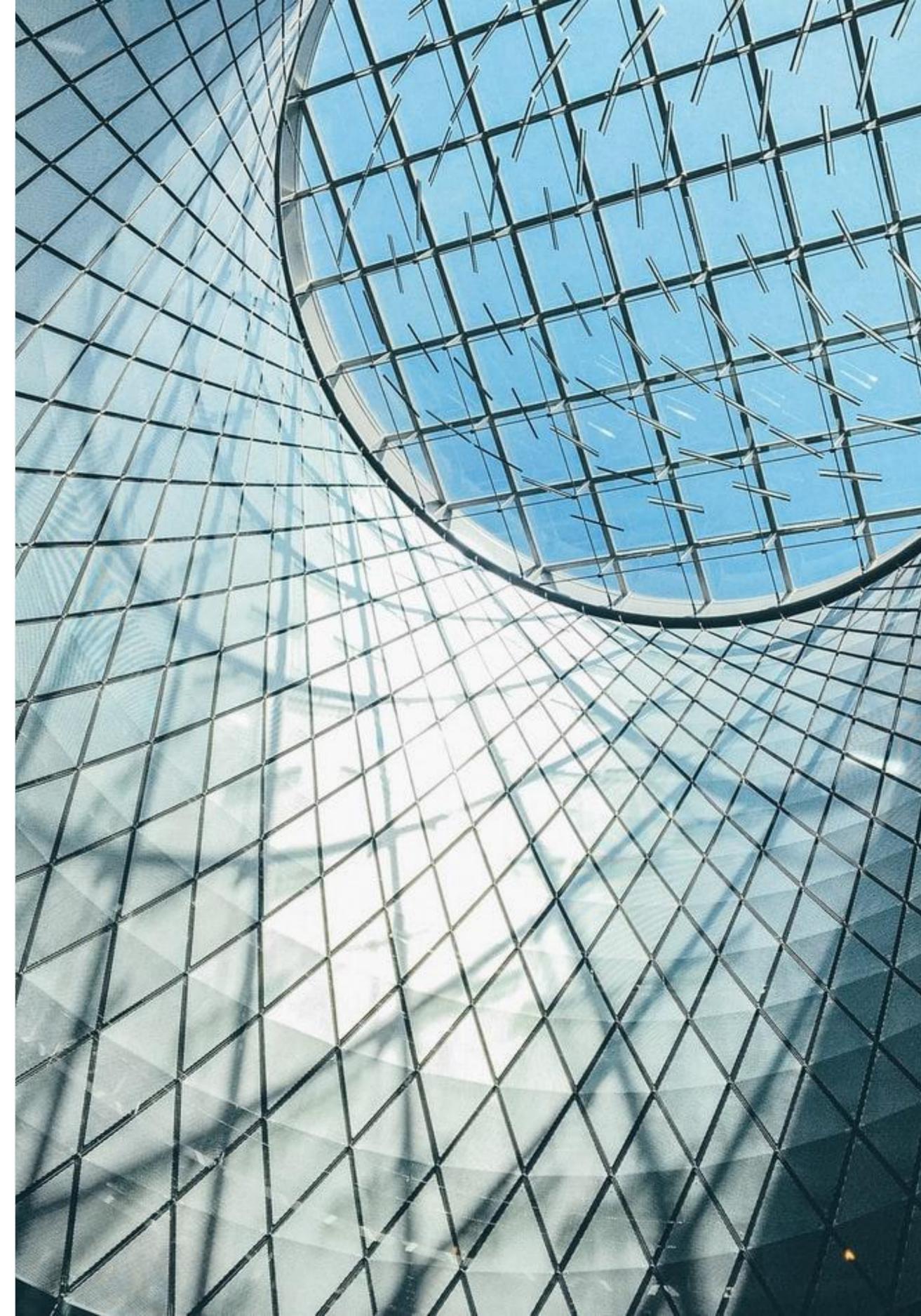


ActiveState

Python Workshop

Improving Productivity and Security with the
ActiveState Platform

Sep 16th, 2021



Housekeeping

Team: Avery & Dana

- 40 minutes **Follow Along** workshop
- 20 minutes for **Q & A** in the end

Keep an eye on the **Chat** for links and instructions

Post workshop, head to the **ActiveState Community Forum** for support -
<https://community.activestate.com/>

Agenda: Python Project + real world ML use case

Topics:

- Package management
- Reproducibility
- Security
- API

Tools:

- ActiveState Platform
- State Tool
- Windows or Linux machines

What the Goal of the Project is:

We want to determine which open source licenses are included in the software we use. The problem is, unless we were to read every license file one by one we can't be sure the information stated on PyPI is accurate.

PKG-INFO

```

Metadata-Version: 1.1
Name: configobj
Version: 5.0.6
Summary: Config file reading, writing and validation.
Home-page: https://github.com/DiffSK/configobj
Author: Rob Dennis, Eli Courtwright (Michael Foord & Nicola Larosa)
Author-email: rdennis+configobj@gmail.com, eli@courtwright.org
License: UNKNOWN
Description: ConfigObj is a simple but powerful config file round tripper. Its main feature is that it is straightforward programmer's interface and a simple. It has lots of other features though :

```

Validate.py

```

# validate.py
# A Validator object
# Copyright (C) 2005-2014:
# (name) : (email)
# Michael Foord: fuzzyman AT voidspace DOT org DOT uk
# Mark Andrews: mark AT la-la DOT com
# Nicola Larosa: nico AT tekNico DOT net
# Rob Dennis: rdennis AT gmail DOT com
# Eli Courtwright: eli AT courtwright DOT org

# This software is licensed under the terms of the BSD license.
# http://opensource.org/licenses/BSD-3-Clause

# ConfigObj 5 - main repository for documentation and issue tracking:
# https://github.com/DiffSK/configobj

```

Meta

License: UNKNOWN

Author: [Jeremy Dobbins-Bucklad](#)

Getting Started: Creating our Python Runtime

1. Install tools: State Tool
2. Determine packages needed.
3. Use the ActiveState Platform to create Python runtime.
4. Install the runtime using the State Tool & spin up a virtual environment
5. Create license report by running newly created script.
6. Eliminate the README, share ActiveState runtime with a github project attached.

Natural Language Processing: Text Comparison using Cosine Similarity

Steps:

1. Connect to the ActiveState Platform and download source code from a list of packages.
2. Get important information such as name and version
3. Extract compressed files
4. Find license text and PKG-INFO (for comparison)
5. Find copyright holders
6. Transform the documents into vectors using TF-IDF
7. Measure their Cosine Similarity
8. Compare to a list of known open source licenses
9. Choose closest match
10. Generate CSV report

Let's Go!

Summary

- Created a secure Python runtime from source code.
- Works on multiple operating systems
- Managed multiple Python versions on one machine
- Used the ActiveState Platform, State Tool & GraphQL API
- Learned which licenses we are using in our open source software using NLP
- Had a great time! :)

Q & A

Resources + Next Steps

- Keep the conversation going at:
<https://community.activestate.com/t/python-devs-webinar-sep-16-2021/7225>
- Try it with other Python distributions
 - See how vulnerable you are
 - See what licenses you're using

Resources:

- Platform Documentation: <https://docs.activestate.com/platform/>
- StateTool Documentation: <https://docs.activestate.com/platform/state/>
- Build Graph API Documentation: <https://buildgraph-docs.activestate.build/>
- Video Tutorials: <https://www.activestate.com/resources/product-demos/>
- ActiveState Blogs: <https://www.activestate.com/blog/>