
What's new in GrimoireLab?

CHAOSCon Europe - Brussels - Feb. 1st, 2019

J. Manrique Lopez
@bitergia
jsmanrique@bitergia.com

A long time ago in a university not far,
far away ...

Counting Potatoes: the Size of Debian 2.2

Jesús M. González-Barahona, Miguel A. Ortuño Pérez, Pedro de las Heras Quirós, José Centeno González and Vicente Matellán Olivera

Debian is the largest Free Software distribution, with well over 2,800 source packages in the latest stable release (Debian 2.2) and more than 4,000 source packages in the release currently in preparation. But, how large is “the largest”? In this paper, we use David Wheeler’s sloccount system to determine the number of physical source lines of code (SLOC) of Debian 2.2 (aka Potato). We show that Debian 2.2 includes over 56,000,000 physical SLOC (almost twice than Red Hat 7.1, released about 8 months later), showing that the Debian development model (based on the work of a large group of voluntary developers spread around the world) is at least as capable as other development methods (like the more centralized one, based on the work of employees, used by Red Hat or Microsoft) to manage distributions of this size.

*of employees, used by Red Hat or Microsoft) to manage distributions of this size.
world) is at least as capable as other development methods (like the more centralized one, based on the work
Debian development model (based on the work of a large group of voluntary developers spread around the
20,000,000 physical SLOC (almost twice than Red Hat 7.1, released about 8 months later), showing that the
physical source lines of code (SLOC) of Debian 2.2 (aka Potato). We show that Debian 2.2 includes over*

What has changed since last year?



Marcos



Marcos

What have we done in 1 year?



Numbers,
numbers,
numbers...

The screenshot shows the CHA OSS dashboard for the 'Git' project. The top navigation bar includes 'Overview', 'Git', 'GitHub PRs', 'GitHub Issues', 'Mailing Lists', 'Community', 'Data Status', and 'About'. The main content area is divided into three panels:

- Git:** Displays summary statistics: 2,127 # Commits, 21 # Authors, and 16 # Repositories.
- Organizations (Left):** A donut chart showing the distribution of data across organizations. A legend lists: Bitergia (green), Universidad Rey Juan ... (orange), Unknown (purple), BBVA Data & Analytics (blue), and Inocybe (red).
- Organizations (Right):** A table with columns: Organization, Commits, Authors, Touched Files, Added Lines, Removed Lines, Projects, Repositories, and Avg. Lines/Commit.

Organization	Commits	Authors	Touched Files	Added Lines	Removed Lines	Projects	Repositories	Avg. Lines/Commit
Bitergia	2,099	12	5,523	876784	216837	1	16	521.02
Universidad Rey Juan Carlos	18	1	41	841	359	1	2	66.667
Unknown	8	7	12	59	31	1	4	11.25
BBVA Data & Analytics	1	1	3	61	7	1	1	68
Inocybe	1	1	1	1	1	1	1	2

Numbers,
numbers,
numbers...



CHAOS

Overview **Git** GitHub PRs GitHub Issues Mailing Lists Community Data Status About

Dashboard / Editing Git (unsaved) Save Cancel Add Options Share < January 29th 2017, 14:25:46.371 to January 29th 2018, 14:25:46.372 >

* Uses lucene query syntax

project: "GrimoireLab" NOT Empty Commits NOT Bots Add a filter +

Git Organizations Organizations



CHAOS

Overview **Git** GitHub PRs GitHub Issues Mailing Lists Community Data Status About

Dashboard / Editing Git (unsaved) Save Cancel Add Options Share < Last 1 year >

* Uses lucene query syntax

project: "GrimoireLab" NOT Empty Commits NOT Bots Add a filter +

Git Organizations Organizations

2,215 # Commits 42 # Authors

16 # Repositories



- Bitergia
- GSoC
- Unknown
- Universidad Rey Juan ...
- Independent
- Mozilla
- Samsung
- CloudBees

