

федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

О Т Ч Е Т **по исследовательскому проекту**

(«Кино-граф: актёры, фильмы и связи между ними»)

Содержание

1. Общее описание проекта
2. Содержательная часть
3. Заключение
4. Результат проекта
5. Приложения

1. Общее описание проекта

1.1. Инициатор, заказчик проекта

Инициатором и заказчиком проекта является образовательная программа НИУ ИТМО в рамках учебной дисциплины, ориентированной на проектную и исследовательскую деятельность студентов

1.2. Тип проекта

Проект является исследовательским, так как направлен на анализ реальных данных, построение модели и получение выводов на основе обработки и интерпретации результатов.

2. Содержательная часть

2.1. Цели и задачи проекта

Цель проекта — построение и анализ графа связей между актёрами и фильмами для выявления ключевых участников и структурных особенностей киноиндустрии.

Задачи проекта:

1. Сбор и предварительная обработка данных из датасета.
2. Проектирование структур данных для представления графа.
3. Построение графа с двумя типами узлов.
4. Анализ центральности и связности.

5. Визуализация результатов.
6. Подготовка аналитического отчёта.

Источник данных: датасет *IMDB Top 1000*

2.2. Описание данных и их обработка

Исходный CSV-файл содержит сведения о фильмах из рейтинга IMDB Top 1000, включая название, год выпуска, жанры, рейтинг и четырёх главных актёров.

2.2.1. Производные файлы

Файл	Назначение
edges.csv	Рёбра графа (актёр–фильм, актёр–актёр)
stats.csv	Статистика по актёрам (films_cnt, actors_cnt)
gra.json	Граф в формате списка смежности
spri.csv	Распределение фильмов по годам выпуска

Файлы формируются автоматически с использованием Python-скриптов (sceery.py, tops.py, JSONtable.py, spread.py).

2.3. Модель графа

2.3.1. Узлы

- **Актёры** — вершины вида a_Имя_Фамилия.
- **Фильмы** — вершины вида f_Название_фильма.

2.3.2. Рёбра

- **актёр — фильм:** участие актёра в съёмках;
- **актёр — актёр:** совместное участие в одном фильме.

2.4. Используемые структуры данных и алгоритмы

2.4.1. Структуры данных

- list — хранение списков фильмов и актёров;
- set — хранение уникальных рёбер;
- dict — представление графа в виде списка смежности;
- networkx.Graph — анализ и визуализация графа.

2.4.2. Алгоритмы

- генерация рёбер для каждого фильма;
- подсчёт степени вершин актёров;

- анализ компонент связности;
- визуализация с использованием алгоритма `spring_layout`.

2.5. Результаты анализа

2.5.1. Топ-актёры по связности

Анализ выполнен на основе файла `stats.csv`, содержащего автоматически рассчитанные показатели количества фильмов и уникальных актёрских связей.

Актёр	Кол-во фильмов	Кол-во связей
Robert De Niro	17	45
Tom Hanks	14	38
Brad Pitt	12	36
Al Pacino	13	35
Clint Eastwood	12	33

Вывод: наиболее центральными вершинами графа являются Robert De Niro и Tom Hanks, обладающие максимальными значениями по количеству фильмов и числу уникальных совместных актёрских связей. Это указывает на их высокую степень интеграции в крупнейший компонент связности графа.

2.5.2. Кластеризация и связность

На основе анализа графа и визуализации выявлено 8 компонент связности. Наиболее крупная компонента включает преимущественно голливудских актёров и фильмы. Остальные компоненты соответствуют региональным киноиндустриям (Индия, Япония и др.).

2.5.3. Распределение фильмов по годам выпуска

Анализ файла `spri.csv` показал, что пик количества фильмов в рейтинге IMDb Top 1000 приходится на 2014 год — 32 фильма, что является максимальным значением за весь период наблюдений.

2.6. Визуализация

Граф связей построен с использованием библиотек `networkx` и `matplotlib`. Узлы актёров выделены синим цветом, узлы фильмов — красным.

3. Заключение

В результате выполнения проекта были достигнуты поставленные цели. В ходе работы были развиты навыки анализа данных, построения графовых моделей и интерпретации результатов. Проект способствовал формированию компетенций в области работы со

структурами данных, алгоритмами и инструментами визуализации.

4. Результат проекта

Результатом проекта являются:

- аналитический отчёт;
- граф связей актёров и фильмов;
- набор автоматически сформированных CSV- и JSON-файлов;
- визуализации и материалы для защиты проекта.

5. Приложения

В приложениях представлены дополнительные материалы, подтверждающие получение результатов проекта:

- файлы edges.csv, stats.csv, spri.csv, gra.json;
- Python-скрипты обработки данных;
- визуализации графа;
- презентационные материалы для защиты проекта.