LUCA ZECCHINI

PhD Student at the University of Modena and Reggio Emilia

Place and date of birth: Pavullo nel Frignano (Modena, Italy), August 6, 1994 · **Citizenship:** Italian
City of residence: Lama Mocogno (Modena, Italy) with a support apartment in Modena (Italy)
Office address: Via Pietro Vivarelli, 10 – 41125 Modena (Italy) – Building 27, 1st Floor, Room MO27-01-020
Telephone number: +39 333 2821524 · **Skype ID:** live:zecchini.luca
Email address: zecchini.luca@libero.it (personal address), luca.zecchini@unimore.it (institutional address)
Web page: http://personale.unimore.it/rubrica/dettaglio/190439

I am a second-year PhD student at the University of Modena and Reggio Emilia, where I got my Master's Degree in Computer Engineering in 2020. My research, supervised by Prof. Sonia Bergamaschi and Dr. Giovanni Simonini, deals with the broad and challenging theme of big data integration, with a main interest in data preparation and cleaning, especially in the task of entity resolution. Right on this topic, I recently achieved my first publication in a top venue (VLDB 2022) and I was runner-up at the SIGMOD 2020 Programming Contest (then also part of the organizing team for the two following editions). Finally, I had the opportunity to spend a month as a visiting student at the Hasso Plattner Institute (Potsdam, Germany), to carry out my research activities under the supervision of Prof. Felix Naumann.

EDUCATION

Ongoing — **PhD in INFORMATION AND COMMUNICATION TECHNOLOGIES (ICT)**
DBGroup @ University of Modena and Reggio Emilia (UniMoRe) - "Enzo Ferrari" Engineering Department, Modena (Italy)
Curriculum: Computer Engineering and Science · Research Topic: Task-Driven Big Data Integration
Advisor: Prof. Sonia Bergamaschi · Co-Advisor: Dr. Giovanni Simonini

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

- **Visiting Student @ Hasso Plattner Institute (HPI), University of Potsdam**, Potsdam (Germany)
1 month (from May to June 2022) · Information Systems Group · Supervisor: Prof. Felix Naumann

- **6th International School on Deep Learning (DeepLearn 2022 Summer)**, Las Palmas de Gran Canaria (Spain)
1 week (July 2022)

July 16, 2020 — **MASTER'S DEGREE in COMPUTER ENGINEERING** (110/110 cum laude)
University of Modena and Reggio Emilia (UniMoRe) - "Enzo Ferrari" Engineering Department, Modena (Italy)
Curriculum: Intelligent Cyber Systems · Advisor: Prof. Sonia Bergamaschi · Co-Advisor: Dr. Giovanni Simonini
Thesis: RICONCILIAZIONE PROGRESSIVA DI ENTITÀ GUIDATA DALLE INTERROGAZIONI (PAY-AS-YOU-GO QUERY-DRIVEN ENTITY RESOLUTION)

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

- **Internship @ DBGroup (UniMoRe)**
6 months (from January to June 2020) · Supervisor: Prof. Sonia Bergamaschi · Co-Supervisor: Dr. Giovanni Simonini

October 24, 2017 — **BACHELOR'S DEGREE in COMPUTER ENGINEERING** (99/110)
University of Modena and Reggio Emilia (UniMoRe) - "Enzo Ferrari" Engineering Department, Modena (Italy)
Advisor: Prof. Rita Cucchiara · Co-Advisor: Riccardo Gasparini · Thesis: GRAPHICS PROCESSING UNITS

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

June 27, 2013 — **HIGH SCHOOL DIPLOMA in FOREIGN LANGUAGES** (100/100)
Liceo Linguistico @ Istituto Magistrale Paritario "Francescane dell'Immacolata", Palagano (Modena, Italy)
Studied Languages: English, French, German · Best results in the history of the high school (from 1992/1993 to 2012/2013)

