CHAPTER 5

CLUSTERING OF PROJECT PROPOSALS THROUGH VIRTUAL BUBBLE KIN PROCESS

5.1 INTRODUCTION

In this chapter, the private, government and research funding agencies are having the more number of problems to manage the grouping process of creating matching groups that had the same arrangement in terms of similar requirement proposals. In this consideration, MLTPGRPS assure and fulfill the problems followed by the Research and Development agencies. Aliane(2006) provides the verification and comparison regarding multilingual comparison and text mining method of text comparison. The group might be containing the more related text in to the similar group. In this regard introduce ME VS MLM Matching Estimation based Virtual spread shuffle Similarity of Multi-Lingual Text Mining method.

Every sorting group produces the performance comparison graph demonstrating that every group includes more number of the same proposal of the individual group. The group should include the relevant language and text. In this consideration the proposed ME VS MLM method efficiently minimizes unmatched items in the group and increase the accurate relevant text group formation in mining of the language in addition to the text input research proposal.
5.2 VIRTUAL BUBBLE KIN PROCESS

Bubble Kin visualizes the Research Proposal of a Researcher. The larger bubbles denote the researcher proposal with more relevance to the funding agency’s requirements. These requirements should group in to the database and its related data is to be verified for selection of the individual. The bubble illustrating the group in that proposal is stored in the database table. When it is needed to verify the virtual bubble kin, the database is updated to the database table of reviewer group up to the last entering Research Proposal which is posted to the Virtual Bubble Kin. The Virtual Bubble Kin process is required to take the following steps to find out the important data, or word or sentence in the Research Proposal which will be grouped for creating the proposal’s review.

Figure 5.1 Diagram for group of related proposal (before sorting)
1) First process allocates the time period of expiry for the Research Proposal Process. As the time period ends, the database receives the final group of the proposals generation. Each and every time period updates the virtual bubbles kin automatically by including the existing proposals. The proposal should be verified with the content which is already stored in the database. Due to the termination of the already existing content, it helps the database to manage the individual record of the research project proposals.

2) The second process is that flexible size for bubble is assigned for the identification of each and every proposal which illustrates the grouping of matching data to the reviewer group.

3) The next process is verifying the attributes of every requirement which is relevant to the government or private sector. Then all grouped proposals are verified and assigned to the individual reviewer of the research.

4) Simultaneously the proposal database for the government or private sector verify data with the help of the existing relevant data. If the data is already contained, time process checks the time period for expiry time of the process. Otherwise the bubble is created depending on the user’s requirement of matching the research proposal.

5) This process will continue until the prediction of data or words may be verified with the help of Prediction Checking Algorithm. The algorithm authorization depends on the already existing terms and trails.
Each and every prediction is used as the detail of the individual word or sentence which is taken to the language checking tool. From this each phase of the word prediction is verified with the group and send to the reviewer or it is send to the preliminary requirement of the research proposal process which can be helpful for eliminating the duplications of the same research proposals.

**ALGORITHM FOR FOCAL GROUP ALGORITHM (FGA)**

This algorithm mainly focuses on helping to identify, retrieve and store the matching group optimism that includes the same properties to mingle together with the already contained matching proposal from the database. The database contains useful proposals with the updated data of the
focal group. Then checking of the following terms and scenario which are relevant to the process is made.

The text words that are already contained in the research proposal is checked and, the next step is to check for another condition. At last all the conditions should be checked and the result of the performance is displayed through database, from which the throughput performance getting maximum must be noted and saved to the database.

This process is executed when the last text was compared with the present proposal and it continues till the final text was compared with the present proposal and an existing database. The prediction checking algorithm is used for checking the database with the current proposal and then the proposed proposals are rescheduled.

FOCAL GROUP ALGORITHM

Read Text ← input

Input ← User (Research Proposal)

Classify Proposal weather  (Research Proposal || other than this)

If the proposal! = (database)

Go proposal not relevant. i.e., fake relevant Txt

else

Research Proposal contains the = User (Research Proposal)

Validate || (Bottom up approach) || and

Validate || (top down approach) ||

Compare with Research Proposal.

Research Proposal (Bottom up approach)

for (text= Input; ;Group Comparison — level - -)

result as decremented by next group

arranged list value is Bottom up approach list value

AND

Research Proposal (top down approach)

for (text=Input; ;x)

x←Group Comparison (Result (Bottom up approach))

x←Successive Group (Relevant Text)

else Find text Is the fake suggestion.

Result is determined to next group

Arranged list value is top down approach list value.

