

Fine-grained exploration of the GitHub workflow ecosystem

Laurent Bossart

Département d'informatique
Université de Mons

Directeur : Dr. Tom Mens
Rapporteurs : Dr. Alexandre Decan et Mr. Sebastien Bonte

5 septembre 2022

1. Introduction

2. Réalisation

3. Résultats



Figure 1: GitHub

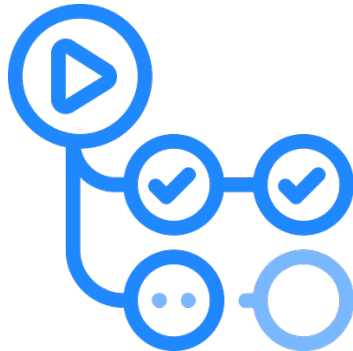


Figure 2: GitHub Actions

Introduction - GitHub Actions et Workflows

The screenshot shows the GitHub Actions workflow configuration page for the repository 'pykeen'. The breadcrumb path 'pykeen / .github / workflows /' is highlighted with a red box. Below the repository name 'cthoyt', the workflow 'Darglint forever (#985)' is selected. The workflow status is '3e9e973 on Jun 20' with a 'History' link. A table lists the workflow files, with the first column highlighted by a red box:

File	Workflow	Last Run
auto-cancellation.yml	Update GHA (#959)	3 months ago
common.yml	Run CI when auto-merge is enabled	3 months ago
lint.yml	Darglint forever (#985)	2 months ago
pr_welcome.yml	Update Welcome GHA	3 months ago

Introduction - Exemple de fichier *workflow*

```
name: Welcome

on:
  pull_request:
    types:
      - opened
    branches:
      - master

jobs:
  welcome:
    runs-on: ubuntu-latest
    steps:
      - uses: actions/first-interaction@v1
        with:
          repo-token: \${{ secrets.GITHUB_TOKEN }}
          pr-message: |-
            Congrats on making your first Pull Request and thanks for taking the time
            to improve PyKEEN!!
```

How Do Software Developers Use GitHub Actions to Automate Their Workflows?

- Timothy Kinsman
- Mairieli Wessel
- Marco A. Gerosé
- Christoph Treude

On the rise and fall of CI services in GitHub

- Mehdi Golzadeh
- Alexandre Decan
- Tom Mens

Introduction - Motivation

1

De plus en plus populaires

2

Intégrés à GitHub

3

Très facile à utiliser

4

Marketplace avec une grande diversité

Introduction - Questions de recherche

1

À quelle vitesse le Marketplace de GitHub se développe-t-il au fil du temps ?

2

Quelles catégories d'Actions sont le plus souvent proposées sur le Marketplace ?

3

À quelle vitesse les Actions évoluent-elles ?

4

Est-ce que les *workflows* utilisent plusieurs Actions ?

5

Quelle sont les Actions les plus populaires ?

6

Y a-t-il beaucoup d'Actions avec des *issues* non résolues ?

7

Quelle proportion d'Actions sont développées par des utilisateurs vérifiés ?

8

Comment les Actions sont-elles *triggered* ?

Introduction - Réalisation de l'objectif

```
graph LR; A[Identifier les données nécessaires] --> B[Identifier comment et où récupérer ces données]; B --> C[Implémentation d'un programme permettant de récupérer les données]; C --> D[Analyse des données récupérées];
```

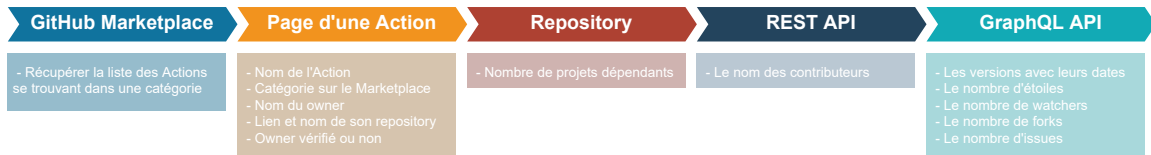
Identifier les données nécessaires

Identifier comment et où récupérer ces données

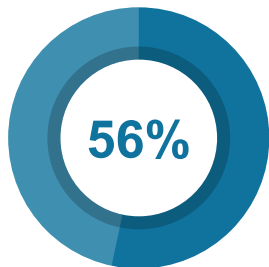
Implémentation d'un programme permettant de récupérer les données

Analyse des données récupérées

Réalisation - Récupération des données



Réalisation - Taille du dataset (Juillet 2022)



Actions récupérées (8 114)



Actions disponibles (~ 14 000)

Résultats - À quelle vitesse le Marketplace de GitHub se développe-t-il au fil du temps ?

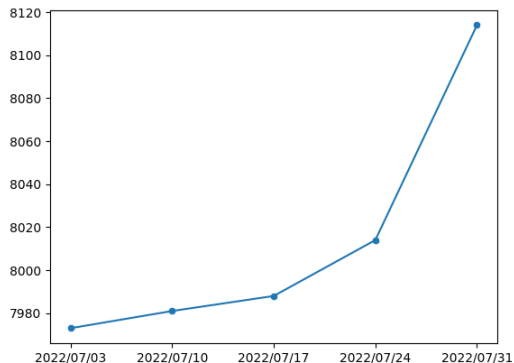
Motivation

- Vérifier si les GitHub Actions continuent à gagner en popularité
- Comparer avec l'observation faite par Golzadeh et al.

