

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

**«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»**

**ПОЯСНИТЕЛЬНАЯ ЗАПИСКА
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ**

**«Реализация эффективного взаимодействия между платформой
для анализа экспрессии генов Morpheus и библиотекой
вычислительных методов R/Bioconductor»**

Автор: Зенкова Дарья Михайловна _____

Направление подготовки (специальность): 01.03.02 Прикладная математика и
информатика

Квалификация: Бакалавр

Руководитель: Сергушичев А.А., канд. техн. наук _____

К защите допустить

Зав. кафедрой Васильев В.Н., докт. техн. наук, проф. _____

« ____ » _____ 20 ____ г.

Санкт-Петербург, 2017 г.

Студент Зенкова Д.М. **Группа** М3436 **Кафедра** компьютерных технологий **Факультет** информационных технологий и программирования

Направленность (профиль), специализация Математические модели и алгоритмы разработки программного обеспечения

Квалификационная работа выполнена с оценкой _____

Дата защиты « _____ » _____ 20 _____ г.

Секретарь ГЭК _____

Листов хранения _____

Демонстрационных материалов/Чертежей хранения _____

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»**

УТВЕРЖДАЮ

Зав. каф. компьютерных технологий
докт. техн. наук, проф.
_____ Васильев В.Н.
« ____ » _____ 20 ____ г.

**ЗАДАНИЕ
НА ВЫПУСКНУЮ КВАЛИФИКАЦИОННУЮ РАБОТУ**

Студент Зенкова Д.М. **Группа** М3436 **Кафедра** компьютерных технологий **Факультет** информационных технологий и программирования
Руководитель Сергушичев Алексей Александрович, канд. техн. наук, программист кафедры информационных систем

1 Наименование темы: Реализация эффективного взаимодействия между платформой для анализа экспрессии генов Morpheus и библиотекой вычислительных методов R/Bioconductor

Направление подготовки (специальность): 01.03.02 Прикладная математика и информатика

Направленность (профиль): Математические модели и алгоритмы разработки программного обеспечения

Квалификация: Бакалавр

2 Срок сдачи студентом законченной работы: «31» мая 2017 г.

3 Техническое задание и исходные данные к работе.

Разработать веб-приложение для анализа экспрессии генов, интегрирующее возможности визуального анализа morpheus.js и методы анализа библиотек R/Bioconductor. Веб-приложение должно быть легко дополняемо новыми методами для исследования и анализа экспрессии генов.

4 Содержание выпускной квалификационной работы (перечень подлежащих разработке вопросов)

- а) Обзор предметной области
- б) Архитектура проекта
- в) Практическая реализация и результаты

5 Перечень графического материала (с указанием обязательного материала)

Не предусмотрено

6 Исходные материалы и пособия

- а) Joshua Gould. Morpheus.js. JavaScript matrix visualization and analysis. [Электронный ресурс]. URL: <https://github.com/cmap/morpheus.js/>;
- б) Arora Sonali, Carlson Marc, Hayden Nate [и др.]. Bioconductor is an open source, open development software project to provide tools for the analysis and comprehension of high-throughput genomic data. [Электронный ресурс]. URL: <https://www.bioconductor.org/>;
- в) Ooms Jeroen. OpenCPU is a system for embedded scientific computing and reproducible research. [Электронный ресурс]. URL: <https://www.opencpu.org/>;
- г) Docker. Docker is the software container platform. [Электронный ресурс]. URL: <https://www.docker.com/>;

7 Календарный план

№№ пп.	Наименование этапов выпускной квалификационной работы	Срок выполнения этапов работы	Отметка о выполнении, подпись руков.
1	Ознакомление с предметной областью	30.09.2016	
2	Изучение исходного кода morpheus.js	31.10.2016	
3	Проектирование метода взаимодействия	30.11.2016	
4	Внедрение и тестирование нового функционала	31.03.2017	
5	Запуск веб-приложения в публичное пользование	28.04.2017	
6	Обработка результатов, написание пояснительной записки	31.05.2017	

8 Дата выдачи задания: «01» сентября 2016 г.

