

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Бойко Елена Григорьевна

Должность: Ректор

Дата подписания: 07.05.2024 10:12:38

Уникальный программный ключ:

e69eb689122030af7d22cc354bf0eb9d453ecf8f

Министерство науки и высшего образования РФ

ФГБОУ ВО Государственный аграрный университет Северного Зауралья

Инженерно-технологический институт

Кафедра математики и информатики

«Утверждаю»

И.о. заведующего кафедрой



Н.Н. Мальчукова

«25» мая 2023 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Сбор, предобработка и хранение данных

для направления подготовки 35.03.06 Агроинженерия

образовательные программы: *Технические системы в агробизнесе, Технический сервис в агропромышленном комплексе*

Уровень высшего образования – бакалавриат

Форма обучения: очная, заочная

Тюмень, 2023

При разработке рабочей программы учебной дисциплины в основу положены:

- 1) ФГОС ВО по направлению подготовки (специальности) 35.03.06 Агроинженерия, утвержденный Министерством образования и науки РФ 23 августа 2017 г., приказ № 813.
- 2) Учебные планы основных образовательных программ Технический сервис в агропромышленном комплексе, Технические системы в агробизнесе одобрены Ученым советом ФГБОУ ВО ГАУ Северного Зауралья от 25 мая 2023 г. Протокол № 10.

Рабочая программа учебной дисциплины (модуля) одобрена на заседании кафедры математики и информатики от 25 мая 2023 г. Протокол № 9.1.

И.о. заведующего кафедрой _____  Н.Н. Мальчукова

Рабочая программа учебной дисциплины (модуля) одобрена методической комиссией института от 25 мая 2023 г. Протокол № 7а.

Председатель методической комиссии института _____  О.А. Мелякова

Разработчик:

Ерёмина Д.В., к.с.-х.н., доцент кафедры математики и информатики

Директор института:

_____ 

Н.Н. Устинов

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Код компетенции	Результаты освоения	Индикатор достижения компетенции	Перечень планируемых результатов обучения по дисциплине
УК-1	Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	ИД-2ук-1 Осуществляет анализ и оценку информации, с использованием различных информационных ресурсов для решения поставленных задач	<p><i>знать:</i></p> <ul style="list-style-type: none"> - механизмы и методики поиска, предобработки и хранения данных, включающие системный подход в области образования; <p><i>уметь:</i></p> <ul style="list-style-type: none"> - находить, хранить и критически анализировать информацию, необходимую для решения поставленной задачи; <p><i>владеть:</i></p> <ul style="list-style-type: none"> - механизмами поиска и хранения информации, в том числе с применением современных информационных и коммуникационных технологий.

2. Место дисциплины в структуре образовательной программы

Данная дисциплина относится к *Блоку 1* части, формируемой участниками образовательных отношений.

Для изучения дисциплины необходимы знания в области: математики, информатики и цифровых технологий.

Сбор, предобработка и хранение данных является предшествующей дисциплиной для дисциплин: *Алгоритмы обработки и анализа данных, Машинное обучение и аналитика Big Data для бизнеса.*

Дисциплина изучается на 4 курсе в 7 семестре по очной форме обучения, на 4 курсе в 8 семестре по заочной форме.

3. Объем дисциплины и виды учебной работы

Общая трудоемкость дисциплины составляет 108 часов (3 зачетных единицы).

Вид учебной работы	Форма обучения	
	очная	заочная
Аудиторные занятия (всего)	48	12
<i>В том числе:</i>	-	-
Лекционного типа	24	6
Семинарского типа	24	6
Самостоятельная работа (всего)	60	96
<i>В том числе:</i>	-	-
Проработка материала лекций, подготовка к занятиям	30	72
Самостоятельное изучение тем	6	
Курсовой проект (работа)	-	-
Расчетно-графические работы	-	-
Контрольная работа	-	24
Реферат	24	-
Индивидуальные задания	-	-
Вид промежуточной аттестации:	зачет	зачет
Общая трудоемкость:		
часов	108	108
зачетных единиц	3	3

