

4. Algorithm

Take each pattern from the equivalence class list. Initialize regular expression class for accessing regular expression tool. Internally verify whether the search pattern is present or not. If the search pattern is present, then the corresponding searched pattern is displayed and provides download facility. Not only searching can be done directly but also through Author community. Author Community is generated based on the text clustering algorithm Kmean. Kmean text clustering algorithm is used to find similar document and then list their author into a clusters.

5. Evaluation

Based on the experimental evaluation, out of 20 documents NFA based MPBTM lists more documents as search result than MPBTM .The detailed description is shown in Table 4 and figure 2.

In this evaluation, check the performance of MPBTM and NFA based MPBTM in terms of hit rate. To facilitate, input same query and then compare how much documents is obtained in MPBTM and NFA based MPBTM.

Table 4: Evaluation

| Data Set | MPBTM | NFA based MPBTM |
|----------|-------|-----------------|
| 5 | 1 | 3 |
| 10 | 4 | 7 |
| 15 | 9 | 13 |
| 20 | 11 | 17 |

Table 4 shows the performance of MPBTM and NFA based MPBTM in terms of data hit rate.

When 5 data set was uploaded, MPBTM showed 1 and NFA based MPBTM showed 3 documents as search result. Continue this evaluation up to 20 data set .Based on Table 4, Figure 2 is plot.

In Figure 2, X coordinate shows data set that is the number of document uploaded .Y coordinate shows document hit rate that is number of documents retrieved based on the search.

Efficiency of NFA based MPBTM is higher than MPBTM in terms of document hit rate. User can search document based on Author also. That is there are two possible way to search document. One is based on the input query of user other is user interested authors. In both case meaningful and expression based document is retrieved. So that while comparing with previous model, proposed model give efficient searching result.

From the graph, in the case of NFA based MPBTM data hit rate is increasing while increasing the data set. That is data hit rate is directly proportional to data set. That is effective semantic check is done. But in the case of MPBTM we cannot predict that efficient data hit rate obtained for corresponding data set.

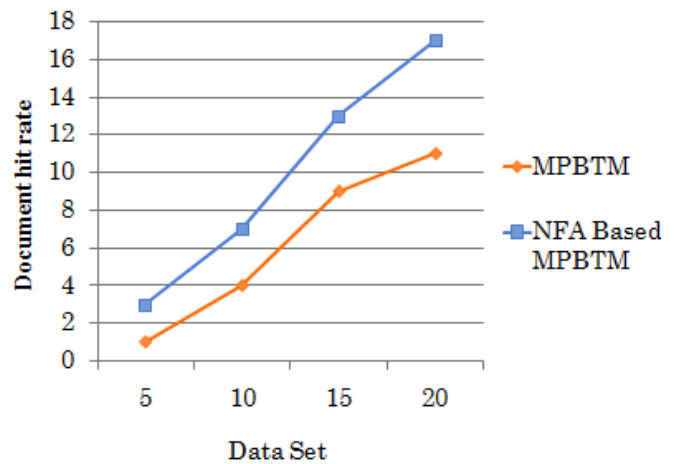


Figure 2: Comparison of NFA based MPBTM and MPBTM in terms of hit rate

From the evaluation, it can be stated that performance of NFA based MPBTM and author community is efficient than MPBTM. Search results are also relevant.

6. Conclusion

Comparing with previous model MPBTM, NFA based MPBTM and author community can give relevant search result that is searching document in terms of the author as well as the semantic meaning of user input.

MPBTM consists of topic distributions, describing topic preferences of documents or collection of documents and structured pattern based topic representation , representing semantic meaning of the topics in a document .To improve the searching efficiency of document , NFA based MPBTM and Author community are proposed. From the evaluation, it is clear that proposed model overcome the limitation of MPBTM.So declare that efficiency of document search of proposed model higher than that of early model.

NFA based MPBTM does not support case sensitivity and it takes more processing time. In future methods for identifying case sensitive topics as well as methods to reduce the processing time for generating the model can be introduced. Then it will be the one of the efficient model in information filtering.

References

- [1] Y. Bastide, R. Taouil, N. Pasquier, G. Stumme, and L. Lakhal, "Mining frequent patterns with counting inference," ACM SIGKDD Explorations Newslett., vol. 2, no. 2, pp. 66–75, 2000
- [2] H. Cheng, X. Yan, J. Han, and C.-W. Hsu, "Discriminative frequent pattern analysis for effective classification," in Proc. IEEE 23rd Int. Conf. Data Eng., 2007, pp. 716–725
- [3] X. Wei and W. B. Croft, "LDA-based document models for ad-hoc retrieval," in Proc. 29th Annu. Int. ACM SIGIR Conf. Res. Develop. Inform. Retrieval, 2006, pp. 178–185.
- [4] C.Wang and D.M.Beli,"Collaborative topic modelling for recommending scientific articles," in Proc.17th

- ACM SIGKDD Int. Conf. knowl. Discov. Data Min., 2011, pp. 448-456
- [5] D.M. Blei, A. D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," J. Mach. Learn. Res., vol. 3, pp. 993-1022, 2003.
- [6] T. Hofmann, "Probabilistic latent semantic indexing," in Proc 22nd Annu. Int. ACM SIGIR Conf. on Res. Develop. Inform. Retrieval 1999, pp. 50-57.
- [7] Y. Gao, Y. Xu, and Y. Li, "Pattern-based topic models for information filtering," in Proc. Int. Conf. Data Min. Workshop SENTIRE, 2013, pp. 921-928.
- [8] Yang Gao, Yue Xu, and Yuefeng Li, "Pattern-based Topics for Document Modelling in Information Filtering," IEEE Transaction on Knowledge And Data Engineering, June 2015.
- [9] Tapas Kananga, David M. Mount, Nathan S. Netanyahu, Christine D. Piatko, Ruth Silverman, and Angela Y. Wu, "An Efficient k-Means Clustering Algorithm: Analysis and Implementation," IEEE Transaction on Pattern Analysis and Machine Intelligence, VOL. 24, NO. 7, JULY 2002
- [10] https://en.wikipedia.org/wiki/User_modeling

Author Profile



Neena G. Krishnan is pursuing M. Tech in Computer Science and Engineering from Mangalam College of Engineering, Mahatma Gandhi University. She received B.Tech degree in 2013 from Mangalam College of Engineering, Mahatma Gandhi University, Kottayam, Kerala, India. Her research interests are .NET, Java etc.



Neena Joseph received the B.tech Degree in Computer Science & Engineering from Mangalam College of Engineering in 2008 and M. Tech Degree from Achariya College of Business & technology in 2012. She is presently working as Assistant Professor at Mangalam College of Engineering. Her Research areas of interest are Relational Database and Compilers.