Руководитель _____

Задание принял к исполнению _____ «01» сентября 2016 г.

Министерство образования и науки Российской Федерации
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
**«САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ ИНФОРМАЦИОННЫХ
ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ»**

**АННОТАЦИЯ
ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ**

Студент: Зенкова Дарья Михайловна

Наименование темы работы: Реализация эффективного взаимодействия между платформой для анализа экспрессии генов Morpheus и библиотекой вычислительных методов R/Bioconductor

Наименование организации, где выполнена работа: Университет ИТМО

ХАРАКТЕРИСТИКА ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ

1 Цель исследования: Создать веб-приложение, интегрирующее существующие возможности веб-приложения morpheus.js и методы анализа, реализованные в Bioconductor.

2 Задачи, решаемые в работе:

- а) разработка способа взаимодействия между js-клиентом и R и встраивание его в morpheus.js;
- б) создание графического интерфейса в js-клиенте и серверной реализации в R-пакете;
- в) объединение всех составляющих в единое веб-приложение phantassus;
- г) запуск веб-приложения в открытый доступ для исследователей.

3 Число источников, использованных при составлении обзора: _____

4 Полное число источников, использованных в работе: 0

5 В том числе источников по годам

Отечественных			Иностранных		
Последние 5 лет	От 5 до 10 лет	Более 10 лет	Последние 5 лет	От 5 до 10 лет	Более 10 лет

6 Использование информационных ресурсов Internet: _____

7 Использование современных пакетов компьютерных программ и технологий:

Были использованы следующие программы и технологии: язык программирования JavaScript, фреймворк Node.js, веб-приложение morpheus.js, язык программирования R, библиотека биоинформатических алгоритмов Bioconductor, система интеграции R OpenCPU, механизм для сериализации данных Protocol Buffers, репозитория геномных данных Gene Expression Omnibus, программное обеспечение для запуска приложений в контейнерах Docker, веб-сервер Apache, среда разработки WebStorm, среда разработки RStudio, система контроля версий git, система компьютерной верстки L^AT_EX.

8 Краткая характеристика полученных результатов: Реализовано веб-приложение phantusus, отвечающее всем поставленным требованиям. Веб-приложение было запущено в публичный доступ, используется в лаборатории Максима Артемова в Washington University in St. Louis. Демонстрация приложения входит в программу семинара по системной биологии в Сиднее (10-13 апреля 2017) и в Санкт-Петербурге (14-19 мая 2017).

9 Гранты, полученные при выполнении работы: Работа над данной инженерной разработкой велась без поддержки грантами.

10 Наличие публикаций и выступлений на конференциях по теме работы: По данной инженерной разработке не имеется публикаций и она не была представлена на конференциях.

Выпускник: Зенкова Д.М. _____

Руководитель: Сергушичев А.А. _____

« ____ » _____ 20 ____ г.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
1. Обзор предметной области	6
1.1. Биоинформатика	6
1.1.1. Анализ экспрессии генов	6
1.1.2. Используемые методы	6
1.2. Существующие решения для анализа экспрессии генов	6
1.2.1. R/Bioconductor	6
1.2.2. GENE-E	6
1.2.3. morpheus.js	6
1.3. Инструменты, которые могут быть применены	6
1.3.1. Язык R и библиотека Bioconductor	6
1.3.2. JavaScript	6
1.3.3. R shiny	6
1.3.4. OpenCPU	6
1.3.5. Gene Expression Omnibus	6
1.3.6. Docker	6
1.3.7. JSON	6
1.3.8. Protocol Buffers	6
1.3.9. Apache	6
1.3.10. HTML	6
1.4. Постановка задачи	6
1.4.1. Цель работы	6
1.4.2. Основные задачи	6
1.4.3. Требования к веб-приложению phantasus	6
Выводы по главе 1	6
2. Архитектура проекта phantasus	7
2.1. morpheus.js	7
2.1.1. Чтение данных	7
2.1.2. Класс Dataset	7
2.1.3. Класс SlicedDatasetView	7
2.1.4. Класс HeatMap	7
2.1.5. Реализованные метод	7

