**Пропозиція проекту**

| **Аспект пропозиції** | **Суть аспекту пропозиції** |
| --- | --- |
| Назва програмної системи | Fair Online Judge |
| Призначення системи. | Призначенням розроблюваної системи є **визначення близькості текстових файлів** (наприклад, програмних кодів) у структурному та семантичному розумінні, **візуалізація означеної близькості.** Система буде розгорнута у вигляді веб-серверу.  Сценарій користування наступний:   1. Надати довільну кількість текстових файлів для подальшої обробки; 2. Отримати посилання від сервісу на результати обчислення - користувачі, перейшовши за посиланням, можуть виконати дії з пунктів 3-4; 3. Отримати діаграму близькості завантажених корпусів тексту; 4. Детально вивчити різницю між двома підозрілими на схожість корпусами тексту через веб-інтерфейс. |
| Порівняння з існуючими аналогічними системами. Ефект від її впровадження. | Авторам не відомий сервіс, що надавав би можливість порівняння та візуалізації близькості двох та більше файлів з можливістю збереження результату на сервері після обрахунків.  Ефектом від впровадження є надання унікального сервісу; також відкритий код на репозиторії сприятиме зацікавленості у вдосконаленні або наданні схожого рішення. |
| Загальна характеристика задачі, що вирішується. | Вирішення прикладної задачі є необхідною частиною для демонстрації існуючого успіху у вирішенні дослідницьких задач у галузі роботи з текстовими даними на структурному та семантичному рівні. Характеристика задачі подана у вигляді основних підзадач:   * Пошук близькості між текстами може базуватися на статистичних та інших методах розпізнавання та обробки людської мови (NLP) - насамперед TF-IDF, Word Embedding. * Візуалізація близькості текстових корпусів може бути виконана у вигляді:   1. теплових карт (зображення 3D на площині із розподілом насиченості кольору на ній);   2. двовимірного зображення відстані між однією точкою до усіх інших;   3. t-sne діаграма   4. повноцінне 3D-зображення та інші підходи. |
| Мотивація вибору задачі | * Стимулювання зацікавленості до досліджень у напрямку методів роботи із текстовими даними на семантичному рівні, впровадження успіхів у цьому напрямку досліджень в програмну систему; * Бажання вирішення проблеми справедливого оцінювання робіт шляхом візуалізації подібності рішень; * Бажання розібратися у наявних рішеннях задач NLP на прикладі їх практичного застосування. |
| Аналіз здійсненності. Необхідні для виконання проекту ресурси. Основний ризик для проекту. | Складність полягає у конструктивному оцінюванні необхідного часу для розробки або інтеграції готових рішень задачі визначення схожості текстів у розроблювану систему.  Ризиком є дослідницька складова проекту - пошук вирішення задачі визначення схожості текстів. |
| Модель виробничого циклу. Обгрунтування вибору. Відмінні риси процесу розробки обраної системи. | В якості моделі була вибрана ітеративна модель. З вибором такої моделі на кожному кроці буде змога повенутися на попередні для редагування аспектів проекту. |

**Команда: Геворгян Артем(Developer, Project Manager), Пилипець Гліб(Main Developer).**