



GrimoireLab Workshop

A platform for doing Analytics

Alvaro del Castillo <acs@bitergia.com>

Workshop Summary

The Python History Dashboard

GrimoireLab Environment

Perceval

GrimoireELK: Retrieval and Enrichment Experimental Arch

Based on ElasticSearch (ES)

Arthur and Merlin: Paving the complete platform

Kibana: Analytics Dashboards based on ES

«GOAL: Create new backends by the community»

Playing with Analytics in Open Source projects

The Python Language Repository: <https://github.com/python/cpython> (mirror from official Mercurial)

Uploaded git information with Perceval and p2o to ES:

https://bitergia:pycones2016@pycones2016.biterg.io/data

Enriched with p2o in the same ES

Created a basic dashboard with Kibana at:

<https://pycones2016.biterg.io/app/kibana#/dashboard/GitPython>

Time to play and to MOTIVATE

Ready to DIY?

GitPythonMetrics

104,324

Commits

202

Authors

127

Domains

10

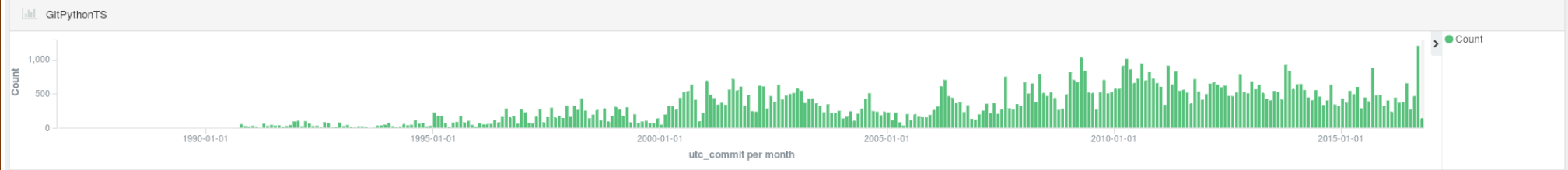
Lines changed

GitPythonAuthors

Author ↕ Q	Commits ↕
Guido van Rossum	11,266
Benjamin Peterson	6,208
Fred Drake	6,173
Georg Brandl	6,153

GitPythonDomains

author_domain: Descending ↕ Q	Count ↕
python.org	27,574
gmail.com	26,399
acm.org	7,051
rcn.com	4,334



GitPython

1 2 3 4 5 ...10 »

Time ↕	Author	lines_changed	message
▶ October 6th 2016, 16:51:29.000	Ned Delly <nad@python.org>	2	Fix vestigial spelling.
▶ October 6th 2016, 08:32:20.000	Benjamin Peterson <benjamin@python.org>	0	merge 3.6
▶ October 6th 2016, 08:32:15.000	Benjamin Peterson <benjamin@python.org>	2	merge 3.5
▶ October 6th 2016, 08:32:09.000	Benjamin Peterson <benjamin@python.org>	2	ensure read size is initialized
▶ October 6th 2016, 08:29:22.000	Benjamin Peterson <benjamin@python.org>	0	merge 3.6
▶ October 6th 2016, 08:29:16.000	Benjamin Peterson <benjamin@python.org>	4	merge 3.5
▶ October 6th 2016, 08:29:07.000	Benjamin Peterson <benjamin@python.org>	4	do not leak buffer if mmap is not writable
▶ October 6th 2016, 08:26:24.000	Benjamin Peterson <benjamin@python.org>	6	make 'where' Py_ssize_t
▶ October 6th 2016, 08:22:28.000	INADA Nanki <sonnogatandw@gmail.com>	3	Issue #28201: Dict reduces possibility of 2nd conflict in hash table

Review you have installed GrimoireLab

Perceval

```
git clone https://github.com/grimoirelab/perceval.git  
cd perceval && sudo python3 setup.py install (virtualenv)
```

GrimoireELK

```
git clone https://github.com/grimoirelab/GrimoireELK.git
```

You can also use your own ElasticSearch and Kibana if you want. In the workshop we will use (**HTTPS**):

ElasticSearch: **https://pycones2016.biterg.io/data**

Kibana: **https://pycones2016.biterg.io/edit**

→ **While you do it let's create some context about Analytics** ←

Perceval

“Just” get the data and generate JSON items with it

19 data sources supported and growing: git, gerrit, bugzilla, github, stackoverflow ...

GPLv3, github based, pull request development model, unit testing with Travis and a life of 1 year

Development doc at <http://perceval.readthedocs.io>

perceval <backend> <repository>

perceval git <https://github.com/grimoirelab/perceval.git>

Perceval (API)

fetch

from_date and offset

category

fetch_from_cache

```
from perceval.backends.git import Git

git = Git('https://github.com/grimoirelab/perceval.git', '/tmp/perceval')

for commit in git.fetch():
    print(commit)
```

Perceval: Git Backend

Time to go to the code: common classes in all backends

- Git class

- GitClient class

- GitCommand class

Metadata

Testing the backend

GrimoireELK

Prototype to be dropped (reusing code if possible)

Data retrieval using perceval

Publishing in Elastic Search

The Enrichment Process

- Sorting Hat

- Projects Mapping

- Copying, renaming and real enrichment

Perceval 2 Ocean (p2o)

Ocean is the ElasticSearch with “all” the raw JSON items

Basic tool to read options from command line and use gelk libraries for:

- Feed data to ES from Perceval: JSON items to ES, easy!

- Enrich data from ES to use it in Kibana

GrimoireELK: p2o with git

Upload and enrich the commits from
<https://github.com/grimoirelab/perceval>

```
GrimoireELK/utils $ ./p2o.py -g --enrich -e  
https://bitergia:pycones2016@pycones2016.biterg.io/data  
git https://github.com/grimoirelab/perceval
```

GrimoireELK: p2o with mailing lists

Upload and enrich the emails from
<https://mail.python.org/mailman/listinfo/flask>

```
GrimoireELK/utils $ ./p2o.py -g --enrich -e  
https://bitergia:pycones2016@pycones2016.biterg.io/data  
pipermail https://mail.python.org/pipermail/flask/
```

Arthur and Merlin*

Arthur: A service for scheduling data retrieval tasks based in Python RQ

<https://github.com/grimoirelab/arthur> (Going to production now)

Merlin*: The enrichment platform based on Dask

Working on the first release (Planned for next FOSDEM 2017)

(*) Merlin name could change

Kibana (and Bitergia's fork Kibiter)

Create an index pattern to access the enriched items

Use Discover to review the data and do searches

Create widgets for doing analytics with specific metrics, reports and so on

Create dashboards integrating the widgets and doing filtering to all the widgets at the same time for a integrated view

- Date filtering

- Terms filtering

Thank you very much!

GrimoireLab Workshop

**A platform for doing
Analytics**