4. Содержание дисциплины

4.1. Содержание разделов дисциплины

№ п/п	Наименование раздела дисциплины	Содержание раздела
1.	Методы сбора и обработки данных из сети Интернет	Основы компьютерных сетей, клиент-серверного взаимодействия. Работа с HTTP, открытыми данными. Парсинг. Архитектура системы обработки больших данных (прием, сбор, анализ и представление результатов).
2.	Обзор технологий хранения Больших данных	Свойства больших данных и ограничения RDBMS. ACID требования. CAP-теорема, BASE архитектура. NoSQL. Интерфейсы. Технология распределённых вычислений MapReduce. Типы NoSQL: ключ-заключение, колоночные, документо-ориентированные, графовые.
3	Программные платформы и системы для Больших данных	Системы управления потоками данных. Системы хранения Больших данных. Платформы Больших данных (экосистема Hadoop, Spark). Обработка данных в реальном времени. Системы управления Большими данными. Аналитические платформы. Оборудование для Больших данных. Центры обработки Больших данных.
4.	Основы реляционных баз данных. MySQL	Проектирование реляционных баз данных на основе принципов нормализации: первые шаги нормализации, дальнейшая нормализация. Проектирование реляционных баз данных с использованием семантических моделей: ER-диаграммы. Хранилища данных для анализа. MySQL (Управление БД. Язык запросов SQL. Введение в проектирование БД. Операторы, фильтрация, сортировка и ограничение. Агрегация данных. Сложные запросы).

4.2. Разделы дисциплины и виды занятий

очная форма обучения

№ п/п	Наименование раздела дисциплины	Лекционного типа	Семинарского типа	СР	Всего, часов
1	2	3	4	5	6
1.	Методы сбора и обработки данных из сети Интернет	8	4	15	27
2.	Обзор технологий хранения Больших данных	8	-	15	23
3.	Программные платформы и системы для Больших данных	4	2	15	21
4.	Основы реляционных баз данных. MySQL	4	18	15	37
	Итого:	24	24	60	108

заочная форма обучения

№ п/п	Наименование раздела дисциплины	Лекционного типа	Семинарского типа	СР	Всего, часов
1	2	3	4	5	6
1.	Методы сбора и обработки данных из сети Интернет	-	2	24	26
2.	Обзор технологий хранения Больших данных	2	-	24	26
3.	Программные платформы и системы для Больших данных	2	-	24	26
4.	Основы реляционных баз данных. MySQL	2	4	24	30
	Итого:	6	6	96	108

4.3. Занятия семинарского типа

№ п/п	№ раздела дисциплины	Тема	Трудоемкость (час)	
			очная	заочная
1	2	3	4	5
1.	1	MS Excel. Сбор и подготовка данных.	2	2
2.	1	MS Excel. Бинарная и множественная классификация данных.	2	-
3.	3	Интернет. Знакомство с программными платформами и системами для Больших данных.	2	-
4.	4	Основы работы в MySQL. Проектирование базы данных. Нормализация.	2	2
5.	4	MySQL. Выборка данных - оператор SELECT.	2	-
6.	4	MySQL. Запросы. Уточнения запросов.	2	2
7.	4	MySQL. Агрегированные функции. Строковые и числовые функции.	2	-
8.	4	MySQL. Условные выражения с оператором CASE .	2	-
9.	4	MySQL. Простые и связанные подзапросы.	2	-
10.	4	MySQL. Операции соединения.	2	-

11.	4	MySQL. Представления.	2	-
12.	4	Знакомство с дополнительными функциями MySQL.	2	-
		Итого:	24	6

4.4. Примерная тематика курсовых проектов (работ) - Не предусмотрено ОПОП

5. Организация самостоятельной работы обучающихся по дисциплине

5.1. Типы самостоятельной работы и её контроль

Тип самостоятельной работы	Форма обучения		Текущий контроль
	очная	заочная	
Проработка материала лекций, подготовка к занятиям	30	- 72	Тестирование
Самостоятельное изучение тем	6		Тестирование
Контрольная работа	-	24	Защита контрольной работы
Реферат	24	-	Защита реферата
всего часов:	60	96	

5.2. Учебно-методические материалы для самостоятельной работы:

1. Сбор, предобработка и хранение данных [Электронный ресурс]: Методические указания к выполнению самостоятельной работы для направления подготовки 35.03.06 Агроинженерия / Автор-сост. Д.В. Ерёмкина. - Тюмень: ГАУ Северного Зауралья, 2023 – 12 с.

