Dashboard

Type / to search

Top Repositories | New

Find a repository...

Home | Send feedback | Filter 8

Finding your digital sibling

"Can we group GitHub projects based on interactions and activities?"

rowandebruin/Discussion

> Research Question 1:
> We remove the most correlated metrics from the list in the results.
> We then get the following list of most expressive metrics: *Number of pull requests, Number of Releases, Number of branches, Number of Forks, Time between pull requests, Time between commits, Size of pull requests, Size of commits*

> Research Question 2:
> Analysis indicated inconclusive differences between group and project similarities. Meaning that our algorithm leads to different results as the CrossSim algorithm.
> Time constraints noted as a limitation, leading to the use of a smaller dataset for analysis.

rowandebruin/Conclusion

> Study confirms the feasibility of grouping GitHub projects based on interactions and activities.
> Identifies distinct project clusters with attributes related to interactions and activities.
> Emphasizes the importance of interactions and activities in project categorization.

rowandebruin/Subquestions

- RQ1: Which attributes that are related to interactions and activities would be suitable when comparing two open-source projects?
- RQ2: Does grouping GitHub repositories based on inter-action and activities bring the same or different results as compared to grouping on different metrics?

Shell · 2 · Updated on Jan 27, 2011

rowandebruin/Research Method

- Main Algorithm Building:
  - Read repository names.
  - Pull repositories.
  - Calculate similarity.
  - Group repositories.
- Metrics Selection
  - Find most expressive metrics
  - Kendall Tau Correlation on 80 projects
- Define similarity.
- Compare grouping methods.
- Compare group similarities.
- Determine similarity between different grouping algorithms.

Java · 1.1k · Updated 10 hours ago

rowandebruin/Results

Similarity matrix between groups

	Group 1	Group 2	Group 3
Group 1	1,00	0,97	0,91
Group 2	0,97	1,00	0,93
Group 3	0,91	0,93	1,00

Similarity matrices of the results of the CrossSim tool after running our own algorithm on it.

Similarity matrix inside group 1

	1	2	3	4	5	6
1	1,00	0,94	0,90	0,95	0,90	0,90
2	0,94	1,00	0,91	0,90	0,86	0,92
3	0,90	0,91	1,00	0,93	0,85	0,85
4	0,95	0,90	0,93	1,00	0,91	0,84
5	0,90	0,86	0,85	0,91	1,00	0,74
6	0,90	0,92	0,85	0,84	0,74	1,00

Shell · 2 · Updated on Jan 27, 2011

Metrics with the highest correlation values

Metrics	Correlation
Issues vs. Pull Requests	0.77655
Collaborators vs. Pull Requests	0.71070
Collaborators vs. Issues	0.69158
Forks vs. Stars	0.66745
Commits vs. Issues	0.66678
Forks vs. Watchers	0.65558
Commits vs. Pull Requests	0.60765

Latest repositories:

- rowan/CanWeGroupEm?
  2.6k · Python
- rowan/OfCourseWeCan!
  346 · JavaScript

rowandebruin/WhatsNext?

- Offers potential for businesses to find relatable projects easily, facilitating adoption of successful practices and enhancing efficiency.
- Calls for additional research in combining grouping algorithms to develop a comprehensive project search engine.
- Highlights the potential of exploring relations between interaction/activity metrics and source code/dependency metrics for deeper insights into project dynamics.

Rowan de Bruin
CSE, TUDELFT
4701801

E-mail:
R.W.deBruim@student.tudelft.nl
Phone number:
+31 (0) 6 22 36 32 34