

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Уральский федеральный университет имени первого Президента России Б.Н. Ельцина»

УТВЕРЖДАЮ
Директор по образовательной
деятельности


«30» 08 / С.Т. Князев
2020 г.



РАБОЧАЯ ПРОГРАММА МОДУЛЯ

Код модуля	Модуль
1154387	Большие данные

Екатеринбург, 2020

Перечень сведений о рабочей программе модуля	Учетные данные
Образовательная программа Цифровая медицина и биоинформатика	Код ОП 30.05.03/22.01
Направление подготовки 1. Медицинская кибернетика	Код направления и уровня подготовки 30.05.03

Программа модуля составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Емельянов Виктор Владимирович	к.м.н., доцент	доцент	Кафедра медицинской биохимии и биофизики
2	Зимницкая Светлана Анатольевна	кандидат биологических наук, доцент	доцент	департамент биологии и фундаментальной медицины
3	Ушенин Константин Сергеевич	-, -	руководитель образовательной программы	Школа бакалавриата института естественных наук и математики

Согласовано:

Учебный отдел



1. ОБЩАЯ ХАРАКТЕРИСТИКА МОДУЛЯ

Большие данные

1.1. Аннотация содержания модуля

Модуль «Большие данные» относится к вариативной части учебного плана и направлен на формирование общепрофессиональных и профессиональных компетенций в сфере анализа и обработки больших данных, в том числе математическим методам и техническим средствам обеспечивающим хранение и передачу информации. Объем модуля 6 з.е.

Целью изучения модуля является получение учащимися знаний о структуре и принципах работы современных информационных систем, позволяющих осуществлять обработку больших объемов информации. Курс «Структуры данных» посвящен алгоритмам и структурам представления данных, обеспечивающих возможность их быстрой обработки. Курс «Системы обработки больших данных» посвящен распределенным хранилищам информации, высоконагруженным сервисам итд.

1.2. Структура и объем модуля

Таблица 1

№	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах
1	Структуры данных	3
2	Системы обработки больших данных	3
ИТОГО по модулю:		6

1.3. Последовательность освоения модуля в образовательной программе

Пререквизиты модуля	нет
Постреквизиты и кореквизиты модуля	

1.4. Распределение компетенций по дисциплинам модуля, планируемые результаты обучения (индикаторы) по модулю

Таблица 2

Перечень дисциплин модуля	Код и наименование компетенции	Планируемые результаты обучения (индикаторы)
1	2	3
Структуры данных	ПК-5 - Способен проводить системный анализ объектов	32 Демонстрировать знание характеристик и принципов работы технических и

	<p>исследования в медицине и здравоохранении, разрабатывать модели и стандарты информационного взаимодействия, разрабатывать системы информационной поддержки управления знаниями в медицине, биологии и здравоохранении</p>	<p>программных средств в области здравоохранения.</p>
<p>Системы обработки больших данных</p>	<p>ПК-5 - Способен проводить системный анализ объектов исследования в медицине и здравоохранении, разрабатывать модели и стандарты информационного взаимодействия, разрабатывать системы информационной поддержки управления знаниями в медицине, биологии и здравоохранении</p>	<p>32 Демонстрировать знание характеристик и принципов работы технических и программных средств в области здравоохранения.</p> <p>У1 Уметь оценивать объект информатизации, определять характеристики необходимого комплекса технических и программных средств в области здравоохранения, и разрабатывать техническую документацию.</p>

1.5. Форма обучения

Обучение по дисциплинам модуля может осуществляться в очной формах.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Структуры данных

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Емельянов Виктор Владимирович	к.м.н., доцент	доцент	Кафедра медицинской биохимии и биофизики
2	Зимницкая Светлана Анатольевна	кандидат биологических наук, доцент	доцент	департамент биологии и фундаментальной медицины
3	Ушенин Константин Сергеевич	-, -	руководитель образователь ной программы	Школа бакалавриата института естественных наук и математики

Рекомендовано учебно-методическим советом института

1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

Авторы:

1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Разноуровневое (дифференцированное) обучение
 - o Продвинутый уровень

**Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания; Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.*

1.2. Содержание дисциплины

Таблица 1.1

Код раздела, темы	Раздел, тема дисциплины*	Содержание
P1	Алгоритмы поиска и сортировки	Бинарный поиск. Сортировки за $O(n \ln n)$: быстрая сортировка (Quick Sort), сортировка на куче, сортировка слиянием. Красно-черное дерево и AVL дерево.
P2	Пространственные индексы и алгоритмы обработки пространственных данных	Хранение облаков точек, поверхностных и объемных сеток. Пространственные индексы: kdTree, BallTree. Визуализация результатов сегментации медицинских данных (алгоритм CubeMarching). Алгоритмы построения поверхностей по данным трехмерного сканирования.
P3	Структуры данных для биоинформатики	Расстояния между последовательностями, эволюционные расстояния, матрицы замен, глобальное и локальное выравнивание. Алгоритмы поиска подстроки в тексте, суффиксное дерево и массив, построение сжатых индексов по геному. Префиксное дерево. De Bruijn граф. Дерево палиндромов, палиндромный поиск.
P1	Алгоритмы поиска и сортировки	Бинарный поиск. Сортировки за $O(n \ln n)$: быстрая сортировка (Quick Sort), сортировка на куче, сортировка слиянием. Красно-черное дерево и AVL дерево.

1.3. Программа дисциплины реализуется на государственном языке Российской Федерации.

2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

ДИСЦИПЛИНЫ

Программирование на языке Python

Электронные ресурсы (издания)

- Асанов, Магаз Оразкимович. Учебно-методический комплекс дисциплины "Комбинаторные алгоритмы" [Электронный ресурс] / М. О. Асанов, В. В. Расин ; Федер. агентство по образованию, Урал. гос. ун-т им. А. М. Горького, ИОНЦ "Информационная безопасность" [и др.]. — Электрон. дан. (1,54 Мб). — Екатеринбург : [б. и.], 2008.
- Bioinformatics Algorithms - <https://www.bioinformaticsalgorithms.org/>

Печатные издания

- Кормен Т. и др. Алгоритмы. Построение и анализ:[пер. с англ.]. – Издательский дом Вильямс, 2009.

Профессиональные базы данных, информационно-справочные системы

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Структуры данных

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
1	Лекции	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Подключение к сети Интернет	Не требуется
2	Практические занятия	Мебель аудиторная с количеством рабочих мест в	Не требуется

		<p>соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Доска аудиторная</p> <p>Персональные компьютеры с доступом в сеть Интернет на каждого учащегося</p>	
3	Текущий контроль и промежуточная аттестация	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Рабочее место преподавателя</p> <p>Доска аудиторная</p>	Не требуется
4	Самостоятельная работа студентов	<p>Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов</p> <p>Подключение к сети Интернет</p>	Не требуется

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

Вопросы к зачету/экзамену по дисциплине

1. Бинарный поиск.
2. Сортировки за $O(n \ln n)$: быстрая сортировка (Quick Sort), сортировка на куче, сортировка слиянием.
3. Красно-чёрное дерево и AVL дерево.
4. Хранение облаков точек, поверхностных и объемных сеток.
5. Пространственные индексы: kdTree, BallTree.
6. Визуализация результатов сегментации медицинскз данных (алгоритм CubeMarching).
7. Алгоритмы построения поверхностей по данным трехмерного сканирования.
8. Расстояния между последовательностями, эволюционные расстояния, матрицы замен, глобальное и локальное выравнивание.
9. Алгоритмы поиска подстроки в тексте, суффиксное дерево и массив, построение сжатых индексов по геному.
10. Префиксное дерево.
11. De Bruijn граф.
12. Дерево палиндромов, палиндромный поиск.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Системы обработки больших данных

Рабочая программа дисциплины составлена авторами:

№ п/п	Фамилия Имя Отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Емельянов Виктор Владимирович	к.м.н., доцент	доцент	Кафедра медицинской биохимии и биофизики
2	Зимницкая Светлана Анатольевна	кандидат биологических наук, доцент	доцент	департамент биологии и фундаментальной медицины
3	Ушенин Константин Сергеевич	- , -	руководитель образователь ной программы	Школа бакалавриата института естественных наук и математики

Рекомендовано учебно-методическим советом института

1. СОДЕРЖАНИЕ И ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ

1.1. Технологии реализации, используемые при изучении дисциплины модуля

- Традиционная (репродуктивная) технология
- Разноуровневое (дифференцированное) обучение
 - o Продвинутый уровень

