

## DAFTAR PUSTAKA

- Abedjan, Z., Golab, L., & Naumann, F. (2015). Profiling relational data: a survey. *VLDB Journal*, 24(4), 557–581. <https://doi.org/10.1007/s00778-015-0389-y>
- Abedjan, Z., & Naumann, F. (2016). CONFERENCE: Data Profiling, 1432–1435.
- Academy, K. (2016). Interquartile range review. Retrieved January 5, 2018, from <https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/interquartile-range-iqr/a/interquartile-range-review>
- Acronis, A., & Paper, W. (2009). How Deduplication Benefits Companies of All Sizes, 2000–2009.
- Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, 27(5), 1834–1839. <https://doi.org/10.1016/j.chb.2011.04.004>
- Bleuel, J. (2014a). Fuzzy match. Retrieved May 3, 2018, from <https://wiki.pentaho.com/display/EAI/Fuzzy+match>
- Bleuel, J. (2014b). String operations. Retrieved May 3, 2018, from <https://wiki.pentaho.com/display/EAI/String+operations>
- Bleuel, J. (2014c). Unique Rows (HashSet). Retrieved May 3, 2018, from [https://wiki.pentaho.com/display/EAI/Unique+Rows+\(HashSet\)](https://wiki.pentaho.com/display/EAI/Unique+Rows+(HashSet))
- Casters, M. (2014). Add sequence. Retrieved May 5, 2018, from <https://wiki.pentaho.com/display/EAI/Add+sequence>
- Casters, M. (2017a). Calculator. Retrieved May 5, 2018, from <https://wiki.pentaho.com/display/EAI/Calculator>
- Casters, M. (2017b). Select Values. Retrieved May 3, 2018, from <https://wiki.pentaho.com/display/EAI/Select+Values>
- Casters, M. (2017c). Sort rows. Retrieved May 5, 2018, from

<https://wiki.pentaho.com/display/EAI/Sort+rows>

Casters, M. (2017d). Table Input. Retrieved May 3, 2018, from <https://wiki.pentaho.com/display/EAI/Table+Input>

Casters, M. (2017e). Table Output. Retrieved May 3, 2018, from <https://wiki.pentaho.com/display/EAI/Table+Output>

Casters, M., & Moran, D. (2014). Add Constants. Retrieved May 5, 2018, from <https://wiki.pentaho.com/display/EAI/Add+Constants>

Corporation, O. (2009). Oracle ® Warehouse Builder User ' s Guide, 2(April), 1–1.

Cunningham, S. J., & Holmes, G. (1991). Developing innovative applications in agriculture using data mining. *The Proceedings of The*, 12. Retrieved from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Developing+innovative+applications+in+agriculture+using+data+mining#0>

Cupoli, P., Earley, S., Henderson, D., & Deborah Henderson. (2014). DAMA-DMBOK2 Framework, 26. Retrieved from <https://www.dama.org/sites/default/files/download/DAMA-DMBOK2-Framework-V2-20140317-FINAL.pdf%5Cnhttp://www.dama.org/content/body-knowledge%5Cn>

Dasu, T., Loh, J. M., & Srivastava, D. (2014). Empirical glitch explanations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '14*, 572–581. <https://doi.org/10.1145/2623330.2623716>

Dwiandriani, F. (2017). Fingerprint Clustering Algorithm for Data Profiling using Pentaho Data Integration, 358–362.

Dye, J. F., Schatz, I. M., Sout, N., Rosenberg, B. A., & Coleman, S. T. (2000). Constant Comparison Method: A Kaleidoscope of Data. *The Qualitative Report*, 4(1), 1–10. <https://doi.org/10.1017/CBO9781107415324.004>

