

РОЗРОБКА ВЕБ-ЗАСТОСУНКУ ДЛЯ АНАЛІЗУ ТА ВІЗУАЛІЗАЦІЇ БІОЛОГІЧНИХ ДАНИХ

Лахтін В. В.

Науковий керівник – д. т. н., проф. Саваневич В. Є.

Харківський національний університет радіоелектроніки, каф. СТ

м. Харків, Україна

e-mail: vitalii.lakhtin@nure.ua

This report discusses the development of a web application designed for the sophisticated analysis and visualization of biological data. It highlights the integration of cutting-edge web technologies and interactive data visualization methods to assist in the comprehensive examination of genomes datasets. Crucially, the application leverages cloud computing technologies to ensure scalability, reliability, and accessibility, allowing users to process and analyze large volumes of biological data efficiently. This cloud-based architecture not only enhances computational power and storage capacity but also enables seamless collaboration among researchers across the globe.

У докладі розглядається процес проектування та розробки веб-застосунку для аналізу та візуалізації біологічних даних, який включає в себе створення комплексної платформи, призначеної для дослідників та спеціалістів у галузі біотехнологій. Застосунок має на меті спростити процес аналізу великих об'ємів біологічної інформації, яка походить з різноманітних джерел, таких як геномні послідовності, забезпечуючи при цьому високий рівень точності та ефективності обробки даних. Основну увагу приділено розробці інтуїтивно зрозумілого користувацького інтерфейсу та реалізації алгоритмів для обробки та візуалізації біологічних даних, які дозволяють користувачам легко ідентифікувати важливі біологічні патерни та тенденції.

Зареєстрованому користувачу надається можливість додавати геномні послідовності, щоб отримувати прогнози та розуміння наслідків генетичних варіацій [1]. Додаток має інструменти візуалізації даних для спрощення інтерпретації геномної інформації.

Зареєстрований користувач має доступ до функцій:

– створення облікових записи для збереження своїх аналізів і результатів;

– інструменти для обрізання послідовностей для видалення низькоякісних або неоднозначних даних. Фільтрування послідовностей, які не відповідають певним пороговим значенням або критеріям якості. Нормалізація, попередньої обробка, щоб підготувати дані для аналізу;

– аналіз послідовності, вирівнювання послідовності, обчислення GC-вмісту;

– візуалізація послідовностей з анотаціями, генами. Теплові карти, щоб показати присутність, відсутність або частоту певних ознак послідовності в зразках або умовах;

– пошук за допомогою фільтрів: тип послідовності (ДНК, РНК), організм (людина, миша), джерело даних (загальнодоступні бази даних, такі як NCBI, спеціальні набори даних) або конкретні анотації;

– експорт своїх результатів та візуалізації в різних форматах (наприклад, PNG, CSV, TSV) для звітності та подальшого аналізу.

Система створена в середовищі IntelliJ IDEA, використовуючи Java Development Kit версії 11 для розробки бекенд-частини. Задіяні такі інструменти та технології як Gradle, Spring Boot та Hibernate. Gradle слугує інструментом для ефективної автоматизації збірки проєктів, Spring Boot спрощує процес створення високоефективних застосунків на Java, а Hibernate допомагає в роботі з базами даних, використовуючи об'єктно-орієнтований підхід для автоматизації генерації SQL-запитів [3].

Розробка клієнтської частини системи здійснена на мові TypeScript із використанням фреймворка Angular [2] та бібліотеки RxJS для асинхронної роботи компонентів. Для створення та стилізації веб-сторінок також застосовано HTML і CSS.

Для розгортання та управління інфраструктурою проєкту обрано хмарну платформу Amazon Web Services (AWS). Серед використаних ресурсів: Amazon EC2 для запуску серверних інстансів, Amazon S3 для ефективного зберігання даних та розподілених ресурсів, а також Amazon RDS, що дозволило оптимізувати управління базами даних. Використання AWS як основи для хмарної інфраструктури забезпечило високий рівень надійності, доступності та безпеки проєкту, а також спростило процес управління ресурсами [4].

Список використаних джерел:

1. Kim J. Genome Data Analysis. 1th ed. Berlin : Springer, 2019. 75 p.
2. Документація до Angular. URL: <https://metanit.com/web/angular> (дата звернення: 02.03.2024).
3. Документація до Spring. URL: <https://docs.spring.io> (дата звернення: 02.03.2024).
4. Richards M., Ford N. Fundamentals of Software Architecture: An Engineering Approach. 1th ed. London : O'Reilly, 2020. 345 p.