Organization	Commits	Authors	Touched Files	Added Lines	Removed Lines	Projects	Repositories	Avg. Lines/Commit
Bitergia	2,046	11	5,147	215486	193880	1	16	200.081
GSoC	63	1	234	90283	64901	1	3	2,463.238
Unknown	49	27	65	1121	2696	1	9	77.898
Universidad Rey Juan Carlos	47	1	144	4822	594	1	8	115.234
Independent	5	1	8	153	23	1	1	35.2
Mozilla	2	1	2	2	2	1	1	2
Samsung	2	1	2	2	0	1	2	1
CloudBees	1	1	6	194	0	1	1	194

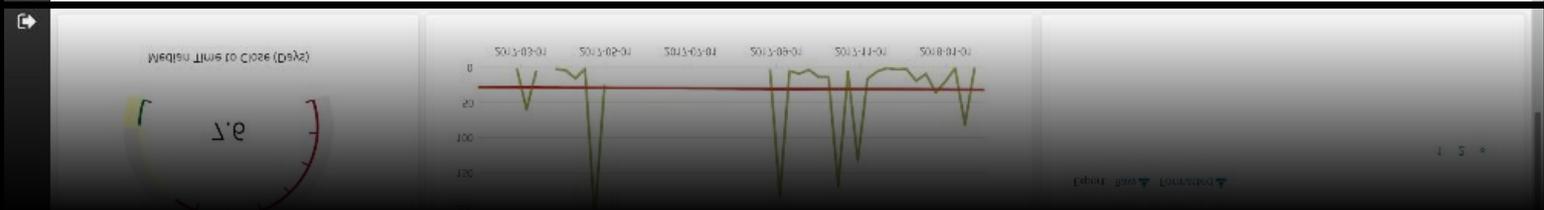
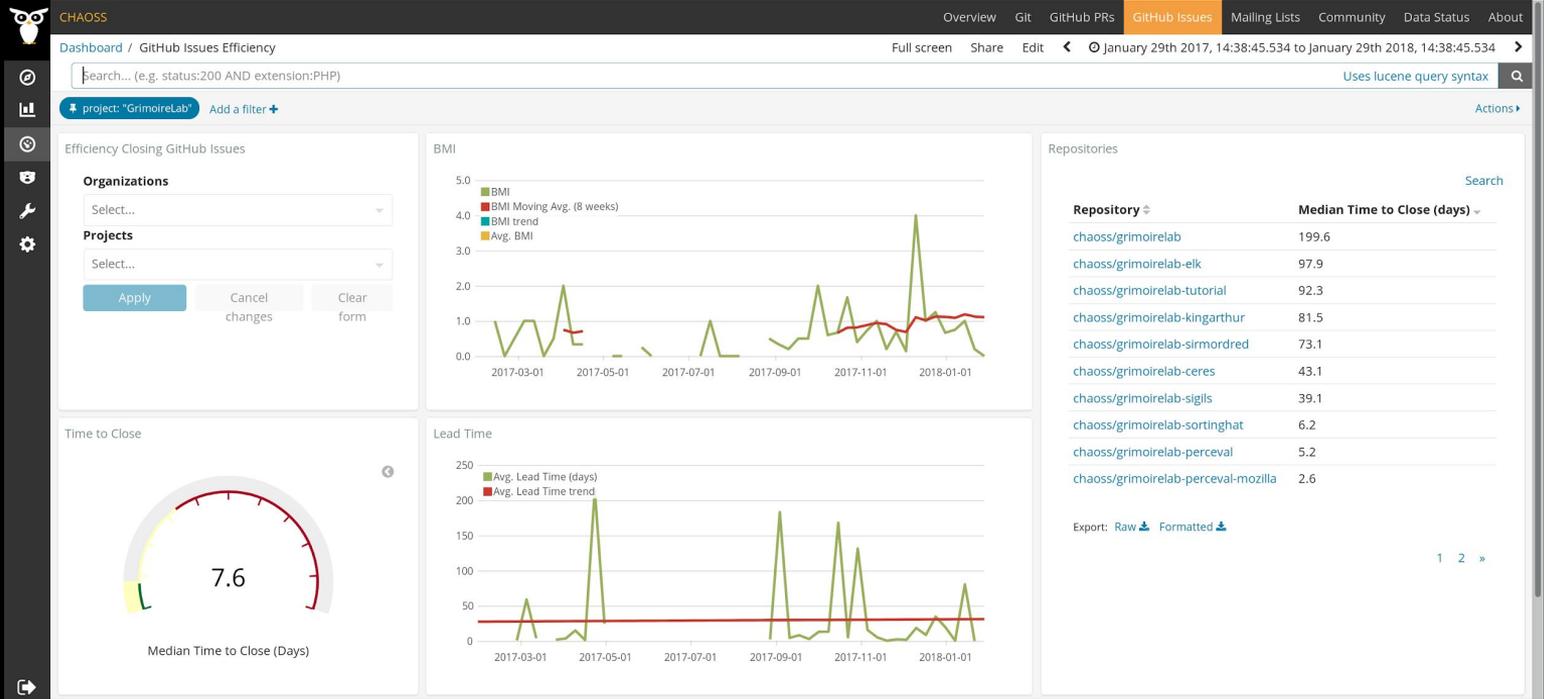
Thank you
GSOC!!

