# REFERENCES

[1] E. Ulken  'A Question of Balance: Are Google News Search Results Politically Biased?', *USC Annenberg School for Communication*. Available: http://ulken.com/thesis/googlenews-bias-study.pdf, 2005.

[2] K. Rajan, V. Ramalingam, M. Ganesan, S. Palanivel, B. Palaniappan "Automatic classification of Tamil documents using vector space model and artificial neural network" *Expert Systems with Applications: Elsevier*, Vol. 36, pp.10914–10918, 2009.

[3] V.G.S. Sharma, "Recent Developments in Text Clustering Techniques," *International Journal of Computer Applications*, vol. 37, no. 6, pp. 14-19, 2012.

[4] P. Nanayakkara and S. Ranathunga, "Clustering sinhala news articles using corpus-based similarity measures," in *2018 Moratuwa Engineering Research Conference (MERCon).* IEEE, 2018, pp. 437– 442.

[5] T. K. Landauer, P. W. Foltz and D. Laham, "An introduction to latent semantic analysis," *Discourse processes,* vol. 25, no. 2-3, pp. 259-284, 1998.

[6] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," *lcml*, vol. 97, pp. 412-420, 1997.

[7] A. Karima, E. Zakaria, T. G. Yamina, "Arabic text categorization: A comparative study of different representation modes", *Journal of Theoretical and Applied Information Technology*, vol. 38, no. 1, pp. 1-5, 2012.

[8] Z Elberrichi.  ,A. Rahmoun, and M. Bentaalah, "Using WordNet for Text Categorization", *International Arab Journal of Information Technology(IAJIT)*, 5(1): 16-24, 2008.

[9] H. Schütze, D. Hull,  and  A Pedersen,  "A Comparison of Classifiers and Document Representations for the Routing Problem," in *Proceedings of SIGIR-95, 18th ACM International Conference on Research and Development in Information Retrieval,* New York, pp. 229-237, 1995.

[10] A. Amine, Z. Elberrichi and M. Simonet, "Evaluation of Text Clustering Methods Using WordNet," *The International Arab Journal of Information Technology*, vol. 7, no. 4, pp. 349-357, 2010.

[11] J Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," in *Proc..of the eleventh international conference on Information and knowledge management,* McLean, Virginia, USA, 2002, pp. 515-524

[12] C. Bouras and V. Tsogkas, "A clustering technique for news articles using WordNet," *Knowledge-Based Systems*, vol. 36, pp. 115-128, 2012.

[13] Marcus Lönnberg, Love Yregård, "Large scale news article clustering", *M.S. thesis, Dept. of Comput. Sci. and Eng., Chalmers University of Technology, Gothenburg*, Sweden, 2013

[14] D. Shen, J. Liu J, C. Nicholas and E. Miller, "Performance and Scalability of a Large-Scale N-gram Based Information Retrieval System," *Computer Journal of Digital Information*, vol. 1, no. 5, pp. 257-265, 1999.

[15] T Hofmann, "Probmap: A Probabilistic Approach for Mapping Large Document Collections", *Journal for Intelligent Data Analysis*, vol. 4, pp. 149-164, 2000.

[16] A. Huang, "Similarity measures for text document clustering," in *Proc. New Zealand Computer Science Research Student Conf. (NZCSRSC 2008)*, Christchurch, New Zealand, pp. 49-56, 2008.

[17] W. H. Gomaa and A. A. Fahmy, "A survey of text similarity approaches," *International Journal of Computer Applications*, vol. 68, no. 13, 2013.

[18] A. Hotho, A. Maedche and S. Staab, "Ontology-based text document clustering," *KI*, vol. 16, no. 4, pp. 48-54, 2002.

[19] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *In Proceedings of ICLR Workshops Track*, 2013.

[20] T. Mikolov, I Sutskever, K. Chen, G. Corrado and J. Dean.Distributed representations of words and phrases and their compositionality. *In Proc. Advances in Neural Information Processing Systems* 26 3111–3119 (2013).

[21] Y. Goldberg, O. Levy. word2vec explained: deriving Mikolov et al.'s negative-sampling word-embedding method, 2014.

