

Measuring Python adoption by CMS physicists using GitHub API

Tuesday, August 11, 2020 10:30 AM (8 minutes)

In the pipeline from detector to published physics results, the last step, “end-user analysis,” is the most diverse. It can even be hard to discover what tools are being used, since the work is highly decentralized among students and postdocs, many of whom are working from their home institutes (or their homes).

However, GitHub offers a window into CMS physicists’ analysis tool preferences. For the past 7 years, CMSSW has been hosted on GitHub, and GitHub’s API allows us to query the public repositories of users who have forked CMSSW, a sample dominated by CMS physicists and consisting of 19,400 user-created (non-fork) repositories.

In these 7 years, we see a clear reduction in the use of C++ and increase in the use of Python and Jupyter notebooks. 2019 marks the first year in which CMS physicists have created more Python repositories (excluding Jupyter) than C or C++. Finally, we can also search the code in these repositories for substrings that quantify the adoption of specific physics, plotting, and machine learning packages.

Understanding how physicists do their work can help us make more informed decisions about software development, maintenance, and training.

Primary author: PIVARSKI, Jim (Fermilab)

Presenter: PIVARSKI, Jim (Fermilab)

Session Classification: Community Feedback