CHARACTERISTICS

In this phase, when the number of proposals in one cluster is still very large (e.g., more than 20), the applicants' characteristics (e.g. affiliated universities) are considered. The proposal group composition should be diverse. Reviewers may feel confused and uncomfortable when evaluating proposals that may have poor group composition, so it is advisable that the applicant's characteristics in each proposal group should be as diverse as much as possible. Furthermore, the group size in each group should be similar. This may be very complex optimization problem and one solution method that could be use is Genetic Algorithm [16].

- **Feasibility of GA**

GA is used for optimization of clusters and optimization problems generally come under NP-Hard category. NP-Hard problems are more complex and more than simple polynomial.
GA is based on mechanics of biological evolution. GA provides solution for high complex search space.

GA Operators:

Population	set of solutions
Fitness	quality of solution
Chromosome	encoding for solution
Gene	part of encoding solution
Reproduction	crossover

Offspring	parent's child
Mutation	change is genetic structure results in variant form

- **Genetic algorithm use for OTMM**

Input: Fitness function f(), maximum number of iteration max_tier
Output: best found solution
begin
Generate at random initial population of solution;
i:=0;
while i<= max_tier and stop_cond.= false
do
begin
– evaluate each solution with f();
– apply crossover on selected solution;
– mutate some of the new obtained solutions
– add new solution to population;
– remove less adopted solutions according to f() from population;
–i:= i+1;
end;
– return best found solution;
end;

- **Encoding**

The process of representing a solution in the form of a string that conveys the necessary information is encoding. Each bit in the string represents a characteristic of the solution. Most common method of encoding is binary coded. Chromosomes are strings of 1 and 0 and each position in the chromosome represents a particular characteristic of the problem.

- **Fitness Function**

A fitness function value quantifies the optimality of a solution. The value is used to rank a particular solution against all the other solutions. A fitness value is assigned to each solution depending on how close it is actually to the optimal solution of the problem.
Ƒ(d,h)=c((ɪd2/2)+ɪdh) … Fitness equation

- **Crossover**

The crossover operator is used to create new solutions from the existing solutions. This operator exchanges the gene information between the solutions in the mating pool. The most popular crossover selects any two solutions strings randomly from the mating pool and some portion of the strings is exchanged between the strings. The selection point is selected randomly.

•Simple Crossover

It is similar to binary crossover.

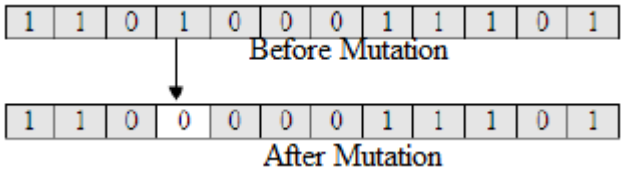
P1 = [8 6 3 7 6] C1 = [8 6 3 7 6]
P2 = [2 9 4 8 9] C2 = [2 9 4 8 9]

•Linear Crossover

Parents: (x1,…,xn) and (y1,…,yn)
Select a single gene (k) at random
Three children are created as,
(x1, ...,xk , 0.5*yk+0.5*xk , ...,xn)
(x1, ...,xk , 1.5*yk-0.5*xk , ...,xn)
(x1, ...,xk ,- 0.5*yk+1.5*xk , ...,xn)
From the three children, best two are selected for the next generation.

- **Mutation**

Mutation is the occasional introduction of new features in to the solution strings of the population pool to maintain diversity in the population. Though crossover has the main responsibility to search for the optimal solution, mutation is also used for this purpose. Mutation operator changes a 1 to 0 or vice versa.



G. ASSIGN TO EXPERT

Finally, the proposal is assigned for expert review. This decision making task will be done by server itself. Server may notify to particular expert about their assignment for proposal in the form of mail or message. This is the atomization done in assignment of proposal for expert review.

IV. RESULTS

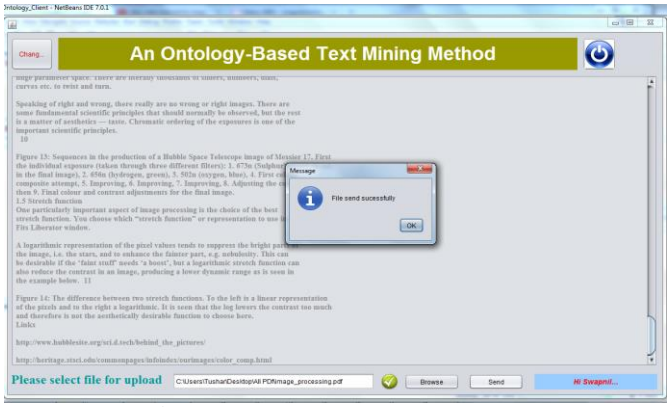


Fig.11Snapshot of scanning the PDF and sending it to domain wise guide

The above snapshot for scanning the PDF .In it the client will browse the pdf file and then it will be scan.

