

ТИЖДЕНЬ #5

ВИЗНАЧЕННЯ ВИМОГ ДО ПРОГРАМИ

Цього тижня було остаточно вирішено зосередити свою роботу на обробці та аналітиці даних, отриманих із судового реєстру.

Я визначила функціональні та нефункціональні вимоги до програми, яку реалізовуватиму. Кілька своїх ідей щодо можливого аналізу текстів судових рішень і зведення інформації до статистики описала у вимогах, та все ж в мене є ще декілька ідей щодо функціоналу, зокрема розділу “Аналітика”, але я поки що не впевнена в можливості їх реалізації на основі тих даних і знань, які в мене є.

Почала дивитись за літературою чи інформацією, пов’язаною з тематикою, поки що зосередилася на публікаціях Daniel Martin Katz і в процесі виникла ідея ще прогнозувати рішення певного судді на основі аналізу його попередніх рішень в певному розділі, скажімо, кримінального кодексу. Але це, мабуть, малоймовірно в рамках курсової.

Виникло питання щодо того, чи спершу завантажувати дані а потім групувати і аналізувати їх, чи організовувати роботу програми так, щоб все це робилося відповідно до запиту і в реальному часі. Я схилилася до першого варіанту, оскільки об’єми даних великі. Можливо варто було б створити свою базу даних на основі MySQL чи MariaDB, тим більше мені завжди було цікаво дізнатись як працюють ці бази даних.

Щодо аналітики – я думаю, що треба вибрати якусь частину з усіх цих судових рішень і працювати з нею, оскільки загалом їх зараз є більше 70 мільйонів. Можна аналізувати лише за останні роки, або ж обрати якісь цікаві ключові слова і шукати рішення тільки за ними (як, наприклад, “OLX” – такі рішення справді цікаво навіть просто читати, можна дізнатись як хтось через OLX купляє чистого білого паперу на суми, які вимірюються в сотнях тисяч гривень, а потім залишається ні з чим).

Між тим ознайомилась із документацією стандартних пакунків Python для роботи з xml та спробувала ознайомитись з документацією lxml, поки не знаю наскільки це було ефективним, тому що об’єм інформації дуже великий.