5.3 FOCAL GROUP ALGORITHM (FGA)

Aliane(2006) uses ME VS MLM as a Matching Estimation based Virtual spread shuffle Similarity of Multi-Lingual text Mining. Based on this following algorithm and tools which are applied for the comparison is performed by Balamurugan & Pushpa(2015), Sergio Bolasco et al. (2002) checks and the relevant concern for the proposal is directed for mining from the group of words and compare it with the existing as follows.

a. Virtual Reality Simulation for the Virtual Spread Shuffle Similarity Matching minimizes the fake suggestion.
b. Simulacrum Tool is used for verifying and evaluating the language of any type of stream and creating the maximum to the group.

c. Focal Group based on the relevant content (Virtual Reality Simulation)

Focal Group based on the relevant content contains the words that are virtually relevant to the similar domain. That means contents other than fake should be compared with the database of virtual reality simulation. Here, focal group algorithm is used to separate the relevant text from the separate group.

d. Sub group based on the lingual contents (Simulacrum Tool)

Sub Group contains the sentences that are relevant to the already existing language. That means the input concerning the language may be checked based on proposals and it must be compared with the database of Simulacrum Tool and with the virtual reality simulation which is helpful for the grouping and sub group based on the language and Research proposals.

5.4 ME VS MLM

5.4.1 Virtual - Reality Model Algorithm (VRMA)

That means other than fake suggestion, the proposals must compare the database with virtual reality Model. Here we separate the relevant Text and assign to the individual group are separated based on Focal Group algorithm. The group already contains some nominal terms in database, which means nominal terms and relative database will be related to the each one of its own text of proposal.
Total number of checking criteria must be verified and shuffled by using the Virtual Reality Model Algorithm which gives the resultant performance to improve the group certainty of the individual proposal of each and every terms and scenario of the Virtual Reality Model Algorithm (VRMA). This algorithm is based on the Sorting tool and checking phase of individual classes of data that is, each data will be checked based on the following group of elements like Grammar, Architecture, Inner Structure, Exterior Structure, Technical, Referential, Web Collection, Report based verification and comparison. This procedure is useful to calculate the hybrid common factor for relevant data to the concern group. The common maximum influences are directed to the concern group and delivered to the reviewer. So the reviewer must know about the particular group common factor, which is very much useful for the relevant group formation as well as review for individual domain.

5.4.2 Sorting tool (Mingle Verification Comparison)

Sorting tool is used for one of the most important checking and comparison of Mingle verification of the individual terms and text or words used in the proposal. The most relevant data should be stored in the concern group after verification with the mingled words.

5.4.3 Algorithm Sorting Tool (Mingle Verification Comparison)

ComparisonTokenSort(Start [R1…RN], End[Condition Result …Next Result])

Input: summary Start [R1…RN] of orderable values

Output: summary End [Condition Result …Next Result] of start elements sorted in non-decreasing order
Assign ↔ Start element ← i;
Assign ↔ Next element ← j;
Assign ↔ Final element ← n;
for Start element ← 0 to n − 1 do
    Count[Start element] ← 0
    for Start element ← 0 to n − 2 do
        for Next element ← Start element + 1 to Final element − 1 do
            if Start[Start element] < A[Next element]
                Count[Next element] ← Count[Next element] + 1
            else
                Count[Start element] ← Count[Start element] + 1
        for Start element ← 0 to Final element − 1 do
            End[Count[Start element]] ← A[Start element]
    return End.

If each type contains a variety of proposals, then the proposals must be checked for matching of the individual text with one another. If one target point of the combinational text is reached that text related proposal must be grouped in a separate group with the help of database. Subsequently the database must update its own data, based on the sorting tool. Each time the database will work in the direction of the output of sorting. If the output ends, it is indicated by the zero. That means all the text has been compared with the proposal and the matching result must be send to the reviewer.
Virtual Reality Model Algorithm

Virtual reality model algorithm is used to check each and every consideration of the maximum text appearance for checking of the following terms and scenario that consists of comparison with individual elements like Sorting tool, Grammar centered, Numerical Logical centered, Inner
structure centered, Structural centered, Technical centered, Referential centered, Web collection centered and Report centered.

The overview of the Virtual reality model algorithm is described based on the input term and it checks every text as described by the following algorithm.