Google Summer Of Code

Improvements in **GrimoireLab Manuscripts**

Getting ready for GSOC 2019!!

Numbers,
numbers,
numbers...



Numbers,
numbers,
numbers...

CHAOSS Overview Git GitHub PRs **GitHub Issues** Mailing Lists Community Data Status About

Dashboard / GitHub Issues Efficiency Full screen Share Edit January 29th 2017, 14:38:45.534 to January 29th 2018, 14:38:45.534

Search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax

project: "GrimoireLab" Add a filter + Actions

Efficiency Closing GitHub Issues BMI Repositories

CHAOSS Overview Git GitHub PRs **GitHub Issues** Mailing Lists Community Data Status About

Dashboard / GitHub Issues Efficiency Full screen Share Edit Last 1 year

Search... (e.g. status:200 AND extension:PHP) Uses lucene query syntax

project: "GrimoireLab" Add a filter + Actions

Efficiency Closing GitHub Issues

Organizations
Select...

Projects
Select...

Apply Cancel changes Clear form

BMI

5.0
4.0
3.0
2.0
1.0
0.0

2018-03-01 2018-05-01 2018-07-01 2018-09-01 2018-11-01 2019-01-01

- BMI
- BMI Moving Avg. (8 weeks)
- BMI trend
- Avg. BMI

Repositories

[Search](#)

Repository	Median Time to Close (days)
chaoss/grimoirelab-kibiter	160.9
chaoss/grimoirelab-tutorial	48.1
chaoss/grimoirelab-sirmordred	20.4
chaoss/grimoirelab-sigils	19
chaoss/grimoirelab-kingarthur	17
chaoss/grimoirelab-manuscripts	14.9
chaoss/grimoirelab-elk	13.9
chaoss/grimoirelab-hatstall	11.9
chaoss/grimoirelab-perceval-opnfv	8.2
chaoss/grimoirelab	7.3

Export: [Raw](#) [Formatted](#)

1 2 »

Time to Close

11.3

Median Time to Close (Days)

