

Open Source Licensing Course

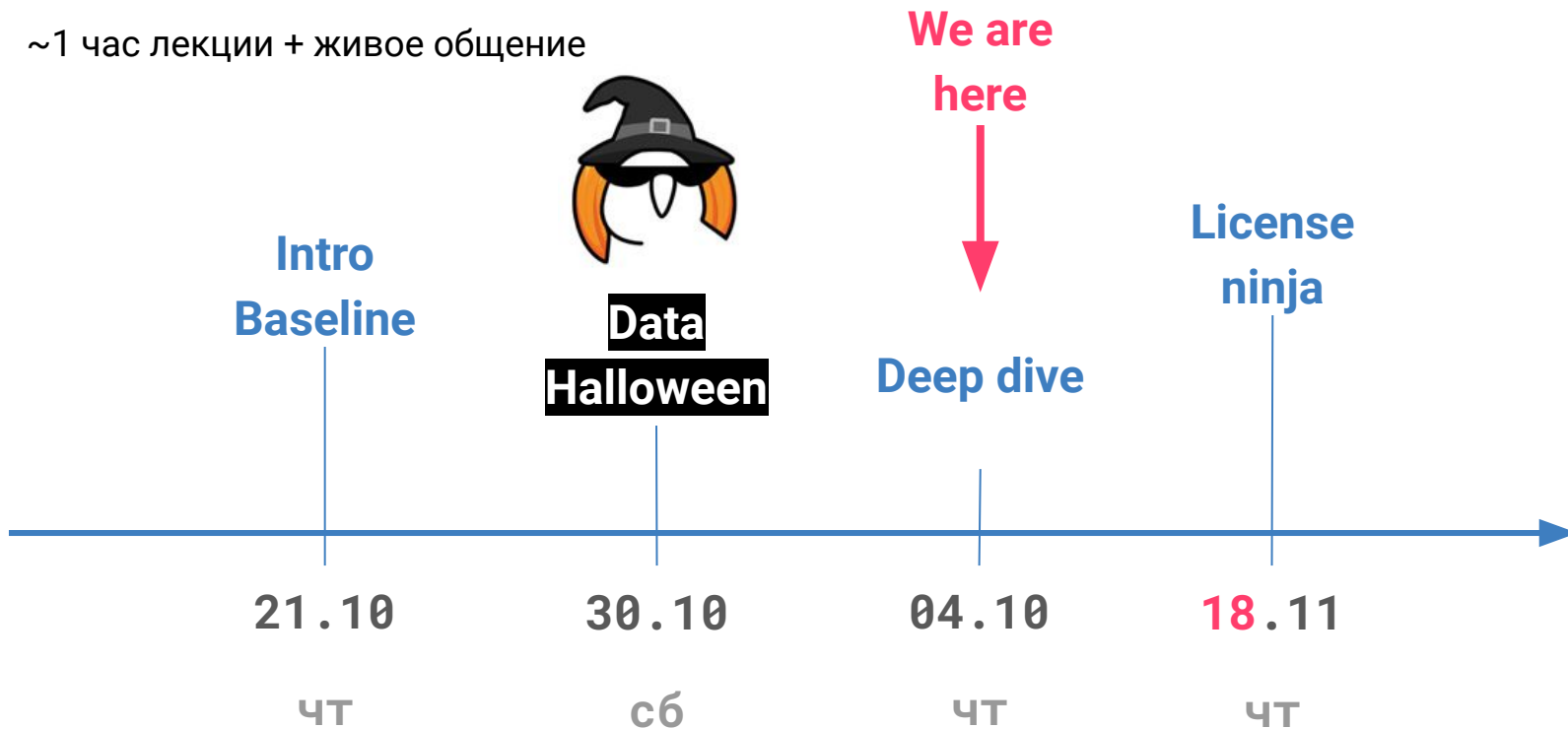
ODS Course Fest #1, 2021

ods.ai | profiscope.io
Code Mining track



Timeline

~1 час лекции + живое общение



В этой серии

- разбор лицензионной чистоты на примере пакета requests
- что делать когда лицензия отсутствует
- какие бывают кастомные наклеи к лиц. соглашениям
- как учитывать патенты
- как учитывать окружения зависимостей
- какие бывают случаи смены лицензии



Лицензионная
несовместимость
на примере пакета
requests

Как отслеживать лицензионную чистоту?

- ручками `^\\(ツ)^/`
- open source-решения
- платные решения

Лицензионная чистота на примере

Requests

(apache 2.0)

46k ★★★★★

github.com/psf/requests | pypi.org/project/requests

Не всегда зависимости указаны в манифесте

Разработчики Requests ведут разработческие зависимости в манифесте `requirements-dev.txt`:

```
pytest>=2.8.0,<=6.2.5
```

```
pytest-cov
```

```
pytest-httpbin==1.0.0
```

```
pytest-mock==2.0.0
```

```
httpbin==0.7.0
```

```
Flask>=1.0,<2.0
```

```
trustme
```

```
wheel
```

Зависимости продукта в `requests/setup.py`

```
requires = [  
    'charset_normalizer~=2.0.0; python_version >= "3"',  
    'chardet>=3.0.2,<5; python_version < "3"',  
    'idna>=2.5,<3; python_version < "3"',  
    'idna>=2.5,<4; python_version >= "3"',  
    'urllib3>=1.21.1,<1.27',  
    'certifi>=2017.4.17'  
]
```

