

[IAXO, VM@N, standalone] Инструменты для валидации научного софта

Задача:

Разработать фреймворк (вероятно на Питоне), в котором можно тестировать стабильность работы и и выдаваемых результатов для разных научных программ. Надо иметь возможность задавать для каждого конкретного теста данные и модуль с описанием теста (может быть много модулей для одного софта). При запуске вручную или при помощи CI эти тесты должны запускаться (по возможности параллельно). Должна быть настройка, позволяющая в случае сбоя одного из тестов продолжать выполнение остальных. По результату тестов должен генерироваться отчет (вероятно в html).

Помимо обычных тестов на выполнение, должны быть дата тесты, где сравниваются результаты. При этом надо учесть возможную статистическую природу результатов и считать не только полное совпадение, но и учитывать вероятность результата (то есть делать отрезку по p-value). Теоретическое обоснование может быть сделано на основе теории принятия статистических решений.

Применение:

Потребность в таком валидаторе есть во всех крупных коллаборациях. Непосредственное внедрение предполагается в экспериментах IAXO и VM@N.

Прототип:

Пример как можно сделать реализацию на основе unittest: <https://github.com/mipt-npm/scikit-validation>