[22] R. Nalawade, A. Samal and K. Avhad, "Improved Similarity Measure For Text Classification And Clustering," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 5, pp. 214-219, 2016.

[23] M.Thangarasu and Dr.R.Manavalan, "Stemmers for Tamil Language: Performance Analyses," *International Journal for Computer Science & Engineering Technology*, Vol. 4 No. 07 Jul 2013

[24] S. Bhatia, "Adaptive K-Means Clustering," in *FLAIRS Conference*, 2004. pp. 695-699

[25] N. Friburger and D. Maurel, "Textual similarity based on proper names," *in Proc. of the workshop Mathematical/Formal Methods in Information Retrieval,* 2002, pp.155-167.

[26] G. Furnas and W. Jones, "Pictures of Relevance: A Geometric Analysis of Similarity Measures*," Computer Journal of the American Society for Information Science*, vol. 38, no. 6, pp. 420-442, 1987.

[27] M. Steinbach, G. Karypis and V. Kumar, "A comparison of document clustering techniques," *KDD workshop on text mining*, vol. 400, no. 1, pp. 525-526, 2000.

[28] Rai, P., and Singh, S. "A Survey of Clustering Techniques," *International Journal of Computer Applications,* 7(12), 1-5, (2010, October).

[29] A. Moore and D. Pelleg, "X-means: Extending K-means with Efficient Estimation of the Number of Clusters," in *Proc. of the Seventeenth International Conference on Machine Learning*, pp. 727-734, 2000.

[30] Y. He, Q. Chen, X. Wang, R. Xu, X. Bai and X. Meng, "An adaptive affinity propagation document clustering," *The 7th International Conference on Informatics and Systems (INFOS)*, Cairo, 2010, pp. 1-7, 2010.

[31] Shailendra Kumar Shrivastava, J L Rana and R C Jain. Article: "Text Document Clustering based on Phrase Similarity using Affinity Propagation". *International Journal of Computer Applications* 61(18):38-44, January 2013.

[32] G. Leban, B. Fortuna and M. Grobelnik, "Using News Articles for Real-time Cross-Lingual Event Detection and Filtering," *in Proc. of the NewsIR'16 Workshop at ECIR*, Padua, Italy, 2016. pp. 33-38

[33] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern recognition letters*, vol. 31, no. 8, pp. 651-666, 2010.

[34] M. F. Porter, "An algorithm for suffix stripping," *Program*, vol. 14, no. 3, pp. 130-137, 1980.

[35] G. A. Miller, "WordNet: a lexical database for English," *Communications of the ACM,* vol. 38, no. 11, pp. 39-41, 1995.

[36] David D. Lewis, "An evaluation of phrasal and clustered representations on a text categorization task," *Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval,* p.37-50, June 21-24, 1992, Copenhagen, Denmark

[37] G. Salton , A. Wong , C. S. Yang, "A vector space model for automatic indexing", *Communications of the ACM*, v.18 n.11, p.613-620, Nov. 1975

[38] M.Thangarasu and Dr.R.Manavalan, "Design and Development of Stemmer for Tamil Language: Cluster Analysis," *International Journal of Advanced Research in Computer Science and Software Engineering*, Vol. 3 No. 07, 2013

[39] D. Pfitzner, R. Leibbrandt and D. Powers, "Characterization and evaluation of similarity measures for pairs of clusterings," *Knowledge and Information Systems,* vol. 19, no. 3, pp. 361-394, 2009.

[40] Farook Fathima Farhath, " Sinhala - Tamil Statistical Machine Translation (SMT) for Official Documents," MPhil. thesis, Dept. of Comput. Sci. and Eng., Univ. of Moratuwa, Sri Lanka, 2018.

[41] D. A. Meedeniya and A. S. Perera, "A Comparative Study on Data Representation to Categorize Text Documents," in *Twentieth International Conference on Software Engineering and Knowledge Engineering*,2008.

[42] J. L. Fleiss, "Measuring nominal scale agreement among many raters," *Psychological Bulletin*, vol. 76, no. 5, pp. 378–382, 1971.