

Отзыв научного руководителя

на диссертационную работу Карпулевича Евгения Андреевича на тему:

“Построение программного конвейера для выравнивания последовательностей в приложениях биоинформатики”,

представленную на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 - “Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей”.

Карпулевич Е.А. (1990 г.р.) окончил факультет вычислительной математики и кибернетики Московского государственного университета имени М.В. Ломоносова по специальности «Прикладная математика и информатика».

С 2015 года Карпулевич Е.А. является сотрудником ИСП РАН, с 2017 года - сотрудником отдела “Информационные системы” и занимается вопросами разработки методов и программных инструментов для автоматизированной обработки больших биомедицинских данных. В представляемой диссертации исследование сфокусировано на применении математических методов выравнивания последовательностей в приложениях биоинформатики.

Тема диссертационного исследования Карпулевича Е.А. имеет высокую актуальность в связи с отсутствием методов выравнивания сочетающих в себе преимущества методов выравнивания на линейный референсный геном и выравнивания на граф, составленный по ДНК нескольких организмов. При этом, например, в такой области как биоинформатика выравнивание генетических последовательностей является одной из важных задач, от результатов решения которой зависят все последующие этапы анализа. В результате работы Карпулевича Е.А. был разработан метод выравнивания генетических последовательностей на референсный геном с использованием данных об известных генетических вариантах, алгоритмы в составе метода выравнивания генетических последовательностей и аналитические оценки их вычислительной и пространственной сложности через доказательство соответствующих теорем. Была предложена архитектура и реализация системы анализа данных NGS на базе программного конвейера для обработки данных секвенирования ДНК человека с использованием модифицированного индекса.

Современные методы для решения исследуемой задачи опираются на использование инструментов в основе которых лежат алгоритмы выравнивания на линейный референсный геном или алгоритмы выравнивания на пангеномный граф. Недостатком алгоритмов выравнивания на линейный референсный геном является их сравнительно низкая точность, недостатком методов выравнивания на пангеномный граф является их большая вычислительная и пространственная сложность. В рамках работы Карпулевича Е.А. предлагается метод выравнивания генетических

последовательностей на референсный геном с использованием данных об известных генетических вариантах основанный на понятии минимизатора и применимый к генетическим данным.

При разработке метода выравнивания генетических последовательностей на референсный геном с использованием данных об известных генетических вариантах нужно учитывать требования к вычислительным ресурсам необходимым для работы алгоритмов в составе метода. Оценки, полученные в результате доказательства теорем, показывают, что вычислительная сложность алгоритмов построения создания модифицированного индекса референсной генетической последовательности остается линейной, а вычислительная сложность алгоритмов выравнивания генетических последовательностей на модифицированный индекс не изменяется по сравнению с выравниванием на индекс референсного генома. Теорема об оценке пространственной сложности позволяет оценить количество оперативной памяти необходимой для работы реализации алгоритмов.

Результаты диссертации имеют практическую значимость с точки зрения анализа данных секвенирования ДНК. Разработанный программный конвейер для обработки данных секвенирования ДНК человека с использованием модифицированного индекса превосходит существующие реализации по полноте (Recall) идентификации однонуклеотидных полиморфизмов, что позволяет использовать его для решения практических и научных задач. Кроме того, представленные результаты могут быть полезны в других областях, в частности в задачах анализа естественных языков.

Считаю, что диссертация Карпулевича Е.А. носит законченный характер и вносит значимый вклад в область выравнивания последовательностей в приложениях биоинформатики. Работа отвечает требованиям “Положения о порядке присуждения ученых степеней”, предъявляемых к кандидатским диссертациям по специальности 2.3.5 - “математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей”, а ее автор, Карпулевич Евгений Андреевич, заслуживает присуждения ему ученой степени кандидата физико-математических наук.

Кандидат физико-математических наук,
заведующий отделом
Информационных систем ИСП РАН

Д.Ю. Турдаков

11 сентября 2023 г.