

Министерство науки и высшего образования Российской Федерации
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Институт биологии, экологии, почвоведения, сельского и лесного хозяйства
(БИОЛОГИЧЕСКИЙ ИНСТИТУТ)

УТВЕРЖДАЮ:

Директор Биологического института

Д.С. Воробьев

« 22 » 03 20 22 г.

Рабочая программа дисциплины

Компьютерный анализ биологических данных

по направлению подготовки

06.04.01 Биология

Направленность (профиль) подготовки:

«Биологическое образование (преподавание биологических дисциплин в учреждениях общего и профессионального образования)»

Форма обучения

Очная

Квалификация

Магистр

Год приема

2022

Код дисциплины в учебном плане: Б1.В.ДВ.05.01

СОГЛАСОВАНО:

Руководитель ОП

А.С. Ревушкин

Председатель УМК

А.Л. Борисенко

1. Цель и планируемые результаты освоения дисциплины (модуля)

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-6 – Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок;

ОПК-8 – Способен использовать современную исследовательскую аппаратуру и вычислительную технику для решения инновационных задач в профессиональной деятельности;

ПК-2 – Способен представлять результаты научных исследований и методических разработок.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-6.1 – Описывает разнообразие, пути и перспективы применения компьютерных технологий в современной биологии;

ИОПК-6.2 – Использует компьютерные технологии и профессиональные базы данных при планировании профессиональной деятельности, обосновывает их выбор;

ИОПК-6.3 – Профессионально оформляет и представляет результаты новых разработок;

ИОПК-8.2 – Применяет современную исследовательскую аппаратуру и вычислительную технику при решении стандартных и инновационных задач в профессиональной деятельности;

ИПК-2.1 – Описывает, обобщает и делает выводы на основе результатов исследования, в том числе с помощью современных компьютерных технологий.

2. Задачи освоения дисциплины

– Освоение аппарата концепций, понятий, терминов, определяющих современное состояние и направления развития информационных технологий в области биологии.

– Умение применять понятийный аппарат современных методов и практических приемов обработки естественнонаучных данных и представления результатов с использованием компьютерных технологий для решения практических задач профессиональной деятельности.

– Владение теоретическими знаниями и практическими приемами компьютерной обработки данных с учетом особенностей биологических объектов; освоение методик массовой обработки полевого зоологического, флористического, геоботанического, экологического материала с использованием интегрированных информационных систем на основе реляционных баз данных; изучение методов специальной статистической обработки биологической информации.

– Получение систематизированных знаний о спектре методов и приемов обработки естественнонаучных данных и представления результатов с использованием компьютерных технологий. Многообразные методы, реализованные в программном обеспечении, позволяют значительно повысить эффективность научных исследований в различных разделах биологической науки – фаунистике, флористике, систематике, фитоценологии, экологии растений и животных.

– Приобретение навыков самостоятельной исследовательской деятельности, связанной с выполнением законченного естественнонаучного исследования: от стадии планирования эксперимента, через получение, обработку и анализ данных, до этапа интерпретации и презентации результатов.

3. Место дисциплины (модуля) в структуре образовательной программы

Дисциплина относится к части, формируемой участниками образовательных отношений.

4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине

Семестр 3, зачет.

5. Входные требования для освоения дисциплины

Для успешного освоения дисциплины необходимо иметь достаточные знания в области математики, информатики и математических методов в биологии на уровне, соответствующем программам бакалавриата. Для реализации предметно-ориентированной части дисциплины необходимым условием является успешное прохождение курсов биологического цикла.

6. Язык реализации

Русский

7. Объем дисциплины (модуля)

Общая трудоемкость дисциплины составляет 3 з.е., 108 часов, из которых:

– лекции: 8 ч.;

– практические занятия: 22 ч.;

Объем самостоятельной работы студента определен учебным планом.

8. Содержание дисциплины (модуля), структурированное по темам

Тема 1. *Введение в предмет. Периоды становления биометрии до наших дней. Ключевые персоналии в истории биометрии. Современные программные средства для обработки статистических данных. Сбор данных в различных областях биологической науки и последующая статистическая обработка. Основные методы статистического анализа биологических данных.*

1. Ключевые моменты истории информатизации наук биологического цикла.
2. Краткая история зарубежной и отечественной биометрии.
3. Ключевые персоналии в истории развития биометрии.
4. Современные компьютерные программные средства для обработки статистических данных.
5. Основные типы данных, применяемые для регистрации характеристик растительного покрова, флор, фаун.
6. Специфика сбора данных в различных областях биологической науки и последующая статистическая обработка.
7. Основные методы статистического анализа биологических данных.
8. Deskриптивные статистики: основные меры уровня, изменчивости и репрезентативности признаков, используемых для описания биологических объектов.
9. Параметрические и непараметрические критерии различия, применяемые при статистических сравнениях.
10. Ограничения, накладываемые параметрическими критериями на характер и свойства анализируемых выборочных совокупностей.
11. Дисперсионный анализ. Схемы однофакторного и двухфакторного дисперсионного анализа.

