

The 24th Belgium-Netherlands Software Evolution Workshop  
17–18 November 2025, Enschede, The Netherlands



# ghmap: A Python tool for mapping GitHub events to contributor activities

**Youness Hourri**, Alexandre Decan, Tom Mens  
Software Engineering Lab, University of Mons, Belgium

**BENEVOLES**

**UMONS**  
Université de Mons

GitHub logs **contributor activity** through low-level events



uhourri commented on Feb 20

Owner ...

- This PR fixes an issue where all `CreateIssueComment`, `CreatePullRequestComment`, etc. actions were incorrectly grouped into a single activity, even when related to different issues or PRs.
- Now, comments are only grouped if they belong to the same `issue.number` or `pull_request.number`.

## Changes

- Updated validation logic in `_validate_gathered_actions` to enforce issue/PR consistency.

fix incorrect grouping of issue/PR comments

751b24f

Merge pull request

uhourri linked an issue on Feb 20 that may be closed by this pull request

**Incorrect grouping of comments across different issues/PRs #3**

uhourri merged commit `47d4d0e` into `main` on Feb 20

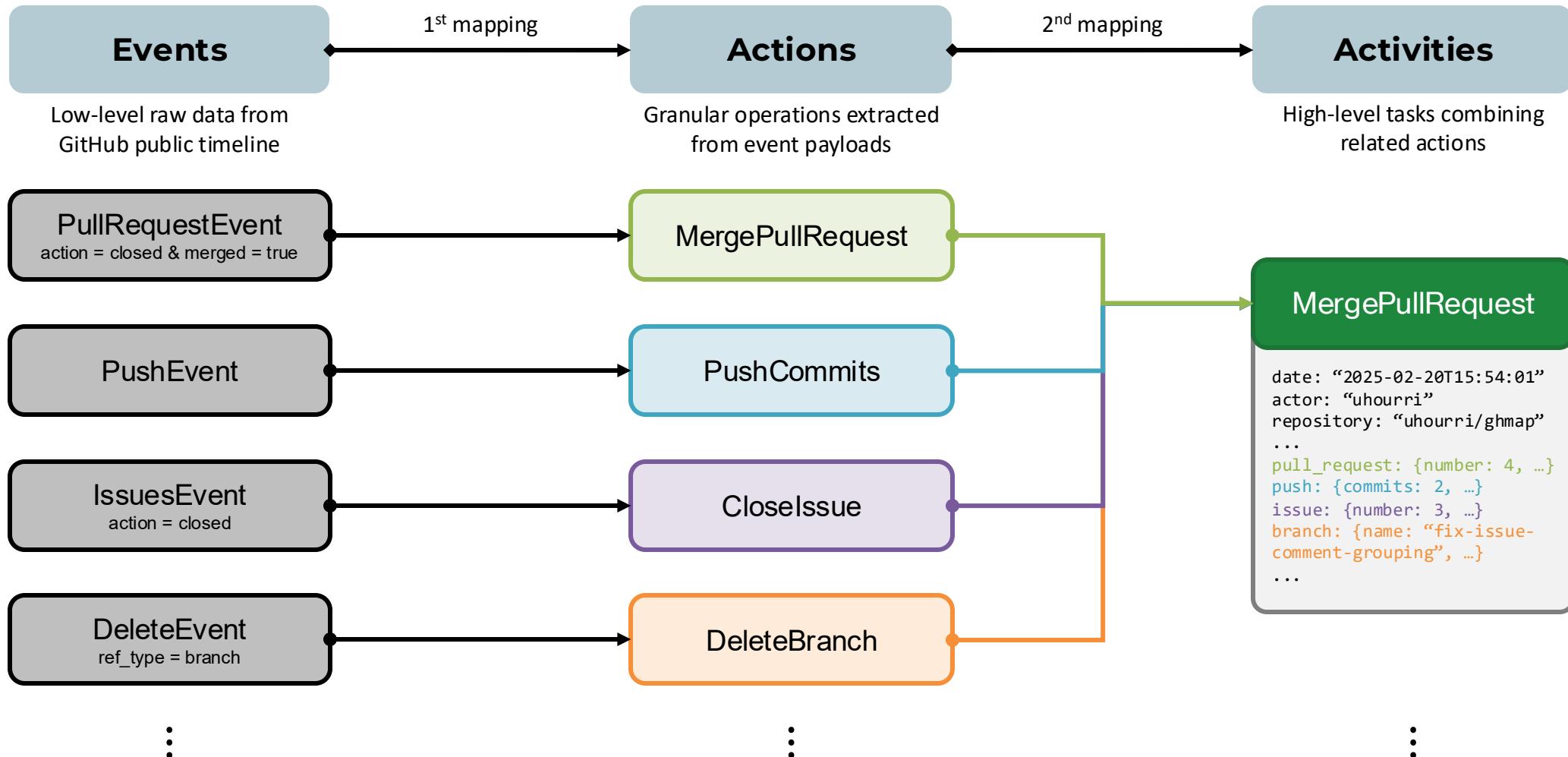
uhourri deleted the `fix-issue-comment-grouping` branch 2 months ago

Closed

**ghmap** is an open-source Python tool that converts raw GitHub event data into structured contributor actions and activities



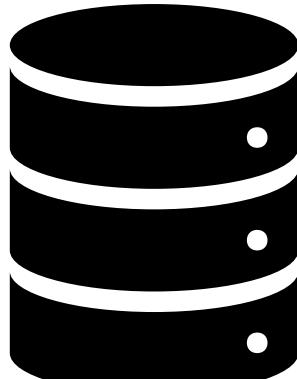
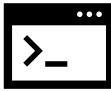
`pip install ghmap`



**Dataset** of contributor activities for the NumFocus community,  
built using *GH Archive* & *ghmap* command-line tool



58 projects  
2,851 repositories



3 years

+180k contributors



+2.7M actions  
+2.2M activities

Version 2025-02-24

