

Проверка теории "6 рукопожатий" на примере локального git-репозитория

Выполнила:

Студентка ВМК МГУ, 241гр

Куулар Валерия

Цель и задачи

Написать программу на Python, которая вычисляет наибольшую длину из всех минимальных длин путей "рукопожатий" между программистами.

Будем считать, что между программистами есть "рукопожатие", если они отправляли коммиты с изменением одного и того же файла.

Задачи:

1. Собрать данные о коммитах локального git-репозитория
2. Построить граф программистов
3. Найти наибольшую длину "рукопожатий" между программистами

Реализация программы

- Сбор данных о коммитах локального репозитория:
- **data_dict = collect_data(repo_path)**
 - Для выполнения данной задачи используется библиотечка GitPython;
 - Из списка всех коммитов создаем словарь, где key – source, value – set of programmers;
 - Фильтруем словарь. Избавляемся от файлов, которые имеют только одного программиста-коммитера;
- Построение графа:
- **G = gr(data_dict)**
 - Используя библиотеку NetworkX, создаем граф с вершинами - программисты, ребра - рукопожатие между ними;
- Нахождение пути:
- **max_handshakes(G)**
 - Используя библиотеку NetworkX, создаем словарь расстояний между друг другом всех вершин;
 - Перебором значений словаря находим максимальное расстояние среди наименьших;

Результаты тестирования

Git-repo	handshakes	commits	All programmers	*Filtered programmers	time
cli	3	1.787	174	174	18s
paramiko	2	4.020	181	181	39s
keras	3	11.397	1389	1388	129s
matplotlib	3	52.025	1818	1818	576s
pytorch	-	85.290	4872	-	-
tensorflow	-	177.104	4495	-	-

*filtered programmers- отфильтрованный список программистов, который не включает в себя программистов, работающих над одним файлом в одиночку и не фигурирующих в других коммитах.

all programmer - полный список программистов - авторов коммитов.

Результаты

- Поставленная цель была достигнута: мы выяснили, что максимальная длина кратчайших путей "рукопожатий" между программистами < 6 .
- Данная реализация задачи тратит на обработку таких больших git-репозиторий, как tensorflow и pytorch, огромное количество времени.