

Rita Torres Sousa

Rita Torres Sousa holds a Bachelor's degree in Health Sciences, a Master's degree in Bioinformatics and Computational Biology, and a Ph.D. degree in Informatics from the University of Lisbon. Since her master's thesis, her research has focused on integrating life sciences expertise with computational skills to develop new approaches to learn from complex biomedical data and discover new knowledge. More specifically, her research interests include biomedical knowledge graphs, semantic similarity, machine learning and explainable artificial intelligence.



✉ rita.sousa@uni-mannheim.de | 🏠 ritatsousa.github.io | 📄 ritatsousa | 🐦 @RitaTorresSousa

Education

PhD Student in Informatics, Approved with Distinction and Prizes

FACULTY OF SCIENCES OF THE UNIVERSITY OF LISBON

Lisbon, Portugal

September 2019 - July 2024

- Relevant Courses: Big Data, Intelligent Computer Systems, Research Topics, Technical, Scientific Communication
- Thesis: Semantic perspectives for learning over biomedical knowledge graphs

International Semantic Web Research Summer School

STI INTERNATIONAL

Bertinoro, Italy

July 2023

- Topics: F.A.I.R. Knowledge Graphs, Machine (and Deep) Learning, Linked Data, Knowledge Graph Engineering with Ontology Design Patterns, Human in the Loop, Querying and Reasoning, Knowledge Graph learning and production, Natural Language Processing, Distributed, Decentralised Semantic Web.

Master's Degree in Bioinformatics and Computational Biology (BBC), 19/20

FACULTY OF SCIENCES OF THE UNIVERSITY OF LISBON

Lisbon, Portugal

September 2017 - July 2019

- Relevant Courses: Web Applications, Data Processing Technologies, Introduction to Databases, Computer Programming Complements
- Dissertation: Evolving meaning - using Genetic Programming to learn similarity perspectives for mining biomedical data (19/20)

Bachelor's Degree in Health Sciences, 17/20

UNIVERSITY OF LISBON

Lisbon, Portugal

September 2014 - July 2017

Professional Experience

University of Mannheim, Germany

RESEARCHER AT THE KI-DIABETESDETEKTION PROJECT

Mannheim, Germany

January 2024 - Present

- The goal of the KI-DiabetesDetektion project is to integrate indication data from various sources in a knowledge graph and apply machine learning methods to improve the early-stage detection of Diabetes.

Faculty of Sciences, University of Lisbon, Portugal

TUTOR OF MASTER THESIS

Lisbon, Portugal

September 2020 - Present

- Predicting Gene-Disease Associations with Knowledge Graph Embeddings over Multiple Ontologies (Susana Nunes, MSc. in BBC)
- Combining Deep Learning and Knowledge Graphs in PPI Network Prediction Tasks (Laura Balbi, MSc. in BBC)
- Learning with uncertainty: improving supervised learning of protein-protein interactions with lower quality examples (André Mendes, MSc. in Informatics Engineering)

LASIGE, Faculty of Sciences, University of Lisbon, Portugal

RESEARCHER

Lisbon, Portugal

September 2019 - 2023

Católica Lisbon School of Business and Economics

TEACHING ASSISTANT OF DATABASE MANAGEMENT

Lisbon, Portugal

February 2022 - July 2023

Scholarships

Knowledge At the Tip of Your fingers: Clinical Knowledge for Humanity (KATY)

PHD STUDENT FELLOWSHIP

Lisbon, Portugal

November 2023 - December 2023

Evolving Meaning for Supervised Learning in Complex Biomedical Domains using Knowledge Graphs (SFRH/BD/145377/2019)

PHD STUDENT FELLOWSHIP

Lisbon, Portugal

September 2019 - October 2023

DAAD Funding program: STIBET-Doktoranden

CONTACT FELLOWSHIP FOR VISITING DOCTORAL STUDENTS

Mannheim, Germany

September 2022 - November 2022

SMiLax: Semantic Mining with Linked Data (PTDC/EEI-ESS/4633/2014)

MASTER STUDENT FELLOWSHIP

Lisbon, Portugal

November 2018 - August 2019

Publications

Relevant Full Papers:

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Explaining protein-protein interactions with knowledge graph-based semantic similarity.** Computers in Biology and Medicine, 2024. (Q1 Scimago)

Nunes, Susana; **Sousa, Rita T.;** Pesquita, Catia. **Multi-domain Knowledge Graph Embeddings for Gene-Disease Association Prediction.** Journal of Biomedical Semantics, 2023. (Q2 Scimago)

Sousa, Rita T.; Silva, Sara; Paulheim, Heiko; Pesquita, Catia. **Biomedical Knowledge Graph Embeddings with Negative Statements.** International Semantic Web Conference, 2023. (Core A)

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Explainable representations for relation prediction in knowledge graphs.** International Conference on Principles of Knowledge Representation and Reasoning, 2023. (Core A*)

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Supervised Biomedical Semantic Similarity.** IEEE Access, 2023. (Q1 Scimago)

Cardoso, Carlota; **Sousa, Rita T.;** Köhler, Sebastian; Pesquita, Catia. **A collection of benchmark data sets for knowledge graph-based similarity in the biomedical domain.** Database: The journal of biological databases and Curation, 2020. (Q1 Scimago)

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Evolving knowledge graph similarity for supervised learning in complex biomedical domains.** BMC Bioinformatics 21 1, 2020. (Q1 Scimago)

Relevant Workshop Papers and Posters:

Sousa, Rita T.; Paulheim, Heiko. **Integrating Heterogeneous Gene Expression Data through Knowledge Graphs for Improving Diabetes Prediction.** Semantic Web Solutions for Large-Scale Biomedical Data Analytics Workshop, 2024.

