Thoth

A recommendation engine for Python applications

Fridolin Pokorny <fridolin@redhat.com>
2020-02-01
FOSDEM 2020
$ whoami

https://fridex.github.io

- Fridolín “fridex” Pokorný
- Senior Software Engineer at Red Hat
- Distributed systems, AI/ML and (of course) Python fan
- Projects:
  - Reverse engineer RetDec (AVG)
  - Linux kernel TLS/DTLS module **AF_KTLS**
  - **Selinon** - distributed task flows scheduler on top of Celery
  - Project **Thoth**
What is Thoth?

Why Thoth?
Why Thoth?

- PyPI - Python Package Index
  - [https://pypi.org/](https://pypi.org/)
  - 215,218 projects
  - 1,645,362 releases (approx. 7 releases per project)
Why Thoth?

```python
import tensorflow as tf
from flask import Flask
application = Flask()
sess = tf.Session()
```
Why Thoth?

Python application

Direct Python dependencies

Transitive Python dependencies

Native dependencies

Python interpreter

Kernel modules

Operating System

Hardware
Why Thoth?

- Python application
  - Direct Python dependencies
  - Transitive Python dependencies
  - Native dependencies
  - Python interpreter
  - Kernel modules
    - Operating System
  - Hardware
Transitive dependencies

- Flask (33)
  - click, itsdangerous, jinja2, markupsafe, werkzeug

Estimated number of combinations: 54,395,000
Transitive dependencies

- TensorFlow (85)
  - absl-py, astor, backports-weakref, bleach, enum34, gast, google-pasta, grpcio, h5py, html5lib, keras, keras-applications, keras-preprocessing, markdown, mock, numpy, pbr, protobuf, pyyaml, scipy, setuptools, six, tensorboard, tensorflow-estimator, tensorflow-tensorboard, termcolor, tf-estimator-nightly, werkzeug, wheel

Estimated number of combinations: $139,740,802,927,165,440,000$

approx. $1.39 \times 10^{20}$

- number of possible game positions is around:
  - $10^{172}$

Flask, TensorFlow and mathematics

- Number of possible Python software stacks is around
  - $54,395,000 \times 1.39 \times 10^{20} = 7.6 \times 10^{20}$ (rough estimate)
Why Thoth?

- Python application
  - Direct Python dependencies
  - Transitive Python dependencies
  - Native dependencies
  - Python interpreter
  - Kernel modules
  - Operating System
  - Hardware
Why Thoth?

How does Thoth work?

1. Python application
2. Direct Python dependencies
3. Transitive Python dependencies
4. Native dependencies
5. Python interpreter
6. Kernel modules
7. Operating System
8. Hardware
How good is my software stack?

simplelib
anotherlib
Overall stack score

Different versions of “simplelib”

Different versions of “anotherlib”
Why Thoth?

- Create knowledge base
  - What packages in which versions should I use?
    - Application builds correctly
    - Application runs correctly
    - Application behaves and performs well
- Create an advanced Python resolver which uses knowledge base to resolve software stacks

*Latest versions are not always greatest choices.*
Thoth’s adviser

- Server side resolution

- Multiple iterations on implementation

- Pure Python implementation
  - Memory consumption
  - N-ary graph with transactional operations

- Rewritten into C/C++
  - Too many queries to database
  - Cca. 2.5k queries just to obtain TensorFlow dependency graph
  - The main database changed 2 times
Thoth’s adviser

- Later stochastic approaches - Operations Research
  - Hill climbing
  - Adaptive Simulated Annealing

- Implementation split into two parts
  - Resolver
    - Resolve software stacks respecting Python ecosystem
  - Predictor
    - Guide resolver in resolution
Thoth’s adviser

- Reinforcement Learning - Gradient-based methods
  - Not responsive enough

- Reinforcement Learning - Gradient-free methods
  - Temporal Difference, Monte Carlo Tree Search

- Reconfigurable pipeline made out of units
  - Units define scoring function (units of type step)
  - Units define action space (units of type sieve)

- Dependency Monkey
  - Sample state space to gather “observations”
Running application from script (app.sh) ...