2. Сбор, предобработка и хранение данных [Электронный ресурс]: Методические указания к выполнению контрольной работы обучающимися заочной формы по направлению подготовки 35.03.06 Агроинженерия / Автор-сост. Д.В. Ерёмкина. – Тюмень: ГАУ Северного Зауралья, 2023. – 12 с.

3. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу: учебное пособие / Ф. О. Федин, Ф. Ф. Федин. - Москва: Московский городской педагогический университет, 2012. - 204 с. - ISBN 2227-8397. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/26444.html>- Режим доступа: для авторизир. пользователей.

5.3. Темы, выносимые на самостоятельное изучение:

Раздел 2. Обзор технологий хранения Больших данных.

1. Реляционные хранилища данных.
2. Многомерные хранилища данных.
3. Гибридные хранилища данных.
4. Виртуальные хранилища данных.

5.4. Темы рефератов:

1. Ключевые элементы технологии Data Warehousing:
 - 1.1. Концептуальная архитектура систем на основе Data Warehousing.
 - 1.2. Практическое использование понятия «интеллект бизнеса» в технологии Data Warehousing.
 - 1.3. Концепция целевого распространения информации.

- 1.4. Основные свойства хранимых и обрабатываемых данных.
- 1.5. Компоненты Хранилища данных.
- 1.6. Разные категории программных систем (хранилище данных, витрина данных).
- 1.7. Связь технологии Хранилища данных.
- 1.8. Связь технологии Хранилищ данных с ERP- и CRM-системами.
2. Концептуальное моделирование информационных потребностей в технологии Хранилищ данных:
 - 2.1. Общие методы структурирования информации в процессе принятия решений.
 - 2.2. Информационное моделирование корпоративных данных.
 - 2.3. Бизнес-размерности и бизнес-метрики.
 - 2.4. Многомерный анализ размерностей.
 - 2.5. Использование методологии совместной разработки.
 - 2.6. Способы анализа существующей документации.
 - 2.7. Выработка требований к отдельным компонентам.
 - 2.8. Хранилища данных на основе информационных потребностей.
3. Обзор архитектуры системы принятия решений:
 - 3.1. Архитектура СППР на основе технологии Хранилища данных.
 - 3.2. Понятие интеграции приложений предприятия.
 - 3.3. Назначение и принципы интеграции основных подсистем: исходные данные, трансформация данных, хранение данных, распространение данных.
 - 3.4. Особенности архитектуры распределительных СППР.
 - 3.5. Принципы выбора программных реализаций для подсистем.
 - 3.6. Примеры существующих программных решений.
4. Представление и обработка метаданных в системах поддержки принятия решений:
 - 4.1. Понятие метаданных и их роль в системе поддержки принятия решений.
 - 4.2. Разновидности метаданных по функциональным областям и их использование в деловой деятельности.
 - 4.3. Пассивные и активные метаданные.
 - 4.4. Извлечение и использование метаданных в процессе разработки системы.
 - 4.5. Стандарты на представление и обработку метаданных.
 - 4.6. Инициатива CWM – объектно-ориентированная стандартизация метаданных на основе UML.
5. Информационное моделирование в технологии Data Warehousing:
 - 5.1. Специфические характеристики аналитических моделей данных.
 - 5.2. Взаимосвязь с реляционной моделью данных.
 - 5.3. Таблицы фактов и таблицы размерностей.
 - 5.4. Процесс перехода от информационных требований к разработке модели данных.
 - 5.5. STAR-схема.
 - 5.6. Методы информационного моделирования событий и темпоральных характеристик.
 - 5.7. Методы определения ключевых атрибутов.
 - 5.8. Реализация обновлений в таблицах размерностей.
 - 5.9. SNOWFLAKE-схема.
 - 5.10. Таблицы агрегированных фактов.
6. Загрузка, верификация и очистка данных в системах поддержки принятия решений:
 - 6.1. Назначение и основные требования к подсистеме загрузки, верификации и очистки данных (ETL).
 - 6.2. Способы извлечения данных из операционных БД и внешних и информационных источников.
 - 6.3. Методы трансформации данных.
 - 6.4. Известные программные средства верификации и очистки данных.
 - 6.5. Стандартизация загрузки, верификации и очистки данных.