Lead Time

200
150
100
50
0

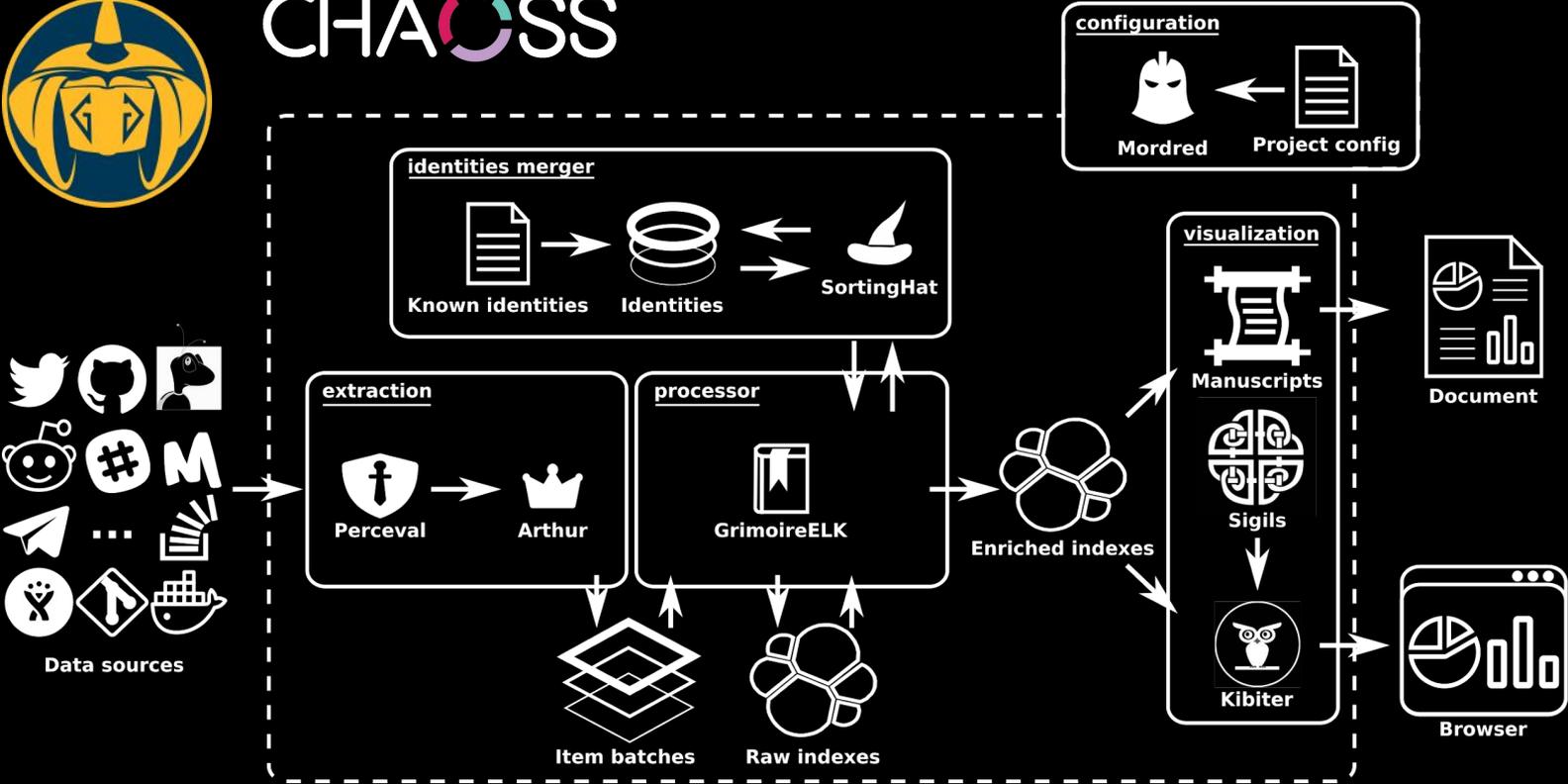
2018-03-01 2018-05-01 2018-07-01 2018-09-01 2018-11-01 2019-01-01

- Avg. Lead Time (days)
- Avg. Lead Time trend

GrimoireLab as a platform



CHA OSS



GrimoireLab is not a data collector

GrimoireLab is not a dashboard

GrimoireLab is not a product

GrimoireLab is not an “academic” project

It's not a data collector, but...

GrimoireLab beyond data collection



Source Code

Git, Baazar*, Mercurial*
(*exporting to git)



Tickets / issues

Bugzilla, GitHub Issues,
GitLab Issues, JIRA,
Launchpad, Phabricator
Maniphest, Redmine



Code Review

Gerrit, GitHub Pull
Requests, GitLab Merge
Requests



Continuous Integration

Jenkins



Mailing lists / Forums

Askbot, Discourse, Gmane,
Hyperkitty, mbox files,
NNTP, Pipermail, Stack
Exchange



Chat

IRC (Supybot), Slack,
Telegram



Downloads

Apache http logs



Wiki

Confluence, MediaWiki



Meetings

Meetup, Mozilla Reps
events



Others

crates.io, Docker Hub, RSS,
Twitter

New data sources:

Mattermost (development sponsored by Red Hat & Bitergia)

GitLab (development sponsored by GitLab & Bitergia)

Groups.io (development sponsored by The Linux Foundation & Bitergia)

SortingHat 1.0 (72% complete)

New GraphQL API

```
1 query identities {
2   identities {
3     uuid
4     profile {
5       name
6       email
7       gender
8     }
9     identities {
10      id
11      name
12      email
13      username
14      source
15    }
16  }
17 }
18
```

QUERY VARIABLES

```
1 null
```

```
{
  "data": {
    "identities": [
      {
        "uuid": "0f9949815a8e8ef771cfcfe60aedbc142d7405be",
        "profile": {
          "name": "Luis Cañas",
          "email": "lcanas@bitergia.com",
          "gender": "M"
        },
        "identities": [
          {
            "id": "0f9949815a8e8ef771cfcfe60aedbc142d7405be",
            "name": null,
            "email": "lcanas@bitergia.com",
            "username": null,
            "source": "mls"
          },
          {
            "id": "5ca99d0e8e4618d9454e93fe0cf5d281a1ae4990",
            "name": "Luis Cañas",
            "email": null,
            "username": null,
            "source": "mls"
          }
        ]
      },
      {
        "uuid": "fac2242f069766037e8bc611e6a9fc3c795215a2",
        "profile": {
          "name": "Santiago Dueñas",
          "email": "sduenas@bitergia.com",
          "gender": "M"
        },
        "identities": [
          {
            "id": "4225149aa4e51f4b97a1053d871664f06d057128",
            "name": "Santiago Dueñas",
            "email": "sduenas@bitergia.com",
            "username": null,
            "source": "mls"
          }
        ]
      }
    ]
  }
}
```

SortingHat 1.0 (72% complete)

New GraphQL API

New Hatstall is coming! (based in React)

It's not a dashboard, but...

GrimoireLab Sigils

chaoss.github.io/grimoirelab-sigils

GrimoireLab Sigils by CHAOSS

CHAOSS GMD WG panels for Code Development

- **Lines of Code Changed:** panel focused on the number of lines of code changed.
- **Maintainer Response to Merge Request Duration:** panel focused on the time to response after a pull request took place.
- **Pull Requests Merged:** panel focused on the number of Pull Requests merged.

GrimoireLab Sigils Panels

- **Community Structure:** find who are your core, regular and casual contributors.
- **Data Status:** data freshness information.
- **Demographics:** attraction and retention of contributors.
- **Gerrit Review Efficiency:** efficiency closing reviews in Gerrit.
- **Git Demographics:** attraction and retention of developers specifically for Git.
- **Git:** metrics focused on Git commits.
- **GitHub Backlog:** focused on pending (open) tasks.
- **GitHub Issues Efficiency:** efficiency closing issues on GitHub.
- **GitHub Issues Timing:** metrics focused on how long issues remain open.
- **GitHub Issues:** activity and community metrics focused on Issues.
- **GitHub Pull Requests Efficiency:** efficiency closing Pull Requests.
- **GitHub Pull Requests Timing:** metrics focused on how long pull requests remain open.
- **GitHub Pull Requests:** activity and community metrics focused on Pull Requests.
- **GitLab Issues Efficiency:** efficiency closing issues on GitLab.
- **GitLab Merge Requests Efficiency:** efficiency solving (merging or closing) MRs on GitLab.
- **Lifecycle:** Level of activity in git repositories
- **Mediawiki:** contains metrics focused on reviews, including editions, revisions and editors.
- **Overview:** summary of basic metrics on all analyzed sources.

