

F22.003: Task-Driven Skill Identification in OpenSource Project Issue Tracking Systems

Faculty mentor: Marco Aurelio Gerosa

Overview

OSS projects engage a massive, globally distributed community of contributors and establish a relevant economic motivation. The sustainability of some projects depends on their capacity to engage and retain a motivated group of developers. The first step towards contributing is challenging for some developers who pick up a task from a list of open issues, which may have varying complexity levels and require different skills to be completed. Additionally, the skills required for a task are hard to presume based on the available task data. Thus, identifying the required skills in a project is still a challenge. In this sense, the research aims to model which technical skills are required for a given OSS project by mining information from different artifacts (e.g., source code, pull requests, commits). New contributors usually do not have historical contributions to the project and do not know the project structure and libraries. Identifying the skills related to a project and its issues may help these contributors finding appropriate tasks.

What the student will DO and LEARN

The student will implement small applications related to the research embracing but not limited to:

- Machine Learning (ML)
- Natural Language Processing (NLP)
- Data Mining
- Statistic
- Data Engineering

Also, the intern may integrate the research team defining methods, planning, and executing user studies.

Additional benefits

Learn how to work in a research team.

Learn new programming languages, libraries, and frameworks.

Learn how to write scientific articles.

Additional qualifications

- Required:
 - Teamwork
 - Know a programming language (better Python or Java)
- Desirable:
 - Databases is a plus
 - R and statistic are a plus

Time commitment

6 hrs/week for 30 weeks

