

# Тематический анализ корпуса текстов

Выполнил: Андрей Орлов

Курс ДПО по направлению  
«Компьютерная лингвистика», НИУ ВШЭ

Москва, 2023 г.

## Задача проекта:

- нормализовать данные в корпусе текстов;
- построить таблицу сопряженности тем корпуса;
- выполнить статистические тесты на выделение временного тренда по определенной теме

Материал для работы – новостные публикации на сайте [SecurityLab.ru](https://securitylab.ru)

Объем корпуса: 8000 текстов

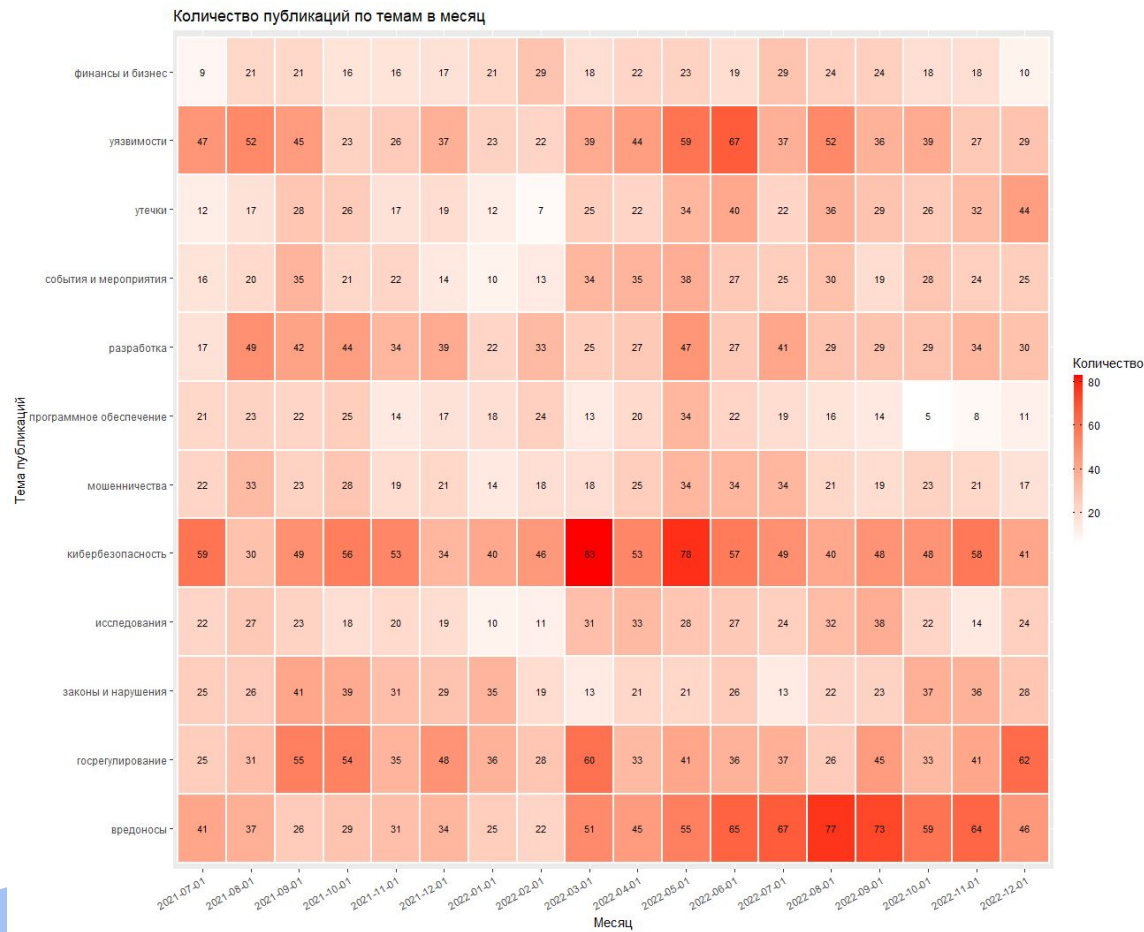


Таблица сопряженности по темам

Количество публикаций по теме "Утечки"

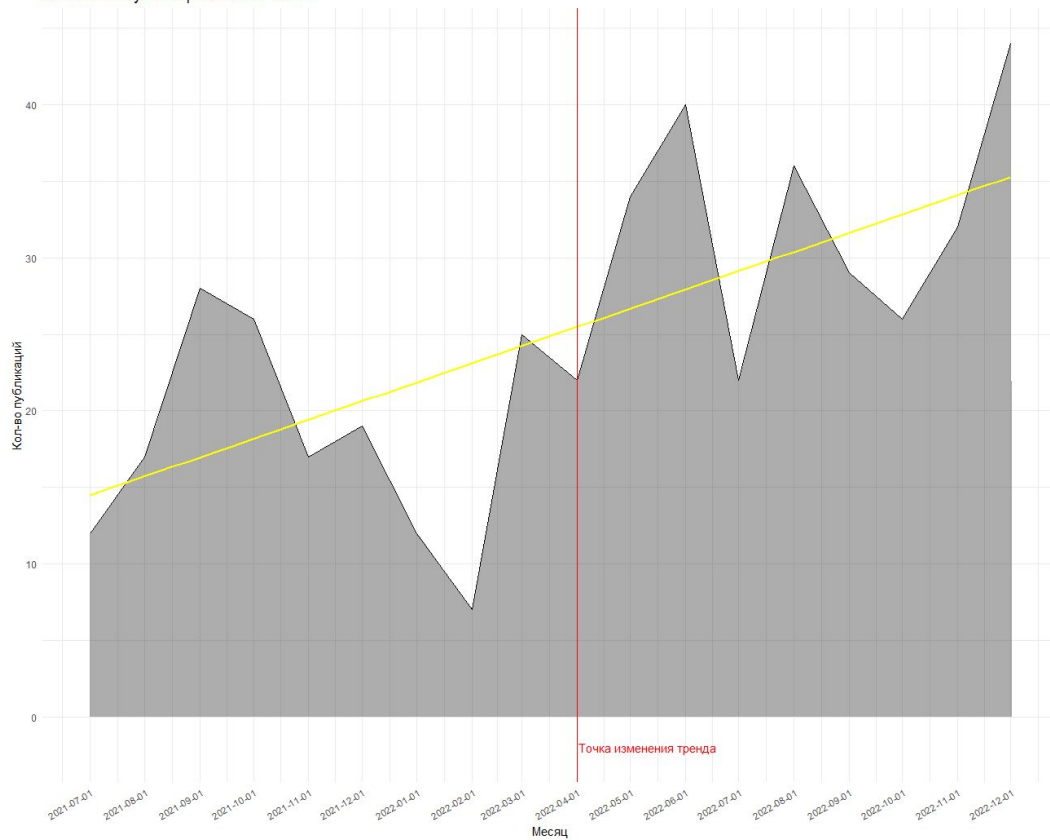


График количества публикаций по теме и  
выделение временного тренда