

Towards Mining OSS Skills from GitHub Activity

Jenny T. Liang¹, Thomas Zimmermann², Denae Ford²

University of Washington¹, Microsoft Research²

jliang9@cs.washington.edu, [tzimmer, denae]@microsoft.com

Summary

Problem: Open source software (OSS) development requires contributors with diverse skill sets. For example, maintainers help lead the project and promote its longevity, technical writers assist with documentation, bug reporters identify defects in software, and developers program the software. However, to our knowledge, there are no tools that detect OSS-related skills, which includes both soft skills (e.g., communication) and hard skills (e.g., programming expertise).

Contribution: We present a novel method to detect OSS skills and prototype it in DISKO, a tool which detects and rates four types of OSS skills from GitHub data.

DISKO: Detecting Skills in OSS

Key insight: Identify relevant measurable GitHub activities or cues (i.e., *signals*) that correspond to OSS-related skills.

Teaches others to be involved in project

Contributes to a pull request with a newcomer at least 3 times

Contributes at least 5 changes to community health files at least once



Commitment to OSS project

At least 12 months where there is at least one contribution a month across all projects

Is at the 75 percentile by # of commits to a repository across users



Familiar with OSS practices

Has made a commit

Has made a comment on another's issue

Has been assigned to an issue, closed an issue, or merged a pull request



Programming language knowledge*

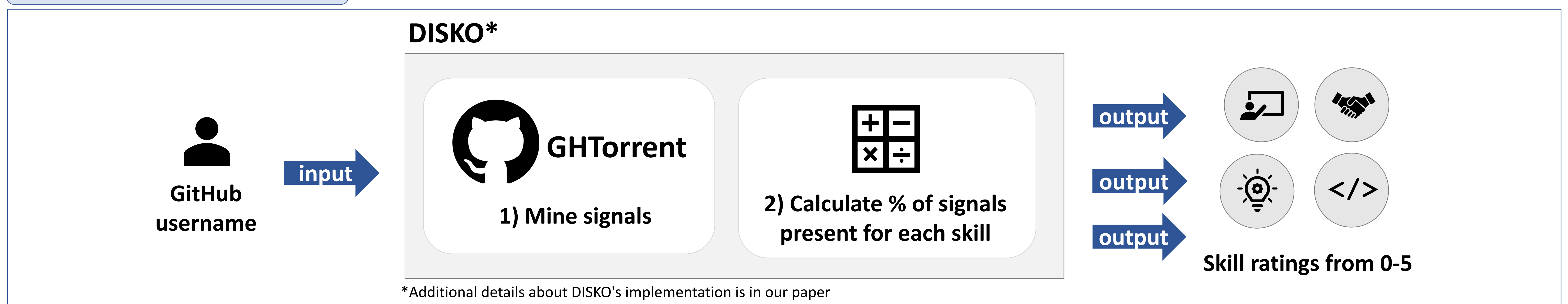
Has made a commit in the language at least once

The lines changed in the language is at the 60th percentile across users



* Languages: JavaScript (JS), Python, Java, PHP, C#, TypeScript (TS), Shell, C, Ruby

System overview



Evaluation & Results

We surveyed 455 OSS contributors who rated the importance of OSS skills & whether they'd show them publicly. 316 contributors self-evaluated their skills and provided GitHub usernames to evaluate DISKO's precision.

	Survey		DISKO Precision	
	Importance	Would Display	Skill Presence (> 0)	Skill Proficiency (> 3)
Teaches others to be involved in project	64%	69%	97%	73%
Commitment to OSS project	67%	67%	96%	63%
Familiar with OSS practices	56%	79%	97%	74%
Programming language knowledge	45%	54% (PHP) - 80% (JavaScript)	77% (Ruby) - 96% (Python)	56% (PHP) - 100% (C#)

Conclusion

Takeaway: OSS skills detection for soft and hard skills is a promising direction for SE researchers.

- OSS contributors find the selected OSS skills from DISKO important and would display them publicly.
- Detected skills can enhance existing SE / OSS experiences, e.g., project matching and mentor-mentee pairing.
- Our simple and intuitive skills detection method yields a promising baseline that future research can easily improve.