**Базовый I уровень – сохраняет логику самой науки и позволяет получить упрощенное, но верное и полное представление о предмете дисциплины, требует знание системы понятий, умение решать проблемные ситуации. Освоение данного уровня результатов обучения должно обеспечить формирование запланированных компетенций и позволит обучающемуся на минимальном уровне самостоятельности и ответственности выполнять задания; Продвинутый II уровень – углубляет и обогащает базовый уровень как по содержанию, так и по глубине проработки материала дисциплины. Это происходит за счет включения дополнительной информации. Данный уровень требует умения решать проблемы в рамках курса и смежных курсов посредством самостоятельной постановки цели и выбора программы действий. Освоение данного уровня результатов обучения позволит обучающемуся повысить уровень самостоятельности и ответственности до творческого применения знаний и умений.*

1.2. Содержание дисциплины

Таблица 1.1

Код раздела, темы	Раздел, тема дисциплины*	Содержание
P1	Введение в дисциплину	Понятие больших данных. Требования к хранилищам больших данных. Закон о хранении персональных данных на территории РФ.
P2	Распределенные базы данных, NoSQL	Отличия баз Microsoft SQL, Oracle, Postges, MySql. Распределенные реляционные базы данных. Индексация в базах данных. Транзакции в базы данных. Методы разрешения конфликтов в базы данных.
P3	NoSQL решения	NoSQL решения. Mongo DB. Google BigTable, Azure Cosmos DB, Oracle NoSQL Database,
P4	Технологии доставки вычислений до данных	Концепция MapReduce. Apache Hadoop, Spark.

1.4. Программа дисциплины реализуется на государственном языке Российской Федерации.

2. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Машинное обучение и статистика

Электронные ресурсы (издания)

1. Dean J., Ghemawat S. MapReduce: simplified data processing on large clusters //Communications of the ACM. – 2008. – Т. 51. – №. 1. – С. 107-113.

2. Boicea A., Radulescu F., Agapin L. I. MongoDB vs Oracle--database comparison //2012 third international conference on emerging intelligent data and web technologies. – IEEE, 2012. – С. 330-335.
3. Györödi C. et al. A comparative study: MongoDB vs. MySQL //2015 13th International Conference on Engineering of Modern Electric Systems (EMES). – IEEE, 2015. – С. 1-6.
4. Chodorow K. MongoDB: the definitive guide: powerful and scalable data storage. – " O'Reilly Media, Inc.", 2013.

Печатные издания

Профессиональные базы данных, информационно-справочные системы

Материалы для лиц с ОВЗ

Весь контент ЭБС представлен в виде файлов специального формата для воспроизведения синтезатором речи, а также в тестовом виде, пригодном для прочтения с использованием экранной лупы и настройкой контрастности.

Базы данных, информационно-справочные и поисковые системы

1. Cloud BigTable - <https://cloud.google.com/bigtable>
2. База данных SQL Azure - <https://azure.microsoft.com/ru-ru/products/azure-sql/database/>

3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Системы обработки больших данных

Сведения об оснащённости дисциплины специализированным и лабораторным оборудованием и программным обеспечением

Таблица 3.1

№ п/п	Виды занятий	Оснащённость специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
1	Лекции	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная Подключение к сети Интернет	Не требуется
2	Практические занятия	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная	Не требуется

		Персональные компьютеры с доступом в сеть Интернет на каждого учащегося	
3	Текущий контроль и промежуточная аттестация	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Рабочее место преподавателя Доска аудиторная	Не требуется
4	Самостоятельная работа студентов	Мебель аудиторная с количеством рабочих мест в соответствии с количеством студентов Подключение к сети Интернет	Не требуется

**Приложение
к рабочей программе дисциплины**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ

Вопросы к зачету/экзамену по дисциплине

1. Понятие больших данных. Требования к хранилищам больших данных.
2. Отличия баз Microsoft SQL, Oracle, Postges, MySql.
3. Распределенные реляционные базы данных.
4. Индексация в базах данных
5. Транзакции в базы данных. Методы разрешения конфликтов в базы данных.
6. NoSQL. Концепции и причины появления подхода.
7. Mongo DB. Google BigTable, Azure Cosmos DB, Oracle NoSQL Database
8. Концепция MapReduce. Распределенные вычисления.
9. Apache Hadoop
10. Apache Spark