- **Internship @ C.M.F. Technology S.p.A.**, Pavullo nel Frignano (Modena, Italy) · 1 month (from May to June 2011)
- **1 Week Junior English Cultural Course** by **Language in Dublin**, Dublin (Ireland) · 1 week (March 2012)
- **European Computer Driving Licence (ECDL)** achieved in 2011
- **Preliminary English Test (PET)** (B1 English Qualification by the University of Cambridge) achieved in 2012

TEACHING APPOINTMENTS

- **BASI DI DATI E LAB.** (DATABASES AND LAB.)
Bachelor's Degree in Computer Engineering, University of Modena and Reggio Emilia - "Enzo Ferrari" Engineering Department, Modena (Italy)
Language: Italian · Teaching Assistant: 2020/2021, 2021/2022
- **BIG DATA MANAGEMENT AND GOVERNANCE**
Master's Degree in Computer Engineering, University of Modena and Reggio Emilia - "Enzo Ferrari" Engineering Department, Modena (Italy)
Language: English · Teaching Assistant: 2021/2022
- **HUMAN RESOURCE INFORMATION SYSTEMS E DATA SCIENCE** (HUMAN RESOURCE INFORMATION SYSTEMS AND DATA SCIENCE)
Master's Degree in Labor Relations, University of Modena and Reggio Emilia - "Marco Biagi" Economics Department, Modena (Italy)
Language: Italian · Teaching Assistant: 2020/2021, 2021/2022

SERVICES FOR THE SCIENTIFIC COMMUNITY

- Member of the organizing team for the **SIGMOD 2021 Programming Contest**, in collaboration with the **Roma Tre University** (Rome, Italy), and for the **SIGMOD 2022 Programming Contest**, in collaboration with the **Georgia Institute of Technology** (Atlanta, Georgia, USA)
- Subreviewer (external reviewer) for the following journals:
  - **TKDE** (IEEE Transactions on Knowledge and Data Engineering): 2022, 2021

- Subreviewer (external reviewer) for the following conferences and workshops:
  - **VLDB** (International Conference on Very Large Data Bases): 2023 (Research Papers), 2022 (Application Papers)
  - **ICDE** (IEEE International Conference on Data Engineering): 2023 (Research Papers), 2022 (Research Papers), 2021 (Demo Papers)
  - **SEA Data** (International Workshop on Search, Exploration, and Analysis in Heterogeneous Datastores): 2021 (@ VLDB)
- Subreviewer (external reviewer) for two calls for funding applications by the **Academy of Finland** (2020)

RESEARCH PROJECTS

- **BrewER: Entity Resolution On-Demand** (since 2020)
  Entity Resolution (ER) plays a fundamental role in data integration processes, aiming to identify and merge records that refer to the same real-world entity. ER is typically employed as an expensive offline cleaning step performed on the entire data before consuming it. This batch approach constitutes a significant limitation when the task at hand is characterized by an information need and/or time constraints (e.g., dealing with Web data or ELT pipelines in data lakes). BrewER is a framework which allows to evaluate SQL SP queries on dirty data while progressively returning results as if they were issued on cleaned data, addressing the issues of traditional batch solutions through its query-driven and progressive approach, respectively guided by the WHERE clauses and the ORDER BY clause of the performed query.
  - References: **R1** (Research Paper @ VLDB 2022)

Application Projects:

- **DXP: Digital Experience Platform** (since 2020)
  DXP manages the billing data of the users of different companies operating in several areas (electricity and gas, telephony, etc.). The goal of the platform is to acquire the billing data provided by these companies, then to process it in order to produce analytics for the companies (e.g., churn prediction or customer segmentation) and reports for their users (e.g., interactive billing). The project, commissioned and supervised by Doxee, is funded by Regione Emilia-Romagna.