12. Изучение взаимной связи признаков. Корреляционный анализ: линейная, нелинейная корреляция. Коэффициенты ранговой корреляции. Область их применения, характер данных.
13. Изучение зависимости изменения одного признака от изменений другого. Регрессионный анализ. Модели регрессии.

Тема 2. *Введение в базы данных и информационные системы биологических данных. Интегрированная ботаническая информационная система IBIS. Биологические коллекции. Количественная оценка биоразнообразия.*

1. Базы данных и типы моделей, используемые при построении баз данных.
2. Типы атрибутов, используемых при описании биологических объектов, принципы их формализации. Ограничения.
3. Шкалы для регистрации количественного участия популяций таксонов при описании биологических сообществ.
4. Базы данных, системы управления базами данных, информационные системы, банки данных: общие свойства и отличия.
5. Принципы создания и ведения компьютерных таксономических библиотек.
6. Типы пользовательского интерфейса биологических информационных систем.
7. Общая характеристика программ специализированной биологической и экологической статистики.
8. Общая характеристика универсальных информационных систем для хранения и обработки биологических данных.
9. Возможности тандема программ TurboVeg и Juice для обработки геоботанических и экологических данных.
10. Принципы обмена данными о растительном покрове между ботаническими информационными системами.
11. Краткая функциональная характеристика интегрированной ботанической системы IBIS.
12. Пути решения проблемы информационной неопределенности и неполноты в биологических данных.
13. В чем выражается интегрированность биологических информационных систем.
14. Таксономическая идентификация и решение вопросов таксономической сопоставимости. Таксономическая синхронизация данных.
15. В чем проявляется комплексный характер ботанических, зоологических и экологических данных.
16. Типы и структуры данных в системе IBIS.
17. Иерархия данных и принципы рациональной группировки описаний растительности и зообиоты.
18. Атрибуты, используемые для описания растительности в системе IBIS
19. Типы шкал обилия и проективного покрытия, их преобразование.
20. Принципы управления эталонной библиотекой таксонов и библиотекой синонимов
21. Для чего необходимы дополнительные цифровые поля в дескрипторах описаний.
22. Возможные стратегии поиска и селекции описаний по параметрам дескриптора и списка таксонов.
23. Вариант преобразования первичных данных: слияние списков таксонов. Атрибуты сводных описаний.
24. Вариант преобразования первичных данных: создание сводных таблиц. Особенности табличного представления данных.
25. Базовые приемы работы со сводными таблицами.
26. Возможности системы IBIS по обмену первичными и вторичными данными с другими программами.

27. Основные форматы для представления биологических данных и их конвертирование.
28. Особенности функционирования системы IBIS в режиме коллективного пользования.
29. Реализация защиты данных в системе IBIS.
30. Концепция фактор-множеств применительно к флористическим данным. Что такое классы эквивалентности. Типы фактор-множеств (спектров).
31. Учет весовых характеристик при преобразовании первичных описаний биоты в спектры.
32. Количественные показатели биоразнообразия.
33. Информационные меры оценки таксономической сложности биологических коллекций.

Тема 3. *Ординация биологических данных. Прямая ординация данных. Опосредованная (биоиндикационная) ординация. Элементы факторного анализа. Непрямая ординация данных: СА, ССА, DCA.*