- 6.6. Понятие качества данных.
- 6.7. Методы и средства повышения качества исходных данных.
- 7. Распространение и представление информации в системах поддержки принятия решений:
 - 7.1. Основные черты принципов целевого распространения информации в СППР на основе технологии Хранилищ данных.
 - 7.2. Виды информационного содержания.
 - 7.3. Категории пользователей системы и их информационные требования.
 - 7.4. Типы программных средств, используемых для распространения информации.
- 8. Пост роение и использование систем на основе технологии OLAP:
 - 8.1. Основные отличия технологии OLAP (On Line Analytical Processing) от традиционных способов анализа данных.
 - 8.2. Инструменты и методы аналитической обработки данных: гиперкуб, свертка, поворот, конкретизация, получение слоя.
 - 8.3. Различные модели аналитической обработки данных: MOLAP, ROLAP, HOLAP.
 - 8.4. Методы реализации OLAP-систем.
- 9. Технология DataMining (часть 1):
 - 9.1. Основные понятия технологии.
 - 9.2. Специфика анализируемых данных: разреженные, отсутствующие и неточные данные.
 - 9.3. Разведочный анализ данных, анализ тенденций.
 - 9.4. Предсказательные и описательные модели.
 - 9.5. Преимущества использования технологии DataMining при выработке стратегических решений и примеры из реальных прикладных областей.
- 10. Технология DataMining (часть 2):
 - 10.1. Математические основы основных классов методов: таблицы и деревья решений, ассоциативные правила, иерархическая, агломеративная и дивизионная кластеризация, иерархические методы классификации, нейросетевые алгоритмы, генетические алгоритма.
 - 10.2. Методы применения статистического моделирования.
 - 10.3. Кластеризация с использованием теории множеств.
- 11. Технология DataMining (часть 3):
 - 11.1. Известные схемы и алгоритмы предварительной подготовки и преобразования данных: выбор атрибутов, дискретизация, многофакторный анализ.
 - 11.2. Основные этапы поддержки принятия решений при использовании деревьев решений, ассоциативных правил и алгоритмов кластеризации.

6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

6.1 Перечень компетенций и оценочные средства индикатора достижения компетенций

Код компетенции	Индикатор достижения компетенции	Перечень планируемых результатов обучения по дисциплине	Наименование оценочного средства
УК-1	ИД-2 _{УК-1} Осуществляет анализ и оценку информации, с использованием различных информационных ресурсов для решения поставленных задач	<i>знать:</i> - механизмы и методики поиска, предобработки и хранения данных, включающие системный подход в области образования; <i>уметь:</i> - находить, хранить и критически анализировать информацию, необходимую для решения поставленной задачи; <i>владеть:</i> - механизмами поиска и хранения информации, в том числе с применением современных информационных и коммуникационных технологий.	Тест

6.2. Шкалы оценивания

Шкала оценивания тестирования на зачете

% выполнения задания	Результат
50 – 100	зачтено
менее 50	не зачтено

6.4. Типовые контрольные задания или иные материалы:

Указаны в приложении 1.

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

а) основная литература

1. Железнов, М. М. Методы и технологии обработки больших данных: учебно-методическое пособие / М. М. Железнов. - Москва: МИСИ-МГСУ, ЭБС АСВ, 2020. - 46 с. - ISBN 978-5-7264-2193-3. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/101802.html> - Режим доступа: для авторизир. пользователей.

2. Кожевникова, П. В. PHP и MySQL: учебное пособие / П. В. Кожевникова. - Ухта: УГТУ, 2020. - 51 с.- Текст: электронный // Лань: электронно-библиотечная система. - URL: <https://e.lanbook.com/book/209591> - Режим доступа: для авториз. пользователей.

