ScanCode Overview
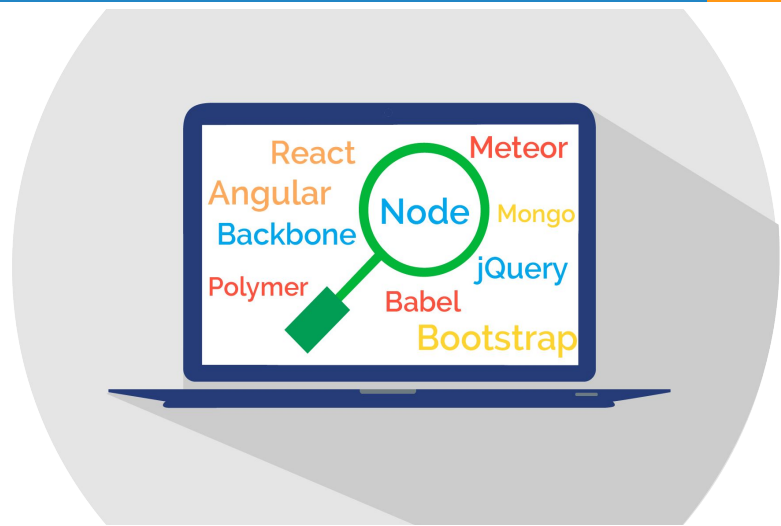
Q2 2019

Topics

▷ What is ScanCode?
  ○ ScanCode Toolkit
  ○ ScanCode Workbench
▷ ScanCode Community
▷ ScanCode Roadmap
▷ About nexB

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)

ScanCode Toolkit

▷ Best in class license & copyright scan accuracy

▷ Scan files or packages

▷ Plugin architecture

▷ Command line tool

▷ Linux, Mac OSX and Windows

ScanCode Toolkit

Detect provenance (origin and license) data from files, packages or package manifests

▷ Copyright detection based on natural language processing
▷ License detection based on automatons, inverted indexes and multi-diffs
   ○ Public repository of license rules and samples
   ○ Add/correct detections by adding/correcting rule or samples - not code
▷ JSON, CSV, SPDX and other output formats

ScanCode Toolkit [2]

Other features

▷ Detect authors, URLs and email addresses

▷ Report copyright holders to summarize copyright notices

▷ Plugin architecture for "pre" or "post" extensions - good for filters, summarization or other…….

▷ DeltaCode to compare Scans

▷ "Universal" archive extractor

ScanCode Workbench

▷ Visualize Scan data

▷ Document license conclusions

▷ Electron-based desktop application

▷ Linux, Mac OSX and Windows
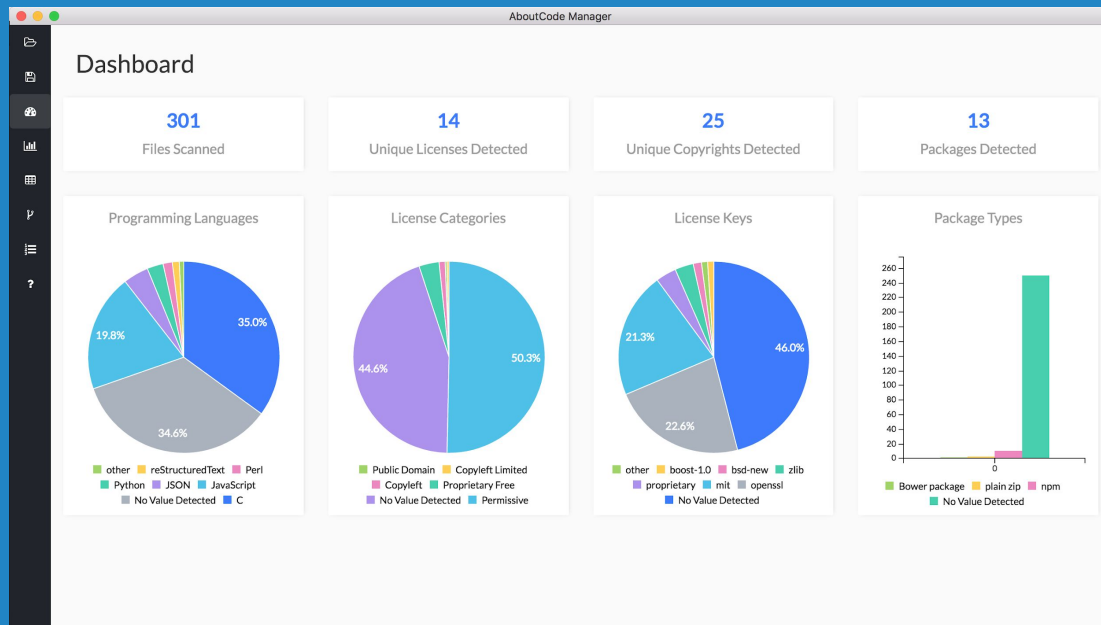
ScanCode Workbench

▷ Tree View - see and navigate codebase hierarchy

▷ Dashboard View - visualizations showing the number of Files Scanned and Licenses, Copyrights and Programming Languages detected

▷ BarChart View - bar charts showing summary data for Copyrights, Licenses and other file data

▷ Table View - DataTable for all Scan data
- ○ Configure columns displayed by set or individually
- ○ Set filters on any column

ScanCode Workbench [2]

▷ Conclusions View
- ○ Option to record your concluded license and copyright holder or other fields - very useful for summarisation
- ○ Most fields are pre-filled from Scan data
- ○ Export as draft Inventory to other systems

▷ Other
- ○ WB converts JSON file to SQLite database for use within WB
- ○ Some users use a SQL-DB tool to query the data separately from WB

ScanCode Community

▷ 69+ contributors, 642+ stars, 191+ forks

▷ Used at major tech companies - Amazon, Facebook, Google, Red Hat and others

▷ Used at top FOSS orgs - ClearlyDefined, Debian, Eclipse, FSF, Linux kernel, ORT, Quartermaster, Bitergia/CHAOSS and others.

▷ Google Summer of Code organization: Evaluating several projects for GSoC 2019

ScanCode Release 3.1

New major release planned for May

▷ Feature complete for both TK and WB
▷ Completed new package manifest scanning for Bower, Composer, Maven, npm, NuGet, RubyGems and many others
▷ Many improvements in TK license detection
▷ Many improvements in Workbench Views and Filters

ScanCode Roadmap

▷ Deduction and inference for complex license expressions
  ○ Traceable Machine Learning
  ○ Starting with extensive prototyping
▷ More plugins!
▷ Open source scancode.io - server configuration for ScanCode TK
▷ Other
  ○ Complete port to Python3

About nexB

▷ Our mission is to make it easier to reuse FOSS
  ○ Open source solutions for open source compliance
  ○ ScanCode, AboutCode, TraceCode and other projects
▷ Bootstrap company based in Silicon Valley
  ○ DejaCode enterprise compliance system (commercial)
  ○ Acquisition and product audit/analysis services
  ○ Working on FOSS compliance since 2007

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