If input← Language (language checker)
Language checker (index term (Vowels), database, null)
Do
Similar language (grouped)
if summery (GRAMMER)
do sorting(R1..RN)← comparison result (GRAMMER)
else
summery (NumLog)
do sorting(R1..RN)← comparison result (NumLog)
else
summery (INNER STRUCTURE)
do sorting(R1..RN)← comparison result (INNER STRUCTURE)
else
summery (STRUCTURAL)
do sorting(R1..RN)← comparison result (STRUCTURAL)
else
summery (TECHNICAL BASED)
do sorting(R1..RN)← comparison result (TECHNICAL BASED)
else
summery (REFERENTIAL)
do sorting(R1..RN)← comparison result (REFERENTIAL)
else
summery (WEB COLLECTION)
do sorting(R1..RN)← comparison result (WEB COLLECTION)
else
summery (REPORT BASED)
do sorting(R1..RN)← comparison result (REPORT BASED)
else
summery (semantic)
do sorting(R1..RN)← comparison result (semantic)
end if

SORTING (R1..RN)← COMPARISON RESULT (REQUIREMENT)

Sorting based the grammar verifies and compares the individual text in terms of verbal and non-verbal.

if element (GRAMMAR)
Assign token (individual)
Check (verbal|Non-verbal)
From Dictionary (Redirect: language dictionary 1.COM)
if available
for (input=first proposal token1;input>token+1;next input)

    Next input ← Increment with token for each

Token assignment
Read (token) = result;

sorting(R1..RN) ← comparison sorting (RESULT

else

    sorting(R1..RN) ← comparison sorting (RESULT

where Expire time(hour)>time(now())
quits(tokenizing)
token.trim().length(token) == 0
Create sub group (language dictionary 2)

Sorting based on the numerical and logical terms are checked and compared
with the individual text.

if element (NumLog)

Assign token (individual)

Check (Logical| Numerical)

From Dictionary (Redirect: language dictionary 1.COM)
if available
for (input=first proposal token 1; input>token+1; next input)

    Next input ← Increment with token for each

Token assignment

read (token 1, token +1) = result (Logical, Numerical);

sorting (R1..RN) ← comparison sorting (RESULTlogical)

else

sorting (R1..RN) ← comparison sorting (RESULTNumerical)

where Expire time (hour) > time (now())

quits (tokenizing)

token.trim().length (token) == 0

Create sub group (language dictionary 1)

Sorting based on the inner structure; in this case the individual text can be verified and compared in terms of index term and documentation.

if element (INNER STRUCTURE)

Assign token (individual)

Check (index|document)

From Dictionary (Redirect: Document.doc)

if available

for (input= token 1; input>token+1; next input)

    Next input ← Increment with token for each

Token assignment
read (token) = result || (index, document)||;

sorting (R1..RN) ← comparison sorting (RESULT_{index})

e else

sorting (R1..RN) ← comparison sorting (RESULT_{DOC})

where Expire time(hour) > time(now())

quits(tokenizing)

token.trim().length(token) == 0

Create sub group (language dictionary 3)

Sorting based on the structure; in this case the individual text can be verified and compared in terms of starting, middle and end text.

if element (STRUCTURAL)

Assign token (start || middle || end)

Check (token ↔ (start && middle && end))

From Dictionary (Redirect: Research proposal input)

if available

do (token ↔ (start && middle && end))

for (input = token(start || middle || end); input > token + 1; nextr input)

	Next input ← Increment with token for each

Token assignment

read (token) = result (start, middle, end ||);

sorting (R1..RN) ← comparison sorting (RESULT_{start})
else

sorting(R1..RN)← comparison sorting (RESULT_end)

while(comparison sorting (RESULT_Middle))

where Expire time(hour)>time(now())

quits(tokenizing)

token.trim().length(token) == 0

Create sub group (language dictionary 4)

Sorting is based on the technical centered; in this case the individual text can be verified and compared in terms of report, reference, and knowledge.

if element (TECHNICAL BASED)

Assign token (individual)

Check (report||reference||knowledge)

From Dictionary (Redirect: Document.doc (report_{r1}, reference_{r2}))

if available

for (input= token ;input>token+(r1,r2) ;nextr input)

    Next input ← Increment with token for each

Token assignment

read (token(r1,r2)) =result ||( report||reference)||;

sorting (R1..RN)← comparison sorting (RESULT_{report(r1)})

else

sorting(R1..RN)← comparison sorting (RESULT_{ref(r2)})
where \( \text{Expire time(hour)} > \text{time(now())} \)

\text{quits(tokenizing)}

\text{token.trim().length(token) == 0}

Create sub group (language dictionary 3).