б) дополнительная литература

3. Григорьев, Ю. А. Реляционные базы данных и системы NoSQL: учебное пособие / Ю. А. Григорьев, А. Д. Плутенко, О. Ю. Плужникова. - Благовещенск: Амурский государственный университет, 2018. - 425 с. - ISBN 978-5-93493-308-2. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/103912.html> - Режим доступа: для авторизир. пользователей.

4. Федин, Ф. О. Анализ данных. Часть 1. Подготовка данных к анализу: учебное пособие / Ф. О. Федин, Ф. Ф. Федин. - Москва: Московский городской педагогический университет, 2012. - 204 с. - ISBN 2227-8397. - Текст: электронный // Электронно-библиотечная система IPR BOOKS: [сайт]. - URL: <https://www.iprbookshop.ru/26444.html> - Режим доступа: для авторизир. пользователей.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

<http://www.intuit.ru/> - Национальный открытый университет «Интуит».

<http://www.planetaexcel.ru> – сайт о возможностях Excel.

<http://office.microsoft.com/ru-ru> - сайт фирмы Microsoft.

9. Методические указания для обучающихся по освоению дисциплины

1. Сбор, предобработка и хранение данных [Электронный ресурс]: Методические указания к практическим занятиям для направления подготовки 35.03.06 Агроинженерия / Автор-сост. Д.В. Ерёмина. – Тюмень: ГАУ Северного Зауралья, 2022. - 48 с.

10. Перечень информационных технологий

1. Microsoft Windows 8 Professional; OpenSUSE Linux;
2. Microsoft Office Standard 2010;
3. Система электронного обучения Moodle;
4. MySQL 5.6.

11. Материально-техническое обеспечение дисциплины

Для проведения практических занятий по дисциплине «Сбор, предобработка и хранение данных» используется компьютерный класс (14-15 компьютеров) с установленным программным обеспечением.

Лекционные занятия проводятся в учебных аудиториях ФГБОУ ВО ГАУ Северного Зауралья с мультимедийным оборудованием (проектор и/или интерактивная доска).

Для выполнения самостоятельной работы обучающиеся могут пользоваться читальными залами библиотеки ФГБОУ ВО ГАУ Северного Зауралья, в том числе оснащёнными компьютерами с локальной сетью и выходом в интернет.

12. Особенности освоения дисциплины для инвалидов и лиц с ограниченными возможностями здоровья

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

В целях освоения учебной программы дисциплины инвалидами и лицами с ограниченными возможностями здоровья обеспечивается:

- для инвалидов и лиц с ограниченными возможностями здоровья по зрению: размещение в доступных для обучающихся, являющихся слепыми или слабовидящими, местах и в адаптированной форме справочной информации о расписании учебных занятий; присутствие ассистента, оказывающего обучающемуся необходимую помощь; выпуск альтернативных форматов методических материалов (крупный шрифт или аудиофайлы), использование версии сайта для слабовидящих ЭБС IPR BOOKS и специального мобильного приложения IPR BOOKS WV-Reader (программы не визуального доступа к информации, предназначенной для мобильных устройств, работающих на операционной системе Android и iOS, которая не требует специально обученного ассистента, т.к. люди с ОВЗ по зрению работают со своим устройством привычным способом, используя специальные штатные программы для незрячих людей, с которыми IPR BOOKS WV-Reader имеет полную совместимость);

- для инвалидов и лиц с ограниченными возможностями здоровья по слуху: надлежащими звуковыми средствами воспроизведение информации;

- для инвалидов и лиц с ограниченными возможностями здоровья, имеющих нарушения опорно-двигательного аппарата: возможность беспрепятственного доступа обучающихся в учебные помещения, туалетные комнаты и другие помещения кафедры, а также пребывание в указанных помещениях.

Образование обучающихся с ограниченными возможностями здоровья может быть организовано как совместно с другими обучающимися, так и в отдельных группах или в отдельных организациях.

