



Tommaso Soru

Data Scientist, Researcher

London, United Kingdom

tom@tommaso-soru.it

+44 7894 456 626

<http://tommaso-soru.it>

I'm an AI enthusiast with the dream of giving my small contribution to the quest for intelligence. My research topics include Entity Resolution, Link Prediction, and Question Answering over Knowledge Graphs, with a focus on the Semantic Web.

EXPERIENCE

Scalpel Ltd., London, UK

SPEECH & NLP ENGINEER > AUGUST 2020 – PRESENT

- Deep learning model for distant Automatic Speech Recognition in a surgery room. **Python, Torch, fairseq**
- Design & implementation of real-time Medical Information Extraction from text. **Python, AWS, Docker**

Data Lens Ltd. and Semantic Integration Ltd., London, UK

DATA SCIENTIST & ARCHITECT > NOVEMBER 2018 – AUGUST 2020

- NLP Pipeline for scalable Information Extraction from unstructured data. **Python, Java, AWS, Docker**
- Development of a Named Entity Recognition algorithm for the Fintech domain with 90% accuracy.
- Design & implementation of a semantic reasoner for an international retail company. **Node.js, SPARQL**
- Design of a data architecture for knowledge discovery for a global firm in the FTSE 250 Index. **Python, RDF(S), SPARQL, AWS, Docker**

AKSW, University of Leipzig, Leipzig, Germany

DOCTORAL RESEARCHER > JULY 2012 – OCTOBER 2018

- PhD thesis titled "Metric Learning for Link Discovery" under the supervision of [Prof. Dr. Axel Ngonga](#).
- R&D of two scalable methods for Link Prediction in directed graphs. **Python, Java, SPARQL, RDF(S)**
- Design and development of [ROCKER](#), algorithm for RDF Key Discovery, and its web-based GUI. **Java, MySQL, RDF(S), AWS, JS, Docker**
- Realisation of a scalable Knowledge Graph Embedding approach. **Python, Gensim, Keras**
- Application of Machine Learning techniques (SVMs, MLP) for Entity Resolution. **Java, Weka ML**
- Contributions in research projects [LinkLion](#), [LIMES](#) 1.0, [ML-Schema](#). **Java, JSP, OWL, SPARQL, RDF(S)**
- 30+ publications in Web conferences (WWW, ISWC, ESWC) and AI workshops (WSDM, ICML, AAAI).

BuzzTweet, Berlin, Germany

LEAD DEVELOPER > APRIL 2015 – JUNE 2017

- R&D of a real-time Sentiment Analysis framework to predict KPIs. **Java, Weka ML, AWS, Docker**

GroopLab srl, Berlin, Germany

DATA SCIENTIST, SHAREHOLDER > JUNE 2014 – JUNE 2015

- R&D of Machine-Learning methods for event recommendation and NER. **Python, scikit-learn, Java, AWS**

ADDITIONAL PROJECTS

Liber AI, London, UK

CO-FOUNDER, RESEARCHER > OCTOBER 2017 – PRESENT

-
- Open-source research project [Neural SPARQL Machines](#): Question Answering over Knowledge Graphs using Recurrent Neural Networks (160+ stars and 80+ forks on GitHub). **Python, TensorFlow, SPARQL**

DBpedia Association, Leipzig, Germany

MENTOR, ORGANISATION ADMIN (REMOTE) > FEBRUARY 2017 – PRESENT

- Org Admin for [Google Summer of Code](#) (GSoC) at [DBpedia](#), a Multilingual Semantic Knowledge Graph.
- Mentor of 8 students in open-source projects on Knowledge Base Embeddings and Question Answering. 100% of them joined the programme as co-mentors the next year. **Python, TensorFlow, Keras, SPARQL**

EDUCATION

University of Leipzig, Leipzig, Germany

PHD IN COMPUTER SCIENCE > JULY 2012 – OCTOBER 2018 (ON HOLD)

University of Milan-Bicocca, Milan, Italy

MASTER OF SCIENCE IN ARTIFICIAL INTELLIGENCE > NOVEMBER 2008 – MARCH 2012

- "Semi-automated identification of Italian lemmas using multilingual WordNet" **Java, CoreNLP, RDF(S)**
- "Topologies comparison in a Public Goods Game" and others. **Java, C++, Repast, OpenCV**

Polytechnic of Milan, Milan, Italy

BACHELOR IN COMPUTER ENGINEERING > SEPTEMBER 2003 – SEPTEMBER 2008

- "A Geographic Ontology for a Semantic Wiki", "RDF-based Settlers of Catan". **Java, RDF(S), OWL**

ABOUT ME

- Team player. Self-starter. Lateral thinker.
- [Drummer](#) since 1996. 200+ live shows, 3 albums and 3 EPs published.
- Interests: progressive music, genealogy, quantified self, association football, science fiction.
- Languages: Italian (native), English (full proficiency), French & German (basic).

SELECTED PUBLICATIONS & AWARDS

Anand Panchbhai, **Tommaso Soru**, Edgard Marx, [Exploring Sequence-to-Sequence Models for SPARQL Pattern Composition](#). In proceedings, KGSWC, 2020.

Faisal Alshargi, Saeedeh Shekarpour, **Tommaso Soru**, Amit P. Sheth, [Metrics for Evaluating Quality of Embeddings for Ontological Concepts](#). In proceedings, AAAI-MAKE Symposium, 2019.

Diego Moussallem, **Tommaso Soru**, Axel Ngonga, [Neural Translation and Enrichment of Knowledge Graphs](#). In proceedings, ISWC, 2019.

Muhammad Saleem, Alexander Potocki, **Tommaso Soru**, Olaf Hartig, Axel Ngonga, [CostFed: Cost-Based Query Optimization for SPARQL Endpoint Federation](#). In proceedings, SEMANTICS, 2018. **Best paper award**

André Valdestilhas, **Tommaso Soru**, Markus Nentwig, Edgard Marx, Muhammad Saleem, Axel Ngonga, [Where is My URI?](#). In proceedings, ESWC, 2018. **Honourable mention**

Tommaso Soru, Edgard Marx, André Valdestilhas, Diego Esteves, Diego Moussallem, Gustavo Publico, [Neural Machine Translation for Query Construction and Composition](#). ICML NAMPI Workshop, 2018.

Tommaso Soru, Edgard Marx, Diego Moussallem, Gustavo Publico, André Valdestilhas, Diego Esteves, Ciro Baron Neto, [SPARQL as a Foreign Language](#). SEMANTICS, 2017. **Best poster nominee**

Tommaso Soru, Edgard Marx, Axel Ngonga, [ROCKER: A Refinement Operator for Key Discovery](#). In proceedings, WWW, 2015. **14% acceptance rate**

Tommaso Soru, Axel Ngonga, [A comparison of supervised learning classifiers for link discovery](#). In proceedings, SEMANTICS, 2014.

Complete list on [Google Scholar](#).