5.5 EXPERIMENTAL ANALYSIS

MLTP Grouping for Research Project Selection method efficiently reduces unmatched items in the group and also increases the accurate relevant text group formation by mining of language as well as text input research proposal. With respect to the section IV, V and VI the grouping based on the domain and language are described below.

![Pie chart showing distribution of languages](image)

**Figure 5.4 Lingual based grouping**
The above Figure 5.4 illustrates that sorting different language based proposals are done by using Virtual Bubble Kin Process. Here the more relevant proposals came from the natural language English and also all other languages such as Mandarin, French, Hindi, Japanese, Spanish, Turkish, Italian etc. These results are sorted based on the subject domain. The figure 5.4 shows that the Comparison for domain and language based on selection and grouping. It is useful to organize the actual keyword that is compared to the domain database. It creates the individual key words appeared in the proposal using key word comparison with the domain keyword. It is tabulated in the following table which includes the number of key words appeared to any proposal that are grouped under the specific proposal.

Figure 5.5  Comparison of domain and language based selection and grouping
The above Figure 5.5 illustrates the comparison for domain and language based on selection and grouping. It is helpful to organize the actual keyword by comparing it with the domain database. It creates the individual key words appeared in the proposal using keyword comparison with the domain keyword. It is tabulated in the following table which includes the number of key words appeared in any proposal which are grouped under the specific proposal.

**Table 5.1 Keywords and its count for different domains**

<table>
<thead>
<tr>
<th>Domain Name</th>
<th>Networking</th>
<th>Data Mining</th>
<th>Image Processing</th>
<th>Mobile Computing</th>
<th>Cloud Computing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>keywords</td>
<td>keyword</td>
<td>keywords</td>
<td>keyword</td>
<td>keywords</td>
</tr>
<tr>
<td></td>
<td>count</td>
<td>count</td>
<td>count</td>
<td>count</td>
<td>count</td>
</tr>
<tr>
<td>Network</td>
<td>757</td>
<td>1736</td>
<td>1062</td>
<td>489</td>
<td>1062</td>
</tr>
<tr>
<td>Packet</td>
<td>319</td>
<td>282</td>
<td>107</td>
<td>161</td>
<td>Encryption</td>
</tr>
<tr>
<td>Communication</td>
<td>101</td>
<td>71</td>
<td>27</td>
<td>44</td>
<td>Latency</td>
</tr>
<tr>
<td>Routing</td>
<td>141</td>
<td>108</td>
<td>56</td>
<td>94</td>
<td>Integrity</td>
</tr>
<tr>
<td>Server</td>
<td>291</td>
<td>268</td>
<td>94</td>
<td>88</td>
<td>Audit</td>
</tr>
<tr>
<td>Throughput</td>
<td>102</td>
<td>72</td>
<td>28</td>
<td>57</td>
<td>Costs</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>489</td>
<td>Node</td>
<td>123</td>
<td>Channel</td>
<td>Cloud</td>
</tr>
<tr>
<td>Ict</td>
<td>116</td>
<td>Learning</td>
<td>33</td>
<td>Encryption</td>
<td>Cluster</td>
</tr>
<tr>
<td>User</td>
<td>378</td>
<td>Support</td>
<td>179</td>
<td>GPS</td>
<td>Packet</td>
</tr>
<tr>
<td>Local</td>
<td>92</td>
<td>Score</td>
<td>70</td>
<td>GSM</td>
<td>SaaS</td>
</tr>
<tr>
<td>Mac</td>
<td>104</td>
<td>Pattern</td>
<td>29</td>
<td>Infrastructure</td>
<td>Outsourcing</td>
</tr>
<tr>
<td>Client</td>
<td>106</td>
<td>outlier</td>
<td>30</td>
<td>IP</td>
<td>IP</td>
</tr>
<tr>
<td>Internet</td>
<td>120</td>
<td>Outliers</td>
<td>33</td>
<td>Location</td>
<td>Qos</td>
</tr>
<tr>
<td>Wireless</td>
<td>335</td>
<td>Mean</td>
<td>130</td>
<td>MAC</td>
<td>User</td>
</tr>
<tr>
<td>Node</td>
<td>408</td>
<td>Model</td>
<td>290</td>
<td>Mobile</td>
<td>Bandwidth</td>
</tr>
<tr>
<td>Ad</td>
<td>166</td>
<td>Transformation</td>
<td>122</td>
<td>SaaS</td>
<td></td>
</tr>
<tr>
<td>Map</td>
<td>105</td>
<td>Classification</td>
<td>80</td>
<td>Presence</td>
<td></td>
</tr>
<tr>
<td>Traffic</td>
<td>197</td>
<td>Rule</td>
<td>88</td>
<td>Source</td>
<td>Public</td>
</tr>
<tr>
<td>Ip</td>
<td>84</td>
<td>Median</td>
<td>23</td>
<td>Throughput</td>
<td>API</td>
</tr>
<tr>
<td>Protocol</td>
<td>226</td>
<td>Accuracy</td>
<td>76</td>
<td>Wireless</td>
<td>Client</td>
</tr>
</tbody>
</table>