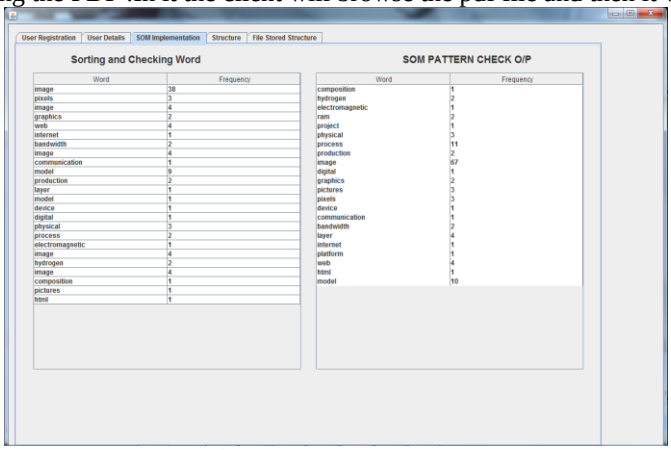


Fig 12 Snapshot of showing sorting and SOM clustering results.

The above snapshot to showing the output result of Sorting and Clustering using SOM.
The following table represents result Sorting and SOM Clustering:

A. Sorting:

We extract all words from PDF by filtering all special character, rare words and most common words. Sorting done by on the basis of Frequency. Frequency is the number of times the word occurs in PDF. With Sorting, we get most semantic words (words that actually carry meaning) from PDF.

Example:-

Word	Frequency
Network	230
Bandwidth	12
Bits	10
Wave length	2

This sorted list become input to SOM.

B. SOM Clustering :

The self-organizing map (SOM) is especially suitable for data survey because it has prominent visualization properties. Clustering are hierarchical and partitioning approaches. The clustering algorithms usually have the following Steps.

1) Initialize: Assign each vector to its own cluster.
2) Compute distances between all clusters.
3) Merge the two clusters that are closest to each other.
4) Return to step 2 until there is only one cluster left.

Table.no.1 sorting and SOM clustering result.

Word	frequency TMM(sorting)	Frequency Modify OTMM(SOM)
Composition	1	1
Hydrogen	2	2
Electromagnetic	1	1
Ram	0	2
Project	0	1
Physical	3	3
Process	2	11
Production	2	2
Image	50	67
Digital	1	1
Graphics	2	2
Pictures	1	3
Pixels	3	3
Device	1	1
Communication	1	1
Bandwidth	2	2
Layer	1	4
Internet	1	1
Platform	0	1
Web	4	4
Html	1	1
Model	10	10

The above table it is result TMM (sorting) and modify OTMM(SOM).

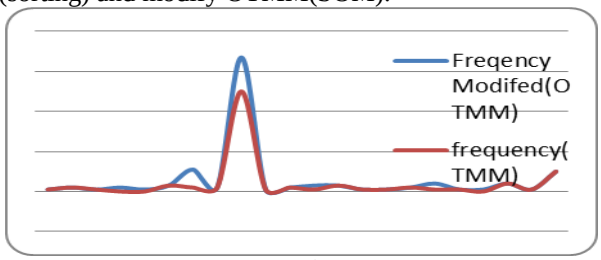


Fig 13 Graph represents result of Sorting and SOM Clustering

Form above graph it is clarify that modify OTMM improvement results better than the results in TMM.

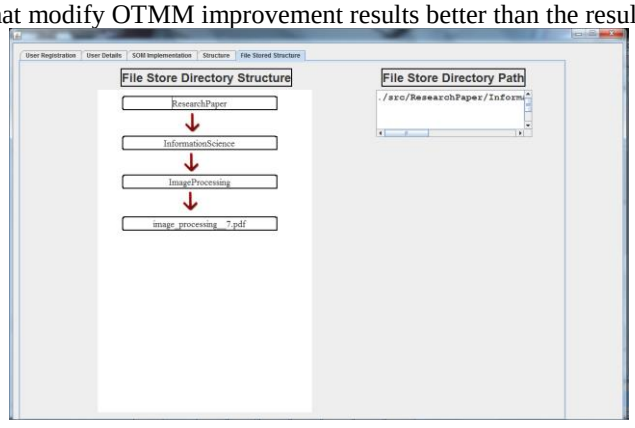


Fig 14 File stored structure form.

The above snapshot to showing the domain wise file store structure and showing the file store Directory path.

V. CONCLUSION

This Concept has presented an OTMM for grouping of research proposals. Research ontology is built to classify the concept terms in various discipline areas and to create association among them. It facilitates TMM and optimization procedures to cluster research proposals depend on their similarities and then to balance them according to the applicants' attributes. The experimental results at the NSFC showed that the proposed method upgrade the equality in group proposal, as well as analysis the applicants 'attributes. Also, the proposed system promotes the efficiency in the proposal grouping process. The proposed method can also be used in other government research funding agencies that face information overload problems. Future work is needed to cluster external reviewers based on their research areas and to assign grouped research proposals to reviewers systematically. Also, there is a need to empirically compare the results of manual classification to text-mining classification. Finally, the method can be expanded to help in finding a better match between proposals and reviewers.

VI. REFERENCES

[1] 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 Transaction On System,man, and cybernetics May,2012.

