



Tommaso Soru

Data Scientist at Semantic Integration Ltd.

London, United Kingdom

tom@tommaso-soru.it

<http://tommaso-soru.it>

EXPERIENCE

Semantic Integration Ltd., London, UK - *Data Scientist & Architect*

NOVEMBER 2018 – PRESENT

- Design and implementation of a semantic reasoner.
- Design of a semantic data architecture for RDF graph virtualisation of relational databases and unstructured data.
- Development of an algorithm for Entity Recognition and Linking.

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.
- Funded by projects LinkingLOD 1-2 (DFG), DIESEL (Eurostars).
- Research and development of two scalable methods for Link Prediction in directed graphs using rule mining and Markov logics.
- Design and development of ROCKER, state-of-the-art algorithm for Key Discovery in RDF datasets.
- Realisation of a fast Knowledge Graph Embedding approach.
- Development and application of Machine Learning techniques (SVMs, Multilayer Perceptrons) for instance matching.
- Development of LinkLion and LIMES 1.0, Link Discovery projects.
- May 2018: Exchange period at the Smart Data Analytics (SDA) Research group, Bonn University, Germany.
- 25+ scientific publications, including top venues such as WWW, ISWC and ESWC as well as WSDM and ICML workshops.

DBpedia, Leipzig, Germany - *Member, GSoC Admin and Mentor*

JANUARY 2019 – PRESENT

- Org Admin for Google Summer of Code (GSoC) at DBpedia, a Large Multilingual Semantic Knowledge Graph.

FEBRUARY 2017 – DECEMBER 2018

- Google Summer of Code mentor in 2017, 2018 and 2019.
- Mentored 5 students on the open-source projects "Knowledge Base Embeddings for DBpedia" and "A Neural Question Answering Model for DBpedia".

-
- Represented DBpedia at the Google Summer of Code Summit 2017 in Sunnyvale, CA, United States.
 - Represented GSoC mentors at the 11th DBpedia Community Meeting, Apple Inc., Cupertino, CA, United States.

LiberAI, Leipzig, Germany - *Co-founder*

OCTOBER 2017 – PRESENT

- Co-founder of LiberAI, part of the AKSW research group.
- Leader of the open-source project [Neural SPARQL Machines](#): Answering natural-language questions over Knowledge Graphs using Recurrent Neural Networks (80+ stars and 30+ forks on GitHub).

BuzzTweet, Berlin, Germany - *Lead Developer*

APRIL 2015 – JUNE 2017

- Design and development of a real-time Sentiment Analysis framework.
- Design and deployment of RESTful APIs.

GroopLab srl, Berlin, Germany - *Data Scientist, Shareholder*

JUNE 2014 – JUNE 2015

- Startup aimed at the development of a mobile app.
- Development of Machine-Learning methods for the categorisation and recommendation of events by type and musical genre.
- Design and development of a Named Entity Recognition technique to identify artist names in texts.
- Design and deployment of RESTful APIs.

SS&C srl, Varese, Italy - *Web Developer and Designer*

JANUARY 2007 – JULY 2011, APRIL 2012 – JUNE 2012

- Website and web application development.

InfoVadis, Laveno, Italy - *Backend Developer*

MARCH 2006 – DECEMBER 2006

- Development of applications for data migration.

EDUCATION

University of Leipzig, Germany - *PhD in Computer Science (Dr. rer. nat.)*

JULY 2012 – PRESENT (EXPECTED LATE 2019 – EARLY 2020)

- See “Experience”.

University of Milan-Bicocca, Italy - *MSc in Computer Science*

NOVEMBER 2008 – MARCH 2012

- Major in “Artificial Intelligence and Complex Systems”.

-
- Thesis and Projects: “Learning Domain-Specific Metrics for Link Discovery” (at the University of Leipzig), “Semi-automated identification of Italian lemmas using multilingual WordNet”, “Topologies comparison in a Public Goods Game”, “Vision-based Autonomous Driving”.
 - Main Courses: Artificial Intelligence, Machine Learning, Computer Vision, Soft Computing, Semantics, Information and Knowledge Representation.