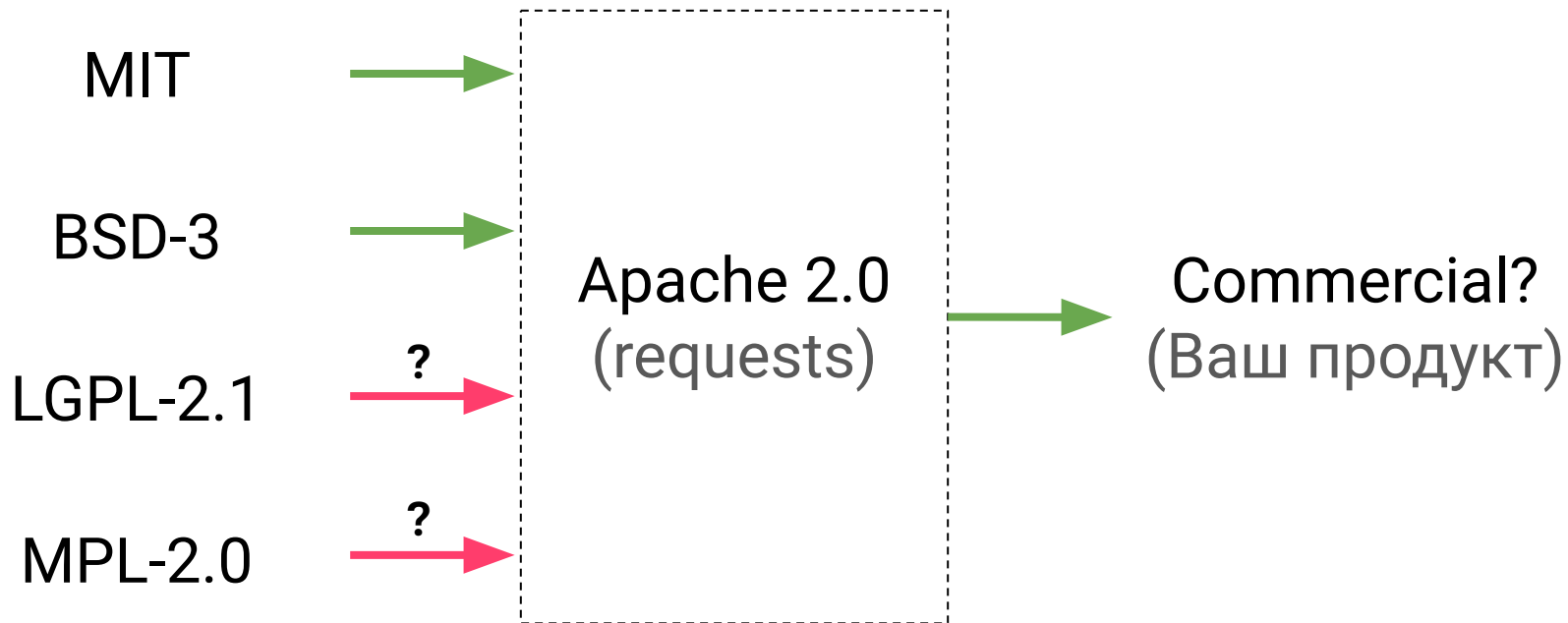

Зависимости requests табличкой

Пакет	Версия	Интерпретатор Python
charset_normalizer	~=2.0.0	>3
chardet	>=3.0.2,<5	<3
idna	>=2.5,<3	<3
idna	>=2.5,<4	>=3
urllib3	>=1.21.1,<1.27	both
certifi	>=2017.4.17	both

Зависимости requests табличкой + лицензии

Пакет	Версия	Интерпретатор Python	Лицензия
charset_normalizer	~=2.0.0	>3	MIT
chardet	>=3.0.2,<5	<3	LGPL 2.1
idna	>=2.5,<3	<3	BSD-3
idna	>=2.5,<4	>=3	BSD-3
urllib3	>=1.21.1,<1.27	both	MIT
certifi	>=2017.4.17	both	MPL-2.0

Совместимость лицензий (license compliance)



Совместимость LGPL-2.1 → Apache 2.0 (py<3)

Apache прямо запрещает использование LGPL-{2, 2.1, 3}:

<https://www.apache.org/legal/resolved.html#category-x>

Одно из обсуждений проблемы в проекте: <https://github.com/psf/requests/issues/5871>

Полезное про лицензию: <https://fossa.com/blog/open-source-software-licenses-101-lgpl-license/>

Пример качества текстов лицензий пакетов

This package contains a modified version of ca-bundle.crt:

ca-bundle.crt -- Bundle of CA Root Certificates

Certificate data from Mozilla as of: Thu Nov 3 19:04:19 2011#

This is a bundle of X.509 certificates of public Certificate Authorities (CA). These were automatically extracted from Mozilla's root certificates file (certdata.txt). This file can be found in the mozilla source tree:

<http://mxr.mozilla.org/mozilla/source/security/nss/lib/ckfw/builtins/certdata.txt?raw=1#>

It contains the certificates in PEM format and therefore can be directly used with curl / libcurl / php_curl, or with an Apache+mod_ssl webserver for SSL client authentication. Just configure this file as the SSLCACertificateFile.#

certifi
MPL-2.1

***** BEGIN LICENSE BLOCK *****

This Source Code Form is subject to the terms of the Mozilla Public License, v. 2.0. If a copy of the MPL was not distributed with this file, You can obtain one at <http://mozilla.org/MPL/2.0/>.

***** END LICENSE BLOCK *****

@(#) \$RCSfile: certdata.txt,v \$ \$Revision: 1.80 \$ \$Date: 2011/11/03 15:11:58 \$

Know the license, это бесплатно



Know the license

[Text](#) [Git repository \(coming soon\)](#)

[Mozilla Public License 2.0](#) Confidence Score **100.0**

Input text here

|This package contains a modified version of ca-bundle.crt:

ca-bundle.crt -- Bundle of CA Root Certificates

knowthelicense.codescoring.com

Совместимость MPL-2.0 → Apache 2.0

- Если (внезапно) модифицируете код **certifi**, то его модификацию придется открыть под MPL-2.0.
- Тем не менее, совместимость **есть**.

Полезное: <https://fossa.com/blog/open-source-software-licenses-101-mozilla-public-license-2-0/>

Совместимость лицензий requests

Python <3

Лиц. совместимости нет!

Что делать:

попробовать не использовать
chardet или перейти на py3.

Python >3

Лиц. совместимость есть!

Что делать:

жить с open source дальше.



Тонкости и кейсы совместимости лицензий

Отсутствие лицензии

Внимание!

Если в коде проекта отсутствует лицензия, его использовать нельзя!

Что делать:

1. проверить что указания лицензии от автора нигде нет;
2. связаться с автором и попросить добавить лицензию;
3. искать альтернативные проекты `¯_(\ツ)_/¯`.

Модификации лицензий

Внимание!

Важно проверять, что используемые библиотеки не имеют дополнительных требований и условий, которые вас могут обязать и/или ограничить.

