

Julia for AGC

Atell-Yehor Krasnopolski

Taras Shevchenko National University of Kyiv, Ukraine

Mentors: Jerry Ling, Alexander Held

May 2023

Project Overview

The idea of the project is to provide a new implementation of the IRIS-HEP Analysis Grand Challenge (AGC) [1] (version 1.0) using the Julia programming language.

Analysis Grand Challenge

The Analysis Grand Challenge (AGC) is about performing the last steps in an analysis pipeline at scale to test workflows envisioned for the HL-LHC. This includes columnar data extraction from large datasets, processing of that data (event filtering, construction of observables, evaluation of systematic uncertainties) into histograms, statistical model construction and statistical inference, relevant visualisations for these steps, all done in a reproducible & preservable way that can scale to HL-LHC requirements. [2]

The Julia Language

Julia provides a powerful set of features for efficient scientific computing while incorporating that into a flexible interactive environment. This project aims to sustain the ongoing exploration [3, 4] driven by the growing enthusiasm for Julia as a viable language option in the realm of high-energy physics (HEP). Currently, C++ and Python stand as the dominant languages in this field. Julia already provides efficient tools and packages for high-energy physics, such as LiteHF.jl, ROOT.jl, UnROOT.jl, JetReconstruction.jl, etc. Some of them may be relevant to this project. More can be found on the JuliaHEP GitHub organisation's page [5].

Project Goals

The goal is to have the AGC 1.0 pipeline as well as some supplementary features implemented in Julia. Additional “stretch” goals might include implementing tools for ML inference from beyond the version 1.0.

Proposed Timeline

Starting June, 22nd:

- **Weeks 1-2:** Diving into the workings of the existing AGC implementations, studying the related algorithms.
- **Weeks 3-4:** By the end of this period there should be a working prototype with the core features that runs locally (columnar data processing).
- **Weeks 5-7:** Design systematics handling in analysis, add statistical inference.
- **Weeks 8-9:** Scaling the project and making it run with real data remotely.
- **Week 10:** This period is dedicated to fixing some possible issues and writing documentation for the existing code.
- **Weeks 11-12:** Machine learning stretch goal and additional fixes, writing more documentation, possibly contributing to the other libraries used.
- **Week 13:** Presentation, reports, etc.

References

1. IRIS-HEP. Analysis Grand Challenge. URL: <https://iris-hep.org/projects/agc.html>
2. Analysis Grand Challenge documentation. URL: <https://agc.readthedocs.io/en/latest/>
3. Julia for HEP Mini-workshop: <https://indico.cern.ch/event/1074269/>
4. CHEP 2023: High-performance end-user analysis in pure Julia programming language <https://indico.jlab.org/event/459/contributions/11560/>
5. Julia for HEP on GitHub: <https://github.com/JuliaHEP>