Министерство науки и высшего образования РФ
ФГБОУ ВО Государственный аграрный университет Северного Зауралья
Инженерно-технологический институт
Кафедра математики и информатики

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по учебной дисциплине «Сбор, предобработка и хранение данных»

для направления подготовки 35.03.06 Агроинженерия

образовательные программы: *Технические системы в агробизнесе,
Технический сервис в агропромышленном комплексе*

Уровень высшего образования – бакалавриат


Форма обучения: очная, заочная

Разработчик:

Ерёмина Д.В., к.с.-х.н., доцент кафедры математики и информатики

Утверждено на заседании кафедры

протокол № 9.1 от 25мая 2023 г.

И.о. заведующего кафедрой  Н.Н. Мальчукова

Тюмень, 2023

**КОНТРОЛЬНЫЕ ЗАДАНИЯ И ИНЫЕ МАТЕРИАЛЫ ОЦЕНКИ
знаний, умений, навыков и (или) опыта деятельности, характеризующие
этапы формирования компетенций в процессе освоения дисциплины
творческих заданий / проектов
СБОР, ПРЕДОБРАБОТКА И ХРАНЕНИЕ ДАННЫХ**

Вопросы к зачёту

Компетенция	Вопросы
<p>УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</p>	<ol style="list-style-type: none"> 1. Определение «источник данных». 2. Понятие «открытые данные». 3. Примеры источников открытых данных. 4. Назначение протокола HTTP. 5. Основные компоненты сети. 6. Технология клиент сервер. 7. Свойства Больших данных. 8. Ограничения RDBMS. 9. ACID требования. 10. NoSQL. Принципиальное отличие от MySQL. 11. Определение «парсинг». 12. Уровни систем обработки Больших данных. 13. Суть приёма данных от источников. 14. Пример полуструктурированных данных. 15. Задачи, которые решает анализ Больших данных. 16. Назначение слоя / уровня компонентов системы «Прием данных». 17. Цели использования слоя / уровня компонентов системы «Сбор данных». 18. Цели использования слоя / уровня компонентов системы «Анализ данных». 19. Цели использования слоя / уровня компонентов системы «Представление результата». 20. Технология распределённых вычислений MapReduce. 21. Типы NoSQL. 22. Языки программирования для работы с фреймворками данных. 23. Позволяет ли лицензия Apache 2.0, под которой выпущены некоторые фреймворки, вносить собственные исправления в код программного обеспечения? 24. Примеры фреймворков, обеспечивающих обработку данных в реальном времени. 25. Примеры фреймворков, обеспечивающих аналитическую обработку данных. 26. Примеры фреймворков, обеспечивающих хранение данных. 27. Примеры фреймворков, обеспечивающих управление потоками данных. 28. Реляционные хранилища данных. 29. Многомерные хранилища данных. 30. Гибридные хранилища данных. 31. Виртуальные хранилища данных. 32. Оборудование для обработки Больших данных. 33. Существенные параметры для вычислительных узлов кластера.

	<p>34. Существенные параметры для сетевой инфраструктуры.</p> <p>35. Существенные параметры для системы хранения данных.</p> <p>36. Максимальное количество узлов, которое может иметь кластер, использующий архитектуру подключения узлов гиперкуб, при наличии четырех быстродействующих сетевых портов в каждом узле.</p> <p>37. Центр обработки данных уровня, обеспечивающего максимальную надежность.</p> <p>38. Центр обработки данных уровня позволяющий проводить обслуживание оборудования одновременно с обработкой данных.</p> <p>39. Центр обработки данных уровня, обеспечивающий резервирование.</p> <p>40. Реляционная модель данных.</p> <p>41. Нормализация данных в MySQL.</p> <p>42. Типы данных в MySQL.</p> <p>43. Понятие «представление» в MySQL.</p> <p>44. Выборка данных. Оператор Select.</p>
--	---

Обучающийся допускается к промежуточной аттестации по дисциплине при условии выполнения всех практических работ, реферата (очная форма), контрольной работы (заочная форма), выполнения на положительные оценки тестов по лекционному материалу и темам, вынесенным на самостоятельное обучение.

Тестовое задание в системе электронного обучения Moodle включает 30 вопросов, в случайном порядке выбранных из банка вопросов. Обучающемуся предоставляется 2 попытки, по 45 минут каждая.

