



# Tommaso Soru

Senior PhD Student at University of Leipzig

Leipzig, Germany

tom@tommaso-soru.it

<http://tommaso-soru.it>

## EXPERIENCE

### University of Leipzig, Leipzig, Germany - *PhD Student by Research*

JULY 2012 - PRESENT (EXPECTED AUGUST 2018)

- PhD topic titled “Learning Metrics for Link Discovery” under the supervision of Prof. Dr. Axel-Cyrille Ngonga Ngomo.
- Research and development of a scalable method 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.
- Realization of a knowledge graph embedding model for RDF graphs.
- Development and application of Machine Learning techniques (SVMs, Multilayer Perceptrons) for instance matching.
- Development of LinkLion and LIMES 1.0, Link Discovery projects.
- 20+ scientific publications.

### DBpedia, Leipzig, Germany - *Mentor for Google Summer of Code*

FEBRUARY 2017 - PRESENT

- Mentored 2 students on different aspects of the project "Knowledge Base Embeddings for DBpedia".
- Represented DBpedia at the Google Summer of Code Summit 2017 in Sunnyvale, CA, United States.

### LiberAI, Leipzig, Germany - *Co-founder*

OCTOBER 2017 - PRESENT

- Open-source project LiberQA: Answering natural-language queries on Knowledge Graphs using Recurrent Neural Networks.

### 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 aiming at the development of a mobile app.

- 
- Development of Machine-Learning methods for the categorization 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

- Website and web application development.

**InfoVadis, Laveno, Italy** - *Backend Developer*

MARCH 2006 - DECEMBER 2006

- Development of applications for data migration.

## EDUCATION

**University of Leipzig, Leipzig, Germany** - *PhD in Computer Science*

JULY 2012 - PRESENT (EXPECTED AUGUST 2018)

See "Experience".

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

NOVEMBER 2008 - SEPTEMBER 2011

Graduated with a major in "Artificial Intelligence and Complex Systems".

Thesis and Projects: "Learning Domain-Specific Metrics for Link Discovery", "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** - *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: Java, Python, PHP, Javascript / jQuery.

Languages: Italian (native), English (full proficiency), French, German (A2).

Excellent teamwork skills. Strong knowledge of maths. Lateral thinker.

---

## ORGANIZATIONS & AWARDS

Member, Agile Knowledge Engineering & Semantic Web (AKSW), 2012–today.

Member, Data Science Research (DICE) Paderborn, 2017–today.

Member, DBpedia Association, 2017–today.

Cofounder, Machine Learning Schema at the World Wide Web Consortium (W3C), 2015–today.

Selected to attend the ESWC Summer School, Kalamaki, Greece, 2013.

Best Student Poster Award, ESWC Summer School, 2013.

Selected to attend the Lisbon Machine Learning School, Portugal, 2016.

Deep Learning Summit attendee, Bilbao, Spain, 2017.

## INTERESTS

Drummer since 1996 with over 180 live shows, 3 Albums, 2 EPs published.

Singularity, genealogy, soccer, sci-fi enthusiast.

## SELECTED PUBLICATIONS

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

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

**Tommaso Soru**, Diego Esteves, Edgard Marx, Axel-Cyrille Ngonga Ngomo, [Mandolin: A Knowledge Discovery Framework for the Web of Data](#). Article, CoRR, 2017.

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

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

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

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

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