2.2. R-пакет phantusus.....	7
2.3. Связь через OpenCPU API	7
2.4. Неотсортировано	7
Выводы по главе 2.....	7
3. Реализация	8
Выводы по главе 3.....	8
3.1. Таблицы	8
3.2. Рисунки.....	8
3.3. Листинги	8
4. Проверка сквозной нумерации	11
Выводы по главе 4.....	11
ЗАКЛЮЧЕНИЕ.....	12
ПРИЛОЖЕНИЕ А. Пример приложения	13
ПРИЛОЖЕНИЕ Б. Еще один пример приложения с невероятно длинным названием для тестирования переносов.....	15

ВВЕДЕНИЕ

В данном разделе размещается введение.

ГЛАВА 1. ОБЗОР ПРЕДМЕТНОЙ ОБЛАСТИ

1.1. Биоинформатика

1.1.1. Анализ экспрессии генов

1.1.2. Используемые методы

1.2. Существующие решения для анализа экспрессии генов

1.2.1. R/Bioconductor

1.2.2. GENE-E

1.2.3. morpheus.js

1.3. Инструменты, которые могут быть применены

1.3.1. Язык R и библиотека Bioconductor

1.3.2. JavaScript

1.3.3. R shiny

1.3.4. OpenCPU

1.3.5. Gene Expression Omnibus

1.3.6. Docker

1.3.7. JSON

1.3.8. Protocol Buffers

1.3.9. Apache

1.3.10. HTML

1.4. Постановка задачи

1.4.1. Цель работы

1.4.2. Основные задачи

1.4.3. Требования к веб-приложению phantassus

Выводы по главе 1

ГЛАВА 2. АРХИТЕКТУРА ПРОЕКТА PHANTASUS

2.1. morpheus.js

2.1.1. Чтение данных

2.1.2. Класс Dataset

2.1.3. Класс SlicedDatasetView

2.1.4. Класс HeatMap

2.1.5. Реализованные метод

2.2. R-пакет phantassus

2.3. Связь через OpenCPU API

2.4. Неотсортировано

Выводы по главе 2

ГЛАВА 3. РЕАЛИЗАЦИЯ

Выводы по главе 3

Пример ссылок на литературные источники: [1–3].

3.1. Таблицы

В качестве примера таблицы приведена таблица 1.

Таблица 1 – Таблица умножения (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

Есть еще такое окружение `tabu`, его можно аккуратно растянуть на всю страницу. Приведем пример (таблица 2).

Таблица 2 – Таблица умножения с помощью `tabu` (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

3.2. Рисунки

Пример рисунка (с помощью `TikZ`) приведен на рисунке 1. Под `pdflatex` можно также использовать `*.jpg`, `*.png` и даже `*.pdf`, под `latex` можно использовать `Metapost`. Последний можно использовать и под `pdflatex`, для чего в стилевике продекларированы номера картинок от 1 до 20.

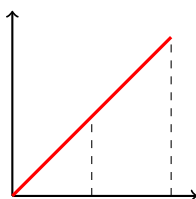


Рисунок 1 – Пример рисунка

3.3. Листинги

В работах студентов кафедры «Компьютерные технологии» часто встречаются листинги. Листинги бывают двух основных видов — исходный код и псевдокод. Первый оформляется с помощью окружения `lstlisting` из пакета `listings`, который уже включается в стилевике и немного настроен. Пример Hello World на Java приведен на листинге 1.