- **ECDP: Energy Community Data Platform** (from 2020 to 2021)
  ECDP is a middleware platform designed to collect and analyze big data about the energy consumption and production inside local energy communities, with the aim of encouraging a more conscious use of energy by the users, at home and in the workplace. ECDP acquires data of different nature in a heterogeneous format from multiple sources. Its modular architecture, designed to support a data integration workflow and a data lake workflow, guarantees flexibility and scalability, allowing the applicability of ECDP to any type of local energy community. The project is funded by the Italian Ministry of Economic Development (MISE) and carried out in collaboration with DataRiver, under the supervision of the Italian National Agency for New Technologies, Energy, and Sustainable Economic Development (ENEA).
  - References: **A1** (Application Paper @ DataPlat 2022)

AWARDS AND GRANTS

2021   **SISAP 2021 Best Doctoral Symposium Paper Award** (with a prize of 500 EUR by Springer)
References: **P1** (PhD Paper @ SISAP 2021)

2020   Three-year scholarship for the PhD program by the Italian Ministry of University and Research (MUR) (2020/2023)

          SIGMOD 2020 Programming Contest Runner-Up (with honorable mention and a prize of 3000 USD by Megagon Labs and Microsoft)
References: **E1** (Experience Paper @ DI2KG 2020)

2013   Grant by the Istituto Magistrale Paritario "Francescane dell'Immacolata" for achieving the best results in the history of the high school (from 1992/1993 to 2012/2013)

2010   Grant by the Lions Club di Pavullo e del Frignano for the results achieved during the high school

2008   Grant by the Lions Club di Pavullo e del Frignano for the results achieved during the junior high school

2007   Qualification for the Italian finals of the International Competition for Mathematical and Logical Games (held at the Bocconi University)

2005   Seventh place in the Italian finals of the Mathematical Kangaroo (held in Mirabilandia)

OTHER WORK EXPERIENCES

2018 – 2022   **WAITER AND BARTENDER @ La Vecchia Scuola**, Lama Mocogno (Modena, Italy)
Traditional restaurant, pizzeria, and bar
From July to September 2018; from November 2018 to February 2022

2011 – 2015   **TRANSLATOR @ C.M.F. Technology S.p.A.**, Pavullo nel Frignano (Modena, Italy)
Manufacturer of machinery for the ceramic industry
From May to June 2011 (internship); from August to September 2011; July 2012; from July to August 2013; from January to February 2014; from August to September 2015

VOLUNTARY EXPERIENCES

2004 – Ongoing   **CLARINET PLAYER @ Corpo Bandistico "Gioacchino Rossini"**, Lama Mocogno (Modena, Italy)
2016 – 2022   **PRESIDENT @ Corpo Bandistico "Gioacchino Rossini"**, Lama Mocogno (Modena, Italy)
Playing alto clarinet in E-flat (previously sopranino clarinet in E-flat)

2013 – 2017   **ASSISTANT YOUTH SOCCER COACH**
A.C. Lama '80, Lama Mocogno (Modena, Italy), Category: Allievi CSI (from 2013 to 2015)
A.S.D.C. Polinago, Polinago (Modena, Italy), Category: Juniores CSI (from 2015 to 2017)

PUBLICATIONS → ORCID: 0000-0002-4856-0838

Research Papers:

R1 ⊛   Giovanni Simonini[1]⊛, <u>Luca Zecchini</u>[1], Sonia Bergamaschi[1], Felix Naumann[4]: **ENTITY RESOLUTION ON-DEMAND**. **VLDB 2022** (48th International Conference on Very Large Data Bases), Proceedings of the VLDB Endowment (PVLDB), vol. 15, no. 7, pp. 1506-1518 (2022). **WWW**

Application Papers:

A1   Luca Gagliardelli[1]⊛, <u>Luca Zecchini</u>[1], Domenico Beneventano[1], Giovanni Simonini[1], Sonia Bergamaschi[1], Mirko Orsini[2], Luca Magnotta[2], Emma Mescoli[2], Andrea Livaldi, Nicola Gessa[3], Piero De Sabbata[3], Gianluca D'Agosta[3], Fabrizio Paolucci[3], Fabio Moretti[3]: **ECDP: A BIG DATA PLATFORM**

FOR THE SMART MONITORING OF LOCAL ENERGY COMMUNITIES. **DataPlat @ EDBT/ICDT 2022** (1st International Workshop on Data Platform Design, Management, and Optimization), CEUR Workshop Proceedings, vol. 3135, CEUR-WS.org (2022). **WWW**

Experience Papers:

