

# Пропозиція проекту

Аспект пропозиції	Суть аспекту пропозиції
Назва програмної системи	Fair Online Judge
Призначення системи.	<p>Призначенням розроблюваної системи є <b>визначення близькості текстових файлів</b> (наприклад, програмних кодів) у структурному та семантичному розумінні, <b>візуалізація означеної близькості</b>.</p> <p>Система буде розгорнута у вигляді веб-серверу.</p> <p>Сценарій користування наступний:</p> <ol style="list-style-type: none"><li>1. Надати довільну кількість текстових файлів для подальшої обробки;</li><li>2. Отримати посилання від сервісу на результати обчислення - користувачі, перейшовши за посиланням, можуть виконати дії з пунктів 3-4;</li><li>3. Отримати діаграму близькості завантажених корпусів тексту;</li><li>4. Детально вивчити різницю між двома підозрілими на схожість корпусами тексту через веб-інтерфейс.</li></ol>
Порівняння з існуючими аналогічними системами. Ефект від її впровадження.	<p>Авторам не відомий сервіс, що надавав би можливість порівняння та візуалізації близькості двох та більше файлів з можливістю збереження результату на сервері після обрахунків.</p> <p>Ефектом від впровадження є надання унікального сервісу; також відкритий код на репозиторії сприятиме зацікавленості у вдосконаленні або наданні схожого рішення.</p>
Загальна характеристика задачі, що вирішується.	<p>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.</p> <p>Характеристика задачі подана у вигляді основних підзадач:</p> <ul style="list-style-type: none"><li>• Пошук близькості між текстами може базуватися на статистичних та інших методах розпізнавання та обробки людської мови (NLP) - насамперед TF-IDF, Word Embedding.</li><li>• Візуалізація близькості текстових корпусів може бути виконана у вигляді:<ol style="list-style-type: none"><li>а. теплових карт (зображення 3D на площині із розподілом насиченості кольору на ній);</li></ol></li></ul>

Аспект пропозиції	Суть аспекту пропозиції
	<ul style="list-style-type: none"> <li>b. двовимірного зображення відстані між однією точкою до усіх інших;</li> <li>c. t-sne діаграма</li> <li>d. повноцінне 3D-зображення та інші підходи.</li> </ul>
Мотивація вибору задачі	<ul style="list-style-type: none"> <li>• Стимулювання зацікавленості до досліджень у напрямку методів роботи із текстовими даними на семантичному рівні, впровадження успіхів у цьому напрямку досліджень в програмну систему;</li> <li>• Бажання вирішення проблеми справедливого оцінювання робіт шляхом візуалізації подібності рішень;</li> <li>• Бажання розібратися у наявних рішеннях задач NLP на прикладі їх практичного застосування.</li> </ul>
Аналіз здійсненності. Необхідні для виконання проекту ресурси. Основний ризик для проекту.	<p>Складність полягає у конструктивному оцінюванні необхідного часу для розробки або інтеграції готових рішень задачі визначення схожості текстів у розроблювану систему.</p> <p>Ризиком є дослідницька складова проекту - пошук вирішення задачі визначення схожості текстів.</p>
Модель виробничого циклу. Обґрунтування вибору. Відмінні риси процесу розробки обраної системи.	В якості моделі була вибрана ітеративна модель. З вибором такої моделі на кожному кроці буде змога повернутися на попередні для редагування аспектів проекту.

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