[16] W. E. Yancey, “Evaluating string comparator performance for record linkage,” Statistical Research Division Research Report,2005.
[17] K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, “Duplicate record detection: A survey,” IEEE Transactions on Knowledge and Data Engineering, 19(1), pp.1-16, 2007.
[18] R. V. Bezu, S. Borst, R. Rijkse, J. Verhagen, D. Vandic, and F. Frasincar, “Multi-component similarity method for web product duplicate detection,” In Proceedings of the 30th Annual ACM Symposium on Applied Computing ACM, 2015 ,p. 761-768.
[19] D. Mrozek, B. Socha, S. Kozielski, and B. Małysiak-Mrozek, “An efficient and flexible scanning of databases of protein secondary structures,” Journal of Intelligent Information Systems, 46(1), 213-23, 2016.
[20] B. Khan, A. R. S. D. Shah, and S. Khusro, “Identification and Removal of Duplicated Records,” World Applied Sciences Journal, 13(5), pp.1178-1184, 2011.
[21] D. Koller and M. Sahami, “Hierarchically classifying documents using very few words,” Technical Report, Stanford InfoLab,1997.
[22] V. S. Verykios, A. K. Elmagarmid, and E. N. Houstis, “Automating the approximate record-matching process,” Information Sciences, 126(1), pp. 83-98, 2000.
[23] X. Wang, “Matching records in multiple databases using a hybridization of several technologies,” Master Dissertation. Department of Industrial Engineering. University of Louisville, KY, USA, 2008.
[24] C. Conrad, “Predicting Political Donations Using Data Driven Lifestyle Profiles Generated from Character N-Gram Analysis of Heterogeneous Online Sources,” Master of Electronic Commerce Thesis, Dalhousie University, Canada, 2015.
[25] J. S. Murray, “Probabilistic Record Linkage and Deduplication after Indexing, Blocking, and Filtering,” Journal of Privacy and Confidentiality, 7(1), pp.3–24, 2016.
[26] P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the EM algorithm.” Journal of the Royal Statistical Society. Series B (methodological), pp. 1-38, 1977.
[27] J. R. Wang and S. E. Madnick, “The inter-database instance identification problem in integrating autonomous systems.” In Proceedings of the Fifth International Conference on Data Engineering, 1989. p. 46-55.
[28] M. Kejriwal and D. P. Miranker, “On the Complexity of Sorted Neighborhood,”1501.01696, Cornell University, 2015.
[29] J. J. Tamilselvi and V. Saravanan, “Detection and elimination of duplicate data using token-based method for a data warehouse: A clustering based approach,” International Journal of Dynamics of Fluids, 5(2), pp. 145-164, 2009.
[30] F. N. Mahamood and A. Ismal, “Semantic Similarity Measurement Methods: State of Art,” Research Journal of Applied Sciences, Engineering and Technology, 26, pp. 415–430, 2014.
[31] S. Ramya and C. Palani Nehr, “An Efficient Duplicate Detection Based on Navie Block Detection Algorithm,” Middle-East Journal of Scientific Research 24 (Techniques and Algorithms in Emerging Technologies), pp. 291-296, 2016.