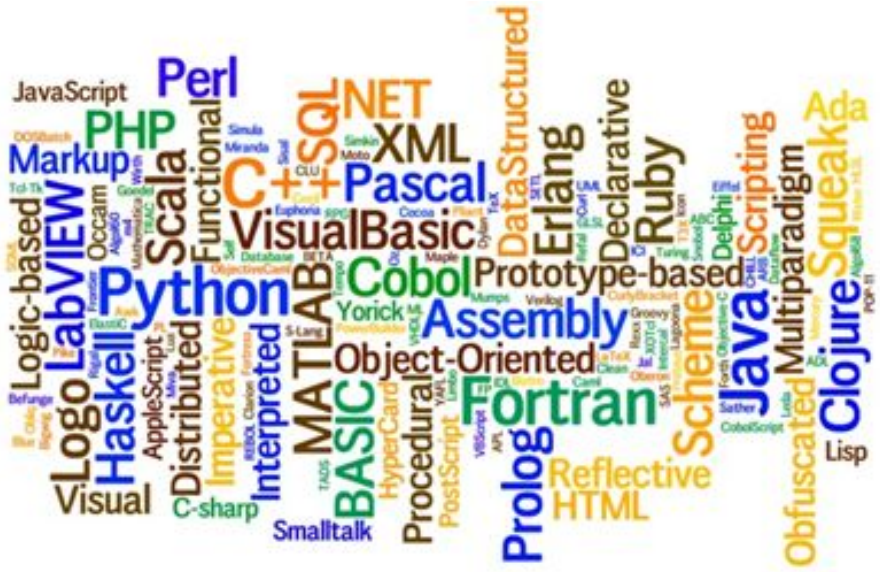
python-inspector: Look Ma No Hands!



About me

- Software Engineer at NexB
- Student at Google Summer of Code 2020
- Long time serving mentor at Google Summer of Code for 3 years
- Speaker at Linux Foundation Open Source Summit North America 2022 and Pycon India 2020

Abstract

python-inspector is a Python library that provides a number of utilities for inspecting Python code and PyPI package manifests. It can be used to:

- Resolve Python dependencies
- Parse various manifests and packages files
- Query PyPI JSON and simple APIs for package information

Agenda

- Why python-inspector
- What does python-inspector do and how does it work
- The internals and the making of python-inspector
- Why python-inspector and not something else ?
- Who is using python-inspector
- Beyond and Next-Steps
- References

Why python-inspector

- Lightweight dependency resolution for Software Composition Analysis (SCA) and Software Bills of Material (SBOMs)
- Build a dependency tree on demand to explore the graph and understand transitive dependencies
- Do What-if? scenarios by tuning the dependency versions and resolve a new graph

What does python-inspector do and how does it work?

- Discussing the type of inputs:
  - Package URL (aka. PURL)
  - requirements.txt / setup.py
- Specifying target OS and python-version
- Understanding the output as JSON tree or pipdeptree compatible mode

DEMO

Internals of python-inspector

- Resolvelib - same as in pip
  - https://github.com/sarugaku/resolvelib
- Package URL spec and library
  - https://github.com/package-url/packageurl-python (I co-maintain)
- pip-requirements-parser (extracted from pip as a standalone lib)
  - https://github.com/nexB/pip-requirements-parser (I co-maintain)
- Config file parsing - using scancode-toolkit's packagecode module
  - https://github.com/nexB/scancode-toolkit (I co-maintain)
- Recursively handling the whole Python distribution tree (sdists and wheels)
- Dynamic dependency resolution (semi-secure evaluation) - allowing dynamic resolution when config files are not parsable

Why python-inspector and not something else to get a dependency graph ?

	python-inspector	**pip**	**pip-audit**
Platform-Independent	Yes	No	No
Requires a build/venv to resolve	No	Yes	Yes
Non Vulnerable Resolution	Yes- By removing vulnerable versions at the time of dependency resolution itself	No	Yes - But by building the tree again and again by removing vulnerable versions every time tree is built

Who is using python-inspector ?

nexB

ScanCode

OSS
Review Toolkit

Package URL aka. PURL

- We created Package URL for VulnerableCode and ScanCode

The critical GLUE between all the software supply chain tools

- Package URL project: https://github.com/package-url
  - Spec is at: https://github.com/package-url/purl-spec
  - Implementations for .NET, Go, Java, JavaScript, PHP, Python, Ruby and Rust
- Recent proposal to add purl to NVD:
  - https://owasp.org/blog/2022/09/13/sbom-forum-recommends-improvements-to-nvd.html

Next steps

- Universal Dependency Resolver
- Better security with Non Vulnerable Dependency Resolution

NVDR: keep the barbarians at the gate!

nexB

- If you could blend
  - **Functional dependency constraints**
  - **Known vulnerable ranges**
- And inject these in a package dependency resolver
- You get

Non Vulnerable Dependency Resolution!

- Working PoC implemented in **python-inspector** tool and paper https://www.tdcommons.org/dpubs_series/5224/

References

- ScanCode
- Package-URL (purl)
- AboutCode

Package-URL

- Package URL project: https://github.com/package-url
  - Spec is at: https://github.com/package-url/purl-spec
  - Implementations for .NET, Go, Java, JavaScript, PHP, Python, Ruby and Rust
- Recent proposal to add purl to NVD:
  - https://owasp.org/blog/2022/09/13/sbom-forum-recommends-improvements-to-nvd.html

ScanCode

- Identify FOSS and other third-party components & packages
- Detect licenses, copyrights and dependencies
- ScanCode Projects include:
  - ScanCode.io: Server system with customizable pipelines
  - ScanCode Toolkit: Scanning engine - use as CLI or library
  - LicenseDB: 2000+ licenses detected by ScanCode
  - ScanCode Workbench: Desktop app to review Scans
- See https://nexb.com/scancode/ for more information

AboutCode

- AboutCode is a platform of FOSS SCA tools
  - Also, the home for our GSoC projects
  - And we are on OpenCollective at: https://opencollective.com/aboutcode
- Our projects are at: https://github.com/nexB
- Documentation for each project is at ReadTheDocs.org
- AboutCode home: https://www.aboutcode.org/

Credits

Special thanks to all the people who made and released these excellent free resources:

- Presentation template by SlidesCarnival at https://www.slidescarnival.com/ licensed under CC-BY-4.0 https://www.slidescarnival.com/terms-of-use#templates-license
- Photographs by Unsplash https://unsplash.com/license licensed under the unsplash license https://scancode-licensedb.aboutcode.org/unsplash.html

All the open source software authors that made VulnerableCode, ScanCode and other AboutCode FOSS projects possible