2020-01-28 21:25:87,912  1 INFO  thoth.adviser:160: Version: 0.7.3
2020-01-28 21:25:87,918  1 INFO  thoth.adviser:423: Starting resolver using 'ApproximatingLatest' predictor with random seed set to 142
2020-01-28 21:25:88,972  90 INFO  thoth.adviser.resolver:729: Hold tight, Thoth is computing recommendations for your application...
2020-01-28 21:25:14,690  90 WARNING thoth.adviser.resolver:68: No solved releases found for 'numpy' which would satisfy version requirements of ('scipy', '1.4.1', 'https://pypi.org/simple')
2020-01-28 21:25:14,690  90 WARNING thoth.adviser.predictors.latest:52: The latest stack couldn't be resolved, performing hops across package versions
2020-01-28 21:25:14,862  90 WARNING thoth.adviser.resolver:68: No solved releases found for 'protobuf' which would satisfy version requirements of ('tensorboard', '2.1.0', 'https://pypi.org/simple')
2020-01-28 21:25:14,862  90 WARNING thoth.adviser.resolver:68: No solved releases found for 'absl-py' which would satisfy version requirements of ('tensorboard', '2.1.0', 'https://pypi.org/simple')
2020-01-28 21:25:14,862  90 WARNING thoth.adviser.resolver:68: No solved releases found for 'grpcio' which would satisfy version requirements of ('tensorboard', '2.1.0', 'https://pypi.org/simple')
2020-01-28 21:25:15,594  90 INFO  thoth.adviser.resolver:834: Pipeline reached 1 final states out of 1 requested in iteration 53 (pipeline pace 8.14 stacks/second)
2020-01-28 21:25:15,928  90 INFO  thoth.adviser.resolver:747: Reached limit of stacks to be generated - 1 (limit is 1), stopping resolver with the current beam size 43 in iteration 53
2020-01-28 21:25:15,929  90 INFO  thoth.adviser.resolver:884: Resolver took 7.82126 seconds in total, pipeline speed 0.127857 stacks/second
2020-01-28 21:25:15,929  90 INFO  thoth.adviser.resolver:854: Pipeline strides discarded 0 and accepted 1 final states in total

End of log.
Resolver took 167.332 seconds in total, pipeline speed 2988.07 stacks/second

2.7 minutes
$ time pipenv install tensorflow flask
Installing tensorflow...
Adding tensorflow to Pipfile's [packages]...
✓ Installation Succeeded
Installing flask...
Adding flask to Pipfile's [packages]...
✓ Installation Succeeded
Pipfile.lock not found, creating...
Locking [dev-packages] dependencies...
Locking [packages] dependencies...
Building requirements...
Resolving dependencies...
✓ Success!
Updated Pipfile.lock (a1a482)!
Installing dependencies from Pipfile.lock (a1a482)...
39/39 - 00:00:23
To activate this project's virtualenv, run pipenv shell.
Alternatively, run a command inside the virtualenv with pipenv run.
114.05s user 11.38s system 196% cpu 1:03.99 total
Thoth parts...

- Bots automating routing tasks
  - Updates of dependencies
  - New releases
- Optimized TensorFlow releases
  - [https://tensorflow.pypi.thoth-station.ninja/](https://tensorflow.pypi.thoth-station.ninja/)
- Topics modeling on Python package metadata
- Dependency Monkey + Adviser
- Static source code analysis
- Container image analysis
- Integration with OpenShift, Jupyter Notebooks, CLI
- ...
Application

Qeb-Hwt

I'm Kebechet bot, goddess of freshness - I will keep your source code fresh and up-to-date.

Read more...

Qeb-Hwt

Qeb-Hwt will use the Thoth Service API to get advises on your application stack.

Build Environment
We have analysed an application stack running on Red Hat Enterprise Linux 8.1 (Ondra), running Python (cpython) 3.6.8. It was Adviser Job ID adviser-4499474, by thoth-analysis.

Our General Assessment
We...

Your Pipfile

`dev-packages = {}`

`[packages]`

`tensorflow = "$"`

`[[requires]]`

`name = "pip installs"`

`url = "https://pypi.org/simple"`

`verify_ssl = true`

Our Pipfile.lock

`
{ 
  "meta": {
    "pip": "20.2.3",
  }
```
Information about Thoth

- **Thoth Bot**

- **Website:**
  - [https://thoth-station.ninja/](https://thoth-station.ninja/)

- **Twitter**
  - [https://twitter.com/thothstation](https://twitter.com/thothstation)

- **GitHub**
  - [https://github.com/thoth-station](https://github.com/thoth-station)
THANK YOU

plus.google.com/+RedHat
facebook.com/redhatinc
linkedin.com/company/red-hat
twitter.com/RedHat
youtube.com/user/RedHatVideos