









The main purposes of the experiments presented in this section is to test the proficiency of GDRM in extracting strong association rules, and to assess its efficiency on several text analysis and text mining tasks. The proposed GDRM method is compared with one of the best graph based text mining algorithms, that is, the Association Rules based on Weighting algorithm (GARW) [2]. Both GARW and GDRM scans the documents only once but GARW concentrates only on the keyword sets that are stored in XML file, while GDRM takes in consideration all words but abbreviations, the file for the proposed technique consists of the terms only together with their frequencies in each document.

the input to the proposed system is the minimum support threshold to extract the frequent terms and then the system requires the minimum confidence to extract only strong rules from the file containing the frequent terms, the output is the time required to get the desired rules.

The experiments have been carried out using a database of documents that contains 250 documents is 1120 KB in size and the total number of single words is about 55000. Each document contained on average 220 single words. After the filtration process, the number of single words is minimized

to 17417. The proposed GDRM algorithm use the same platform as GARW to assure that the comparisons are reasonable. The experiments were performed on a Core i5, 3.8 GHz system running Windows 7 with 8 GB of RAM. The execution time is reduced and the strongness of the extracted rules is increased using the proposed GDRM in comparable to the GARW algorithm.

Execution Time (min)		Min Support %
GARW	GDRM	
19	7	15
15	4	20
14	3	25
12	3	30
9	2	35
7	2	40
6	2	45
5	1	50

Table 1: Comparison between GDRM and GARW  
As shown in figure 1, the time required to mine association rules from text documents is decreased dramatically using the proposed technique.

## References

- [1] Himani Raina, Omais Shafi. *Analysis Of Supervised Classification Algorithms*. International Journal Of Scientific & Technology Research Volume 4, Issue 09, September 2015.
- [2] Hany Mahgoub, Dietmar Rösner, Nabil Ismail and Fawzy Torkey. *A Text Mining Technique Using Association Rules Extraction*. World Academy of Science, Engineering and Technology International Journal of Computer, Electrical, Automation, Control and Information Engineering Vol:2, No:6, 2008.
- [3] Agrawal, R., Imielinski T. & Swami A. 1993. *Mining Association Rules between Sets of Items in Large Databases*. Proceedings of the 1993 ACM SIGMOD Conference Washington DC, USA, pp. 1 – 10.
- [4] G. Salton, A. Wong, and C. S. Yang. *A vector space model for automatic indexing*. Communications of the ACM, 18(11):613{620, 1975.
- [5] S. S. Sonawane, Dr. P. A. Kulkarni. *Graph based Representation and Analysis of Text Document: A Survey of Techniques*. International Journal of Computer Applications (0975 8887) Volume 96 - No. 19, June 2014.
- [6] Surabhi Lingwal, Bhumika Gupta. *A Text Mining Approach for Automatic Classification Of Web Pages*. Proc. of the Second Intl. Conf. on Advances in Electronics, Electrical and Computer Engineering -- EEC 2013. ISBN: 978-981-07-6935-2 doi:10.3850/978-981-07-6935-2\_52
- [7] Inokuchi, A., Washio, T., & Motoda, H. 2000. *An Apriori-based algorithm for mining frequent substructures from graph data*. In Proceedings of the 4th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'00), pp 1 – 11.
- [8] Kuramochi, M. & Karypis, G. 2001. *Frequent subgraph discovery*. In Proceedings of the 2001 IEEE International Conference on Data Mining, pp. 313–320.
- [9] Yan, X. & Han, J. 2002. *gSpan: graph-based substructure pattern mining*. Technical Report UIUCDCS-R-2002-2296, Department of Computer Science, University of Illinois at Urbana-Champaign.
- [10] Srihari, S. 2011. *Principles of data mining*. University at Buffalo. The State University of New York.  
<http://www.cedar.buffalo.edu/~srihari/CSE626/Lecture-Slides/Ch1-Part1-Introduction.pdf>, 2011, pp 1 – 41.
- [11] Wang, J. 2009. *Data warehousing and mining: concepts, methodologies, tools, and applications*. USA: Information Science Reference, pp. 303-335.
- [12] Majeed, S. K. & Abbas, H. K. 2010. *An improved distributed association rule algorithm*. Eng.& Tech. Journal, Vol.28, No.18, 2010, pp. 5695 – 5710.
- [13] S. M. Kamruzzaman, Farhana Haider, Ahmed Ryadh Hasan3. *Text Classification Using Data Mining*. ICTM 2005.
- [14] Ahmed Hamza Osman & Omar Mohammed Barukub. *Graph-Based Text representation and Matching: A Review of the State of the Art and Future Challenges*. IEEE Access. Volume 8, 2020. Pp 87562 -87583.