Например:

1. установление отдельных требований на предмет объемов и вида использования: только для исследовательских целей или только для коммерческих.
2. Введение территориального ограничения. Например использование продукта на спорных территориях.

PS. Информация по этому пункту будет раскрыта на следующей лекции.

Учёт окружения зависимостей

Внимание!

Убедитесь, что вы проверяете зависимости именно runtime, то есть продуктивное окружение, а не тестовое или разработческое.

Разработчики могут использовать разные окружения зависимостей в процессе создания продукта, но в конечном продукте должна сохраняться лицензионная чистота.

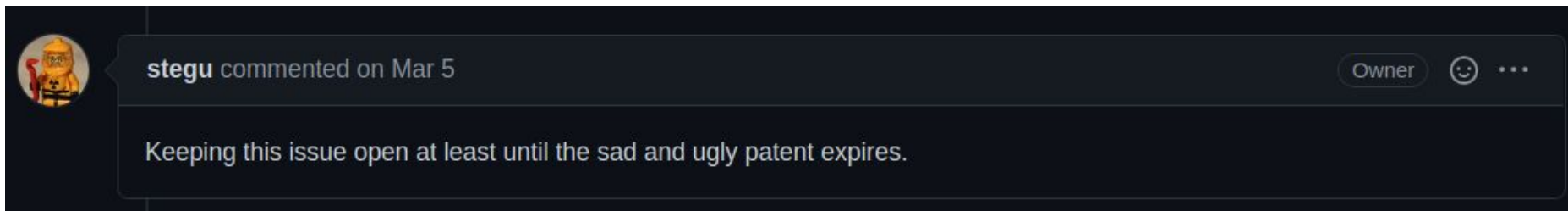
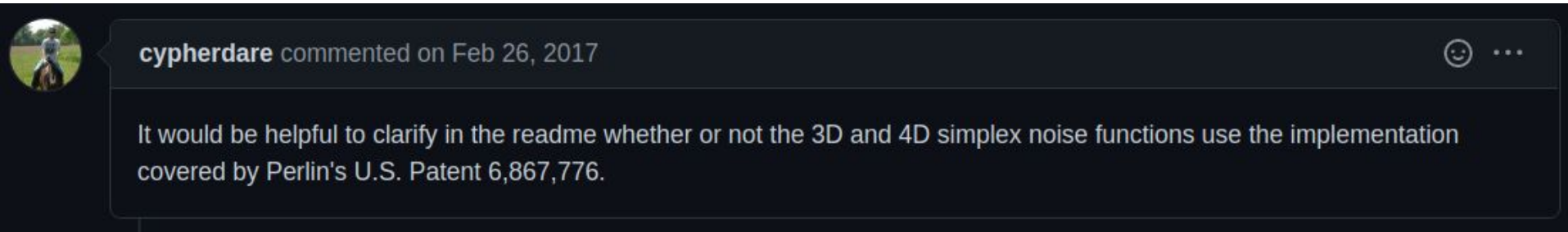
Учёт патентов

Внимание!

1. Важно проверять, что используемые библиотеки не нарушают действующие патенты.
2. Код не может быть запатентован, только изобретение (например алгоритм).

PS. Об этом будет домашнее задание.

Пример “webgl-noise” (MIT)



Проект: <https://github.com/ashima/webgl-noise> (2k ***)

Форк: <https://github.com/stegu/webgl-noise> (267 ***)

Тикет: <https://github.com/stegu/webgl-noise/issues/4>

Патенты и лицензии. Полезная информация



Google Open Source

Google Open Case Book

<https://google.github.io/opencasebook/patents>

<https://github.com/google/opencasebook>

- Patents in Open Source
 - Introduction
 - Statutory Law
 - 17 U.S.C. § 106 - Exclusive rights in copyrighted works
 - 35 U.S.C. 271 - Infringement of patent
 - Explicit Patent Licensing
 - Selections from Open Source Licenses
 - Academic Free License 3.0
 - Apache 2.0
 - Eclipse Public License version 2.0
 - GNU General Public License 2.0
 - GNU General Public License 3.0
 - Mozilla Public License 2.0
 - Discussion
 - Implied Patent Licenses
 - Cases
 - De Forest Radio Telephone Co. v. United States
 - Wang Laboratories, Inc. v. Mitsubishi Electronics America, Inc.
 - A. C. Aukerman Co. v. R. L. Chaides Construction Co
 - Discussion
 - Limitations on Implicit Patent Licenses
 - State Contracting & Engineering Corp. v. Florida
 - Hilgraeve Corp. v. Symantec Corp.
 - Discussion

Где искать информацию о патентах

Есть разные сервисы, но самый простой путь: <https://patents.google.com/>

Google Patents

machine learning

🔍

SEARCH TERMS ⓘ

machine learning x or + Synonym

+ Synonym

SEARCH FIELDS

Date · Priority ▾

YYYY-MM-DD — YYYY-MM-DD

+ Inventor

+ Assignee

Patent Office ▾ Language ▾

Status ▾ Type ▾

Litigation ▾


×

About 150,192 results

Download ▾ Side-by-side

Sort by · Relevance ▾ Group by · None ▾ Deduplicate by · Family ▾ Results / page · 10 ▾

Artificial intelligence and machine learning infrastructure




US · [US10649988B1](#) · Brian Gold · Pure Storage, Inc.

Priority 2017-10-19 · Filed 2018-07-27 · Granted 2020-05-12 · Published 2020-05-12

12. The artificial intelligence and machine learning infrastructure system of claim 11, wherein the query conforms to an application programming interface implemented by the storage system. 13. The artificial intelligence and machine learning infrastructure system of claim 11, wherein the ...

On-the-fly deep learning in machine learning at autonomous machines




US CN · [US11120304B2](#) · Raanan Yonatan Yehezkel Rohekar · Intel Corporation

Priority 2017-05-05 · Filed 2020-07-15 · Granted 2021-09-14 · Published 2021-09-14

21. The non-transitory machine-readable medium as in claim 20, wherein the deep learning framework causes the general-purpose graphics processor to perform operations to merge the first DNN model with the second DNN model into a joint model. 22. The non-transitory machine-readable medium as in

Data intelligence using machine learning




WO US · [US9646262B2](#) · Kelly D. Philipps · Purepredictive, Inc.

