The Register®
Biting the hand that feeds IT

Software

CAST links arms with Software Heritage to tease out your open-source ancestry

Who do you think you are?

By Richard Speed 19 Feb 2019 at 15:57

5 SHARE ▼

with Software Heritage to tackle the sometimes tricky task of identifying the provenance of open-source code in modern projects.

The Register spoke to CAST Software CEO Vincent Delaroche, who told us the aim of the collaboration was to create a "Provenance Index" on code that has been collected in the Software Heritage archive. Essentially, users of its products can fling their source at CAST and be given a list of all components used in the code and, importantly, the original "ancestor" of that component.

"At-risk" components are then automatically flagged and suggestions made on what to do, giving users an opportunity to head off potential legal, IP and compliance nasties before the code seeps out into the hands of users and lawyers.

Behind the Provenance Index is a hookup between CAST Highlight, the company's SaaS platform, which inspects code for iffy practices and vulnerabilities, and the curator of the Software Heritage, which is attempting to collect all publicly available source code along with its development history.

Software Heritage's archive has already "ingested" code from the likes of GitHub, GitLab and the old Google Code archive. Source code archaeologists currently have 5.7 billion source files over 88.3 million projects to cast their eyes over.

CAST emphasises that it doesn't actually slurp any user source code while the process is happening. An agent scans the code and uploads an encrypted file to the CAST Highlight portal. Highlight will cheerfully run on AWS or Azure, or in a private cloud for those organisations needing heightened security or privacy.

How it works

The content of the code archive is open to all, but the process consists of either using a manual search or hitting the service's RESTful web API. That API is still very much a work in progress, and Software Heritage does not consider it to be stable as yet. To cap it off, the API usage is also rate limited, another factor that makes it not ideally suited to inclusion in a modern DevOps pipeline.

Delaroche was keen to position CAST Highlight as an "MRI for software". Yes, we're talking about the technology that involves lying motionless in a narrow tube, fighting off waves of claustrophobia while what sounds like gunfire echoes around you.

To be fair, this hack has been involved in projects that compare unfavourably to the whole MRI experience. (Snark aside, the imagery is meant to explain the insights that can be gleaned from the Highlight software.)

We had a crack at using the free Software Heritage tools to search for code and can confirm that you could get by without using the CAST product at all, but goodness, hunting down code manually takes time.

Since CAST Highlight can be embedded in a DevOps pipeline and used in the CI/CD validation gating process via the command line, making it just another part of the process would certainly bring transparency and reassurance that what is spat out at the other end is sparkly clean from a security and legal perspective.

Of course, this does come at a cost, with 12-month subscriptions starting at $20k for up to 25 applications, rising to $240k for 1,000.

CAST was, however, keen to point out that academics, non-profits and super small businesses (with 10 people of less) can have the thing for free. ®

Tips and corrections

5 Comments

MORE Open Source Devops