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Benchmark datasets for biomedical knowledge graphs with negative statements.** Semantic Web Solutions for Large-Scale Biomedical Data Analytics Workshop, 2023.

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Explaining Protein-Protein Interaction Predictions with Genetic Programming.** Evo* 2022 Late-Breaking Abstracts, 2022. (Core B)

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Towards Supervised Biomedical Semantic Similarity.** Semantic Web Solutions for Large-Scale Biomedical Data Analytics Workshop, 2022.

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **The Supervised Semantic Similarity Toolkit.** Extended Semantic Web Conference Poster and Demo Track, 2022. (Core A)

Nunes, Susana; **Sousa, Rita T.;** Serrano, Filipa; Branco, Ruben; Soares, Diogo F.; Martins, Andreia S.; Auletta, Eleonora; Castanho, Eduardo N.; Madeira, Sara C.; Aidos, Helena; Pesquita, Catia. **Explaining artificial intelligence predictions of disease progression with semantic similarity.** CLEF, 2022.

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **Is there data leakage in protein-protein interaction prediction using knowledge graphs?.** International Semantic Web Conference Poster and Demo Track, 2021. (Core A)

Nunes, Susana; **Sousa, Rita T.;** Pesquita, Catia. **Predicting Gene-Disease Associations with Knowledge Graph Embeddings over Multiple Ontologies.** Bio-Ontologies 2021, 2021.

Sousa, Rita T.; Silva, Sara; Pesquita, Catia. **evoKGsim+: a framework for tailoring Knowledge Graph-based similarity for supervised learning.** Extended Semantic Web Conference Poster and Demo Track, 2021. (Core A)

Cardoso, Carlota; **Sousa, Rita T.;** Köhler, Sebastian; Pesquita, Catia. **A collection of benchmark data sets for knowledge graph-based similarity in the biomedical domain.** Extended Semantic Web Conference Poster and Demo Track, 2020. (Core A)

Skills

- **Experience with Programming Languages** Python, Java, R
- **Experience with Query Languages** SQL, SPARQL
- **Experience with Tools & Technologies** LaTeX, Microsoft Office Suite
- **Experience with working with very large biomedical databases and resources** UniProt, STRING, DisGeNET

Languages

- **Portuguese** Native
- **English**
- **French**

Honors & Awards

2024	Best Researcher Award - PhD Student , LASIGE Workshop 2024
2024	Two Distinguished Publications ("Explainable representations for relation prediction in knowledge graphs" and "Supervised Biomedical Semantic Similarity"), LASIGE Workshop 2024
2023	Best Poster LASIGE Workshop ("Can ChatGPT explain protein interactions better than Knowledge Graphs?"), LASIGE Workshop 2023
2021	Best Poster Award ("evoKGsim*: a framework for tailoring Knowledge Graph-based similarity for supervised learning"), Extended Semantic Web Conference 2021
2021	Two Distinguished Publications ("Evolving knowledge graph similarity for supervised learning in complex biomedical domains" and "A collection of benchmark data sets for knowledge graph-based similarity in the biomedical domain"), LASIGE Workshop 2021
2020	Best Poster Award ("A collection of benchmark data sets for knowledge graph-based similarity in the biomedical domain"), Extended Semantic Web Conference 2020
2019	Best Poster LASIGE Workshop ("Evolving meaning: Using Genetic Programming to learn similarity perspectives for mining biomedical data"), LASIGE Workshop 2019
2018-2019	Merit Scholarship , Direção-Geral do Ensino Superior (DGES)
2017-2018	Maxdata Informatics Award of Excellence , Faculty of Sciences of University of Lisbon

Event Organization and Participation

- Consultant at DataFest Germany **2024**;
- Participant at Mannheim Center for Data Science Academic Speed Dating **2024**;
- Member of the Genetic and Evolutionary Computation Conference Local Team **2023**;

- Organizer of Extended Semantic Web Conference Student-only Gathering **2023**;
- Participant of Ciências's Open Day for Masters and Doctorates **2023**;
- Participant of Girls in ICT **2022 and 2023**;
- Organizer of the 5th LASIGE Workshop **2020**;
- Organizer of Health Sciences Symposiums **2016 and 2017**;
- Organizer of Anatomy Workshops for Health Sciences Students **2015 and 2016**;