

1. **Индекс УДК** – 004.048
2. **Название тезиса доклада** – Предсказание времени жизни строчек кода.
3. **Инициалы и фамилии авторов (название высшего учебного заведения (организации) полностью, название города)** – Петров Д.С. (Федеральное государственное автономное образовательное учреждение высшего образования „Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики“, Санкт-Петербург)
4. **Научный руководитель (название высшего учебного заведения (организации) полностью, название города)** – Ключев А.О. (Федеральное государственное автономное образовательное учреждение высшего образования „Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики“, Санкт-Петербург)

Краткое введение

Известная проблема, возникающая у программистов в повседневной жизни, заключается в постоянном переписывании кода. Известно, что период полураспада кода в репозитории ядра Linux составляет пол года.

Практической целью является увеличение времени жизни строк кода в проекте.

Цель работы

Целью работы является построение pipeline обучения, частью которого является интегрированная среда разработки IntelliJ IDEA. Написать плагин, встроить в него решающую функцию, предсказывающую время жизни строчек кода в репозитории.

Базовые положения исследования

В качестве факторов будут выступать сведения об авторе, число дней, сколько уже прожила строчка, сведения о самой строчке. В качестве репозитория будет выступать официальный репозиторий IntelliJ IDEA Community edition. Таргет – это время жизни строчки кода.

Факторы можно добывать из распределённой системы управления версиями git, а также из абстрактного синтаксического дерева (AST), построенное IntelliJ IDEA.

IntelliJ IDEA представляет из себя интегрированную среду разработки программного обеспечения, имеющую плагинистую архитектуру. Под IDEA разработаны более трех тысяч плагинов, примерно половина из которых имеет открытый исходный код, что позволяет изучать чужие плагины, а также разрабатывать свои. Главная особенность разработки плагина под IDEA заключается в том, что программист работает не со строками кода, а с так называемыми элементами PSI (Program Structure Interface), которые в совокупности представляют из себя обогащенное AST, а это значит, что IDEA можно использовать как источник факторов.

Из IDEA также можно вызывать команды git и получать результат в виде конкретных структур данных, которые в свою очередь легко обрабатывать непосредственно в плагине.

IDEA является основным источником факторов. Более того, можно вычислять факторы в режиме реального времени, а значит, можно встроить обученную на датасете решающую функцию в плагин.

Промежуточные результаты

Написание простого плагина, работающего с AST и git. Сбор датасета. Выбор целевой функции, построение pipeline обучения.

Основной результат

Получение решающей функции. Встраивание решающей функции в плагин.