



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение
высшего образования
**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ИГУ»)**

Институт математики и информационных технологий
Кафедра информационных технологий



Рабочая программа дисциплины (модуля)

Б1.В.14 Обработка больших объемов данных

| | |
|-------------------------------------|--|
| Направление подготовки | 01.03.02 Прикладная математика и информатика |
| Направленность (профиль) подготовки | Системы искусственного интеллекта |
| Квалификация выпускника | бакалавр |
| Форма обучения | очная |

Иркутск 2024 г.

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель

Целями освоения дисциплины являются ознакомление студентов с организацией таких систем хранения больших данных, как озеро данных и хранилище данных, а также процессами сбора и обработки данных.

Задачи:

Задачами освоения дисциплины являются следующие: сформировать компетенции, необходимые для организации систем хранения данных; выработать навыки проектирования хранилищ данных и организации ETL- и ELT-процессов; познакомить с такими подходами к проектированию хранилищ данных, как многомерное моделирование, якорное моделирование и Data Vault.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Учебная дисциплина Б1.В.14 Обработка больших объемов данных относится к части Блока 1 образовательной программы, формируемой участниками образовательных отношений.

3. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс освоения дисциплины направлен на формирование следующих компетенций в соответствии с ФГОС ВО и ОП ВО по направлению подготовки 01.03.02 Прикладная математика и информатика:

ПК-2 Способен осуществлять сбор и подготовку данных для систем искусственного интеллекта;

ПК-4 Способен использовать инструментальные средства для решения задач машинного обучения.

4. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Объем дисциплины составляет 4 зачетных ед., 144 час.

Форма промежуточной аттестации: зачет с оценкой.

4.1. Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

| Раздел дисциплины / тема | Виды учебной работы | | | Самост. работа | Формы текущего контроля; Формы промежут. аттестации |
|---|--|--------------|----------------|----------------|---|
| | Контактная работа преподавателя с обучающимися | | | | |
| | Лекции | Лаб. занятия | Практ. занятия | | |
| Тема 1 Введение в большие данные | 4 | 4 | | 9 | |
| Тема 2 Системы хранения данных | 4 | 4 | | 9 | |
| Тема 3 Хранилище данных | 4 | 4 | | 9 | |
| Тема 4 Многомерное моделирование | 4 | 4 | | 9 | |
| Тема 5 Data Vault | 4 | 4 | | 9 | |
| Тема 6 Якорное моделирование | 4 | 4 | | 9 | |
| Тема 7 Сравнение подходов к хранению и обработке больших данных | 4 | 4 | | 9 | |
| Тема 8 Распределенный анализ данных и машинное обучение | 4 | 4 | | 9 | |
| Итого (8 семестр): | 32 | 32 | | 72 | зач.с оц. |

4.2. Содержание учебного материала

Тема 1 Введение в большие данные

Тема 2 Системы хранения данных

Тема 3 Хранилище данных

Тема 4 Многомерное моделирование

Тема 5 Data Vault

Тема 6 Якорное моделирование

Тема 7 Сравнение подходов к хранению и обработке больших данных

Тема 8 Распределенный анализ данных и машинное обучение

4.3. Методические указания по организации самостоятельной работы студентов

Самостоятельная работа студентов всех форм и видов обучения является одним из обязательных видов образовательной деятельности, обеспечивающей реализацию требований Федеральных государственных стандартов высшего образования. Согласно требованиям нормативных документов самостоятельная работа студентов является обязательным компонентом образовательного процесса, так как она обеспечивает закрепление получаемых на лекционных занятиях знаний путем приобретения навыков осмысления и расширения их содержания, навыков решения актуальных проблем формирования общекультурных и профессиональных компетенций, научно-исследовательской деятельности, подготовки к семинарам, лабораторным работам, сдаче зачетов и экзаменов. Самостоятельная работа студентов представляет собой совокупность

аудиторных и внеаудиторных занятий и работ. Самостоятельная работа в рамках образовательного процесса в вузе решает следующие задачи:

- закрепление и расширение знаний, умений, полученных студентами во время аудиторных и внеаудиторных занятий, превращение их в стереотипы умственной и физической деятельности;
- приобретение дополнительных знаний и навыков по дисциплинам учебного плана;
- формирование и развитие знаний и навыков, связанных с научно-исследовательской деятельностью;
- развитие ориентации и установки на качественное освоение образовательной программы;
- развитие навыков самоорганизации;
- формирование самостоятельности мышления, способности к саморазвитию, самосовершенствованию и самореализации;
- выработка навыков эффективной самостоятельной профессиональной теоретической, практической и учебно-исследовательской деятельности.

