

Department	School of Computing
Supervisors	Dr. Taoxin Peng and Associate Prof. Alistair Lawson
Project Title	Data Quality and Cleaning in Big Data
<p>PROJECT DESCRIPTION</p> <p>The proliferation of Big Data analytics is developing ever more sophisticated models for intelligent data-driven insight and decision making in business and in other areas such as health and social care. However critical issues relating to the data quality that is required for these models to be effective and trustworthy are not getting the attention they deserve. This project will investigate data quality and data cleaning in Big Data, focusing on how the characteristics of Big Data affect the suitability of existing data quality/data cleaning approaches. The successful candidate will be expected to undertake research into current data quality approaches, and then propose and evaluate a novel data quality approach/framework, which can be used in Big Data applications. The area of applications, such as banking, retail, manufacturing, internet of things, or health and social care will be for the successful candidate to determine in conversation with the supervisors.</p> <p>Prospective applicants are encouraged to contact the Supervisor before submitting their applications. Applications should make it clear the project you are applying for and the name of the supervisor(s).</p> <p>Academic qualifications</p> <p>A first degree (at least a 2.1) ideally in Mathematics or Computing with a good fundamental knowledge of Data Science and Algorithms.</p> <p>English language requirement</p> <p>IELTS score must be at least 6.5 (with not less than 6.0 in each of the four components). Other, equivalent qualifications will be accepted. Full details of the University's policy are available online.</p> <p>Essential attributes:</p> <ul style="list-style-type: none"> • Experience of fundamental database applications • Competent in data structures and algorithms • Knowledge of data science • Good written and oral communication skills • Strong motivation, with evidence of independent research skills relevant to the project • Good time management <p>Desirable attributes:</p> <p>A basic understanding of data quality and data cleaning would be beneficial.</p>	
Indicative Bibliography	<p>A. Immonen, P. Paakkonen and E. Ovaska, Evaluating the Quality of Social Media Data in Big Data Architecture, IEEE Access 2015, Vol. 3</p> <p>C. Batini and M. Scannapieco, Data and Information Quality: Dimensions, Principles and Techniques, Springer, 2016.</p> <p>H. Liu, A. Kumar T.K., J. P. Thomas and X. Hou, Cleaning Framework for BigData: An Interactive Approach for Data Cleaning, 2016 IEEE Second International Conference on Big Data Computing Service and Applications (BigDataService), 2016, pp. 174-181, doi: 10.1109/BigDataService.2016.41.</p> <p>F. Ridzuan and Z. Wan, A review on data cleansing methods for big data. Procedia Comput Sci. 2019, doi.org/10.1016/j.procs.2019.11.177</p>

	X. Wang and C. Wang, "Time Series Data Cleaning: A Survey," in IEEE Access, vol. 8, pp. 1866-1881, 2020, doi: 10.1109/ACCESS.2019.2962152 Mayur Kishor Shende, Andrés E. Feijóo-Lorenzo, Neeraj Dhanraj Bokde, cleanTS: Automated (AutoML) tool to clean univariate time series at microscales, Neurocomputing, Volume 500, 2022, Pages 155-176
Enquiries	For informal enquiries about this PhD project, please contact Dr. T. Peng at t.peng@napier.ac.uk
Web page	https://www.napier.ac.uk/research-and-innovation/research-degrees/application-process