



Компонент образовательной программы

Образовательная программа утверждена

Решением Ученого совета

ГБУ «НИИОЗММ ДЗМ»

Протокол от 22.01.2025 г. № 1

Рабочая программа дисциплины

БОЛЬШИЕ ДАННЫЕ И ИХ ОБРАБОТКА

по направлению подготовки

09.04.02 Информационные системы и технологии

направленность (профиль): **Информационные системы и технологии в
здоровоохранении**

(квалификация выпускника: магистр)

Форма обучения: очная

1. Код и наименование дисциплины (модуля): Б1.О.10 Большие данные и их обработка.

2. Уровень высшего образования: магистратура.

3. Направление подготовки: 09.04.02 Информационные системы и технологии, направленность (профиль): Информационные системы и технологии в здравоохранении.

4. Цель изучения дисциплины (модуля): является изучение современных методов и программных средств, используемых при анализе и обработке больших данных.

5. Задачи дисциплины (модуля):

- сформировать у обучающихся представления об информационно-технологическом формате профессиональной деятельности медицинского работника;
- ознакомить с концептуальными основаниями использования Больших данных в медицине и здравоохранении;
- ознакомить с основами международных стандартов сбора, хранения, передачи, преобразования и анализа данных в профессиональной деятельности;
- сформировать представление об информационно-технологическом базисе в современных предиктивном, превентивном, профилактическом и персонифицированном подходах 4П-медицины в работе с пациентами в здравоохранении;
- научить использовать языки программирования, необходимые в работе с Большими данными в профессиональной деятельности;
- развивать коммуникативные умения и навыки, направленные на решение вычислительно прогностических аспектов в профессиональной деятельности;
- способствовать формированию личностной готовности к будущей профессиональной деятельности.

6. Место дисциплины (модуля) в структуре ОПОП: дисциплины (модули), обязательная часть, 2 курс обучения, 3 семестр.

7. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями выпускников):

В результате освоения программы магистратуры у выпускника должны быть сформированы: универсальные, общепрофессиональные, профессиональные компетенции.

В результате освоения указанной программы магистратуры выпускник должен обладать следующими компетенциями:

общепрофессиональными компетенциями:

– способен использовать методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации посредством информационных технологий (ОПК-6).

профессиональными компетенциями:

– способен к управлению развитием баз данных (ПК-1).

8. Планируемые результаты обучения

Магистр должен:

знать: возможности технологий анализа больших данных при проведении научных исследований; существующие в современном мире источники и типы информации; существующие современные технологии высоконагруженных систем хранения и обработки данных, принципы работы высоконагруженных систем.

уметь: оценивать важность применения научных методов для извлечения пользы из больших массивов данных; визуализировать имеющиеся данные, отбрасывать несущественную информацию, структурировать информацию в рамках поставленной задачи; выбирать подходящую технологию хранения и обработки больших данных, использовать современные высоконагруженные системы хранения и обработки больших данных.

владеть: научными методами обработки больших данных, в том числе методами интеллектуального анализа больших данных; современными средствами визуализации, методами предварительной подготовки данных; технологией Map Reduce и ее реализацией Hadoop.

Карта формирующих компетенций (или их частей) дисциплины (модуля)

№ п/п	Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемый результат обучения по дисциплине		
			Знать	Уметь	Владеть
1.	Способен использовать методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации	ОПК-6.1 ОПК-6.2 ОПК-6.3	Знает основные положения системной инженерии в области получения, передачи, хранения, переработки и представления информации посредством	Умеет применять методы и средства системной инженерии в области получения, передачи, хранения, переработки и представления информации посредством информационных технологий	Имеет практический опыт применения методов и средств системной инженерии в области получения, передачи, хранения, переработки и

