

Federico Maria Scafoglieri, PhD

Department of Computer Engineering
University of Rome La Sapienza
Via Ariosto, 25, 00185 Roma RM, Italy

sites.google.com/diag.uniroma1.it/scafoglieri

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 (head of lab Prof. Maurizio Lenzerini). 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 (collaborators Dr. Ronald Fagin, Dr. Lucian Popa, Prof. Phokion Kolaitis, Dr. Yunyao Li, Dr. Kun Qian).

Main Expertise: Knowledge Representation and Automated Reasoning, Ontology Based Data Integration, Information Extraction, Entity Linking/Resolution.

Education

- **University of Rome La Sapienza** Rome, Italy *PhD, Engineering in Computer Science - Artificial Intelligence and Knowledge Representation* Nov 2017 – Sept 2021 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 have shown that the theoretical results obtained are applicable in practice in real financial domains, with a considerable increase in extraction power. I also proposed a formal framework to study the problem of entity resolution in the context of knowledge representation, including defining the complexity of the various services related to it. External reviews of the dissertation identified my accomplishments in the top 5%, with honors (Laude) and merit mention.
- **University of Rome La Sapienza** Rome, Italy *Master's Degree, Engineering in Computer Science* Jun 2013 – Jan 2017 Supervisor: Prof. Domenico Lembo
- **University of Rome La Sapienza** Rome, Italy *Bachelor's degree, Engineering in Computer Science* Sept 2009 – Dec 2012

Appointments

- **Commissione Nazionale per le Società e la Borsa (CONSOB)** Rome, Italy *AI Research Scientist / Research Scholar* April 2020 - Present Collaborators: 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 created system, which combines pattern 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 Advanced Digital Solutions for Professional Food and Nutrition Catering Service: I am working on the formal definition of a domain ontology for food and health sector useful for the integration of various data sources, with the aim of providing a uniform data layer for business intelligence purposes. H2020 ADCATER Project
- **IBM's Almaden Research Center** San Jose, California, USA *Research Scholar* Oct 2019 - April 2021 Mentors: Dr Ronald Fagin, Prof Phokion Kolaitis, Dr. Lucian Popa, Dr. Yunyao Li
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 introduced a new methodology and a formal framework for a joint Entity Resolution and Query Answering over semantically rich and complex knowledge representation systems. (publication to be appear)
- **OBDA Systems** Rome, Italy *Research & Development Engineer* March 2017 - March 2020 Mastro Studio: 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

- Mobility Grant for PhD Student 2019/2020
- IBM Grant 2019/2020

Awards

- [Best Demo ISWC 2020](#)

Academic Teaching

- **University of Rome La Sapienza** Rome, Italy *Professor of Programming Language I* Mar 2022 - Feb 2023 *Teaching Assistant of Programming Language I* Oct 2020 - Dec 2021 *Professor of Multimedia Design II* March 2019 - Sept 2019 *Teaching Assistant of Programming Language I* Oct 2017 - Apr 2018

Thesis Supervisor

Master Thesis Supervision

- 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
- Roberto 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, IBM Almaden Research Center
- Dr. Lucian Popa, IBM Almaden Research Center
- Dr. Valerio Santarelli, CTO OBDA-System Almagest
- Dr. Kun Qian, Amazon

Entity Resolution

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

Program Committee

- In the program committee of the 21th International Conference on Principles of Knowledge Representation and Reasoning (KR 2023) • Reviewer EACL 2023
- Reviewer Journal of Computer in Industry 2022
- Reviewer Journal of Transaction on Services Computing 2022
- In the program committee of the 20th International Conference on Principles of Knowledge Representation and Reasoning (KR 2022)
- Reviewer Journal of Transaction on Services Computing 2021
- Sub-reviewer of the 24th International Conference on Database Theory (ICDT 2021)
- Sub-reviewer of the 2019 IEEE International Conference on Artificial Intelligence and Knowledge Engineering (AIKE 2019)
- Sub-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.

- 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 '20
 - 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. Sofw. 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 Workshops 2017: 173-186

December 12, 2022

Federico Maria Scafoglieri, Phd