1. Ординация списков таксонов в таблицах.
2. Прямая ординация – одно- и двумерная сортировка объектов и признаков на основе прямых измерений абиотических факторов и параметров биоты.
3. Опосредованная ординация сообществ. Базовые постулаты биоиндикации и закономерности влияния экологических факторов.
4. Классификация экологических факторов и их взаимодействие.
5. Понятие об объектах и субъектах биоиндикации. Экологические амплитуды и концепция эврибионтности–стенобионтности.
6. Принципы и преимущества биоиндикации режимов абиотических факторов по сравнению с прямыми измерениями.
7. Численные характеристики растительного покрова, используемые при биоиндикации.
8. Свойства биоиндикационных (экологических) шкал как представителей класса градиентных шкал.
9. Характеристика трех основных типов биоиндикационных шкал.
10. Развитие метода стандартных экологических шкал Л.Г. Раменского и его применение в отечественной экологии.
11. Характеристика наиболее востребованных в России и на Западе фитоиндикационных шкал: набор факторов, тип шкалы и количество градаций, модельная территория.
12. Зачем нужны классификации и ординационные схемы на основе фитоиндикационных шкал.
13. Способы вычисления фитоиндикационных статусов, реализованные в программе IBIS.
14. Понятие об экологическом консенсусе, способы вычисления и примеры использования этого показателя.
15. Различие аутэкологического и синэкологического статусов. Использование для опосредованной ординации синэкологических характеристик.
16. Непрямая ординация сообществ. Базовые постулаты факторного анализа.
17. Анализ главных компонент (РСА).
18. Анализ соответствий и (СА). Канонический анализ соответствия (ССА)
19. Бестрендовый анализ соответствия (DCA). Выполнение анализа DCA в программе PAST
20. Построение и аннотация ордограмм средствами MS Excel и StatSoft Statistica

Тема 4. *Классификация биологических данных. Неиерархический кластерный анализ K-means. Этапы иерархического агломеративного кластерного анализа. Матрицы включения. Матрицы бинарного пересечения. Классификация индексов подобия. Бинарные индексы сходства-различия. Количественные индексы сходства-различия. Трансформация вторичных матриц. Стратегии объединения в кластерном анализе. Классификация на основе вторичных данных. Фактор-множества. Редкоиспользуемые показатели сходства.*

1. Основные параметры флор и фаун, вычисляемые при классическом таксономическом анализе.
2. Информационные индексы сложности флор (комплексности таксономической структуры).
3. Показатели оригинальности (автономности) флор и фаун – индексы автохтонности–аллохтонности.
4. Матрицы мер включения. Количественное и бинарное включение.
5. Способы симметризации мер включения таксономических списков друг в друга.
6. Основные типы бинарных индексов общности, их различия и требования к организации данных.
7. Индексы сходства, учитывающие взаимоотношению таксонов (негативные совпадения). Достоинства и недостатки таких индексов.
8. Использование весовых характеристик таксонов при расчете индексов общности. Трансформация весов таксонов.
9. Что дает учет относительной информативности таксонов при анализе биологических данных.
10. Трансформация вторичных матриц.
11. Общие и отличительные черты таксономических, типологических и градиентных спектры.
12. Критерии адекватности классификаций и типологий.
13. Использование методов ранговой корреляции в сравнительной флористике и фаунистике.
14. Преимущества взвешивания классов эквивалентности обилием и встречаемостью таксонов при получении таксономических и типологических спектров.
15. Алгоритм иерархического агломеративного кластерного анализа. Стратегии объединения в кластерном анализе.
16. Дивизивный алгоритм TWINSPAN.
17. Отличия методов неиерархического кластерного анализа.
18. Сущность метода максимального корреляционного пути Выханду.
19. Классификация на основе вторичных данных.
20. Перспективы флорогенетического анализа с использованием аппарата логического моделирования филогенеза.
21. Основные формы визуализации результатов сравнительного анализа флор и фаун.
22. Анализ сопряженностей в распределении флористических и фаунистических данных.
23. Редкоиспользуемые показатели сходства. Индексы таксономического сходства Е.С. Смирнова.
24. Экологические расстояния и их перспективность для классификации биологических данных.

9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, опросов по лекционному материалу, проверке выполнения домашних тестовых заданий, рефератам студентов (по необходимости) и фиксируется в форме контрольной точки не менее одного раза в семестр.

Успешность выполнения практических занятий в компьютерном классе по обработке наборов данных (учебные примеры) контролируется магистрантами самостоятельно на основе проверки результатов своих действий в программном обеспечении на индивидуальных рабочих местах по параллельной демонстрации правильных действий преподавателем на мультимедийном экране. При необходимости выхода из затруднительных ситуаций и исправления ошибок, корректирующие действия выполняются в интерактивном режиме.

10. Порядок проведения и критерии оценивания промежуточной аттестации

Зачет в третьем семестре проводится в виде индивидуального теста.

Тест предназначен для демонстрации основных практических навыков по работе в освоённом программном обеспечении для ведения баз биологических данных и их обработки и состоит из четырех частей.

Каждая часть при необходимости уточнить качество знаний и уровень владения методами анализа сопровождается дополнительными вопросами теоретического характера, ответы на которые требуют освоения лекционного материала и самостоятельной работы с учебными пособиями в течении семестра.

Время на выполнение теста 60–70 минут. Результаты зачета определяются оценками «зачтено»/«не зачтено». Зачет проводится на зачетной неделе по расписанию.

