WASP Software Engineering and Cloud Computing 2022

Report on the Pynblint Trial Session

Student Name: Romuald Esdras Wandji

It was a fruitful meeting with Luigi Quaranta throughout which I realized the importance of using static analysis for ML-related projects using Jupyter notebook. For Python (.py) files, several static analysis tools such as Pylint, Pyflakes, etc... are available, unfortunately, such features are not available for Notebook files (.ipynb), yet Notebook is a popular data science and ML development platform that allows for cell-by-cell execution of code blocks and therefore convenient for testing of blocks of code.

Pynblint is a great tool for statically analysing Python Jupyter Notebooks and throughout the meeting, we went through several aspects of how it can be used to not only optimize my code but also how it can make it more understandable by others. For instance, it leverages the features of Jupyter Notebook by recommending clarifying the notebook subject by writing an H1 Markdown heading in one of the initial cells of your notebook, suggesting consolidating cells when necessary, checking for invalid syntax in cells, and so forth.

For the time being Pynblint is not relevant to me since my research work is mostly focused on the theoretical aspect of AI for reasoning, and I am not planning to work on any implementation. But as a suggestion for improvement, I think a better integration for notebooks generated by other platforms such as Google collab, Cocalc, etc... will be of great importance. I think some of the limitations of static analysis are the fact it does not support several programming languages, it is likely to produce false positives and false negatives results, it cannot accommodate all possible running contexts, and finally, as it's only based on the statical review of the code, therefore has no knowledge of either the model or the data.