Подготовка к лекции. Качество освоения содержания конкретной дисциплины прямо зависит от того, насколько студент сам, без внешнего принуждения формирует у себя установку на получение на лекциях новых знаний, дополняющих уже имеющиеся по данной дисциплине. Время на подготовку студентов к двухчасовой лекции по нормативам составляет не менее 0,2 часа.

Подготовка к практическому занятию. Подготовка к практическому занятию включает следующие элементы самостоятельной деятельности: четкое представление цели и задач его проведения; выделение навыков умственной, аналитической, научной деятельности, которые станут результатом предстоящей работы. Выработка навыков осуществляется с помощью получения новой информации об изучаемых процессах и с помощью знания о том, в какой степени в данное время студент владеет методами исследовательской деятельности, которыми он станет пользоваться на практическом занятии. Подготовка к практическому занятию нередко требует подбора материала, данных и специальных источников, с которыми предстоит учебная работа. Студенты должны дома подготовить к занятию 3–4 примера формулировки темы исследования, представленного в монографиях, научных статьях, отчетах. Затем они самостоятельно осуществляют поиск соответствующих источников, определяют актуальность конкретного исследования процессов и явлений, выделяют основные способы доказательства авторами научных работ ценности того, чем они занимаются. В ходе самого практического занятия студенты сначала представляют найденные ими варианты формулировки актуальности исследования, обсуждают их и обосновывают свое мнение о наилучшем варианте. Время на подготовку к практическому занятию по нормативам составляет не менее 0,2 часа.

Подготовка к семинарскому занятию. Самостоятельная подготовка к семинару направлена: на развитие способности к чтению научной и иной литературы; на поиск дополнительной информации, позволяющей глубже разобраться в некоторых вопросах; на выделение при работе с разными источниками необходимой информации, которая требуется для полного ответа на вопросы плана семинарского занятия; на выработку умения правильно выписывать высказывания авторов из имеющихся источников информации, оформлять их по библиографическим нормам; на развитие умения осуществлять анализ выбранных источников информации; на подготовку собственного выступления по обсуждаемым вопросам; на формирование навыка оперативного реагирования на разные мнения, которые могут возникать при обсуждении тех или иных научных проблем. Время на подготовку к семинару по нормативам составляет не менее 0,2 часа.

Подготовка к коллоквиуму. Коллоквиум представляет собой коллективное обсуждение раздела дисциплины на основе самостоятельного изучения этого раздела

студентами. Подготовка к данному виду учебных занятий осуществляется в следующем порядке. Преподаватель дает список вопросов, ответы на которые следует получить при изучении определенного перечня научных источников. Студентам во внеаудиторное время необходимо прочитать специальную литературу, выписать из нее ответы на вопросы, которые будут обсуждаться на коллоквиуме, мысленно сформулировать свое мнение по каждому из вопросов, которое они выскажут на занятии. Время на подготовку к коллоквиуму по нормативам составляет не менее 0,2 часа.

Подготовка к контрольной работе. Контрольная работа назначается после изучения определенного раздела (разделов) дисциплины и представляет собой совокупность развернутых письменных ответов студентов на вопросы, которые они заранее получают от преподавателя. Самостоятельная подготовка к контрольной работе включает в себя: — изучение конспектов лекций, раскрывающих материал, знание которого проверяется контрольной работой; повторение учебного материала, полученного при подготовке к семинарским, практическим занятиям и во время их проведения; изучение дополнительной литературы, в которой конкретизируется содержание проверяемых знаний; составление в мысленной форме ответов на поставленные в контрольной работе вопросы; формирование психологической установки на успешное выполнение всех заданий. Время на подготовку к контрольной работе по нормативам составляет 2 часа.

Подготовка к зачету. Самостоятельная подготовка к зачету должна осуществляться в течение всего семестра. Подготовка включает следующие действия: перечитать все лекции, а также материалы, которые готовились к семинарским и практическим занятиям в течение семестра, соотнести эту информацию с вопросами, которые даны к зачету, если информации недостаточно, ответы находят в предложенной преподавателем литературе. Рекомендуются делать краткие записи. Время на подготовку к зачету по нормативам составляет не менее 4 часов.

