

Federico Maria Scafoglieri, PhD

Department of Computer Engineering

University of Rome La Sapienza

Summary

Federico Maria Scafoglieri is an Adjunct Professor and Postdoctoral researcher at the University of Rome La Sapienza. He belongs to the Artificial Intelligence & Knowledge Representation research group. He is also an AI Research Scientist at Commissione Nazionale per le Società e la Borsa (CONSOB). During his Ph.D he spent a time abroad as research scholar at IBM Research Almaden working on Ontology-Based (Knowledge Graph-Based) Information Extraction and Entity Resolution.

Main Expertise: Knowledge Representation and Automated Reasoning, Semantic Technologies, Ontology Based Data Access, Ontology Based Data Integration, Rule-based Information Extraction, Entity Linking/Resolution, Formal Methods, Process Mining.

Education

- **University of Rome La Sapienza** Rome, Italy
PhD, Engineering in Computer Science - Artificial Intelligence and Knowledge Representation 2017 -- 2020
Supervisor: Prof. Domenico Lembo Co-Advisors: Prof. Maurizio Lenzerini & Prof. Massimo Mecella
 - Ontology Mediated Information Extraction: I proposed a new paradigm called Ontology Mediated Information Extraction (OMIE). OMIE is a novel paradigm for Information Extraction that seeks to exploit the semantic knowledge expressed in ontologies to improve query answering over unstructured data (properly raw text). I also proposed a formal framework to study the problem of entity resolution in the context of knowledge representation.
The two External reviews of my dissertation identified my accomplishments in the top 5% (first rev.) and top 25% (second rev.)
- **University of Rome La Sapienza** Rome, Italy
Master's Degree, Engineering in Computer Science (full marks) 2013 -- 2016
Supervisor: Prof. Domenico Lembo
- **University of Rome La Sapienza** Rome, Italy
Bachelor's degree, Engineering in Computer Science 2009 -- 2012

Appointments

- **Commissione Nazionale per le Società e la Borsa (CONSOB)** Rome, Italy
AI Research Scientist / Research Scholar April 2020 - Present
Collaborator: Prof. Francesca Medda
 - Automatic Information Extraction from Key Information Documents : Principal Investigator of the Financial Markets Project (Intermediaries Division) for the automation of information extraction from the Key Information Documents (KIDs), through artificial intelligence techniques, for the purpose of surveillance and data analysis activities. The implemented system, which combines machine based and learning based techniques, allows a significant reduction in the workload of CONSOB operators, otherwise unsustainable through manual extractions (1 million documents per years).
- **University of Rome La Sapienza** Rome, Italy
Postdoctoral Researcher Sept 2021 - Present
 - Ontology-Based OLAP: As part of the Advanced Digital Solutions for Professional Food and Nutrition Catering Service project, I worked on a formal framework and implementation of a new OLAP system from the perspective of semantic technologies. H2020 ADCATER Project
 - Process Mining for Database event extraction: Process Mining by exploiting transactional logs of DBMSs.
- **IBM's Almaden Research Center** (partially remotely) San Jose, California, USA
Research Scholar Oct 2019 - Oct 2020
Mentors: Dr. Yunyao Li, Dr. Lucian Popa
 - Ontology-Mediated Information Extraction: I developed an Ontology-Mediated Information Extraction tool based on the information extraction tool System-T and the Ontology-based data access tool MASTRO. I won the best demo paper at ISWC2020 for the related publication.
 - Entity Resolution: I worked on a novel formal framework for a joint Entity Resolution and Query Answering over semantically rich and complex knowledge representation systems.
- **OBDA Systems** Rome, Italy
Research & Development Engineer March 2017 - March 2020
 - Mastro Studio: I was the team leader and lead developer for the Mastro Studio project. Mastro Studio is a web-based platform for data management through the OBDA paradigm. We built an end-to-end system for the open data publishing through the OBDA paradigm for the Automobile Club d'Italia (ACI). Link to ACI linked Open Data Portal
- **PricewaterhouseCoopers (PWC)** Rome, Italy
Technology Consultant Sept 2016 -- March 2017

Grants

- IBM Research Scholar Grant
- Mobility Grant for PhD Student 2019/2020
- Mobility Grants for various conferences and Summer Schools.

Awards

- Best Demo ISWC 2020

Academic Teaching

• University of Rome La Sapienza	Rome, Italy
<i>Professor of Programming Language (Management Engineering)</i>	2023 - 2024
<i>Professor of Programming Language (Management Engineering)</i>	2022 - 2023
<i>Teaching Assistant of Programming Language I (Engineering in Computer Science)</i>	2020 - 2021
<i>Professor of Multimedia Design II (Architecture)</i>	2018 - 2019
<i>Teaching Assistant of Programming Language (Management Engineering)</i>	2017 - 2018

Thesis Supervisor

Master Thesis Supervision

- Giulia Neccia, t.b.d. (topic: Entity Resolution) 2023 (Excellence Path)
- Gabriele MarguMi, Ontology driven operational and analytical services for ADCATER project, March 2023
- Alessandra Monaco, From Annotation to Data Cleaning, an End to End rule-based and learning-based approach for Information Extraction from Financial documents, March 2022
- Robefio Aducci, Ruled-based Information Extraction: An experience and a tool for evaluating document spanners. October 2019

Bachelor Thesis Supervision

- Giulia Neccia, Classification of Financial Documents through Neural Networks, The CONSOB Experience. July 2021.

