

Master project 2020-2021

Personal Information

Supervisor	Davide Cirillo
Email	davide.cirillo@bsc.es
Institution	BSC - Barcelona Supercomputing Center
Website	https://www.bsc.es/
Group	Computational Biology

Project

Web development & bioinformatic tools

Project Title:

Evolution and crosstalk of biological ontologies

Keywords:

biological ontologies; graph theory; machine learning

Summary:

The research undertaken at the Barcelona Supercomputing Center (BSC) by the Computational Biology group, led by Prof. Alfonso Valencia, covers a wide range of Artificial Intelligence approaches for biomedicine, in particular in the area of biomedical computational graph theory and algorithms with special focus on biological ontologies. Biological ontologies, such as the Human Phenotype Ontology (HPO) (Köhler et al. 2019) and the Gene Ontology (GO) (The Gene Ontology Consortium 2019), are recognized as essential tools in the grand challenge of biomedical data integration and interpretation. In collaboration with the BSC Computer Science Department, we have developed a system for the efficient traversing and exhaustive path enumeration in interconnected biological ontologies (Cirillo et al. 2019) exerting large-scale parallelism and scalability in High-performance computing (HPC). This framework harnesses machine learning to infer a precise mapping between disease-related phenotypic features and distinct molecular processes allowing knowledge discovery. The proposed activity will be centered on the application and extension of this framework to a larger set of biological ontologies with the aim to study aspects such as (1) the dynamics of biological knowledge accumulation across time; (2) the integration and reconciliation of the multiple biological ontologies; (3) the implementation of machine learning approaches for biological knowledge representation and reasoning. The selected candidate will work in a highly sophisticated HPC environment, will have access to systems and computational infrastructures, and will establish collaborations with experts in different areas. What will you learn - Computational biology: biological knowledge representation; resources, formats and tools related to ontologies for use across the biomedical domain; applications and analytical approaches based on ontological information. - Computer Science: basics of High-performance computing; use of BSC supercomputing resources; BSC biology-oriented HPC implementations. - Scientific Dissemination: acquisition of science communication skills through lab meeting presentations and research article writing.

References:

Köhler et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic Acids Res. 2019 Jan 8;47(D1):D1018-D1027. doi: 10.1093/nar/gky1105. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 2019 Jan 8;47(D1):D330-D338. doi: 10.1093/nar/gky1055. Cirillo et al. Graph analytics for

phenome-genome associations inference. bioRxiv. 2019 Jun 26. doi: <https://doi.org/10.1101/682229>.

Expected skills::

- Good statistical and programming skills (Python, R/Bioconductor) - Strong interest in the analysis of biological systems - Basic knowledge of bioinformatics and molecular biology - Ability to access and evaluate scientific literature

Possibility of funding::

Yes

Possible continuity with PhD: :

To be discussed

Comments:

This project will mainly focus on the application and extension of a previously developed tool, which is currently used in the laboratory. The student will be in close contact with collaborators at the BSC Computer Science Department within the groups "Best Practices for Performance and Programmability" led by Javier Teruel Garcia and Marta Garcia Gasulla, and "High Performance Artificial Intelligence" led by Ulises Cortés. The project will be supervised by Davide Cirillo and co-supervised by Alfonso Valencia.