E1   Luca Zecchini[1]®, Giovanni Simonini[1], Sonia Bergamaschi[1]: **ENTITY RESOLUTION ON CAMERA RECORDS WITHOUT MACHINE LEARNING**. **DI2KG @ VLDB 2020** (2nd International Workshop on Challenges and Experiences from Data Integration to Knowledge Graphs), CEUR Workshop Proceedings, vol. 2726, CEUR-WS.org (2020). **WWW**

Discussion Papers:

O1   Giovanni Simonini[1]®, Henrique Saccani[1], Luca Gagliardelli[1], Luca Zecchini[1], Domenico Beneventano[1], Sonia Bergamaschi[1]: **THE CASE FOR MULTI-TASK ACTIVE LEARNING ENTITY RESOLUTION**. **SEBD 2021** (29th Italian Symposium on Advanced Database Systems), CEUR Workshop Proceedings, vol. 2994, pp. 363-370, CEUR-WS.org (2021). **WWW**

O2   Giovanni Simonini[1]®, Luca Gagliardelli[1], Michele Rinaldi[1], Luca Zecchini[1], Giulio De Sabbata[1], Adeel Aslam[1], Domenico Beneventano[1], Sonia Bergamaschi[1]: **PROGRESSIVE ENTITY RESOLUTION WITH NODE EMBEDDINGS**. **SEBD 2022** (30th Italian Symposium on Advanced Database Systems), CEUR Workshop Proceedings, vol. 3194, pp. 52-60, CEUR-WS.org (2022). **WWW**

PhD Papers:

P1   Luca Zecchini[1]®: **PROGRESSIVE QUERY-DRIVEN ENTITY RESOLUTION**. **SISAP 2021** (14th International Conference on Similarity Search and Applications), Lecture Notes in Computer Science (LNCS), vol. 13058, pp. 395-401, Springer (2021). **WWW**
       SISAP 2021 Best Doctoral Symposium Paper

P2   Luca Zecchini[1]®: **TASK-DRIVEN BIG DATA INTEGRATION**. **SEBD 2022** (30th Italian Symposium on Advanced Database Systems), CEUR Workshop Proceedings, vol. 3194, pp. 627-632, CEUR-WS.org (2022). **WWW**

Research Group Activity Surveys:

G1   Domenico Beneventano[1], Sonia Bergamaschi[1]®, Luca Gagliardelli[1], Giovanni Simonini[1], Luca Zecchini[1]: **BIG DATA INTEGRATION & DATA-CENTRIC AI FOR EHEALTH**. **Ital-IA 2022** (2° Convegno Nazionale CINI sull'Intelligenza Artificiale) (2022). **WWW**

® = author presenting the paper at the conference/workshop
Author affiliations: (1) University of Modena and Reggio Emilia; (2) DataRiver; (3) ENEA; (4) Hasso Plattner Institute, University of Potsdam.

SUPERVISED STUDENTS

Bachelor's Degree Students:

2021   **Cristina Ventilati** (Co-Advisor): DATA PREPARATION E WORKFLOW MANAGEMENT CON KNIME. Bachelor's Degree in Computer Engineering, University of Modena and Reggio Emilia - "Enzo Ferrari" Engineering Department, Modena (Italy)
       Enrico De Luca (Co-Advisor): ANALISI DEL TOOL PER LA PREPARAZIONE DEI DATI OPENREFINE. Bachelor's Degree in Computer Engineering, University of Modena and Reggio Emilia - "Enzo Ferrari" Engineering Department, Modena (Italy)

2020   **Lisa Trigiante** (Co-Advisor): STUDIO E SPERIMENTAZIONE DI UN ALGORITMO PER L'ENTITY MATCHING BASATO SULLE ESPRESSIONI REGOLARI. Bachelor's Degree in Computer Engineering, University of Modena and Reggio Emilia - "Enzo Ferrari" Engineering Department, Modena (Italy)

I hereby declare that all the above information is correct and accurate. I authorize the processing of personal data contained in my curriculum vitae based on art. 13 GDPR (EU Regulation 2016/679).