Main Collaborators

Ontology Based Data Management & Information Extraction

- Prof. Francesca Medda, University College London (UCL)
- Prof. Maurizio Lenzerini, University of Rome La Sapienza
- Prof. Domenico Lembo, University of Rome La Sapienza
- Dr. Yunyao Li, Head of Machine Learning, Apple Knowledge, Apple
- Dr. Lucian Popa, Principal Research Staff Member, Research Manager, IBM Almaden Research Center
- Dr. Valerio Santarelli, CTO OBDA-System Almadawave
- Dr. Kun Qian, Senior Machine Learning Researcher and Engineer, Apple

Entity Resolution

- Dr. Ronald Fagin, IBM Fellow, IBM Almaden Research Center
- Prof. Phokion Kolaitis, University of Santa Cruz
- Dr. Lucian Popa, Principal Research Staff Member, Research Manager, IBM Almaden Research Center
- Prof. Domenico Lembo, University of Rome La Sapienza

Process Mining

- Prof. Andrea Marrella, University of Rome La Sapienza
- Prof. Massimo Mecella, University of Rome La Sapienza
- Dr. Simone Agostinelli, University of Rome La Sapienza

Program CommiMee

- Reviewer ACM Journal of Data and Information Quality 2023
- In the program commiMee of the 21th International Conference on Principles of Knowledge Representation and Reasoning (KR 2023)
- Reviewer 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023)
- Reviewer Elsevier Journal of Computer in Industry 2022
- Reviewer IEEE Journal of Transaction on Services Computing 2022
- In the program commiMee of the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022)
- Reviewer Journal of Transaction on Services Computing 2021
- Reviewer of the 24th International Conference on Database Theory (ICDT 2021)
- Reviewer of the 2019 IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE 2019)
- Reviewer of the Russian Federation & Europe Multidisciplinary Symposium on Computer Science and ICT (REMS 2018)

Publications

Authors in alphabetical order as the norm in theoretical computer science.

- A Framework for Combining Entity Resolution and Query Answering in Knowledge Bases, Ronald Fagin, Phokion Kolaitis, Domenico Lembo, Lucian Popa and Federico Scafoglieri. SubmiMed to International Conference 2023.
- Automatic Information Extraction from Financial Documents for Market Supervision, Domenico Lembo, Maurizio Lenzerini, Alessandra Limosani, Francesca Medda, Alessandra Monaco, Giulia Neccia, Federico Maria Scafoglieri. AAAI 2023 (FinST).
- Automatic Information Extraction from Investment Product Documents, Domenico Lembo, Alessandra Limosani, Francesca Medda, Alessandra Monaco, Giulia Neccia and Federico Scafoglieri. SEBD 2022
- Information Extraction through AI techniques: The KIDs use case at CONSOB, Domenico Lembo, Alessandra Limosani, Francesca Medda, Alessandra Monaco and Federico Scafoglieri. Ital-IA 2022
- Boosting Information Extraction through Semantic Technologies: The KIDs use case at CONSOB. Domenico Lembo, Maurizio Lenzerini, Alessandra Limosani, Francesca Medda, Federico Scafoglieri. ISWC 2021
- Ontology Mediated Information Extraction with MASTRO SYSTEM-T. Domenico Lembo, Yunyao Li, Lucian Popa, Kun Qian, Federico Scafoglieri. ISWC 2020 (Best Demo Award)
- Domenico Lembo, Yunyao Li, Lucian Popa, Federico Scafoglieri. Ontology mediated information extraction in financial domain with Mastro System-T. DSMM SIGMOD/PODS 2020
- Ontology-based Document Spanning Systems for Information Extraction. Domenico Lembo, Federico Scafoglieri. International Journal of Semantic Computing. 2020
- Valerio Santarelli, Domenico Lembo, Marco Ruzzi, Giacomo Ronconi, Paolo Bouquet, Andrea Molinari, Flavio Pompermaier, Daniela Caltabiano, Emanuela Catoni, Antonio Fabrizi, Mauro Minenna, Mario Punchina, Federico Scafoglieri. Semantic Technologies for the Production and Publication of Open Data in ACI - Automobile Club d'Italia. ISWC Satellites 2019
- Domenico Lembo, Federico Scafoglieri. Coupling Ontologies with Document Spanners. In Description Logics. 2015.
- Domenico Lembo, Federico Scafoglieri. A Formal Framework for coupling Document Spanners with Ontologies. In Proc. of the 2nd IEEE Int. Conf. on Artificial Intelligence and Knowledge Engineering (AIKE), 2019.
- Ontology population for open-source intelligence: A GATE-based solution. Giulio Ganino, Domenico Lembo, Massimo Mecella, Federico Scafoglieri. Softw. Pract. Exper. 48(12): 2302-2330 (2018)
- Giulio Ganino, Domenico Lembo, Massimo Mecella, Federico Scafoglieri: Ontology Population for Open-Source Intelligence (Discussion Paper). SEBD 2018
- Giulio Ganino, Domenico Lembo, Federico Scafoglieri: Ontology Population from Raw Text Corpus for Open-Source Intelligence. ICWE 2017