Résultat attendu

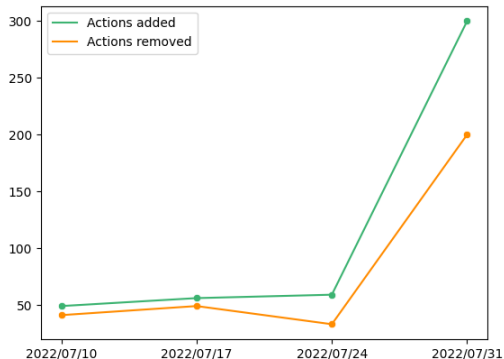
- On s'attend à ce que la popularité des GitHub Actions soit croissante, i.e. le nombre d'Actions disponible est en augmentation.

Résultats - À quelle vitesse le Marketplace de GitHub se développe-t-il au fil du temps ?



Dates	# d'Actions
2022/07/03	7,973
2022/07/10	7,981
2022/07/17	7,988
2022/07/24	8,014
2022/07/31	8,114

Résultats - À quelle vitesse le Marketplace de GitHub se développe-t-il au fil du temps ?



Résultats - Comment les Actions sont-elles *triggered* ?

Motivation

- Comprendre à quelle étape de leur *workflow* les développeurs ont le plus besoin d'utiliser des Actions

Résultat attendu

- On s'attend à avoir une majorité de "push" et "pull requests" étant donné que ce sont les interactions les plus courantes lors du développement d'un software en utilisant GitHub.

Résultats - Comment les Actions sont-elles *triggered* ?



Première approche

Sélection de 50 000 fichiers de *workflows* provenant de *repositories*

- possédants des fichiers *workflows*
- ayant été mis à jour dans les 12 derniers mois

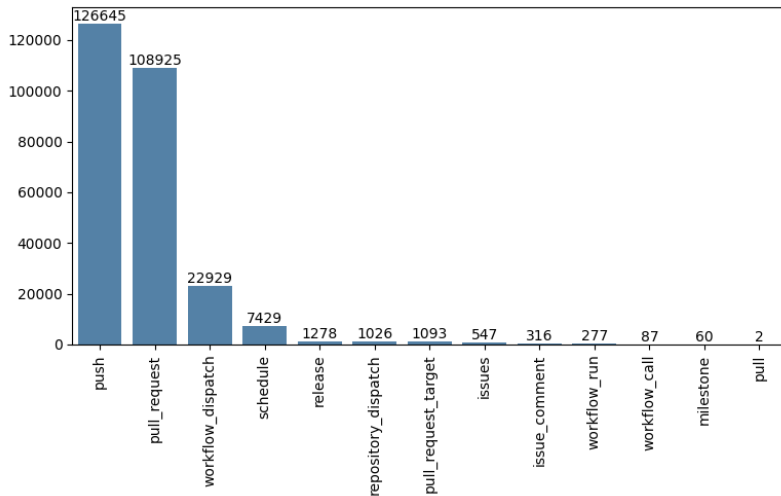


Seconde approche

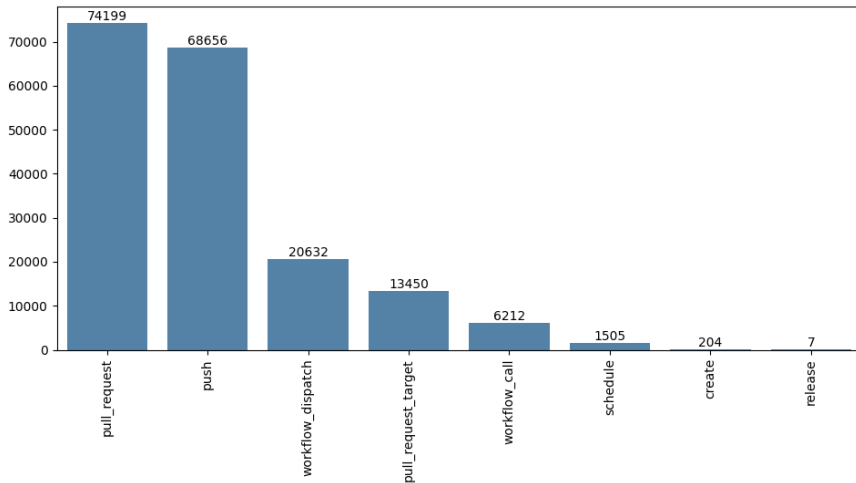
Sélection de 50 000 fichiers de *workflows* provenant de *repositories*

- possédants des fichiers *workflows*
- ayant été mis à jour dans les 12 derniers mois
- n'étant pas un *fork* d'un autre *repository*
- ayant au moins 100 *stars*
- ayant au moins 100 *commits*

Résultats - Comment les Actions sont-elles *triggered* ?



Résultats - Comment les Actions sont-elles *triggered* ?



On the Use of GitHub Actions in Software Development Respositories

- Alexandre Decan
- Tom Mens
- Pooya Rostami Mazrae
- Mehdi Golzadeh

Résultats - Comparaison avec une nouvelle publication

Trigger	% workflows	% workflows nouvelle publication
pull_request	60.7%	56.3%
push	56.2%	63.4%
pull_request_target	11.6%	2.6%
workflow_dispatch	9.5%	15.4%
workflow_call	6.2%	
issues	6.2%	2.0%
schedule	2.3%	16.1%
create	0.2%	
release	0.01%	6.2%

Threats - Threats to validity



Le Marketplace ne donne pas accès à toutes les Actions.

Identification des *release tags*.

Beaucoup de projets n'ont pas effectué l'intégralité de leur développement sur GitHub.



Les *workflows* considérés ne proviennent pas forcément de projets.

La majorité des projets (2/3) sur GitHub sont personnels.

Idées de questions pouvant faire suite

1

Y a-t-il beaucoup de code dupliqué entre *workflows* ?

2

Y a-t-il de potentielles failles de sécurité liées à l'utilisation des Actions ?

3

Est-ce que toutes les Actions utilisées dans les *workflows* sont utilisées ?

Merci pour votre attention