Priority 2013-06-17 · Filed 2014-04-30 · Granted 2017-05-09 · Published 2017-05-09

Apparatuses, systems, methods, and computer program products are presented for performing data analytics using machine learning. An unsupervised learning module is configured to assemble an unstructured data set into multiple versions of an organized data set. A supervised learning module is ...

Machine learning device, numerical control device and machine learning method ...




US CN JP DE · [US10585417B2](#) · Kanta TAKAYAMA · Fanuc Corporation






Priority 2016-06-09 · Filed 2017-05-31 · Granted 2020-03-10 · Published 2020-03-10

wherein the machine learning device is on a fog server. 6. The machine learning device according to claim 1, wherein the machine learning device is connectable to at least one other machine learning device, and mutually exchanges or shares an

Top 1000 results by filing date



Relative count of top 5 values

Assignees	Inventors	CPCs
 Intel Corporation		3.6%
 Fanuc Corporation		2.5%
 Microsoft Technology Licensing, LLC		2.4%
 International Business Machines Corporation		2.4%
 Cisco Technology, Inc.		2.2%
Expand		

24



Смена лицензии (на примере РУПИ)

Лицензия действует на конкретную версию

Внимание!

Если авторы меняют лицензию, то она начинает действовать с момента её установки.

На предыдущие версии ПО она не распространяется.

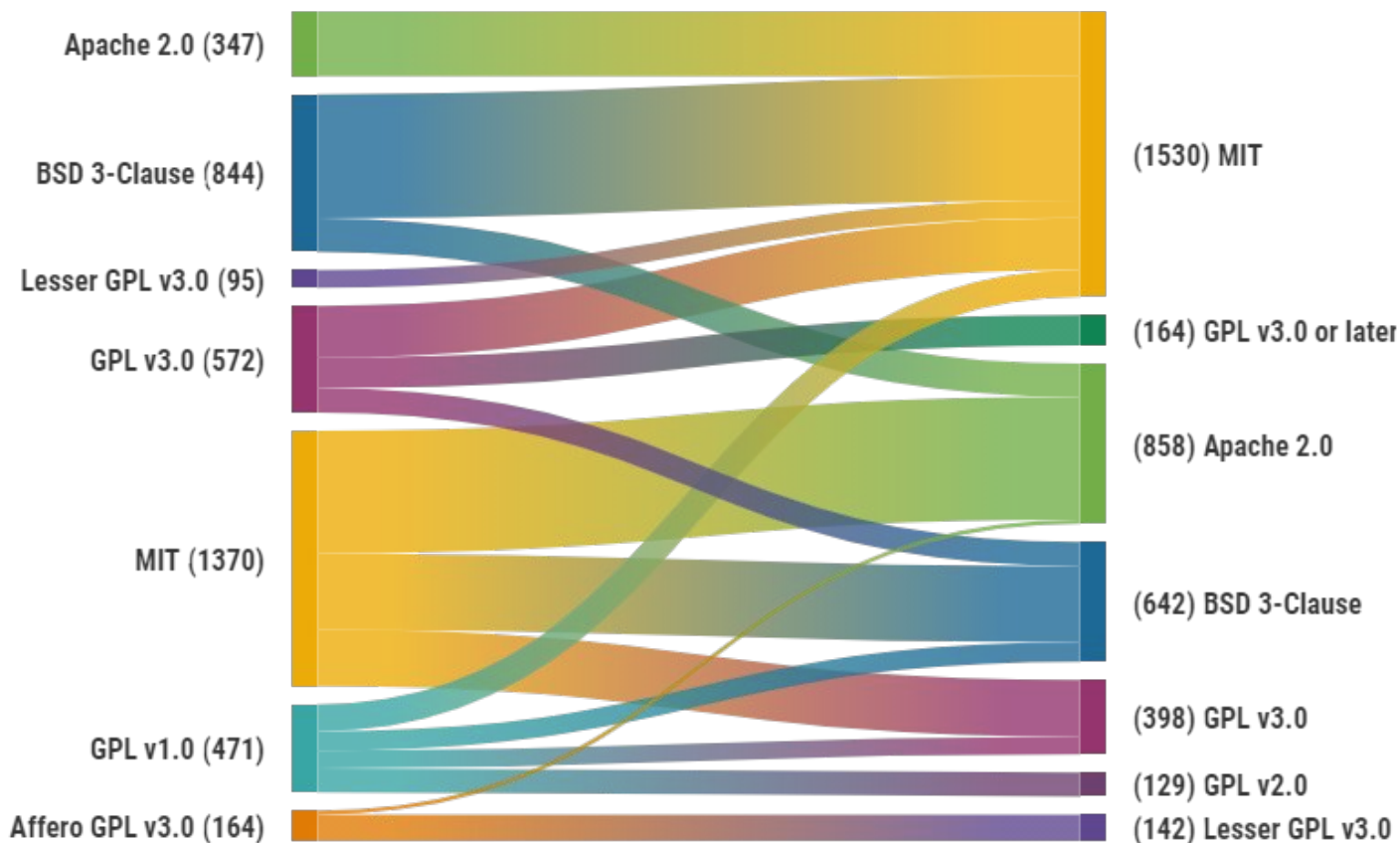
Смена лицензирования пакетов в PyPI

Чистых* случаев смены лицензии — 3842:

- ~30% после выпуска первой версии
- ~70% в последующей жизни (или под её закат ;)