№ п/п	Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемый результат обучения по дисциплине		
			Знать	Уметь	Владеть
	посредством информационных технологии (ОПК- 6)		информационных технологий		представления информации посредством информационных технологий
2.	Способен к управлению развитием баз данных (ПК-1)	ПК-1.1 ПК-1.2 ПК-1.3	Знает основные тенденции развития информационных технологий в области баз данных. Принципы работы, технологии и возможности аппаратного и программного обеспечения баз данных. Принципы построения бизнес-процессов и алгоритмов работы. Требования к подготовке регламентирующих документов. Современные и перспективные технологии в области баз данных. Способы и технологии обновления баз данных, а так же механизмы контроля обновления баз данных. Отличительные особенности современных и перспективных баз данных. Способы и технологии миграции баз данных, а также механизмы контроля успешности выполнения миграции баз данных. Отличительные особенности и функциональность различных версий баз данных. Менеджмент и	Уметь выявлять проблемы организации, связанные с информационным обеспечением и особенностями установленной базы данных. Прогнозировать состояние и осуществлять планирование по развитию баз данных в организации. Разрабатывать и описывать бизнес-процессы. Готовить регламент по обновлению версий программного обеспечения баз данных в соответствии с требованиями нормативных документов. Осваивать новые информационные технологии в области баз данных. Анализировать возможности внедрения новых информационных технологий. Находить информацию, необходимую для выполнения задач по управлению и развитию баз данных. Планировать и осуществлять мероприятия по переходу на новую версию баз данных. Контролировать успешность выполнения работ по обновлению версии баз данных. Планировать и осуществлять	Имеет практический опыт сбора и анализа нерезализованных потребностей пользователей баз данных. Исследования рынка перспективных баз данных, их принципиальных возможностей. Подготовка плана реализации принятых решений по перспективному развитию баз данных. Анализа основных этапов обновления версий программного обеспечения баз данных. Разработки и описания типовых процессов по обновлению версий баз данных. Подготовка регламентирующ их документов по обновлению версий баз данных. Анализа основных этапов миграции баз данных на новые платформы и новые версии программного обеспечения. Разработка и описание типовых процессов миграции баз данных на новые платформы и

№ п/п	Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Планируемый результат обучения по дисциплине		
			Знать	Уметь	Владеть
			<p>управление в информационных технологиях. Перспективы технологического развития организации, современные тенденции развития информационных систем и баз данных.</p>	<p>мероприятия по миграции баз данных. Контролировать успешность выполнения работ по миграции баз данных. Анализировать ситуацию по информационному обеспечению организации и принимать управленческие решения по его развитию.</p>	<p>новые версии программного обеспечения. Подготовка регламентирующих документов по проведению миграции баз данных. Мониторинга новых информационных технологий в области баз данных, появляющихся на рынке. Освоения и внедрения в практику администрирования новых технологий работы с базами данных. Планирование этапов и анализ результатов выполнения каждого этапа обновления версий баз данных. Планирования, проведения и анализа результатов проверки функционирования баз данных после обновления. Планирования этапов миграции баз данных. Анализ результатов тестирования баз данных после миграции. Восстановления баз данных и корректировка действий при обнаружении ошибок миграции.</p>

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

Объем дисциплины (модуля) составляет **3** зачетные единицы, всего **108** часов, из которых **44** часа составляет контактная работа обучающегося с преподавателем (**16** часов - лекционные занятия; **28** часов - практические занятия, мероприятия промежуточной аттестации - зачет), **64** часа составляет самостоятельная работа обучающегося.

10. Форма обучения: очная.

11. Применяемые образовательные технологии для различных видов учебных занятий и для контроля освоения обучающимися запланированных результатов обучения:

В процессе реализации дисциплины применяются следующие образовательные технологии:

- лекционно-практические (учебный материал сконцентрирован в блоки и преподносится как единое целое);
- коммуникативные (обучаемый выступает в роли активного, сознательного, равноправного участника учебного процесса, развивающегося по своим возможностям);
- саморазвивающиеся (основывается на самоорганизации, самоопределении, самоутверждении обучающегося);
- компьютерные (дидактическая система подготовки и трансляции информации обучающемуся, основным средством реализации которой является компьютер).

12. Форма аттестации

Форма промежуточной аттестации по дисциплине (модулю) – зачет.

13. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических или астрономических часов и виды учебных занятий

№ п/п	Наименование разделов (тем) дисциплины, промежуточная аттестация	Всего часов	Из них:			Форма текущего контроля знаний/ промежуточной аттестации	Код компетенции	Литература для самостоятельной работы
			Лекции	Практические занятия (семинарские)	Самостоятельная работа			
1.	Тема 1. Введение в Big Data и медицинские данные	28	4	6	18	Тест, Опрос, Собеседование	ОПК-6 ПК-1	1-8
2.	Тема 2. Технологический стек обработки данных	54	8	16	30	Тест, Опрос, Собеседование	ОПК-6 ПК-1	1-8
3.	Тема 3. Data Mining и прогнозное моделирование	26	4	6	16	Тест, Опрос, Собеседование	ОПК-6 ПК-1	1-8
	ИТОГО	108	16	28	64			

14. Содержание разделов и тем.

Тема 1. Введение в Big Data и медицинские данные.

Концепция Big Data (5V). EHR, DICOM, геномные данные, IoT. Нормативно-правовая база: HIPAA, GDPR, ФЗ-152. Стандарты HL7, FHIR.

Тема 2. Технологический стек обработки данных

Высоконагруженные системы хранения и обработки больших данных. Архитектура Hadoop Экосистема Hadoop (HBase, Hive). Apache Spark: Core, SQL, Streaming. Сравнение с MapReduce. Apache Kafka, data pipelines. Облачные платформы.

Тема 3. Data Mining и прогнозное моделирование

Инструменты Data Mining. Ансамблевые методы. Случайный лес. Градиентный бустинг Разработка и валидация прогнозных моделей. Интерпретация моделей и инженерия признаков. Визуализация данных.

15. Организация самостоятельной работы

При организации самостоятельно работы обучающимся необходимо учитывать следующие особенности взрослых людей:

- осознанное отношение к процессу своего обучения;
- потребность в самостоятельности;
- потребность в осмысленности обучения (для решения важной проблемы и достижения конкретной цели), что обеспечивает мотивацию;
- наличие жизненного опыта – важного источника обучения;
- влияние на процесс обучения профессиональных, социальных, бытовых и временных факторов.

В качестве главного признака самостоятельной деятельности рассматривается не то, что магистр работает без привлечения преподавателя, а то, что каждое действие, выполняемое обучающимся, им осознается, подчиненно цели, которую он сам поставил.

Основной смысл самостоятельной работы состоит в том, чтобы:

- мотивировать магистра к освоению учебного материала;
- повысить ответственность обучающегося за свою учебу;
- формировать у магистра системного мышления на основе самостоятельной работы.

Самостоятельная работа требует упорядочения и системной организации.

Основным видом самостоятельной работы обучающихся при освоении дисциплины является изучение литературы и интернет-ресурсов, рекомендуемых в рабочей программе дисциплины, а также, ответы на вопросы для самопроверки и подготовка в промежуточной аттестации,

При работе с литературой целесообразно придерживаться такой последовательности. Сначала прочитать весь заданный текст в быстром темпе. Цель такого чтения заключается в том, чтобы создать общее представление об изучаемом материале, понять общий смысл прочитанного. Затем прочитать вторично, более медленно, чтобы в ходе чтения понять и запомнить смысл каждой фразы, каждого положения и вопроса в целом.

Чтение приносит пользу и становится продуктивным, когда сопровождается записями. Это может быть составление плана прочитанного текста, тезисы или выписки, конспектирование и др. Выбор вида записи зависит от характера изучаемого материала и целей работы с ним. Если содержание материала несложное, легко усваиваемое, можно ограничиться составлением плана. Если материал содержит новую и трудно усваиваемую информацию, целесообразно его законспектировать.

Различаются четыре типа конспектов:

- план-конспект – это развернутый детализированный план, в котором по наиболее сложным вопросам даются подробные пояснения;

- текстуальный конспект – это воспроизведение наиболее важных положений и фактов источника;
- свободный конспект – это четко и кратко изложенные основные положения в результате глубокого изучения материала, могут присутствовать выписки, цитаты, тезисы; часть материала может быть представлена планом;
- тематический конспект – составляется на основе изучения ряда источников и дает ответ по изучаемому вопросу.