Подготовка к экзамену. Самостоятельная подготовка к экзамену схожа с подготовкой к зачету, особенно если он дифференцированный. Но объем учебного материала, который нужно восстановить в памяти к экзамену, вновь осмыслить и понять, значительно больше, поэтому требуется больше времени и умственных усилий. Важно сформировать целостное представление о содержании ответа на каждый вопрос, что предполагает знание разных научных трактовок сущности того или иного явления, процесса, умение раскрывать факторы, определяющие их противоречивость, знание имен ученых, изучавших обсуждаемую проблему. Необходимо также привести информацию о материалах эмпирических исследований, что указывает на всестороннюю подготовку студента к экзамену. Время на подготовку к экзамену по нормативам составляет 36 часов для бакалавров.

В ФБГОУ ВО «ИГУ» организация самостоятельной работы студентов регламентируется Положением о самостоятельной работе студентов, принятым Ученым советом ИГУ 22 июня 2012 г.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Литература, базы данных, информационно-справочные и поисковые системы

1. Макшанов, А. В. Большие данные. Big Data [Электронный ресурс] / А. В. Макшанов, А. Е. Журавлев, Л. Н. Тындыкарь. - Электрон. текстовые дан. - Санкт-Петербург : Лань, 2021. - 188 с. - ЭБС "Лань". - Неогранич. доступ. - ISBN 978-5-8114-6810-2 - Текст : электронный // Лань : электронно-библиотечная система. - URL:

- <https://e.lanbook.com/book/165835> (дата обращения: 05.05.2022). - Режим доступа: для авториз. пользователей.
2. Грас, Джоэл. Data Science. Наука о данных с нуля [Текст] : учеб. пособие / Д. Грас. - СПб. : БХВ - Петербург, 2019. - 336 с. ; 23 см. - ISBN 978-5-9775-3758-2 (30 экз.).
3. Сенько, А. В. Работа с BigData в облаках. Обработка и хранение данных с примерами из Microsoft Azure [Электронный ресурс] / А. В. Сенько. - Электрон. текстовые дан. - Санкт-Петербург : Питер, 2019. - 448 с. : ил. - ЭБС "Айбукс". - Неогранич. доступ. - ISBN 978-5-4461-0578-6 - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://ibooks.ru/reading.php?short=1&productid=359208> (дата обращения: 05.05.2022). - Режим доступа: для авториз. пользователей.
4. Цильковский, И. А. Методы анализа знаний и данных [Электронный ресурс] : уч.-методич. пособие / И. А. Цильковский. - Новосибирск : Изд-во НГТУ, 2010. - 68 с. - Режим доступа: ЭБС <Руконт>.
5. Сапрыкин, О. Н. Интеллектуальный анализ данных : учебное пособие / О. Н. Сапрыкин. - Самара : Самарский университет, 2020. - 80 с. - ISBN 978-5-7883-1563-8. - Текст : электронный // Лань : электронно-библиотечная система. - URL: <https://e.lanbook.com/book/188906> (дата обращения: 05.05.2022). - Режим доступа: для авториз. пользователей.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебная аудитория для проведения:

- занятий лекционного типа,
- занятий семинарского (практического) типа,
- групповых и индивидуальных консультаций,
- текущего контроля и промежуточной аттестации.

Оснащение:

Учебная аудитория, укомплектованная специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории, для проведения занятий лекционного типа, практических занятий (семинарского типа), курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. Для проведения занятий лекционного типа обучающимся предлагаются наборы демонстрационного оборудования и учебно-наглядные пособия, обеспечивающие тематические иллюстрации.

6.2. Помещения для самостоятельной работы обучающихся.

Оснащение:

Помещения для самостоятельной работы обучающихся, оснащенные учебной мебелью. Рабочие места обучающихся оборудованы компьютерной техникой и подключены в локальную вычислительную сеть, в т.ч. с использованием беспроводного Wi-Fi подключения, с возможностью выхода в глобальную сеть Интернет и с доступом в электронную информационно-образовательную среду.

6.3. Программное обеспечение

Приложение для чтения PDF-файлов, браузер для просмотра интернет контента, приложение для создания PDF-файлов.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Оценочные средства для промежуточной аттестации

Список вопросов для промежуточной аттестации:

1. Понятие больших данных, их основные свойства и применения.
2. OLTP- и OLAP-системы.
3. ETL-процессы, основные принципы организации. Поточковая и пакетная обработка информации.
4. Системы хранения данных. Хранилище данных. Озеро данных.
5. Хранилище данных. Трехслойная архитектура хранилища. Подходы к проектированию по Кимбаллу и Инмону.
6. Многомерное проектирование.
7. Подходы к хранению истории изменения данных. SCD.
8. Data Vault. Типы таблиц. Пример.
9. Якорное моделирование. Типы таблиц. Пример.
10. Распределенные модели и данные. Spark ML.