Листинг 1 – Пример исходного кода на Java

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world!");
    }
}
```

Псевдокод можно оформлять с помощью разных пакетов. В данном стилевике включается пакет `algorithmicx`. Сам по себе он не генерирует флоатов, поэтому для них используется пакет `algorithm`. Пример их совместного использования приведен на листинге 2. Обратите внимание, что флоаты разные, а нумерация — общая!

Листинг 2 – Пример псевдокода

```
function IsPrime( $N$ )
  for  $t \leftarrow [2; \lfloor \sqrt{N} \rfloor]$  do
    if  $N \bmod t = 0$  then
      return false
    end if
  end for
  return true
end function
```

Наконец, листинги из `listings` тоже можно подвешивать с помощью `algorithm`, пример на листинге 3.

Листинг 3 – Исходный код и флоат `algorithm`

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world!");
    }
}
```

ГЛАВА 4. ПРОВЕРКА СКВОЗНОЙ НУМЕРАЦИИ

Листинг 4 должен иметь номер 4.

Листинг 4 – Исходный код и флюид algorithm

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world!");
    }
}
```

Рисунок 2 должен иметь номер 2.

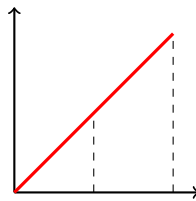


Рисунок 2 – Пример рисунка

Таблица 3 должна иметь номер 3.

Таблица 3 – Таблица умножения с помощью tabu (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

Выводы по главе 4

В конце каждой главы желательно делать выводы. Вывод по данной главе — нумерация работает корректно, ура!

ЗАКЛЮЧЕНИЕ

В данном разделе размещается заключение.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 *Buzdalov M., Shalyto A.* Hard Test Generation for Augmenting Path Maximum Flow Algorithms using Genetic Algorithms: Revisited // Proceedings of IEEE Congress on Evolutionary Computation. — 2015. — P. 2121–2128.
- 2 *Буздалов М. В.* Генерация тестов для олимпиадных задач по программированию с использованием генетических алгоритмов // Научно-технический вестник СПбГУ ИТМО. — 2011. — 2(72). — С. 72–77.
- 3 *Buzdalov M., Doerr B., Kever M.* The Unrestricted Black-Box Complexity of Jump Functions // Evolutionary Computation. — 2016. — Accepted for publication.

ПРИЛОЖЕНИЕ А. ПРИМЕР ПРИЛОЖЕНИЯ

В приложениях рисунки, таблицы и другие подобные элементы нумеруются по приложениям с соответствующим префиксом. Проверим это.

Листинг А.1 должен иметь номер А.1.

Листинг А.1 – Исходный код и флотат algorithm

```
public class HelloWorld {
    public static void main(String[] args) {
        System.out.println("Hello, world!");
    }
}
```

Рисунок А.1 должен иметь номер А.1.

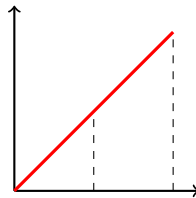


Рисунок А.1 – Пример рисунка

Таблица А.1 должна иметь номер А.1.

Таблица А.1 – Таблица умножения с помощью tabu (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68

Заодно проверим нумерованные и нумерованные перечисления. Нумерованные:

- пункт А;
- пункт Б;
- пункт В.

Нумерованные списки нескольких уровней:

- а) первый элемент;
- б) второй элемент с подэлементами:

- 1) первый подэлемент;
 - 2) второй подэлемент;
 - 3) третий подэлемент.
- в) третий элемент;
 - г) четвертый элемент;
 - д) пятый элемент;
 - е) шестой элемент;
 - ж) седьмой элемент;
 - и) восьмой элемент;
 - к) девятый элемент;
 - л) десятый элемент.

**ПРИЛОЖЕНИЕ Б. ЕЩЕ ОДИН ПРИМЕР ПРИЛОЖЕНИЯ С
НЕИМОВЕРНО ДЛИННЮЩИМ НАЗВАНИЕМ ДЛЯ ТЕСТИРОВАНИЯ
ПЕРЕНОСОВ**

Проверим на примере таблиц, что нумерация в приложениях — по приложениям. Таблица Б.1 должна иметь номер Б.1.

Таблица Б.1 – Таблица умножения с помощью `tabu` (фрагмент)

–	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
2	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
3	3	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51
4	4	8	12	16	20	24	28	32	36	40	44	48	52	56	60	64	68