Приложение № 3А

к регламенту подготовки

образовательных программ

АНО ВО «Университет «Сириус»

Автономная некоммерческая образовательная организация высшего образования

Научно-технологический университет «Сириус»

(АНО ВО Университет «Сириус»)

*Центр генетики и наук о жизни*

Программа утверждена на заседании

Центра генетики и наук о жизни

АНО ВО Университет «Сириус»

Протокол №\_\_\_ от 00 \_\_\_\_\_\_\_\_\_\_\_ 202\_\_ г.
*(заполняется руководителем научного центра*)

**РАБОЧАЯ ПРОГРАММА ОБРАЗОВАТЕЛЬНОГО МОДУЛЯ (ДИСЦИПЛИНЫ)**

***Основы программирования для биологов***

*наименование образовательного модуля (дисциплины)*

**Уровень высшего образования:**

***Магистратура***

*(бакалавриат, специалитет, магистратура, аспирантура)*

**Направление подготовки (специальность):**

***Биология 06.04.01***

*(код и название направления/специальности)*

Форма обучения:

*очная*

*(очная / очно-заочная / заочная)*

**1. Общая характеристика** **образовательного модуля (дисциплины)**

1.1. Аннотация образовательного модуля (дисциплины)

Разработанная программа по новой самостоятельной дисциплине «***Основы программирования для биологов***» предназначена для подготовки магистров-биологов. Ее особенность состоит в изложении современных подходов к прикладному программированию для решения задач в области биоинформатики и биологии. Данная программа нацелена на формирование базовых навыков прикладного программирования на языках Python и R.

Программа включает 4 раздела: работа с языком Python, применение прикладных библиотек для анализа данных, применение прикладных библиотек для визуализации результатов, применение языка R для обработки и визуализации данных. Фокусом первого раздела является знакомство студента с основами применения языка Python. Особое внимание уделяется организации воспроизводимых конвейеров по обработке данных. Второй раздел включает в себя знакомство с прикладными библиотеками Numpy и Pandas. В разделе приведены примеры обработки данных на различных примерах. Третий раздел посвящен прикладным библиотекам для построения графиков и визуализации результатов. В разделе приведены подходы к построению диаграмм и графиков по результатам анализа данных. Четвертый раздел посвящен изучению основ языка R. В разделе обсуждается применение R-studio для обработки и визуализации данных.

1.2. Место образовательного модуля (дисциплины) в структуре основных профессиональных образовательных программ высшего образования (ОПОП ВО):

уровень образования: Магистратура;

направление подготовки: Биология;

наименование ОПОП ВО: Генетика и генетические технологии;

часть ОП: Базовая;

дисциплина по выбору/обязательная; дисциплина общая /межпрограммная; НИР; практика; другое: Обязательная*;*

форма реализации (очная, дистанционная): Очная.

1.3. Описание взаимосвязи образовательного модуля (дисциплины) с другими, реализующимися в АНО ВО Университет «Сириус» образовательными модулями, курсами, образовательными программами, научными исследованиями и проектами:

1.4. Цель образовательного модуля (дисциплины): В результате освоения дисциплины обучающийся должен освоить принципы написания программ на языке Python, уметь самостоятельно реализовывать алгоритмы для решения прикладных задач анализа данных при помощи языка программирования Python; иметь представления о языке программирования R и его применения для решения прикладных задач в области генетики.

1.5. Задачи образовательного модуля (дисциплины): Получение базовых теоретических знаний и освоение методов решения задач программирования на языке Python и языке R. Умение использовать полученные базовые знания.

1.6. Язык преподавания: русский.

**2. Кадровое обеспечение**

2.1. Авторы программы:

к.ф.-м.н., с.н.с. кафдеры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова Армеев Григорий Алексеевич.

к.ф.-м.н., в.н.с. кафдеры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова Шайтан Алексей Константинович.

2.2. Преподаватели:

к.ф.-м.н., с.н.с. кафдеры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова Армеев Григорий Алексеевич.

к.ф.-м.н., в.н.с. кафдеры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова Шайтан Алексей Константинович.

аспирант кафдеры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова Грибкова Анна Кирилловна.

аспирант кафдеры биоинженерии Биологического факультета МГУ имени М.В. Ломоносова Сингх-Пальчевская Лавприт Раджитовна.