- **Overview:** summary of basic metrics on all analyzed sources
- **Mediawiki:** contains metrics focused on reviews, including editions, revisions and editors
- **Lifecycle:** Level of activity in git repositories
- **GitLab Merge Requests Efficiency:** efficiency solving (merging or closing) MRs on GitLab
- **GitLab Issues Efficiency:** efficiency closing issues on GitLab
- **GitHub Pull Requests:** activity and community metrics focused on Pull Requests
- **GitHub Pull Requests Timing:** metrics focused on how long pull requests remain open

GrimoireLab Sigils

chaoss.github.io/grimoirelab-sigils

GrimoireLab Sigils by CHAOSS

CHAOSS GMD WG panels for Code Development

- **Lines of Code Changed:** panel focused on the number of lines of code changed.
- **Maintainer Response to Merge Request Duration:** panel focused on the time to response after a pull request took place.
- **Pull Requests Merged:** panel focused on the number of Pull Requests merged.

GrimoireLab Sigils Panels

- **Community Structure:** find who are your core, regular and casual contributors.

GrimoireLab Sigils by CHAOSS

Lifecycle

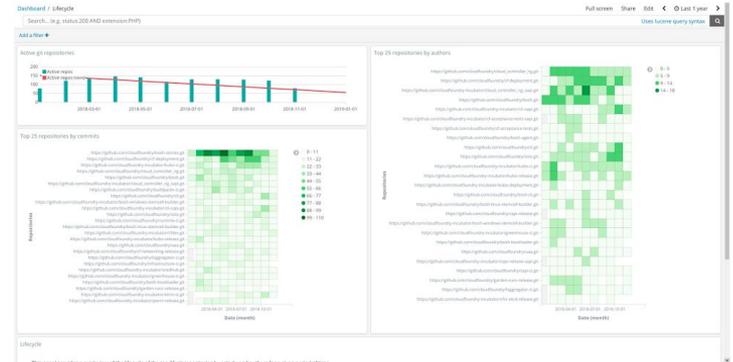
This panel provides a quick view of the lifecycle of the top 25 git repositories by activity and authors for a given period of time.

Definition

Active: at least 1 commit per month

Charts

- Evolution of the number of active repositories during selected period of time, and its trend
- Top 25 git repositories by number of commits for selected period of time, showing level of activity by month
- Top 25 git repositories by number of *active* authors for selected period of time, showing level of active authors by month



Edit [this doc](#)

Edit [this doc](#)

Bitergia Analytics Panels Collection

bitergia.gitlab.io/panel-collections

Bitergia Analytics Panels by Bitergia

This site contains documentation for the set of standard panels included in Bitergia Analytics platform.

Open Source Program Office

Contributors Growth: panel focused on the growth of the community in all of the data sources

Efficiency: Timing overview: this panel focuses on time to first attention and time to close tickets.

Engagement: Contributions overview: panel focused on engagement in terms of contributions.

Organizational Diversity: panel focused on organizational diversity for contributions and contributors

Organizational Diversity by domains: panel focused on organizational diversity for contributions and contributors specifying the domains information.

Sample collection

Sample Documentation File: includes info about a given panel.

Sample Documentation File: includes info about a given panel.

Sample collection

Organizational Diversity by domains: panel focused on organizational diversity for contributions and contributors specifying the domains information.

GrimoireLab
beyond
dashboards

Updated Kibiter!

chaoss.biterg.io

It's not a product, but ...