- Fazel Famili, W. M. Shen, R. Weber, E. S. (1997). Data Pre-Processing and Intelligent Data Analysis. *International Journal on Intelligent Data Analysis*, 18(6), 1–29.
- Fletcher, D. (n.d.). Class InterquartileRange. Retrieved May 23, 2018, from <http://weka.sourceforge.net/doc.dev/weka/filters/unsupervised/attribute/InterquartileRange.html>
- Gueddah, H., Yousfi, A., & Belkasmi, M. (2016). The filtered combination of the weighted edit distance and the Jaro-Winkler distance to improve spellchecking Arabic texts. *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA, 2016–July*, 1–6. <https://doi.org/10.1109/AICCSA.2015.7507128>
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. San Francisco, CA, *itd: Morgan Kaufmann*. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Hassan, A. (2015). Technique using Novel Modified Levenshtein Distance, 204–209.
- Horn, P., Feng, L., Li, Y. M., & Pesce, A. (2002). Effect of Outliers and Nonhealthy Individuals on Reference Interval Estimation. *Clinical Chemistry*, 47, 2137–2145.
- Ikegami, Y., Damiani, E., & Knauf, R. (2017). Flick : Japanese Input Method Editor using N-gram and Recurrent Neural Network Language Model based Predictive Text Input. <https://doi.org/10.1109/SITIS.2017.19>
- Kusumasari, T. F. (2016). Data Profiling for Data Quality Improvement with Openrefine.
- Lehmann, C., Roy, K., & Yo, N. E. (2016). The State of Enterprise Data Quality: 2016 Perception, Reality and the Future of DQM, 1–21. Retrieved from [http://pages.blazent.com/rs/184-CZE-628/images/Blazent\\_State\\_of\\_DataQuality\\_2016.pdf](http://pages.blazent.com/rs/184-CZE-628/images/Blazent_State_of_DataQuality_2016.pdf)

- Matzan, J. (2011). .03 Database Connections. Retrieved May 3, 2018, from <https://wiki.pentaho.com/display/EAI/.03+Database+Connections>
- Naumann, F. (2014). Data profiling revisited. *ACM SIGMOD Record*, 42(4), 40–49. <https://doi.org/10.1145/2590989.2590995>
- Noeng, M. (2000). Metodologi Penelitian Kualitatif. *Rake Sarasin*.
- Olson, J. E. (2003). *Data Quality—The Accuracy Dimension*. Elsevier Science. <https://doi.org/10.1525/jer.2010.5.3.toc>
- Rizza, E. (2011). Recipes Open Refine. Retrieved May 21, 2018, from <https://github.com/OpenRefine/OpenRefine/wiki/Recipes>
- Sai, V., Pulla, V., Varol, C., & Al, M. (2016). Information Technolog: New Generations, 448, 893–902. <https://doi.org/10.1007/978-3-319-32467-8>
- Solberg, H. E., & Lahti, A. (2005). Detection of outliers in reference distributions: Performance of horn’s algorithm. *Clinical Chemistry*, 51(12), 2326–2332. <https://doi.org/10.1373/clinchem.2005.058339>
- Stake, R. E. (1995). *The Art of Case Study Research*. SAGE Publications. Retrieved from <https://books.google.co.id/books?id=ApGdBx76b9kC>
- Walk, K. (1998). How to Write a Comparative Analysis. Retrieved June 12, 2017, from <https://writingcenter.fas.harvard.edu/pages/how-write-comparative-analysis>
- Walsh, C. (2016a). Data and Analytics: Open Source Data Integration Tool Comparison. Retrieved from [https://www.excella.com/wp-content/uploads/2016/03/Open-Source-DI-Tool-Comparison\\_March2016.pdf](https://www.excella.com/wp-content/uploads/2016/03/Open-Source-DI-Tool-Comparison_March2016.pdf)
- Walsh, C. (2016b). Data and Analytics: Open Source Data Integration Tool Comparison.
- Weka. (n.d.). Class InterquartileRange. Retrieved May 1, 2018, from <http://weka.sourceforge.net/doc.dev/weka/filters/unsupervised/attribute/Inter>

quartileRange.html

Yulfitri, A. (2016). Modeling Operational Model of Data Governance in Government.

Zhang, J. (2013). Advancements of Outlier Detection: A Survey. *ICST Transactions on Scalable Information Systems*, 13(1), e2. <https://doi.org/10.4108/trans.sis.2013.01-03.e2>