

### **Отзыв научного руководителя**

на выпускную квалификационную работу Клещина Антона Сергеевича,  
обучающегося по направлению 02.03.03 (Математическое обеспечение и  
администрирование информационных систем)

Тема выпускной квалификационной работы:

“Разработка алгоритмов скаффолдинга при помощи дополнительной геномной  
информации”

Задача ВКР Антона Сергеевича была поставлена в Центре алгоритмической биотехнологии СПбГУ, в рамках проекта по разработке геномного ассемблера SPAdes, одного из самых известных геномных ассемблеров, используемых биоинформатиками. Ассемблер строит геном организма по считанным фрагментам ДНК, при считывании могут возникать ошибки на физическом уровне, длина фрагментов очень невелика в сравнении с длиной всего генома и речь идёт об очень больших объёмах данных, поэтому задача сложная, точное её решение, как правило, невозможно, и все алгоритмы в этой области эмпирические по своей природе. Идея, которую надо было проверить Антону Сергеевичу в ходе ВКР, состоит в том, что можно использовать уже априори известные фрагменты генома для сборки генома конкретного организма — например, “референсный” геном (то есть некоторый усреднённый геном вида, к которому относится исследуемый организм) или результат сборки другим ассемблером.

Интуитивно кажется, что добавление новой информации для сборки только повысит её качество, но первые эксперименты показали, что это совсем не так. Антон Сергеевич, после того, как быстро погрузился в новую для себя и довольно сложную область биоинформатики, разобрался в архитектуре SPAdes и модифицировал используемые там алгоритмы сборки, которые сводятся к поиску пути в огромном графе со строковыми метками, наиболее покрытого считанными фрагментами. Ассемблер стал работать заметно хуже, поскольку априорная информация “запутывала” используемые при построении пути эвристики. Антон Сергеевич проанализировал проблемы и предложил несколько разных модификаций алгоритма сборки, которые помогли эти проблемы решить. Не всё получилось сразу, но в итоге при добавлении референсного генома как априорной информации качество сборки существенно улучшилось, вплоть до того, что некоторые геномы были полностью собраны без структурных ошибок. Также существенно улучшилось качество сборки при совместной работе двух ассемблеров. Результаты доведены до внедрения и будут доступны в версии SPAdes 3.15.

В ходе работы Антон Сергеевич проявил себя как очень ответственный и организованный человек, грамотный исследователь, который способен взять задачу и довести её до окончательного решения. И не остановиться на этом, а продолжить

расширение своих результатов (в тексте ВКР описаны успешные эксперименты с метагеномными сборками, то есть сборками нескольких геномов сразу из образцов, содержащих генетическую информацию нескольких организмов). Антон Сергеевич продемонстрировал высокие навыки владения C++, заранее начал обсуждение текста отчёта (проверка которого на предмет наличия неправомерных зависимостей показала, что работа таких заимствований не содержит), текст написан грамотным языком, хорошо структурирован и очень аккуратно сообщает только те детали из биоинформатики, которые необходимы для понимания работы.

Считаю, что работа, безусловно, заслуживает оценки **“отлично”**, также рекомендую обязательно опубликовать результаты работы, и очень надеюсь, что Антон Сергеевич продолжит обучение в магистратуре.

Литвинов Юрий Викторович,  
доцент кафедры системного программирования СПбГУ, к.т.н.

Дата: 04.06.2020 года

Подпись: 