Rita Torres de Sousa

Rita Torres de 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 in Informatics, Approved with Distinction and Prizes — *Lisbon, Portugal*

Faculty of Sciences of the University of Lisbon — *09/2019 - 07/2024*

- Thesis: Semantic perspectives for learning over biomedical knowledge graphs.

International Semantic Web Research Summer School — *Bertinoro, Italy*

STI International — *07/2023*

Master in Bioinformatics and Computational Biology (BBC), 19/20 — *Lisbon, Portugal*

Faculty of Sciences of the University of Lisbon — *09/2017 - 07/2019*

- Dissertation: Evolving meaning - using Genetic Programming to learn similarity perspectives for mining biomedical data (19/20).

Bachelor in Health Sciences, 17/20 — *Lisbon, Portugal*

University of Lisbon — *09/2014 - 07/2017*

Professional Experience

Data and Web Science Group, University of Mannheim — *Mannheim, Germany*

Researcher — *01/2024 - Present*

Data and Web Science Group, University of Mannheim — *Mannheim, Germany*

Visiting researcher (DAAD Funding program: STIBET-Doktoranden) — *09/2022 - 11/2022*

LASIGE, Faculty of Sciences, University of Lisbon — *Lisbon, Portugal*

Researcher — *09/2019 - 12/2023*

Teaching and Supervision Experience

School of Business Informatics and Mathematics, University of Mannheim — *Mannheim, Germany*

Teaching Assistant of Knowledge Graphs — *09/2024 - Present*

Faculty of Sciences, University of Lisbon — *Lisbon, Portugal*

Tutor of Master Thesis — *09/2020 - 09/2024*

- 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).

Católica Lisbon School of Business and Economics — *Lisbon, Portugal*

Teaching Assistant of Database Management — *02/2022 - 07/2023*

Projects

Serving up-to-date Dynamic Knowledge Graph Embeddings

Mannheim, Germany

Principal Investigator

01/2025 - 06/2025

- The Open Science project is funded by the Open Science Office.
- The goal is to develop a framework capable of periodically collecting new knowledge graph versions, computing embeddings and making them publicly available for further research.

KI-DiabetesDetektion

Mannheim, Germany

Researcher

01/2024 - Present

- The KI-DiabetesDetektion project is funded by BMBF (Bundesministerium für Bildung und Forschung).
- The goal 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.

KATY: Knowledge At the Tip of Your fingers

Lisbon, Portugal

PhD Student Scholarship

11/2023 - 12/2023

- KATY is funded by the European Union's Horizon 2020 research and innovation programme.
- The project aims to build a precise, personalised medicine system empowered by artificial intelligence, which predicts the response of kidney cancer to targeted therapies.

SMiLax: Semantic Mining with Linked Data

Lisbon, Portugal

Master Student Scholarship (PTDC/EEI-ESS/4633/2014)

11/2018 - 08/2019

- SMiLax is funded by Fundação para a Ciência e a Tecnologia (FCT).
- The goal is to improve the state of the art in semantic data mining by establishing novel methods and algorithms for the automated semantic annotation and enrichment of data, whose output will be explored by novel data mining approaches capable of capitalizing on the semantic web.

Publications

Relevant Full Papers:

Sousa, Rita T.; Paulheim, Heiko. **Multi-dataset and transfer learning using gene expression knowledge graphs for patient diagnosis**. Accepted at the Extended Semantic Web Conference (ESWC), 2025. (Core B)

Sousa, Rita T.; Paulheim, Heiko. **Gene Expression Knowledge Graph for Patient Representation and Diabetes Prediction**. Journal of Biomedical Semantics, 2025. (Q2 Scimago)

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)

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 at the Extended Semantic Web Conference, 2024.

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

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)

Skills

- **Experience with programming languages** Python, Java, R
- **Experience with query languages** SQL, SPARQL
- **Experience with machine learning frameworks** scikit-learn, Pytorch, TensorFlow
- **Experience with tools and technologies** LaTeX, Microsoft Office Suite
- **Experience with working with very large biomedical databases and resources** UniProt, STRING, DisGeNET
- **Languages** Portuguese (native), English (full professional proficiency), French (elementary proficiency), German (elementary proficiency)

Honors and Awards

2025	**PhD Thesis FCiências.ID Award**, Dia de CIÊNCIAS 2025
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

- Participant at Open Science Day - Uni Mannheim ***2024***;
- Organizer of the NEGKNOW Challenge at the International Semantic Web Conference ***2024***;
- Consultant at DataFest Germany ***2024***;
- Participant at Mannheim Center for Data Science Academic Speed Dating ***2024***;
- Program Committee of Poster&Demos Track at International Semantic Web Conference ***2023***;
- Member of the Genetic and Evolutionary Computation Conference Local Team ***2023***;
- Organizer of Extended Semantic Web Conference Student-only Gathering ***2023***;
- Program Committee of Poster&Demos Track at Extended Semantic Web Conference ***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***;
- Member of ”Núcleo de Ciências da Saúde da Associação de Estudantes da Faculdade de Medicina da Universidade de Lisboa” ***2015/2016 and 2016/2017***;