Итоговая оценка промежуточной аттестации «зачтено» выставляется по совокупности оценки «зачтено» на тесте и результатов посещаемости: при необходимости каждые два пропущенных занятия должны компенсироваться написанием оформленного по ГОСТУ реферата по пропущенному материалу.

Пример типового тестового задания:

Часть 1.

1. Рассказать о различиях шкал обилия/ проективного покрытия для регистрации количественного участия популяций таксонов при описании растительного растительных сообществ. Объяснить принципы количественной аппроксимации баллов шкал обилия-покрытия процентами проективного покрытия.
2. Дано: заполненный полевой бланк тестового растительного сообщества: Березово-сосновая болотнотравно- кочкарноосоковая согра. В информационной системе IBIS создать новый том описаний, новую группу и ввести геоботаническое описание.
3. Построить семейственно-видовой таксономический спектр тестового геоботанического описания.

Часть 2.

4. Рассказать о различиях усредняющих и структурных методов определения расчетного фитоиндикационного статуса растительного сообщества.
5. Рассчитать средние фитоиндикационные статусы тестового описания растительности методом табличных ограничений и двойного взвешивания по всем зарегистрированным в системе экологическим шкалам и создать экологический паспорт местообитания.
6. Построить экологический профиль тестового описания растительности по фактору увлажнения.

Часть 3.

7. Рассказать об использовании таксономических и типологических фактор-множеств (спектров) при сравнительном анализе флор и фаун различного уровня.
8. Импортировать тестовый набор флористических данных (23 ценофлоры ООПТ "Береговой склон р. Томь") из формата табличного процессора MS Excel в формат интегрированной ботанической информационной системы IBIS.
9. В модуле сравнительного анализа рассчитать матрицу сходства ценофлор (бинарный индекс Czekanowski/Dice/Sorensen). Экспортировать матрицу в программу StatSoft

Statistica и построить дендрограмму сходства ценофлор, используя метод связывания Ward's.

Часть 4.

10. Импортировать в систему IBIS из внешнего формата (MS Excel) таблицу геоботанических описаний (35 описаний). Полученные после импорта родительские описания преобразовать в сводную таблицу.
11. Установить процентную шкалу для отображения веса таксонов. Экспортировать полученную таблицу описаний из системы IBIS в формат CSV. Импортировать ее в программу PAST и провести анализ DCA со стандартными параметрами с построением ордограммы.

11. Учебно-методическое обеспечение

- а) Электронный учебный курс по дисциплине в электронном университете «Moodle» – <https://moodle.tsu.ru/course/view.php?id=19024>
- б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине.
- в) Материалы для практических занятий по дисциплине: учебные наборы данных
- г) Методические указания по организации самостоятельной работы студентов.

12. Перечень учебной литературы и ресурсов сети Интернет

а) основная литература:

- Гашев С.Н., Бетляева Ф.Х., Лупинос М.Ю.* Математические методы в биологии: анализ биологических данных в системе Statistica: Учебно-методическое пособие. Тюмень: Изд-во ТюмГУ, 2014. 208 с.
- Голицына О.Л., Максимов Н.В., Попов И.И.* Информационные системы: учебное пособие для вузов. М.: Форум, 2014. 444 с.
- Ефимов В.М., Ковалева В.Ю.* Многомерный анализ биологических данных: Учебное пособие. Изд. 2-е исп. СПб, 2008 87 с.
- Жукова Л.А., Дорогова Ю.А. Турмухаметова Н.В. и др.* Экологические шкалы и методы анализа экологического разнообразия растений. Йошкар-Ола, 2010. 350 с.
- Песенко Ю.А.* Принципы и методы количественного анализа в фаунистических исследованиях. М.: Наука, 1982. 287 с.

б) дополнительная литература:

- Андреев В.Л.* Классификационные построения в экологии и систематике. М., 1980. 142
- Василевич В.И.* Статистические методы в геоботанике. Л.: Наука, 1969. 232 с.
- Булохов А.Д.* Фитоиндикация и ее практическое применение. Брянск: Изд-во БГУ, 2004. 245 с.
- Гайдышев И.* Анализ и обработка данных: специальный справочник СПб: Питер, 2001. 752 с.
- Джонгман Р.Г.Г., Тер Браак С.Дж.Ф., Ван Тонгерен О.Ф.Р.* Анализ данных в экологии сообществ и ландшафтов. М.: РАСХН, 1999. 306 с.
- Дюран Б., Одделл П.* Кластерный анализ. М., 1977. 128 с.
- Зайцев Г.Н.* Математика в экспериментальной ботанике. М.: Наука, 1990. 296 с.
- Зверев А.А.* Информационные технологии в исследованиях растительного покрова: Учебное пособие. Томск: ТМЛ-Пресс, 2007. 304 с.
- Кендэлл М.* Ранговые корреляции. М.: Статистика, 1975. 214 с.
- Кокс Д., Снелл Э.* Прикладная статистика. Принципы и примеры. Пер. с англ. М.: Мир, 1984. 200 с.
- Компьютерная биометрика / Под ред. Носова В.Н. М.: Изд-во МГУ, 1990. 232 с.