В процессе изучения материала источника и составления конспекта нужно обязательно применять различные выделения, подзаголовки, создавая блочную структуру конспекта. Это делает конспект легко воспринимаемым и удобным для работы.

16. Текущий контроль успеваемости и промежуточная аттестации

16.1. Текущий контроль успеваемости

Текущий контроль направлен на систематическую проверку качества (не менее двух раз за период освоения дисциплины) усвоения разделов (тем) дисциплины. Выбор оценочного средства для проведения текущего контроля успеваемости на усмотрение преподавателя.

Оценочные материалы для проведения текущего контроля успеваемости

Примеры тестовых заданий:

1. Какие основные признаки VVV характеризуют свойства Больших данных?

- Volume,
- Velocity,
- Variety,
- Vocabulary,
- Visuality.

2. Какие компании НЕ используют Hadoop, включая крупные и хорошо известные:

- Facebook;
- Amazon;
- AliExpress;
- Yahoo;
- нет в списке.

3. Какая конкретная функция используется для обработки больших наборов распределенных задач с использованием большого количества компьютеров, образующих кластер?

- Map Reduce;
- Data Mining;
- NET Framework;
- Machine Learning.

4. Какая технология представляет собой большой стек технологий, так или иначе связанных с обработкой больших данных не только при помощи MapReduce?

- Hadoop;
- Deep Learning;
- Big Data.

5. Какая файловая система предназначена для хранения файлов больших размеров с возможностью потокового доступа к информации, поблочно распределённой по узлам вычислительного кластера?

- FAT;
- NTFS;
- HDFS.

Примеры контрольных вопросов:

1. Роль машинного обучения в анализе Больших данных.
2. Какие статистические методы используются в анализе Больших данных.
3. В чем суть принципа горизонтальной масштабируемости в качестве базового принципа обработки Больших данных.
4. В чем специфика статистических методов обработки в применении к Большим данным?
5. Какая файловая система предназначена для хранения файлов больших размеров с возможностью потокового доступа к информации, поблочно распределённой по узлам вычислительного кластера?
6. Из каких основных четырех модулей состоит Архитектурная концепция Hadoop?
7. Опишите назначение и алгоритм работы Загрузчик данных (Data Loader).
8. Опишите алгоритм работы второго этапа обработки больших объемов данных под названием Reduce.
9. Опишите методы обработки больших данных. В чем особенность методов интеллектуального анализа больших данных.
10. В чем состоит основной принцип организации работы fadoopDB?

Материалы для проведения промежуточной аттестации по дисциплине (модулю) представлен в оценочных материалах

17. Ресурсное обеспечение

Материально-техническое обеспечение дисциплины:

Аудитория № 4 (для проведения лекционных, практических занятий, промежуточной аттестации)

Автоматизированное рабочее место (монитор, системный блок, клавиатура, компьютерная мышь) с доступом в Интернет – 1 шт.;

Ноутбук HP EliteBook 840 G3 – 11 шт.;

Доска магнитно-маркерная 100x150 см лаковое покрытие Attache BlackFrame – 1 шт.;

Мультимедийный проектор Epson H311B- 1 шт.;

Экран настенный с электроприводом Digis Electra 160 - 1 шт.;

Веб-камера Logitech HD Webcam C270 со встроенным микрофоном- 1 шт.;

Точка доступа Ubiquiti UniFi AC LR – 1шт.;

Кресло складное с пюпитром СКП-1 – 20 шт.;

Стол – 2 шт.;

Кресло – 2 шт.