Here there are five domains such as Networking, Data Mining, Image Processing, Mobile Computing and Cloud Computing which are taken with the help of different keywords for each domain they are grouped based on the following graph.
Figure 5.6 Keyword based domain grouping - Networking

The above Figure 5.6 illustrates the networking domain by using different keywords for grouping under this domain.

Figure 5.7 Keyword based domain grouping (Data mining)
The above Figure 5.7 illustrates the data mining domain by using different keywords for grouping.

![Image of Figure 5.7](image)

**Figure 5.8 Keyword based domain grouping (Image Processing)**

The above figure 5.8 illustrates Image Processing domain by using different keywords for grouping under this domain.

Based on the individual domain keywords, there are many groups which are formed here depending on the keywords that are grouped under the networking. It forms the network domain. Once the reviewers need to meet the requirement of their own domain this keyword based results are recommended to the reviewer. Also, all the domains are grouped such as Data Mining, Image Processing and networking and so on. The performance of the proposed MLTP Grouping for Research Project Selection is evaluated by calculating Precision, Recall and F-measure values.
The performance of the proposed MLTP Grouping for Research Project Selection is evaluated by calculating Precision, Recall and F measure. Precision is defined as the ratio of the number of retrieved relevant proposals to the number of retrieved.

\[
\text{precision} = \frac{\text{number of relevant proposal}}{\text{number of available proposal}}
\]

Recall is defined as the ratio of the number of grouped relevant proposal to the total number of relevant proposal in the collection.

\[
\text{recall} = \frac{\text{number of grouped relevant proposal}}{\text{total number of relevant proposal}}
\]

\(f_{\text{measure}}\) can be calculated as

\[
f_{\text{measure}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

![Figure 5.9 F-measure for different approach](chart.png)
The above Figure 5.9 illustrates F-measure for five different domain approaches and mentions the performance of the proposed method. Evaluated precision, recall and F-measure values are defined in percentage (%).

![Graph showing F-measure for different domain approaches](image)

**Figure 5.10 Domain based keyword result count**

This above figure 5.10 illustrates domain based keyword result count and performance of the proposed ME VS MLM method evaluates the F-measure values.

Table 5.2: Precision Recall and F-Measure Comparison for Different Domain and Language

<table>
<thead>
<tr>
<th>Domain alone</th>
<th>language alone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>Precision</td>
</tr>
<tr>
<td>Networking</td>
<td>0.61</td>
</tr>
<tr>
<td>Data Mining</td>
<td>0.74</td>
</tr>
<tr>
<td>Internet Programming</td>
<td>0.86</td>
</tr>
<tr>
<td>Mobile Computing</td>
<td>0.90</td>
</tr>
<tr>
<td>cloud computing</td>
<td>0.91</td>
</tr>
</tbody>
</table>
The above Table 5.2 shows improvement in F-measure values for performance metrics of the proposed method while comparing with the existing methods. The proposed ME VS MLM method is using improved accuracy and performance. The proposed method is improved better than all other existing methods.

5.6 CHAPTER SUMMARY

The manual matching of the same proposals and their complexity is minimized by ontology based text mining approach which is working with group of large number of proposals. This method can be used to improve the efficiency and effectiveness of research project grouping. As a replacement for the Keywords based grouping, MLTPGRPS techniques are proposed for Proposal selection processes in R&D agencies. It is useful for classifying a proposal in an efficient manner. Bubble Kin visualizes the Research Proposal of Researcher. The larger bubbles represent the researcher’s proposal with more relevance to the funding agency’s requirement. Total number of checking criteria must be verified and shuffled by using the Virtual Reality Model Algorithm which gives the resultant performance to improve the grouping certainty of the individual proposal of each and every terms and scenario of the Virtual Reality Model Algorithm (VRMA). The proposed method efficiently reduces unmatched items in the group and also increase the accurate relevant text group formation by mining of language as well as text proposal input research proposal.