# Анализ поведения пользователей

## Лабораторный практикум

## Требования

* ОС Windows (либо ОС Linux) с установленным Python-интерпретатором и IDE (IDE опциональна, если вы собираетесь использовать Google Colab или Jupyter Notebook);
* Предварительно загруженный датасет;
* Подключение к сети Интернет;
* Предварительно установленные библиотеки для Python:
  + Numpy, itertools.

## Постановка задачи

Дан набор пользователей и ресурсов, к которым у них есть доступ. Набор ресурсов для всех пользователей одинаков, но разные пользователи могут иметь доступ к разным ресурсам. Требуется произвести т.н. “role-mining”, т.е. разделить пользователей на некоторое число групп, исходя из доступных им ресурсов. Для оценки качества деления используйте число групп, степень наложения доступов к ресурсам между группами и степень покрытия ресурсов группами пользователей.

## Рекомендации по использованию инструментов для решения задачи

Датасет дан в Java-формате .ass, чтобы понять, как прочитать файл, можете открыть его блокнотом и посмотреть на структуру. Для разделения пользователей на группы используйте один из role mining алгоритмов. Рекомендуется попробовать несколько разных алгоритмов. Четкой метрики или ожидаемого результата нет.

## Ожидаемый результат

Четкой цели нет, однако, число групп не должно превышать 25, не должно быть недоступных никому ресурсов и наложение между группами не должно быть слишком большим.

## Рекомендованная литература

1. <https://www.researchgate.net/publication/221366816_Evaluating_role_mining_algorithms> - обзорная статья по решению проблемы role mining
2. <https://www.researchgate.net/publication/221366849_Role_engineering_using_graph_optimisation> - доклад с конференции о решении задачи role mining на графах
3. <http://www.sis.pitt.edu/jjoshi/courses/IS2620/Spring09/schlegelmilch.pdf> - решение задачи role mining с помощью алгоритма ORCA
4. <https://www.hpl.hp.com/techreports/2008/HPL-2008-33.pdf> - статья от экспертов HP, посвященная одному из подходов role mining