Критерии оценки:

- «зачтено» выставляется обучающемуся, если итоговое тестирование в системе электронного обучения Moodle выполнено с результатом 50% и выше;
- «не зачтено» выставляется обучающемуся, если итоговое тестирование в системе электронного обучения Moodle выполнено с результатом 49% и ниже.

Варианты заданий для контрольной работы

Вариант № 1

1. Технология клиент сервер.
2. Задачи, которые решает анализ Больших данных.
3. Системы обработки Больших данных. Этап «Сбор данных (Data Staging)».
4. Платформа Больших данных Hadoop.
5. Нормализация данных в MySQL.

Вариант №2

1. Основные компоненты сети.
2. Технология распределённых вычислений MapReduce.
3. Системы обработки Больших данных. Этап «Представление результатов (Consumption Layer)».
4. Платформа Больших данных Spark.
5. Реляционная модель данных.

Вариант №3

1. Назначение протокола HTTP.
2. Типы NoSQL.

3. Системы обработки Больших данных. Этап «Анализ данных (Analysis Layer)».
4. Платформа Больших данных Solr.
5. Понятие «представление» в MySQL.

Вариант №4

1. Открытые данные. Источники открытых данных.
2. Языки программирования для работы с фреймворками данных.
3. Система управления потоками данных Apache Kafka.
4. Платформа Больших данных Windows Azure HDInsight.
5. Типы данных в MySQL.

Вариант №5

1. Свойства Больших данных.
2. Оборудование для обработки Больших данных.
3. Система управления потоками данных Niagara Files (NiFi).
4. Система обработки данных в реальном времени Apache Storm.
5. Выборка данных в MySQL. Оператор Select.

Вариант №6

1. Ограничения RDBMS.
2. Примеры фреймворков, обеспечивающих управление потоками данных.
3. Система управления потоками данных Flume.
4. Система обработки данных в реальном времени Apache Spark.
5. Компьютерные системы хранения (на примере MySQL).

Вариант №7

1. ACID требования.
2. Примеры фреймворков, обеспечивающих хранение данных.
3. Система хранения Больших данных HDFS (Hadoop Distributed File System).
4. Система управления Большими данными ZooKeeper.
5. Объектно-ориентированные базы данных.

Вариант №8

1. NoSQL. Принципиальное отличие от MySQL.
2. Примеры фреймворков, обеспечивающих аналитическую обработку данных.
3. Система хранения Больших данных OpenStack Swift.
4. Система управления Большими данными Atlas.
5. Объектно-реляционные базы данных.

Вариант №9

1. Уровни систем обработки Больших данных.
2. Примеры фреймворков, обеспечивающих обработку данных в реальном времени.
3. Система хранения Больших данных Cassandra.
4. Аналитическая платформа RapidMiner.
5. Типы данных даты и времени в MySQL.

Вариант №10

1. Парсинг.
2. Системы обработки Больших данных. Этап «Прием данных (Data Ingestion)».
3. Система хранения Больших данных Apache Drill.
4. Аналитическая платформа STATISTICA Data Miner.
5. Логические операторы в MySQL.

Контрольная работа выполняется обучающимися заочной формы. Сдается на проверку преподавателю в период сессии.

При оценке контрольной работы определяются выдержанность требований к оформлению контрольной работы, соответствие содержания заданному варианту, полнота и последовательность изложения, наличие достаточных пояснений, число и характер

ошибок (существенные или несущественные), а также ответы на уточняющие вопросы преподавателя по содержанию контрольной.

Критерии оценки:

- оценка «зачтено» выставляется обучающемуся, если контрольная работа выполнена по своему варианту, допущено по каждому заданию по одной несущественной ошибке и в одном задании допущена одна существенная ошибка, приведены схемы, таблицы и рисунки, требующие эти пояснения по работе.
- оценка «не зачтено» выставляется обучающемуся, если контрольная работа выполнена не по своему варианту, допущено по половине заданий по одной существенной ошибке, не приведены схемы, таблицы и рисунки, требующие эти пояснения по работе.