Polytechnic of Milan, Italy - *Bachelor in Computer Engineering*

SEPTEMBER 2003 – SEPTEMBER 2008

- Thesis and Projects: “A Geographic Ontology for a Semantic Wiki”, “Knowledge-engineering-based Settlers of Catan”.
- Main Courses: Knowledge Engineering, Hypermedia Applications, Theoretical Computer Science, Data Analysis.

SPECIALTIES

- Programming languages: Python, Java; experience with Javascript / Node.js / jQuery, R, Matlab, C, PHP.
- Semantic Technologies: RDF, OWL, SPARQL.
- ML Frameworks: TensorFlow, Keras, Weka, scikit-learn, gensim.
- Languages: Italian (native), English (full proficiency), French & German (basic).
- Excellent teamwork skills. Good knowledge of maths. Lateral thinker.

ORGANISATIONS

- Member, Agile Knowledge Engineering & Semantic Web (AKSW), 2011–today.
- Member, DBpedia Association, 2012–today.
- Member, Data Science Research (DICE) Paderborn, 2017–today.
- Guest Member, Smart Data Analytics (SDA) Bonn, 2018–today.
- Co-founder, Machine Learning Schema community at the World Wide Web Consortium (W3C), 2015–today.

OTHER ACTIVITIES

- Best Lightning Talk Award at CSCUBS Conf., Bonn, Germany, 2018.
- Deep Learning Summit, Bilbao, Spain, 2017.
- Lisbon Machine Learning Summer School, Portugal, 2016.
- Best Student Poster Award at ESWC Summer School, Kalamaki, Greece, 2013.
- Coursera online course on "Game Theory", 2013.

-
- Stanford online course "An Introduction to Artificial Intelligence", 2011.

INTERESTS

- Drummer since 1996 with 180+ live shows, 3 albums and 2 EPs published.
- AI singularity, progressive music, genealogy, association football, science fiction enthusiast.

SELECTED PUBLICATIONS & AWARDS

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**.

Tommaso Soru, Stefano Ruberto, Diego Moussallem, Edgard Marx, Diego Esteves, Axel Ngonga, [Expeditious Generation of Knowledge Graph Embeddings](#). Article Under Review, Presented at ECDA, 2018.

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

Tommaso Soru, André Valdestilhas, Edgard Marx, Axel Ngonga, [Beyond Markov Logic: Efficient Mining of Prediction Rules in Large Graphs](#). Article, Working paper, 2018.

Tommaso Soru, Edgard Marx, André Valdestilhas, Diego Esteves, Diego Moussallem, Gustavo Publio, [Neural Machine Translation for Query Construction and Composition](#). Extended Abstract, ICML Workshop on Neural Abstract Machines & Program Induction v2, 2018.

Edgard Marx, **Tommaso Soru**, André Valdestilhas, [Triple Scoring Using a Hybrid Fact Validation Approach](#). In proceedings, WSDM Cup Workshop, 2017. **Ranked 4th out of 33 participants**.

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

Edgard Marx, Saeedeh Shekarpour, **Tommaso Soru**, Adrian M. P. Brasoveanu, Muhammad Saleem, Ciro Baron, Albert Weichselbraun, Jens Lehmann, Axel Ngonga, Sören Auer, [Torpedo: Improving the State-of-the-Art RDF Dataset Slicing](#). In proceedings, ICSC, 2017. **Best paper nominee**.

Tommaso Soru, Edgard Marx, Axel Ngonga, [Enhancing Dataset Quality Using Keys](#). In proceedings, ISWC (Companion Volume), 2015.

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

Markus Nentwig, **Tommaso Soru**, Axel Ngonga, Erhard Rahm, [LinkLion: A Link Repository for the Web of Data](#). In proceedings, ESWC (Satellite Events), 2014.

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

Tommaso Soru, Axel Ngonga, [Active Learning of Domain-Specific Distances for Link Discovery](#). In proceedings, JIST, 2012.