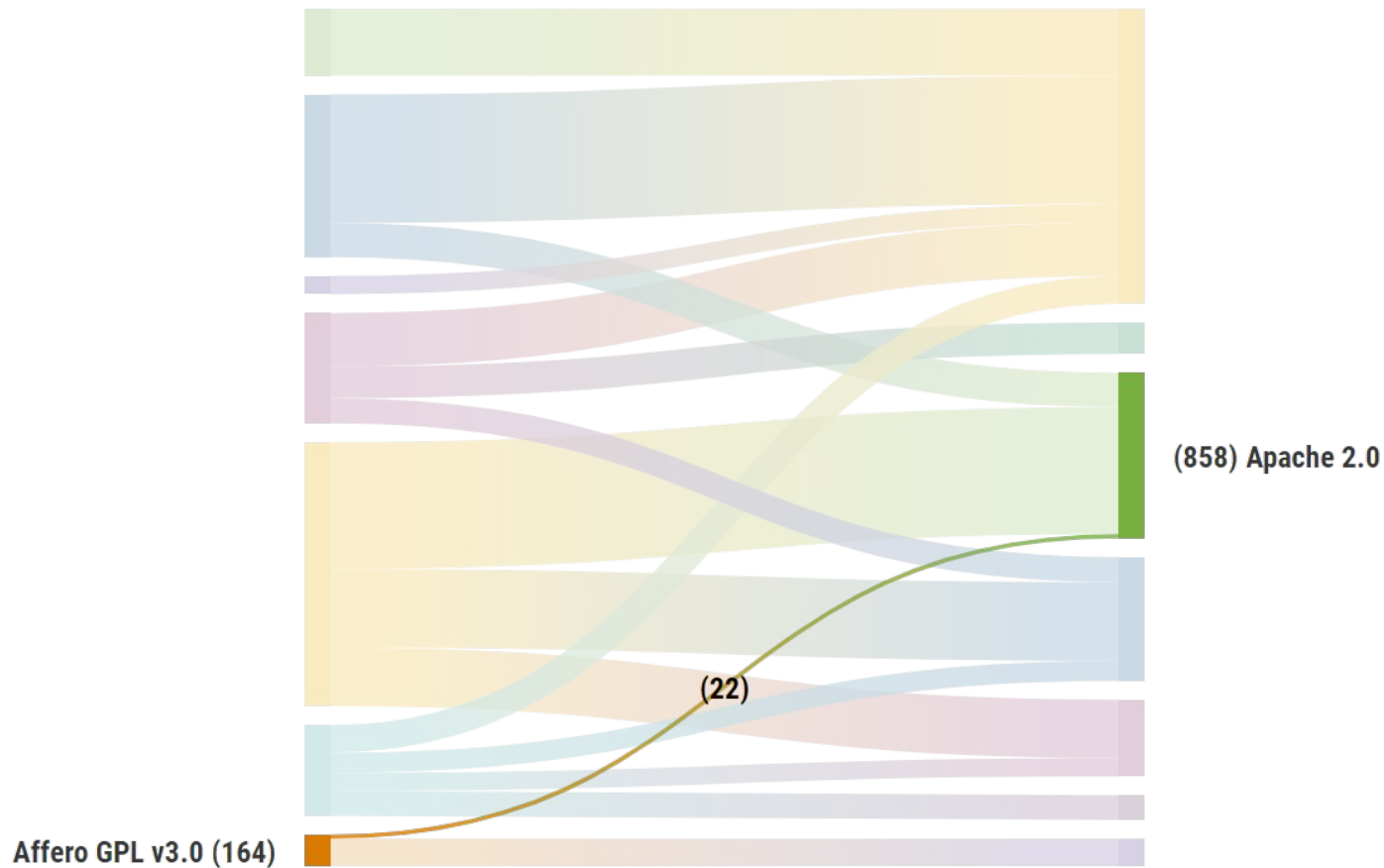
*чистый случай — когда одна лицензия сменяется другой, без двойного лицензирования и подобных прелестей.

Trends of OSS license changes in PyPI packages



Source: codescoring.com, profiscope.io • Licensing magic on PyPI packages data (3863 clean transitions)

