Daniele Foroni | Ph.D.

📞 *+39 348 068 7589* • ✉ *daniele.foroni@gmail.com* • 🌐 *forons.github.io* • 🐦 *forons*
○ *forons* • in *daniele.foroni*

My research interests focus on Big Data Management and Analysis, particularly on the assessment of the quality of the data and to analysis considering sport data. During my PhD, I studied how to improve the quality of the data by considering the involved application and the optimization metric the user has in mind, as well as the intrinsic value of the data. To do so, I built F4U, a framework for measuring the impact of some noise on the accuracy of the results, and Moira, a system to dynamically allocate the right amount of resource to optimize the goal metric defined by the user.

Current position

From 2019 August	**Big Data Senior Research Engineer**, *Huawei ERC (European Research Center)*, Munich, Germany. Research and develop Big Data technologies with special focus on real-time and streaming technologies, and techniques for traffic management optimization in smart city environments.

Education

2014–2019	**PhD in Computer Science**, *University of Trento*, Italy. Thesis: *Putting Data Quality in Context: How to generate more accurate analyses.* Advisor: Yannis Velegrakis
2012–2014	**Master of Science in Computer Science**, *University of Trento*, Italy, *102/110*.
2009–2012	**Bachelor of Science in Computer Science**, *University of Trento*, Italy, *97/110*.

Work Experience

2018 Oct–Present	**CTO and Developer**, *IOHub*, Rovereto, Italy. Development of technologies and solutions for financial invoices for Italian businesses and citizens, providing an integration with the ad-hoc system created by the Italian Revenue Agency.
2017–2018	**PhD Intern**, *Huawei ERC*, Munich, Germany. Visiting researcher working on dynamic rescheduling resources for streaming queries in a distributed environment based on the goal metric defined by the user. Advisors: Stefano Bortoli and Radu Tudoran
2016–2017	**Data Scientist**, *University of Trento and Saidea SRL*, Trento, Italy. Development of XplainBox, an industrial project that from a stream of data, stores it, and then analyzes and applies ML algorithms on it.
2014 Feb–Nov	**Intern Data Engineer**, *Engineering Tributi SPA*, Trento, Italy. Development of a tool for resolving similar entities in a dataset obtained integrating multiple sources from the public administrative domain. Given a set of rules, the proposed solution builds a lattice of operations that efficiently and effectively finds similar entities without running several algorithms over the whole dataset. Advisor: Nicola Mezzetti

Teaching Experience

2015–2019	**TA, Big Data and Social Networks**, *University of Trento*, Trento, Italy. Computer Science and Data Science Master course about big data technologies that covers from the storage of such an amount of data (NoSQL, HDFS) to its analysis (Hadoop MapReduce, Spark, ML models, Streaming, Graph processing). Professor: Yannis Velegrakis
2017	**Big Data Technologies**, *TIM*, Trento, Italy. On site and from remote course about Big Data Technologies for scientists working in TIM. The program covered from the storage of such an amount of data (NoSQL, HDFS) to its analysis (Hadoop MapReduce, Spark, ML models, Streaming, Graph processing).
2015–2016	**TA, Programming for Mathematicians**, *University of Trento*, Trento, Italy. Bachelor course for introducing Math students to programming. Professor: Roberto Zunino

Skills

Programming	Python · Java · Scala · C# · SQL · Javascript
Framework	Apache Spark · Flink · Kafka · Hadoop · NoSQL
Certifications	Scalable Machine Learning with Apache Spark
OS & Tools	Linux · Bash · Git
AI	Machine Learning · Data Mining

Publications

2019 *Putting Data Quality in Context*
D. Foroni, M. Lissandrini, and Y. Velegrakis
Under review in TKDE

2019 *F4U: The Contextualized Data Quality System*
D. Foroni, M. Lissandrini, and Y. Velegrakis
Under review in VLDB

2019 *On Contextualizing Data Quality*
D. Foroni, M. Lissandrini, and Y. Velegrakis
Submitted to Sigmod Record

2018 *Moira: A Goal-Oriented Incremental Machine Learning Approach to Dynamic Resource Cost Estimation in Distributed Stream Processing Systems*
D. Foroni, C. Axenie, S. Bortoli, M. Al Hajj Hassan, R. Acker, R. Tudoran, G. Brasche, and Y. Velegrakis
BIRTE '19
doi:10.1145/3242153.3242160

2018 *STARLORD: Sliding Window Temporal Accumulate-Retract Learning for Online Reasoning on Datastreams*
C. Axenie, R. Tudoran, S. Bortoli, M. Al Hajj Hassan, **D. Foroni**, and G. Brasche
ICMLA '18
doi:10.1109/ICMLA.2018.00181

2018 *Data Management and Smart Cities*
M. Brugnara, C. Consonni, **D. Foroni**, S. Pasupathipillai, G. Preti, P. Sottovia, and Y. Velegrakis
UniTN - White Paper '18
Trento - Smart Cities

Projects

2019 **Exploit data information for Entity Matching**, *Ongoing work*.
Exploit the intrinsic value of the data through a profile to improve the accuracy of the duplicates found.

2017 **Data Quality Measurement**, *Ongoing work*.
Providing the user an estimation of the quality of a dataset for a specific application.

2017 **Creation of analysis from a data lake with Data Mining and ML algorithms, and statistics**, *Industrial work*.
Designed a flow from the reading of the data to the analysis, which is performed with Data Mining and Machine Learning algorithms to provide the user the key insights from the available data.

Other activities

PC member	SEA Data 2020
Reviewer	CSGRID · EDBT · JPDC · ICMLA · VLDB · KDD - SIGMOD Record

Languages

English	Full professional proficiency (C1)
German	Basic knowledge (A1)
Italian	Native