GrimoireLab as
an open
platform

Bitergia Analytics bitergia.com

CHAOSS GMD WG Implementations

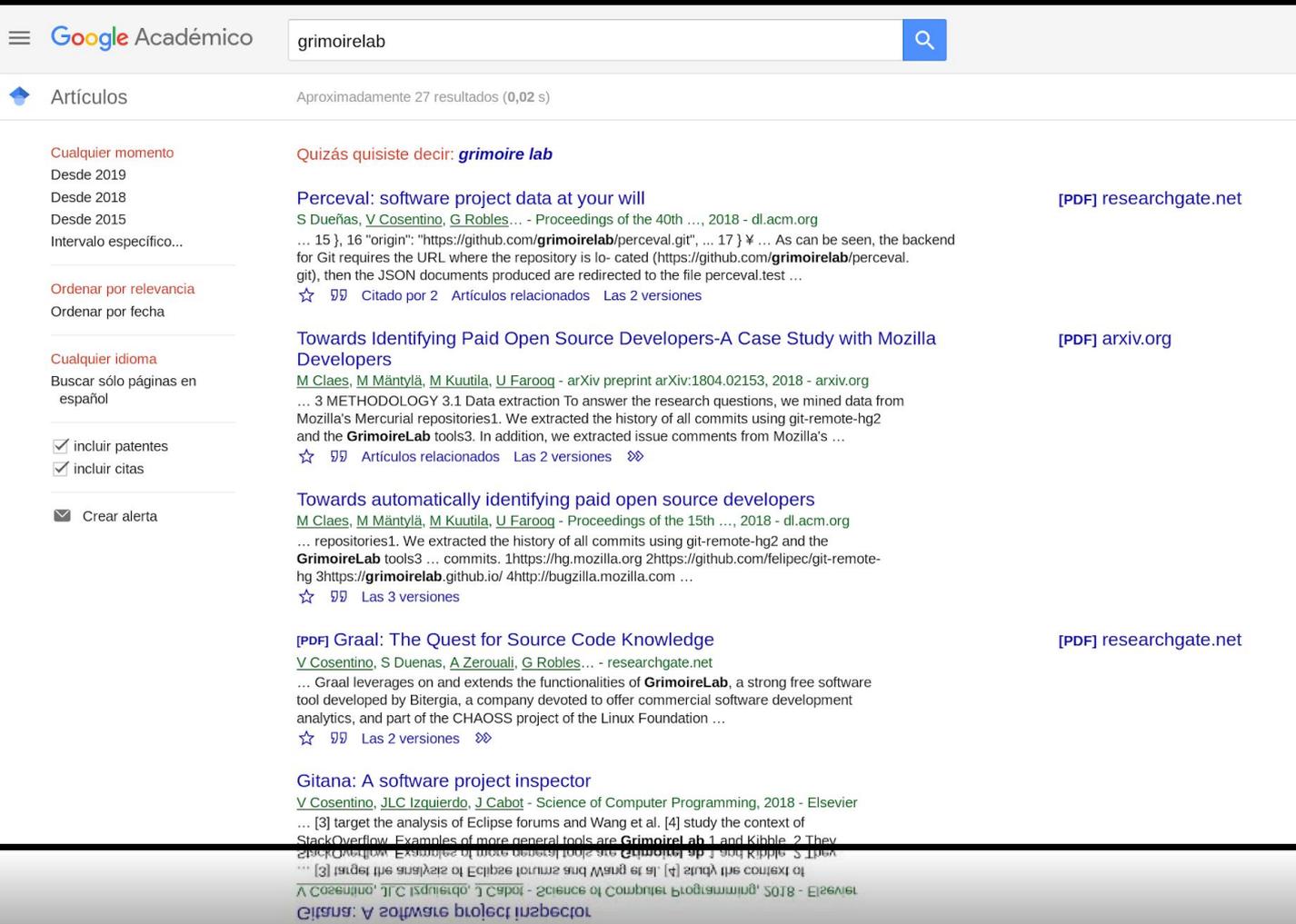
github.com/chaoss/wg-gmd/tree/master/implementations

CROSSMINER www.crossminer.org

Generic Repository AnALyzer github.com/bitergia/graal

VBoard github.com/dlumbrer/VBoard

It's not an “academic” project, but ...



The screenshot shows a Google Académico search page for the term "grimoirelab". The search results are filtered to "Artículos" and show approximately 27 results. The left sidebar contains filters for "Cualquier momento" (with sub-options: Desde 2019, Desde 2018, Desde 2015, Intervalo específico...), "Ordenar por relevancia" (with sub-option: Ordenar por fecha), "Cualquier idioma" (with sub-option: Buscar sólo páginas en español), and checkboxes for "incluir patentes" and "incluir citas". There is also a "Crear alerta" option.

