SPECIES SCHOLARSHIP 2024 PROPOSAL

Alcides Fonseca is an Assistant Professor at the University of Lisbon, and the Reliable Software Systems research line leader at LASIGE. He has a PhD in automatic optimization of parallel programs by the University of Coimbra.

Alcides is the PI of Resource Aware Programming, a project that aims to provide developers with compile-time models of the energy consumption of their programs. He was also a Co-PI of the CAMELOT project, that used Genetic Programming to automate Machine Learning pipelines. He is also a consultant with Genomed, a genetics diagnosis company and a mentor at Decipad, a company that is re-inventing spreadsheets as computable documents.

Selected recent publications:

- Semantically Rich Local Dataset Generation for Explainable AI in Genomics (GECCO'24, to appear)
- Comparing the expressive power of Strongly-Typed and Grammar-Guided Genetic Programming (GECCO'23)
- Domain-Aware Feature Learning with Grammar-Guided Genetic Programming (EuroGP23)
- Data types as a more ergonomic frontend for Grammar-Guided Genetic Programming (GPCE'22)
- Grammatical Evolution Mapping for Semantically-Constrained Genetic Programming (GPTP'21)
- The usability argument for Refinement Typed Genetic Programming (PPSN'20)

More information can be found at http://alcidesfonseca.com

Research Group

LASIGE is the research center for Computer Science and Engineering in the Faculty of Sciences of the University of Lisbon. LASIGE had the highest score in the Portuguese evaluation of research centers with 15/15. Our motto is driven by excellence.

In the *Reliable Software Systems* line we apply several formal methods techniques to guarantee that we can trust software. Many of these techniques (grammars, type systems) have been applied to evolutionary computation (and Genetic Programming in particular).

The research group is led by Alcides Fonseca (Assistant Professor) and is currently comprised of:

- Pedro Barbosa - PhD student on Grammar-guided interpretable surrogate models for Deep Learning-based genetic variant pathogenicity prediction.
- Guilherme Espada - PhD Student on Probabilistic Modeling of Software Resource Usage (using GP to predict energy and time consumption)
- Paulo Santos - PhD Student on Evolutionary Architecture Repair of Robotics System
- Catarina Gamboa - PhD Student on the Usability of Advanced Type Systems
- Eduardo Madeira — MSc Student working on Program Synthesis using Genetic Programming and Liquid Types
- Sara Silva (http://gplab.sourceforge.net/sara/) is a frequent collaborator and office neighbor.

Project 1: Program Synthesis with Liquid Types using Genetic Programming

Liquid Types extend traditional types with a predicate: ({x:Int | x > 0}, {y:Int | y > x}) is a pair where the first element is positive and the second is greater than the first. These types can be useful for detecting more interesting types of bugs during compilation in programming languages.

However, these types are also useful for restricting the search space of programs, especially in contexts where the grammar of the problem is quite large, and there are domain constraints.

The goal of this project is to explore different representation methods (Stack-based, Grammatical Evolution, Tree-based) that encode these restrictions. The student will implement different alternatives on top of GeneticEngine (https://github.com/alcides/GeneticEngine), which already has some initial support, and an evaluation framework. Then, the goal is to evaluate the performance of different implementations in real-world problems with constraints.

This work will have impact three-fold: will help practitioners to use Genetic Programming more efficiently: the exploration of the search space will focus only on possible programs, without the need to invalidate them; and it will make program synthesis more practical for users of programming languages

with Liquid Types; As a subset of dependent types, program synthesis using Liquid Types will allow to synthesize mathematical proofs for theorems.

Feel free to reach out to Alcides (me@alcidesfonseca.com) if you want more information.

Project 2: Automatic Generation of Mathematical Proofs

Interactive Theorem Provers like Agda, Coq or Lean are becoming popular among mathematicians. These tools allow the user to verify that their proof is correct, using dependent types. You can learn more about this approach here: https://www.ams.org/notices/202011/rnoti-p1791.pdf

The Curry-Howard correspondence shows that programs are proofs, and proofs are programs. As such, we can generate proofs using Genetic Programming, just like one can generate programs, with the big difference of requiring support for Full-Dependent Types. In this work, we expect the student to address this limitation. We have a small prototype implemented in Genetic Engine (https://github.com/alcides/GeneticEngine), but it requires deep thought.

For evaluation, we plan to use the Lean programming language (https://leanprover.github.io), and the proof database mathlib (https://leanprover-community.github.io), and the LeanGym environment (https://github.com/openai/lean-gym). We also plan to compare with Neural-Network approaches that use Language Models to predict the next step in a proof (https://arxiv.org/abs/2202.01344).

Feel free to reach out to Alcides (me@alcidesfonseca.com) if you want more information.

Other relevant information:

This work is being conducted as part of the FCT-funded RAP project (https://www.lasige.pt/project/rap). We will fund the travel arrangements to Evo* to present the paper, a complementary grant, as well as necessary hardware. We also have other funded projects solely for High Performance Computing resources that can be used in the context of this work.

About Lisbon

Lisbon was considering the Leading City Destination (and Portugal the leading Destination) by the World Travel Awards in recent years. It is also mentioned as one of the best cities for foreigners to live. https://www.visitlisboa.com/en