ScanCode Overview
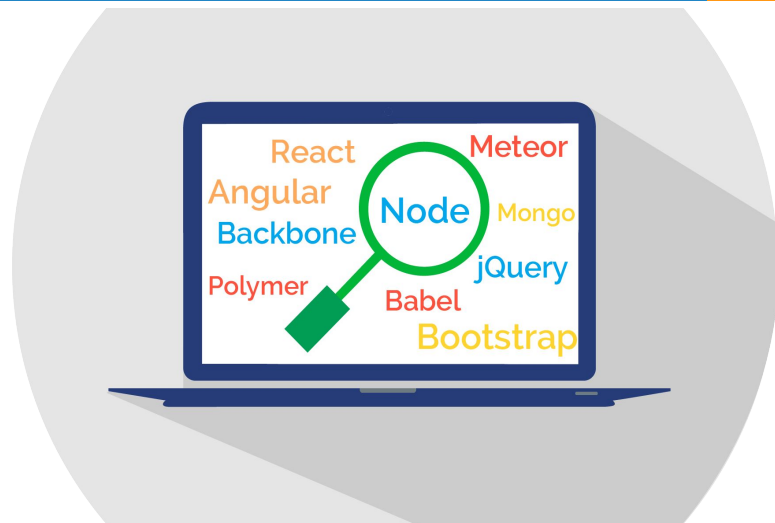
Fall 2019

Topics

▷ Why ScanCode

▷ ScanCode Community

▷ What is ScanCode?

▷ ScanCode Roadmap

Why ScanCode Toolkit

▷ Easy and simple to install and run, self contained

▷ Best in class license & copyright scan accuracy

○ Based on natural language processing

▷ Runs on Linux, Mac OSX and Windows

▷ Find structured package manifest and dependencies

▷ Modern codebase, easy to understand and evolve

Easy to install and run

▷ ScanCode is **easy to install and run**. It runs on Windows, macOS and Linux. It runs on your laptop.
- ○ It can be used by your development teams
- ○ It can be used **by your supply chain partners, big and small**

▷ So easy my mother can install and run it!

Best in class license detection

▷ ScanCode has the most accurate license detection engine
  ○ **Less review needed**
▷ Also collects the full matched notice texts
  ○ Can be used to **automate attribution notice** creation
▷ Other scanning tools miss detecting or misdetect.
  ○ Used by Linux kernel maintainers to clean kernel licensing
  ○ Selected by Here.com for ORT as best in class
  ○ Selected by the Eclipse Foundation for ip due diligence
  ○ Used in ClearlyDefined to scan 6 Million+ FOSS packages

Easy to integrate in your process

▷ Command line tool, minimal dependencies
  ○ Easy to add to your CI/CD process pipeline
▷ Multiple OS support: Windows, macOS and Linux.
▷ SPDX, JSON, CSV outputs

ScanCode community

▷ 750+ stars on GitHub , 200+ forks, 70+ contributors

▷ Reactive bug fixing

▷ Heavily tested with 10,000+ unit and integration tests

▷ Apache-licensed

▷ ScanCode is part of the OpenChain compliance automation tooling group

▷ ScanCode is the referenced scanner in ClearlyDefined, ORT and Quartermaster.

Takeway

▷ ScanCode is **best in class tool** for FOSS scanning and is **easy to install and run**

▷ It **collects all the evidence**: license notices, texts, copyrights and package info

▷ Therefore it should be the tool of choice to use **by your supply chain partners, big and small**

○ Low friction when scanning and sharing scans

Who uses ScanCode

▷ Used at top FOSS orgs and projects
  ○ ClearlyDefined, Debian, Eclipse, FSF, Linux kernel, Object Web, OpenEmbedded.org, Openshift analytics, ORT, Quartermaster, CHAOSS and others.

▷ Used at major companies
  ○ Amazon, Comcast, Facebook, Google, Here.com, Microsoft, Red Hat, VMware and others

Some issues with ScanCode

▷ Like many open source projects the documentation is something that is lacking a bit …. BUT

▷ We have a tech writer starting today courtesy of the **Google Seasons of Docs** and we expect vastly improved documentation over the next three months!

What is ScanCode?

Open source tools for open source compliance

▷ Toolkit - Identify software origin and license from the code

https://github.com/nexB/scancode-toolkit

▷ Workbench - Review scans and conclude licensing

https://github.com/nexB/scancode-workbench

▷ Licenses

○ Software - Apache 2.0 (SPDX id:Apache-2.0)

○ License Data - Creative Commons Public Domain (SPDX id: CC0-1.0)

Other companions tools and projects

▷ DeltaCode - compare two scans

▷ AboutCode toolkit - generate attribution notices

▷ TraceCode - trace your build: what code is used and how

▷ VulnerableCode - The free correlated vulnerabilities DB

▷ conan - Analyze Docker images packages

Next steps

▷ New version 3.x about to be released

▷ We are starting monthly community requirement planning online meeting to plan for 4.x

ScanCode Roadmap

▷ Deduction and inference for scan conclusions
  ○ Traceable Machine Learning
▷ Open source scancode.io - server for scanning
▷ License detection in other languages, beyond English
▷ VulnerableCode - the free vulnerabilities DB is now funded by the EU and NLnet
▷ SPDX Lite support?

Credits

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

▷ Presentation template by [SlidesCarnival](#)

▷ Photographs by [Unsplash](#)

▷ And all the software authors that made ScanCode possible