Quizás quisiste decir: *grimoire lab*

Perceval: software project data at your will [PDF] researchgate.net
S Dueñas, V Cosentino, G Robles... - Proceedings of the 40th ... , 2018 - dl.acm.org
... 15 } , 16 "origin": "https://github.com/grimoirelab/perceval.git", ... 17 } ¥ ... As can be seen, the backend for Git requires the URL where the repository is located (https://github.com/grimoirelab/perceval.git), then the JSON documents produced are redirected to the file perceval.test ...
☆ 📄 Citado por 2 Artículos relacionados Las 2 versiones

Towards Identifying Paid Open Source Developers-A Case Study with Mozilla Developers [PDF] arxiv.org
M Claes, M Mäntylä, M Kuutila, U Farooq - arXiv preprint arXiv:1804.02153, 2018 - arxiv.org
... 3 METHODOLOGY 3.1 Data extraction To answer the research questions, we mined data from Mozilla's Mercurial repositories1. We extracted the history of all commits using git-remote-hg2 and the GrimoireLab tools3. In addition, we extracted issue comments from Mozilla's ...
☆ 📄 Artículos relacionados Las 2 versiones 📄

Towards automatically identifying paid open source developers
M Claes, M Mäntylä, M Kuutila, U Farooq - Proceedings of the 15th ... , 2018 - dl.acm.org
... repositories1. We extracted the history of all commits using git-remote-hg2 and the GrimoireLab tools3 ... commits. 1https://hg.mozilla.org 2https://github.com/felipecg/git-remote-hg 3https://grimoirelab.github.io/ 4http://bugzilla.mozilla.com ...
☆ 📄 Las 3 versiones

[PDF] Graal: The Quest for Source Code Knowledge [PDF] researchgate.net
V Cosentino, S Duenas, A Zerouali, G Robles... - researchgate.net
... Graal leverages on and extends the functionalities of GrimoireLab, a strong free software tool developed by Bitergia, a company devoted to offer commercial software development analytics, and part of the CHAOSS project of the Linux Foundation ...
☆ 📄 Las 2 versiones 📄

Gitana: A software project inspector
V Cosentino, JLC Izquierdo, J Cabot - Science of Computer Programming, 2018 - Elsevier
... [3] target the analysis of Eclipse forums and Wang et al. [4] study the context of StackOverflow. Examples of more general tools are GrimoireLab 1 and Kibble 2. They track OpenFlow. Examples of more general tools are GrimoireLab 1 and Kibble 2. They ... [3] target the analysis of Eclipse forums and Wang et al. [4] study the context of Science of Computer Programming' 2018 - Elsevier
Gitana: A software project inspector

Join us!



[github.com/chaoss/grimoirelab-*](https://github.com/chaoss/grimoirelab-)
(repositories being migrated from
github.com/grimoirelab)



bit.ly/grimoirelab-mailman

#irc

#GrimoireLab in
irc.freenode.net



@GrimoireLab
#GrimoireLab



About

Bitergia

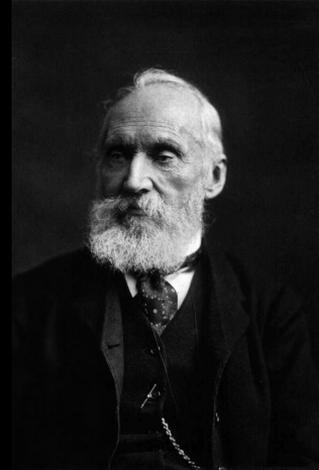
Software Development Analytics

For Your Peace of Mind

“To measure is to know”

“If you can not measure it,
you cannot improve it”

Lord Kelvin



Software Development Monitoring Made Easy

Bitergia solves any question you might have about your software development processes



Activity

What? How many?



Size

How big? How many?



Performance

How fast? Average time?



Demography

How old? New one? Inactive ones?



Diversity

Where from? Organizations? Gender?



[LEARN MORE](#)

[LEARN MORE](#)

Customers



openstack
CLOUD SOFTWARE



midokura

Engine Yard™



Joomla!™

Pivotal



The Document
Foundation



Core OS

moz://a



Couchbase



OPEN CONTAINER
INITIATIVE



THE
LINUX
FOUNDATION

BBVA

CLOUD FOUNDRY



AUTODESK™



FINOS

confluent



CHEF™



Santander

UBER



GitLab





DevOps Engineer

Sales Manager (North America)

Let's go for questions!

Manrique López

CEO at Bitergia

jmanrique@bitergia.com

