

Пропозиція проекту

Аспект пропозиції	Суть аспекту пропозиції
Назва програмної системи	Fair Online Judge
Призначення системи.	<p>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.</p> <p>Сценарій користування наступний:</p> <ol style="list-style-type: none">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.	<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">• 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:<ol style="list-style-type: none">а. теплових карт (зображення 3D на площині із розподілом насиченості кольору на ній);

Аспект пропозиції	Суть аспекту пропозиції
	<ul style="list-style-type: none"> b. двовимірного зображення відстані між однією точкою до усіх інших; c. t-sne діаграма d. повноцінне 3D-зображення та інші підходи.
Мотивація вибору задачі	<ul style="list-style-type: none"> • 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.	<p>Складність полягає у плануванні розробки проекту, тобто рішенні підзадач, зазначених у відповідних пунктах, проектуванні архітектури, опануванні фронт-енд частини.</p> <p>Ризиком є</p> <ul style="list-style-type: none"> ▪ можливість неіснування на даний момент задовільного вирішення задачі визначення близькості текстів (зокрема програмних кодів), ▪ можлива незадовільність методів структурного аналізу програм для визначення близькості текстів у задовільному для людини розумінні, ▪ можлива недостача в часі для виконання бажаного обсягу роботи, що призведе до урізання функціоналу, ▪ вибрана ітеративна модель, що може сповільнити процес розробки, ▪ відсутність розробника, що на достатньому рівні вміє робити Front-End.
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).