Trends of OSS license changes in PyPI packages



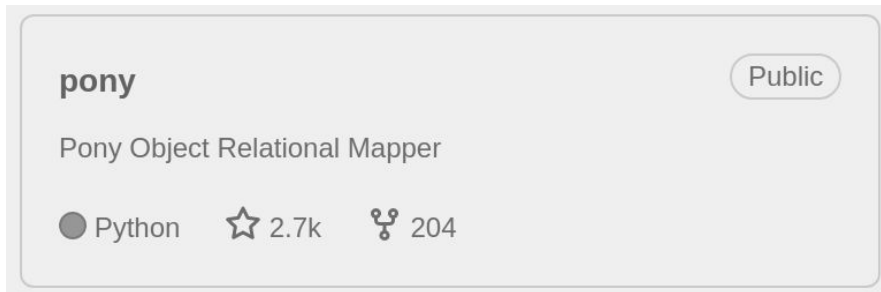


AGPL v3 → Apache 2.0

Very pythonic

Pony Object Relational Mapper

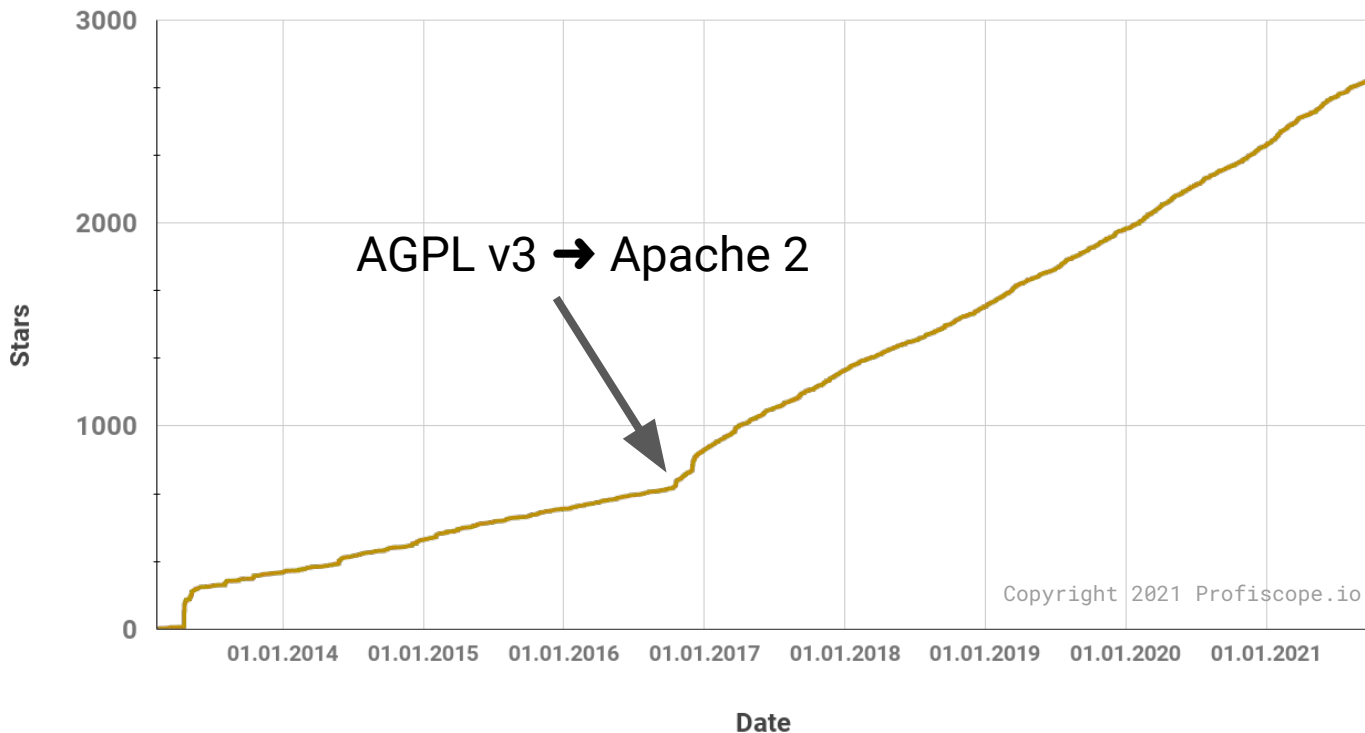
2705 ★★★★★



<https://github.com/ponyorm/pony>

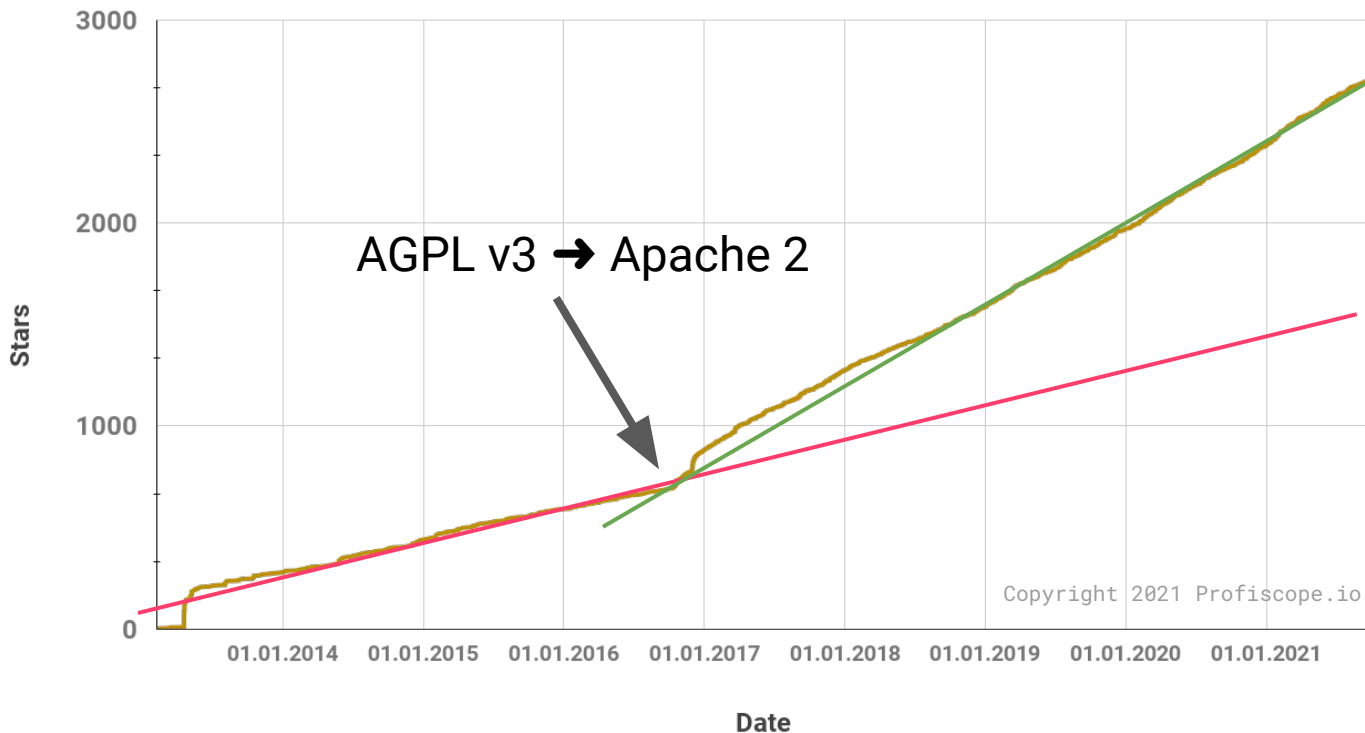
Pony. Популярность до и после смены лицензии

PonyORM stars dynamics



Pony. Популярность до и после смены лицензии

PonyORM stars dynamics



Реакция сообщества на смену лицензии



zalpha314 · 5y

This is good news. The previous license was little restrictive for me.



3



Share

Report

Save



muposat · 5y

I went with SQLAlchemy because of the license less than a month ago. I use dozens of packages in my work and if each and every one milked me with a license fee it would be a death by 1000 cuts.