**4. Структура образовательного модуля (дисциплины)**

4.1. Объем образовательного модуля (дисциплины) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся:

*объем дисциплины (модуля) составляет 72 ак. часов, из которых 54 ак. часа составляет контактная работа студента с преподавателем (18 ак. часов занятий лекционного типа, 36 ак. часов семинарско-практических занятий), 10 ак. часов составляет текущий контроль успеваемости (экзамен), 8 ак. часов составляет самостоятельная работа студента.*

4.2. Учебный план (структура образовательного модуля (дисциплины):

|  |  |  |
| --- | --- | --- |
| **Наименование разделов и тем образовательного модуля (дисциплины),** **форма промежуточной аттестации по образовательному модулю (дисциплине)** | **Всего****(ак. часы**) | В том числе |
| **Контактная работа (работа во взаимодействии с преподавателем), ак. часы**из них | **Самостоятельная работа обучающегося, ак. часы** из них |
| Лекции | Лабораторные занятия | Практические занятия | Семинары | Учебные занятия, направленные на проведение текущего контроля успеваемости, промежуточной аттестации | **Всего**  | Выполнение домашних заданий | Подготовка рефератов и т.п. | **Всего** |
| Основы работы с языком Python, организация работы | 14 | 4 |  | 4 | 4 |  | 12 | 2 |  | 2 |
| Применение библиотек Numpy и Pandas для решения прикладных задач | 26 | 8 |  | 8 | 8 |  | 24 | 2 |  | 2 |
| Библиотеки для построения графиков и визуализации результатов | 18 | 5 |  | 6 | 5 |  | 16 | 2 |  | 2 |
| Основы обработки и визуализации данных на языке R. | 4 | 1 |  |  | 1 |  | 2 | 2 |  | 2 |
| Промежуточная аттестация *Экзамен*  | 10 |  |  |  |  | 10 | 10 |  |  |  |
| **Итого** | 72 | 18 |  | 18 | 18 |  | 64 | 8 |  | 8 |

4.3. Содержание образовательного модуля (дисциплины), структурированное по темам (разделам):

|  |  |  |  |
| --- | --- | --- | --- |
| № п/п | Наименование раздела/темы образовательного модуля (дисциплины) | Вид и содержание занятия | ФИО преподавателя |
|  | Основы работы с языком Python, организация работы. | Содержание темы №1. Запуск программ на языке Python. Знакомство с Anaconda и Jupyter. Синтаксис языка Python, переменные, операторы и типы данных.Содержание темы №2. Структуры в языке Python, определение функций и обработка исключений. Итераторы и ветвление.Содержание темы №3. Процедурное и объектно-ориентированное программирование. Классы, генераторы и регулярные выражения. Модули и пакеты расширения. | Армеев Григорий Алексеевич, Шайтан Алексей Константинович |
|  | Применение библиотек Numpy и Pandas для решения прикладных задач | Содержание темы №4 Отличие структур и типов данных Numpy от встроенных в язык Python. Введение в массивы библиотеки Numpy. Импорт и экспорт данных в массивах Numpy. Слияние и разбиение массивов, изменение размерностей массивов.Содержание темы №5. Выполнение вычислений над массивами, преимущества перед циклами. Операции над массивами, транслирование, сравнения, маски.Содержание темы №6. Продвинутые способы индексации массивов, сортировка массивов и частичная сортировка. Структурирование массивов и неоднородные по типу данных массивы.Содержание темы №7. Применение пакета Pandas для расширения функционала Numpy. Серии, индексы и таблицы Pandas. Особенности индексации в библиотеке Pandas. Операции над данными: объединение наборов данных, математические операции.Содержание темы №8. Агрегирование данных, сводные таблицы, работа с временными рядами. Оптимизация программ.Содержание темы №9. Примеры обработки данных: линейная регрессия, метод главных компонент, смеси Гауссовых распределений, применение статистических критериев.. | Армеев Григорий Алексеевич,Сингх-Пальчевская Лавприт Раджитовна,Грибкова Анна Кирилловна |
|  | Библиотеки для построения графиков и визуализации результатов | Содержание темы 10. Применение библиотеки Matplotlib. Интерфейсы использования: процедурный и ООП интерфейс. Области построения графиков, стили и сохранение рисунков в файл.Содержание темы 11. Построение графиков, диаграмм рассеяния. Визуализация погрешностей. Гистограммы. Построение двумерных гистограмм. Настройка осейСодержание темы 12. Настройка легенд, шкал цветов и задание стилей. Построение множества графиков. Аннотирование графиков. Создание интерактивных графиков | Армеев Григорий Алексеевич,Сингх-Пальчевская Лавприт Раджитовна,Грибкова Анна Кирилловна |
|  | Основы обработки и визуализации данных на языке R. | Содержание темы 13. Установка и запуск R, Rstudio, статистические распределения и тесты, написание функций, установка библиотек, Bioconductor, визуализация данных ggplot2. | Шайтан Алексей Константинович,Сингх-Пальчевская Лавприт Раджитовна,Грибкова Анна Кирилловна |

**5. Входные требования и предварительные условия для освоения образовательного модуля (дисциплины)**

Для изучения образовательного модуля (дисциплины) обучающийся должен:

1. освоить дисциплины базовой части образовательной программы специалитета или бакалавриата обучения.
2. обладать следующими компетенциями: освоить дисциплины математического профиля - «Математическая статистика», «Математические методы в биологии» и уметь применять полученные знания; освоить дисциплины биологического профиля - «Генетика», Молекулярная биология», «Биохимия» и уметь применять полученные знания.