Аудитория № 10 (для проведения лекционных, практических занятий, промежуточной аттестации, итоговой аттестации)

Автоматизированное рабочее место (монитор, системный блок, клавиатура, компьютерная мышь) с доступом в Интернет – 15 шт.;

Доска магнитно-маркерная 100x150 см лаковое покрытие Attache BlackFrame – 1 шт.;

Мультимедийный проектор Mitsubishi XD550U - 1 шт.;

Веб-камера Logitech HD Webcam C270 со встроенным микрофоном- 1 шт.;

Точка доступа Ubiquiti UniFi AC LR – 1шт.;

Экран настенный с электроприводом Digis Electra 160 - 1 шт.;

Стул – 16 шт.;

Кресло – 1 шт.;

Стол – 13 шт.

Аудитория № 8 (для самостоятельной работы)

Автоматизированное рабочее место (монитор, системный блок, клавиатура, компьютерная мышь) с доступом в Интернет – 4 шт.;

Веб-камера Logitech HD Webcam C270 со встроенным микрофоном- 4 шт.;

Точка доступа Ubiquiti UniFi AC LR – 1шт.;

Стул – 6 шт.;

Стол – 6 шт.

18. Учебно-методическое и информационное обеспечение дисциплины:

18.1. Литература

1. Nadoor в действии / Ч. Лэм. - М. : ДМК Пресс, 2019. - 424 с.
2. Базы данных: теория и практика / Б.Я. Советов, В.В. Цехановский. - 2-е изд. - М. : Юрайт, 2013. - 463 с.
3. Введение в системы баз данных / К.Д. Дейт ; Пер. с англ. и ред. К.А.Птицына. - 8-е изд. - М. : Вильямс, 2008. - 1328 с.
4. Интеллектуальное право в условиях развития технологии Big Data. База данных как объект интеллектуальных и иных прав : монография / Е.А. Войниканис, М.А. Кольздорф, В.А. Корнеев [и др.]. - Электронные текстовые данные. - Москва : Проспект, 2022. - 176 с.
5. Информационные технологии анализа данных. Data analysis : учебное пособие / Ю.Ю. Петрунин. - 2-е изд. ; Электронные текстовые данные. - М. : КДУ, 2010.
6. Компьютерная обработка статистических данных : практикум : учебно-методическое пособие / Т.Н. Ледащева, В.Е. Пинаев. - Электронные текстовые данные. - Москва : РУДН, 2021. - 81 с.
7. Конструирование и сопровождение баз данных в среде Access-2016 : учебно-методическое пособие / А.С. Панкратов, Л.В. Александрова, М.Б. Фомин. - Электронные текстовые данные. - М. : РУДН, 2018. - 44 с.
8. Проектирование и разработка баз данных в среде Access : учебно-методическое пособие для студентов технических специальностей / А.С. Панкратов, Л. В. Александрова, М.Б. Фомин. - Электронные текстовые данные. - Москва : РУДН, 2014. - 40 с.

18.2. Программное обеспечение и Интернет-ресурсы

- Microsoft Office Стандартный 2010
- Microsoft Office 2016 Professional Plus
- Научная электронная библиотека elibrary.ru
- Научная электронная библиотека УНИБЦ (НБ) РУДН library@rudn.ru
- Научная электронная библиотека <https://cyberleninka.ru/>
- Сайт Департамента здравоохранения города Москвы. Режим доступа: <https://mosgorzdrav.ru/>, свободный.
- Официальный интернет-портал правовой информации. Государственная система правовой информации. Режим доступа: <http://pravo.gov.ru/ips/>, свободный.
- Сайт Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека. Режим доступа: <https://rospotrebnadzor.ru/documents/documents.php>, свободный.
- Электронный фонд правовой и нормативно-технической документации Режим доступа: <http://docs.cntd.ru/>, свободный.

– Сайт ГБУ «НИИОЗММ ДЗМ». Режим доступа: <http://niioz.ru/>, свободный.

Зарубежные ресурсы:

- Реферативная база научных публикаций Web of Science
<http://www.webofscience.com>
- База Scopus scopus.com
- Всемирная полнотекстовая база PhD диссертаций Proquest
<https://www.proquest.com/>
- Международная база данных научных периодических изданий Jstore
<https://www.jstor.org/>