2



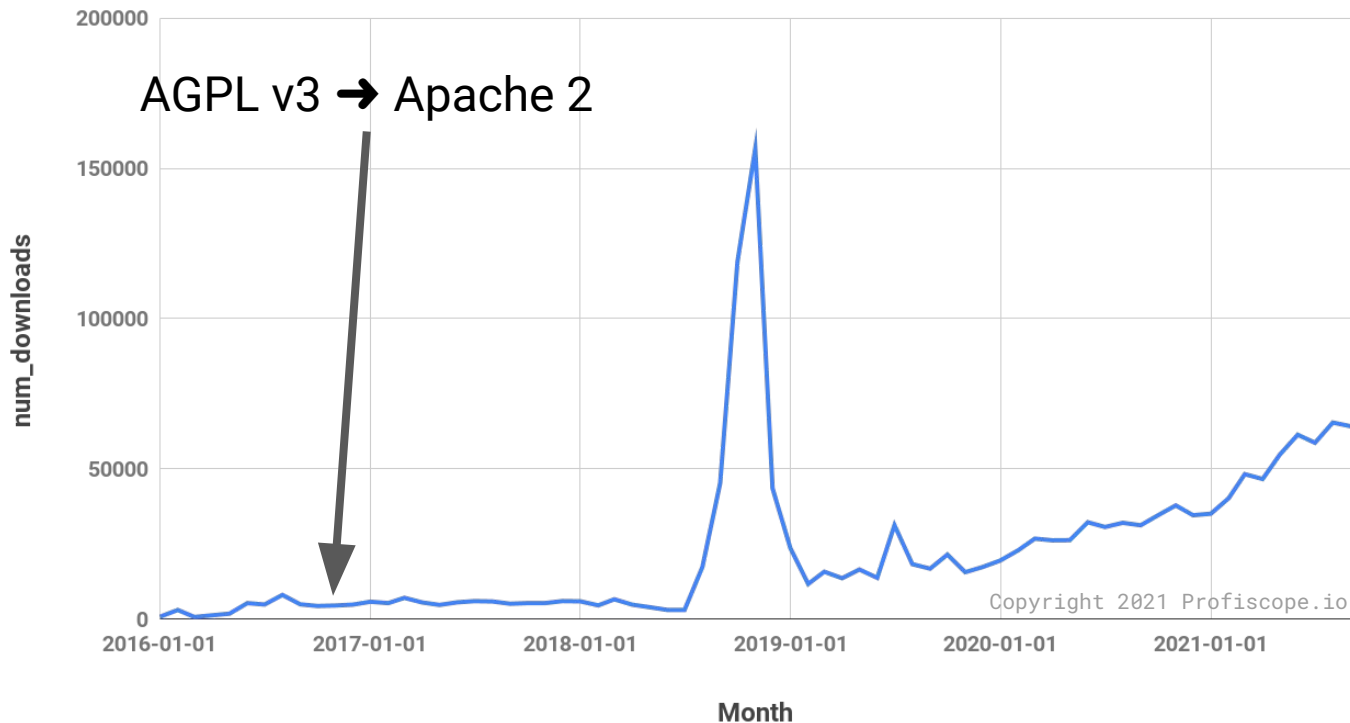
Share

Report

Save

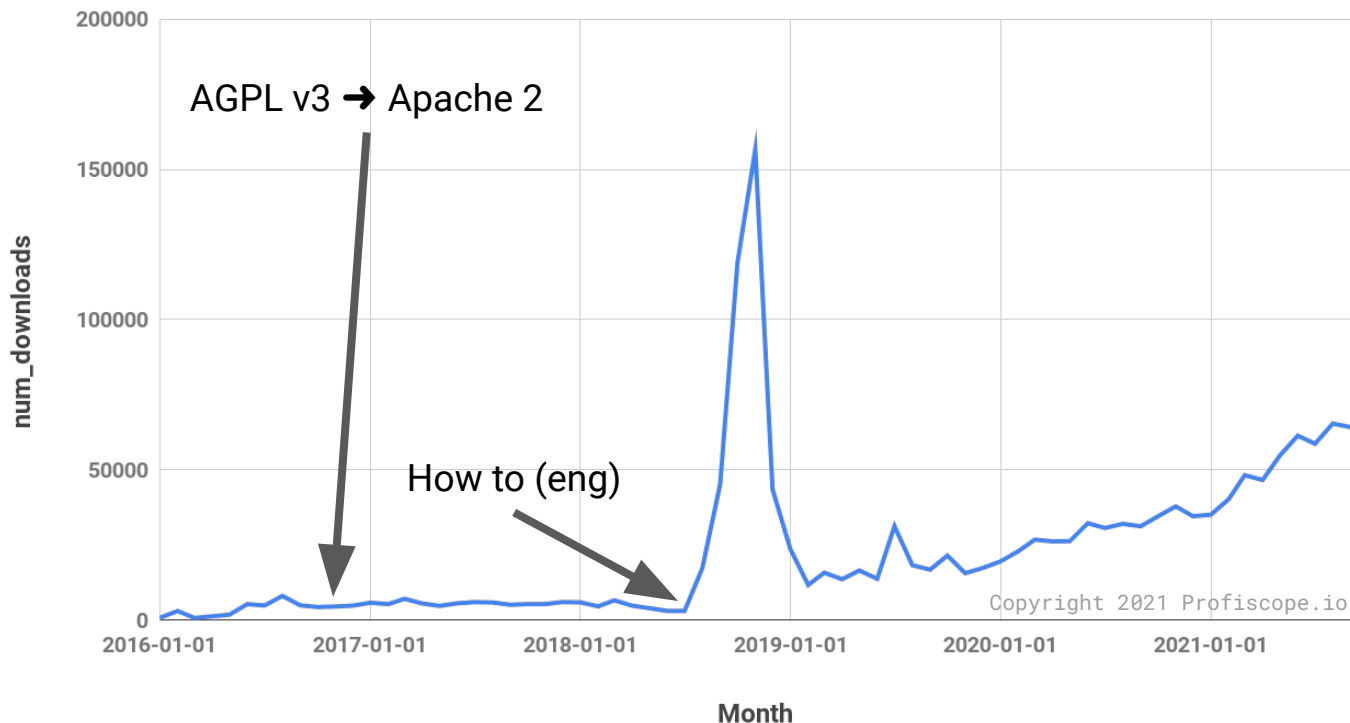
«Используемость»

