Fondamentaux de la recherche reproductible et du logiciel libre
Fundamentals of reproducible research and free software

Miguel Colom

November 13, 2024

TP #2

In this second practical session, you'll study the reproducible aspects of the software project you have chosen. You're free to choose include what you consider is relevant, but some suggestions follow:

- Remind the kind of software: an application, a library, a framework, a platform, . . .

- According to the definitions of ACM, which is the level of reproducibility of the software project? Do we have artifacts to compare with? Is it repeatable? Is it reproducible? How far does its reproducibility go? For example, if the method is based on a large neural network, is it possible for you to re-train? Do you have all the required information from the pseudocodes or the article's description to recode the programs? It'd be very convincing if you perform an experiment yourself showing if you managed to obtain the same or comparable results when compiling and running the method.

- Does the software depend on 3rd-party libraries or services? How does it affect its reproducibility?

- Was the software peer-reviewed? Was it granted any badge or mention about reproducibility? Does the software implement any tests (say, unit tests or others) to ensure that the output artifacts don't change after code updates?

- Is the software well labeled with version and release numbers? Does the software itself have a DOI? Is it archived in Software Heritage and thus it has a SWHID?

- Which artifacts do we have? For example, is the source code available, along with a pseudo code? Any associated data? Output artifacts which we can compare to?

- If the software is not fully reproducible, what do you think might be the cause? (you might speculate a little bit here).

- What is the impact of being this software project reproducible (or non reproducible)? Compare your project to similar ones which are non-reproducible (or reproducible, if yours is not), in

terms of the impact they have (number of users, used in publications, forks in Github, . . .)

Note that these points are simply guidelines and you can include any elements that you consider important to describe the project.

The work is incremental, and therefore this will go together with what you started in TP #1. It is thus recommended that you progressively add content and modify the workinng report progressively.