[2] K. Chen and N. Gorla, "Information system project selection using fuzzylogic," IEEE. Syst., Man, Cybern. A, Syst., Humans, vol. 28, no. 6,pp. 849–855, Nov. 1998.

[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]Muller, H.M., Kenny, E.E., Sternberg, P.W.: Textpresso: An Ontology-Based Information Retrieval and Extraction System for Biological Literature. PLoS Biol. 2(11):e309. doi:10.1371/journal.pbio.0020309 (2004) F

[7]Young, L., Tu, S.W., Tennakoon, L., Vismer, D., Astakhov, V., Gupta, A., Grethe, J.S., Martone, M.E., Das, A.K., McAuliffe, M.J.: Ontology Driven Data Integration for Autism Research. 22nd IEEE International Symposium on Computer Based Medical Systems, pp. 1–7, Albuquerque, NM (2009)

[8]Tu, S.W., Tennakoon, L., Das, A.K.: Using an Integrated Ontology and Information Model for Querying and Reasoning about Phenotypes: The Case of Autism. AMIA Annual Symposium, pp. 727–731, Washington, DC (2008)

[9]Hus, V., Pickles, A., Cook, E.H., Risi, S., Lord, C.: Using the Autism Diagnostic Interview-Revised to Increase Phenotypic Homogeneity in Genetic Studies of Autism. Biol Psychiatry. 61(4), 438–448 (2007)

[10]McGuinness, D.L., van Harmelen, F.: OWL Web Ontology Language Overview. W3C Recommendation, (2004)

[11]Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A Semantic Web Rule Language Combining OWL and RuleML. (2004)

[12] Open Directory Project, http://www.dmoz.org/ Google support on snippets, http://www.google.com/support/ webmasters/bin/answer.py?hl=en&answer=35624

[13]Cooper, W.S.: Fact Retrieval and Deductive Question-Answering Information Retrieval Systems. J. ACM. 11(2), 117–137 (1964)

[14]Miliaraki, S., Androutsopoulos, I.: Learning to Identify Single-snippet Answers to Definition Questions. 20th International Conference on Computational Linguistics, Geneva, Switzerland (2004)

[15]D.E Goldberge ,Genetic Algorithms in search , Optimization and machine Learning. Redwood city :Addision – Wesely,1989

[16]F.M Ham and I,kostanic , Principle of Neurocomputing for Science and Engineering .New york :McGraw-Hill,2001.

[17]J. vesanto and E.Alhoniemi ,"Clustering of the self -organizing map" IEEE Trans. Neural network ,vol. 11,no.3,pp.586-600,May 2000.

[18]L.Razamerita , "An ontology –based framework for modeling user behavior –A case study in knowledge management "IEEE Trans. Syst. Man Cybern, A, Humans ,vol 41,no. 4,pp. 772-783,jul,2011.

[19]Q. Liang, X. Wu. E. K. park ,T.M. Khoshgoftaar ,and C.H Cho,"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

[20] A. D. Henriksen and A. J. Traynor, "A practical R&D project-selection scoring tool," IEEE Trans. Eng. Manag., vol. 46, no. 2, pp. 158–170,May 1999.

[21] F. Ghasemzadeh and N. P. Archer, "Project portfolio selection through decision support," Decis. Support Syst., vol. 29, no. 1, pp. 73–88, Jul. 2000.

[22] L. L. Machacha and P. Bhattacharya, "A fuzzy-logic-based approach toproject selection," IEEE Trans. Eng. Manag., vol. 47, no. 1, pp. 65–73,Feb. 2000.

[23] J. Butler, D. J. Morrice, and P. W. Mullarkey, "A multiple attribute utility theory approach to ranking and selection," Manage. Sci., vol. 47, no. 6,pp. 800–816, Jun. 2001.
[24] 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.
[25] L. M. Meade and A. Presley, "R&D project selection using the analyticnetwork process," IEEE Trans. Eng. Manag., vol. 49, no. 1, pp. 59–66, Feb. 2002.
[26] 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.
[27] 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.
[28] 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.
[29] A. Arya and B. Mittendorf, "Project assignment when budget padding taints resource allocation," Manage. Sci., vol. 52, no. 9, pp. 1345–1358, Sep. 2006.
[30] C. Choi and Y. Park, "R&D proposal screening system based on textmining approach," Int. J. Technol. Intell. Plan., vol. 2, no. 1, pp. 61–72,2006.
[31] 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.
[32] 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.
[33] 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.
[34] 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, 2006, pp. 862–871.
[35] R. Feldman and J. Sanger, The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. New York: Cambridge Univ. Press, 2007.
[36] M. Konchady, Text Mining Application Programming. Boston, MA: Charles River Media, 2006.
[37]S.Bloehdom and p.cimiano et al. "An Ontology –based framework for Text Mining" Institute AIFB,July 28, 2004.
[38] Yiheng Chen and Bing Qin et al." The Comparison of SOM and K-means for Text Clustering "School of computer Science and Technology vol 3,no 2;May 2010
[39]http://en.wikipedia.org/wiki/Mathematical_model

