

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
Кафедра системного анализа и компьютерного моделирования

Лабутин Михаил Дмитриевич

**Разработка сервера веб-платформы анализа
данных полноэкзномного секвенирования**

Аннотация к магистерской диссертации

специальность «**Радиофизика**»

Научный руководитель:
кандидат физико-математических наук,
доцент В.В. Скакун

Допущена к защите

«__» _____ 2021 г.

Зав. кафедрой системного анализа и
компьютерного моделирования

кандидат физико-математических наук,
доцент _____ В.В. Скакун

Минск, 2021

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель работы - исследование и реализация возможности автоматизации процессов исследования генетических данных, полученных при высокопроизводительном секвенировании ДНК. В данной работе предметной областью является поиск вариаций числа копий генов. Поиск этих вариаций у людей различается в зависимости от популяции и выбора алгоритмов поиска. Поиск вариантов занимает много времени и представляет собой многошаговый линейный алгоритм. Разработанная платформа планируется к внедрению в Центре детской онкологии, гематологии и иммунологии и находится в режиме финальной апробации на реальных пациентах.

Задачи:

1. Изучить спецификацию программных модулей, чтобы найти различия в количестве копий.
2. Спроектировать и разработать серверное приложение для автоматизации получения промежуточных результатов поиска.
3. Проанализировать и протестировать разработанную платформу на реальных данных пациентов.

Серверная часть имеет микросервисную архитектуру и разработана на языке программирования Java с использованием фреймворка Spring Boot.

АГУЛЬНАЯ ХАРАКТАРЫСТЫКА ПРАЦЫ

Мэта працы - даследаваць і рэалізаваць магчымасць аўтаматызацыі працэсаў даследавання генетычных дадзеных, атрыманых у ходзе высокапрадукцыйнага секвеніравання ДНК. У дадзенай працы ў якасці прадметнай вобласці з'яўляецца пошук варыяцый колькасці копій генаў. Пошук дадзеных варыяцый у чалавек розны ў залежнасці ад папуляцый і выбары алгарытмаў пошуку. Пошук варыяцый займае працяглы час і ўяўляе сабой шматступенны лінейны алгорымт. Распрацаваная платформа плануецца для ўкаранення ў ДУ "Цэнтр дзіцячай анкалогіі, гематалогіі і імуналогіі" і знаходзіцца ў рэжыме фінальнага тэставання на рэальных пацыентах.

Задачы:

1. Вывучыць спецыфікацыю праграмных модуляў для пошуку варыяцый колькасці копій.
2. Спраектаваць і распрацаваць сервернае прыкладанне для аўтаматызацыі атрымання прамежкавых вынікаў пошуку.
3. Правесці аналіз і апрабацыю распрацаванай платформы на дадзеных рэальных пацыентах.

Серверная частка мае микросервісную архітэкттуру і распрацавана на мове праграмавання Java з выкарыстаннем фреймворка Spring Boot.

ABSTRACT

The aim of this work is to investigate and implement the possibility of automating the processes of studying genetic data obtained during high-throughput DNA sequencing. In this work, as a subject area is the search for variations in the number of copies of genes. The search for these variations in humans differs depending on the populations and the choice of search algorithms. The search for variations takes a long time and is a multi-step linear algorithm. The developed platform is planned for implementation in the State Institution “Center for Pediatric Oncology, Hematology and Immunology” and is in the final testing mode on real patients.

Tasks:

1. Examine the specification of software modules to find variations in the number of copies
2. Design and develop a server application to automate the receipt of intermediate search results
3. Analyze and test the developed platform on real patient data.

The server side has a microservice architecture and is developed in Java using the Spring Boot framework.