**6. Планируемые результаты освоения образовательного модуля (дисциплины), компетенции выпускников**

* владение современными методами молекулярной биологии, биохимии, биоорганической химии и способность их комбинировать для решения задач структурной биологии;
* обладание знаниями о современных методах программирования на языке Python;
* способность использовать современные компьютерные средства и специализированное программное обеспечения
* способность использовать современные методы прототипирования алгоритмов и создания программ на языке Python;
* способность к квалифицированному анализу научных источников;
* способность к критическому анализу, поиску, обобщению и систематизации научной информации в области структурной биологии;
* возможность квалифицированного участия в дискуссиях и обсуждениях научных проблем по структурной биологии;
* способность к дальнейшему самостоятельному повышению уровня образованности в области междисциплинарных и специальных исследований различных аспектов структурной биологии и биотехнологии.

**7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по образовательному модулю (дисциплине)**

* Критерии и процедуры оценивания результатов обучения по образовательному модулю (дисциплине):

|  |  |  |
| --- | --- | --- |
| **РЕЗУЛЬТАТ ОБУЧЕНИЯ****по образовательному модулю (дисциплине)** | **КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТА ОБУЧЕНИЯ****по образовательному модулю (дисциплине) и ШКАЛА оценивания** | **ПРОЦЕДУРЫ ОЦЕНИВАНИЯ** |
| 2Или «неудовлетворительно» | 3Или «удовлетворительно» | 4Или «хорошо» | 5Или «отлично» |
|  | Отсутствие знаний | Общие, но не структурированные знания  | Сформированные, но содержащие отдельные пробелы специальных разделов  | Сформированные систематические знания специальных разделов  | Экзамен  |

* Типовые контрольные задания и/или иные материалы, необходимые для оценки результатов обучения:

**Примерный список заданий для проведения текущей и промежуточной аттестации (темы для докладов, рефератов, презентаций и др.)**

1. Опишите преимущества и недостатки функционального программирования.
2. Опишите преимущества и недостатки объектно-ориентированного программирования.
3. Опишите особенности структур в языке Python, особенности работы с памятью.
4. Обозначьте область применения метода главных компонент, приведите примеры.
5. Опишите способы хранения данных json, csv, hdf5 их назначение, преимущества и недостатки.
6. Напишите программу для обработки статистически данных из открытых источников (ФСГС, открытые данные) и построения графиков.

**Примерный список вопросов для проведения текущей и промежуточной аттестации.**

1. Перечислите типы данных языке Python, опишите структуры в языке Python.
2. Что такое области переменных? Как осуществляется передача значений в/из функций?
3. Что такое объекты, их атрибуты и методы?
4. Какие существуют библиотеки для обработки табулированных данных в Python?
5. Какие существуют библиотеки для визуализации табулированных данных в Python?
6. Каким образом осуществляется индексация и сортировка массивов Numpy?
7. В чем отличия между библиотекой Pandas и библиотекой Numpy?
8. Опишите метод линейной регрессии и область его применения.
9. Каким образом можно ускорить работу программ на языке Python?
10. Какую информацию можно узнать из одномерных и двумерных гистограмм?

**8. Ресурсное обеспечение**

**А. Основная литература.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Марк Лутц | Изучаем Python |  | Москва | Символ-Плюс | 2010 |  |  |  |
|  | Дж. Вандер Плас | Python для сложных задач. Наука о данных и машинное обучение |  | Санкт-Петербург | Издательский дом «Питер» | 2018 |  |  |  |

**Б. Дополнительная литература.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Джоэл Грас | Data Science. Наука о данных с нуля |  | Санкт-Петербург | БХВ-Петербург | 2018 |  |  |  |
|  | Уэс Маккинни | Python и анализ данных |  | Москва | ДМК Пресс | 2015 |  |  |  |

**В.Программное обеспечение и Интернет-ресурсы.**

- Программа менеджер пакетов Python Conda <https://conda.io/en/latest/>

- Справочная информация по основам языка Python <https://github.com/jakevdp/whirlwindtourofpython>

8.4. Описание материально-технической базы:

8.4.1. Аудиторный фонд и оборудование:

|  |  |  |  |
| --- | --- | --- | --- |
| № п.п. | Вид аудитории | Время использования | Оборудование (кол-во) |
| 1. | Лекционная  | 2 час в неделю | Проектор и ноутбук |
| 2. | Семинарская | 2 час в неделю | Ноутбуки по числу студентов |

8.4.2. Оборудование для практических занятий:

|  |  |  |  |
| --- | --- | --- | --- |
| № п.п. | Тип оборудования | Характеристика | Кол-во  |
| 1. | Ноутбуки по числу студентов |  | 20 |