PonyORM PyPI download stats



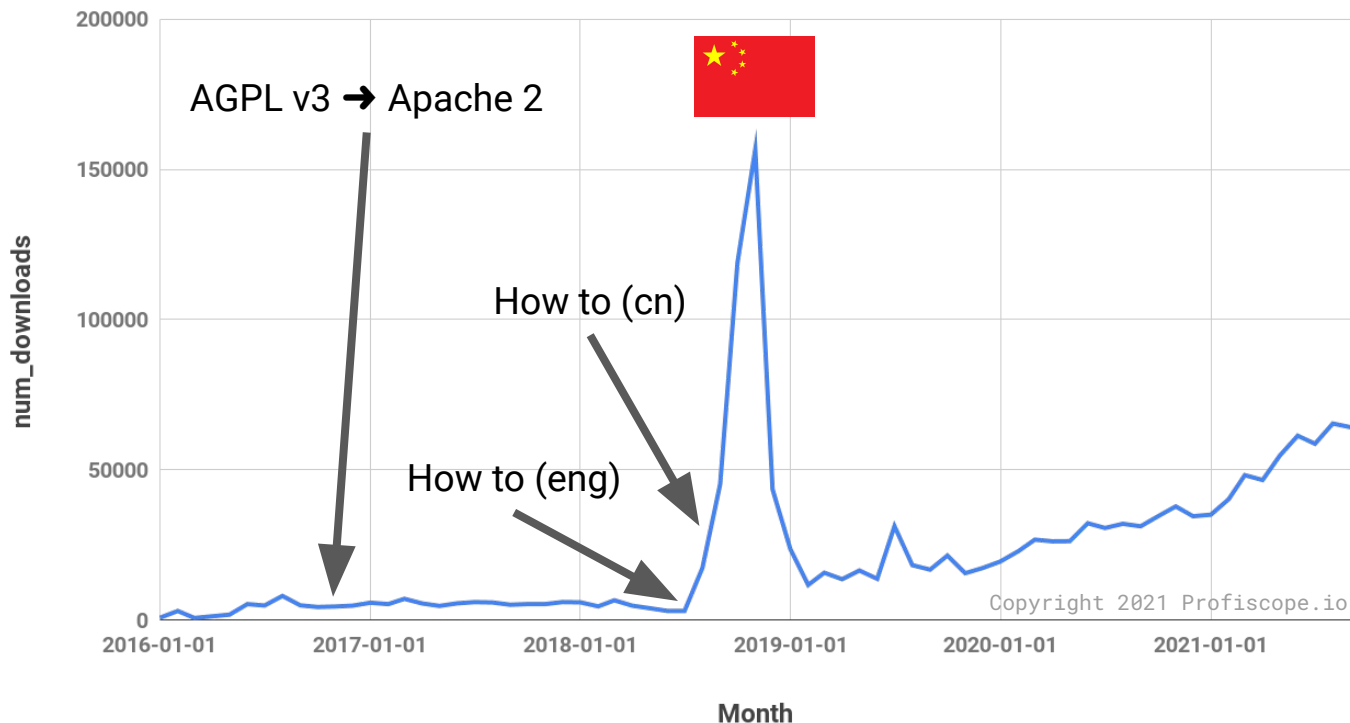
Что было дальше?

PonyORM PyPI download stats



Что было дальше?

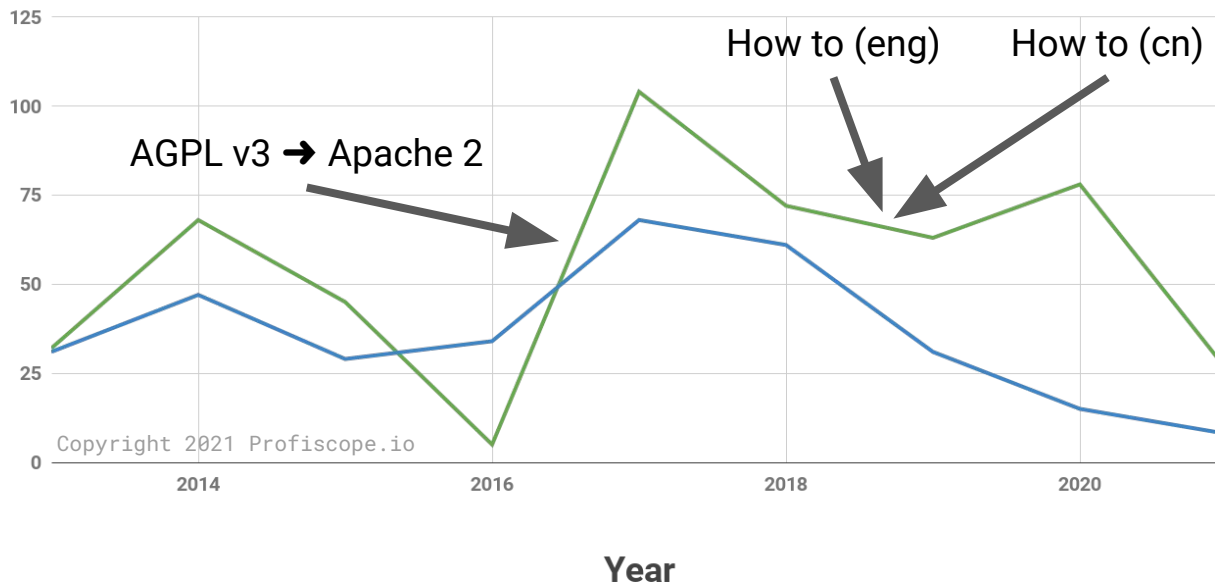
PonyORM PyPI download stats



Включение сообщества в жизнь PonyORM

PonyORM. Issues created vs Issues closed

- Issues created - Issues closed



История PonyORM. Выводы

Лицензия влияет на:

1. принятие сообществом
2. активность и соучастие сообщества
3. осуществление задуманного авторами ;)

Лицензия в прямом смысле не влияет на «популярность», но может являться катализатором.

Да, мы смотрели переход Network copyleft в Permissive.

¬_(\ツ)_/¬

Благодарим PonyORM-авторов Сашу Козловского и Алексея Малашкевича за помощь в подготовке материала!



Домашнее задание #2

Найти проекты с реальной передачей патентных прав

Использовать все доступные средства, чтобы найти open source проекты, в которых действительно происходит передача патентных прав.

Сервисы: Github, Gitlab, StackOverflow, Google.

Для более гибкого поиска может помочь gharchive.org и его агрегация в Clickhouse — ghe.clickhouse.tech.

Пример запроса в ClickHouse

ghe.clickhouse.tech

```
SELECT t.*
```

```
FROM default.github_events t
```

```
where
```

```
event_type='IssuesEvent' and
```

```
created_at > toDateTime('2017-01-01 00:00:00', 'Europe/Moscow') and
```

```
match(body, '.*patent.*US[A-Z0-9]{10}.*')
```

```
ORDER BY created_at DESC
```

Домашнее задание отправить

alexey@profiscope.io

Спасибо за внимание!

alexey@profiscope.io

Канал про анализ кода и всего такого:

t.me/codemining

