

accuracy value of 88%. However, several clusters are not quite right. It should be re-evaluated in the keyword feature extraction process to extract appropriate features for clustering. In the future, this method can be developed further by adding other functions, updates and add models, and several other functions to evaluate the model periodically.

NOMENCLATURE

| | |
|---------------|--|
| w | notation for a word |
| d | notation for a document |
| $f_{w_i d}$ | frequency word-i in a document |
| F_d | feature vector of document |
| th_d | threshold value in a document |
| a | notation for alphabet letter |
| $f_{a_j w_i}$ | frequency of alphabet-j in the word-i |
| S_{min} | minimum shifting value |
| C_k | feature vector of centroid-k |
| O | vector of zero value |
| g_{min} | minimum distance value to ground (zero) |
| r_k | radius of cluster-k |
| α | coefficient for updating centroid |
| β | coefficient for determining threshold radius |

REFERENCES

- [1] A. S. M. Romli, *Jurnalistik Online: Panduan Mengelola Media Online*, I. Kurniawan, Bandung, Indonesia: Nuansa, 2012.
- [2] S. S. Kurnia, *Jurnalisme Kontemporer*, Jakarta, Indonesia: Yayasan Pustaka Obor Indonesia, 2017.
- [3] D. Z. E. Puspitasari, A. R. Barakbah, I. Winarno, "Automatic Representative News Generation using Automatic Clustering", in *Industrial Electronics Seminar (IES) 2012*, 2012.
- [4] AMSI. (2019) Dari 47 Ribu, Baru 2.700 Media Online Terverifikasi Dewan Pers. [Online]. Available: <https://www.amsi.or.id/dari-47-ribu-baru-2-700-media-online-terverifikasi-dewan-pers/>
- [5] J. B. Schmitt, C. A. Debbelt, F. M. Schneider, "Too much information? Predictors of information overload in the context of online news exposure", *Information Communication and Society*, vol. 21, no. 8, pp. 1151-1167, Apr. 2017.
- [6] I. Subašić, B. Berendt, "Peddling or Creating? Investigating the Role of Twitter in News Reporting", in *Proc. ECIR 2011*, 2011, p 207-213, 2011.
- [7] P. Virmani, S. Taneja, "A Text Preprocessing Approach for Efficacious Information Retrieval," *Smart Innovations in Communication and Computational Sciences*, pp.13-22, Jan. 2019.
- [8] J. Asian, H. E. Wiliams dan S. M. M. Tahaghoghi, "Stemming Indonesian", in *Proceedings ACSC '05*, 2005, p. 307-314.
- [9] Z. Aliniya, S. A. Mirroshandel, "A novel combinatorial merge-split approach for automatic clustering using imperialist competitive algorithm", *Expert System with Applications*, vol. 117, p. 243-266, Mar. 2019.
- [10] M. K. Islam, M. M. Ahmed, K. Z. Zamli, "A buffer-based online clustering for evolving data stream", *Information Sciences*, vol. 489, p. 113-135, Jul. 2019.
- [11] M. Adriani, J. Asian, B. Nazief, S. M. M. Tahaghoghi dan H. E. Wiliam, "Stemming Indonesian: A confix-Stripping Approach", *ACM Transactions on Asian Language Information Processing*, vol. 6, no. 4, Dec. 2007.
- [12] L. D. Pratiwi, "Perbandingan Algoritma Nazief Adriani dan Paice Husk untuk Proses Stemming Teks Bahasa Indonesia," B.Sc thesis, UIN Sunan Gunung Djati, Bandung, Indonesia, Oct. 2019.
- [13] A. S. Rizki, A. Tjahyanto dan R. Trialih, "Comparison of Stemming Algorithm on Indonesian Text Processing", *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 17, no. 1, pp. 95-102, Feb. 2019.
- [14] M. S. H. Simarangkir, "Studi Perbandingan Algoritma-Algoritma Stemming untuk Dokumen Teks Bahasa Indonesia," *Jurnal INKOFAR*, vol. 1, no. 1, Jul. 2017.
- [15] P. Prihatin, I. D. Putra, I. Giriantri dan M. Sudarma, "Stemming Algorithm for Indonesian Digital News Text Processing", *International Journal of Engineering and Emerging Technology*, vol. 2, no. 2, pp. 1-7, Mar. 2018.
- [16] M. Sigita, A. R. Barakbah, E. M. Kusumaningtyas, I. W., "Automatic Representative News Generation using On-Line Clustering", *EMITTER International Journal of Engineering Technology*, vol. 1, no. 1, pp. 107-113, Dec. 2013.
- [17] J. Azzopardi, C. Staff, "Incremental Clustering of News Reports", *Algorithms - Open Access Journal*, vol. 5, no. 3, pp. 364 - 378, Dec. 2012.
- [18] A. M. Bakr, N. M. Ghanem dan M. A. Ismail, "Efficient Incremental Density-based algorithm for clustering large datasets," *Alexandria Engineering Journal*, vol. 54, no. 4, pp. 1147-1154, Dec. 2015.
- [19] P. Laban dan M. Hearst, "newsLens: building and visualizing long-ranging news stories," in *Proceedings of the Events and Stories in the News Workshop*, 2017, p. 1-9.
- [20] R. Florence, B. Nogueira dan R. Marcacini, "Constrained Hierarchical Clustering of News Events," in *Proceedings of the 21st International Database Engineering & Applications Symposium (IDEAS) 2017*, 2017, p. 49-56.
- [21] I. Shabirin, "Cluster Based News Representative Generation with Automatic Incremental Clustering", M.Eng. thesis, Politeknik Elektronika Negeri Surabaya, Surabaya, Indonesia, Jul. 2017.
- [22] A. R. Barakbah, K. Arai, "Reversed Pattern of Moving Variance for Accelerating Automatic Clustering", *EEPIS Journal*, vol. 2, no. 9, pp. 15-21, 2004.
- [23] T. Kohonen, *Self-Organizing Maps: Learning Vector Quantization*, ser. Springer Series in Information Sciences, Springer, Berlin, Heidelberg : Springer, 1995, vol 30.