- Красовский Г.И., Филаретов Г.Ф.* Планирование эксперимента. Минск.: Изд-во БГУ, 1982. 302с.
- Мандель И.Д.* Кластерный анализ. М.: Финансы и статистика, 1988. 176 с.
- Перишков В.И., Савинков В.М.* Толковый словарь по информатике. 2-е изд., доп. М.: Финансы и статистика, 1995. 544 с.
- Раменский Л.Г., Цаценкин И.А., Чижиков О.Н., Антипин Н.А.* Экологическая оценка кормовых угодий по растительному покрову. М.: Сельхозгиз, 1956, 472 с.
- Селедец В.П.* Метод экологических шкал в ботанических исследованиях на Дальнем Востоке России. Владивосток, 2000. 245 с.
- Селедец В.П., Пробатова Н.С.* Экологический ареал вида у растений. Владивосток: Дальнаука, 2007. 97 с.
- Цыганов Д.Н.* Фитоиндикация экологических режимов в подзоне хвойно-широколиственных лесов. М.: Наука, 1983. 196 с.
- Шмидт В.М.* Математические методы в ботанике: Учебное пособие. Л.: Изд-во ЛГУ, 1984. 288 с.
- Шмидт В.М.* Статистические методы в сравнительной флористике. Л.: Изд-во ЛГУ, 1980. 176 с.
- Gauch, Hugh G., jr.* Multivariate analysis in community ecology. Cambridge, 1982. 298 pp.
- Kernshaw, Kenneth A.* Quantitative and Dynamic Plant Ecology. London, 1973. 308 pp.

в) ресурсы сети Интернет:

- Сравнительный анализ в биологии [Электрон. ресурс]. URL: <http://biocomparison.ucoz.ru/>
- Biological Software (Software useful for biological studies) [Электрон. ресурс]. URL: <http://www.bioexplorer.net/Software/>
- The International Plant Names Index [Электрон. ресурс]. URL: <http://www.ipni.org/index.html>
- Software for Biological Collection Management [Электрон. ресурс]. URL: <http://www.bgbm.org/tdwg/acc/Software.htm>
- Global Index of Vegetation-Plot Databases [Электрон. ресурс]. URL: <http://www.givd.info/>
- Internet Directory for Botany [Электрон. ресурс]. URL: <http://www.ou.edu/cas/botany-micro/idb-alpha/botany.html>
- Internet Directory for Botany: Software [Электрон. ресурс]. URL: <http://www.bgbm.fu-berlin.de/idb/botsoft.html>
- Информационные ресурсы на платформе БИН РАН [Электрон. ресурс]. URL: <http://www.binran.ru/resursy/informatsionnyye-resursy/>
- Биометрика: журнал для медиков и биологов, сторонников доказательной биомедицины [Электрон. ресурс]. URL: <http://www.biometrica.tomsk.ru/>

13. Перечень информационных технологий

а) лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Office Standart 2013 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, IBIS 7.1, BioStat 1.1, Synap 5.04 (DOS) и Synap 1.1 (Win), TurboVeg 3, Megatab 4.02, Juice 7.0, Decorana, Twinspan, WinTwins 2.3, NTSYS-рc 2.2, PAST 3.x (4.x), Flora Indicativa 1.0.6, StatSoft Statistica 10.0 (12.0), Semware QEdit Advanced 3.00, Punto Switcher 4.2 (2.9, 3.0);
- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ – <http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ – <http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>

- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

14. Материально-техническое обеспечение

Аудитории для проведения занятий лекционного типа с мультимедийным оборудованием.

Аудитории для проведения практических занятий, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации, оснащенные мультимедийным оборудованием, компьютерной техникой (настольные персональные компьютеры) и доступом в сеть Интернет, в электронную информационно-образовательную среду и к информационным справочным системам (компьютерные классы).

15. Информация о разработчике

Зверев Андрей Анатольевич, к.б.н., доцент, доц. кафедры ботаники ТГУ