Темы рефератов

1. Ключевые элементы технологии Data Warehousing:
 - 1.1. Концептуальная архитектура систем на основе Data Warehousing.
 - 1.2. Практическое использование понятия «интеллект бизнеса» в технологии Data Warehousing.
 - 1.3. Концепция целевого распространения информации.
 - 1.4. Основные свойства хранимых и обрабатываемых данных.
 - 1.5. Компоненты Хранилища данных.
 - 1.6. Разные категории программных систем (хранилище данных, витрина данных).
 - 1.7. Связь технологии Хранилища данных.
 - 1.8. Связь технологии Хранилищ данных с ERP- и CRM-системами.
2. Концептуальное моделирование информационных потребностей в технологии Хранилищ данных:
 - 2.1. Общие методы структурирования информации в процессе принятия решений.
 - 2.2. Информационное моделирование корпоративных данных.
 - 2.3. Бизнес-размерности и бизнес-метрики.
 - 2.4. Многомерный анализ размерностей.
 - 2.5. Использование методологии совместной разработки.
 - 2.6. Способы анализа существующей документации.
 - 2.7. Выработка требований к отдельным компонентам.
 - 2.8. Хранилища данных на основе информационных потребностей.
3. Обзор архитектуры системы принятия решений:
 - 3.1. Архитектура СППР на основе технологии Хранилища данных.
 - 3.2. Понятие интеграции приложений предприятия.
 - 3.3. Назначение и принципы интеграции основных подсистем: исходные данные, трансформация данных, хранение данных, распространение данных.
 - 3.4. Особенности архитектуры распределительных СППР.
 - 3.5. Принципы выбора программных реализаций для подсистем.
 - 3.6. Примеры существующих программных решений.
4. Представление и обработка метаданных в системах поддержки принятия решений:
 - 4.1. Понятие метаданных и их роль в системе поддержки принятия решений.
 - 4.2. Разновидности метаданных по функциональным областям и их использование в деловой деятельности.
 - 4.3. Пассивные и активные метаданные.
 - 4.4. Извлечение и использование метаданных в процессе разработки системы.
 - 4.5. Стандарты на представление и обработку метаданных.
 - 4.6. Инициатива CWM – объектно-ориентированная стандартизация метаданных на

основе UML.

5. Информационное моделирование в технологии Data Warehousing:
 - 5.1. Специфические характеристики аналитических моделей данных.
 - 5.2. Взаимосвязь с реляционной моделью данных.
 - 5.3. Таблицы фактов и таблицы размерностей.
 - 5.4. Процесс перехода от информационных требований к разработке модели данных.
 - 5.5. STAR-схема.
 - 5.6. Методы информационного моделирования событий и темпоральных характеристик.
 - 5.7. Методы определения ключевых атрибутов.
 - 5.8. Реализация обновлений в таблицах размерностей.
 - 5.9. SNOWFLAKE-схема.
 - 5.10. Таблицы агрегированных фактов.
6. Загрузка, верификация и очистка данных в системах поддержки принятия решений:
 - 6.1. Назначение и основные требования к подсистеме загрузки, верификации и очистки данных (ETL).
 - 6.2. Способы извлечения данных из операционных БД и внешних и информационных источников.
 - 6.3. Методы трансформации данных.
 - 6.4. Известные программные средства верификации и очистки данных.
 - 6.5. Стандартизация загрузки, верификации и очистки данных.
 - 6.6. Понятие качества данных.
 - 6.7. Методы и средства повышения качества исходных данных.
7. Распространение и представление информации в системах поддержки принятия решений:
 - 7.1. Основные черты принципов целевого распространения информации в СППР на основе технологии Хранилищ данных.
 - 7.2. Виды информационного содержания.
 - 7.3. Категории пользователей системы и их информационные требования.
 - 7.4. Типы программных средств, используемых для распространения информации.
8. Построение и использование систем на основе технологии OLAP:
 - 8.1. Основные отличия технологии OLAP (On Line Analytical Processing) от традиционных способов анализа данных.
 - 8.2. Инструменты и методы аналитической обработки данных: гиперкуб, свертка, поворот, конкретизация, получение слоя.
 - 8.3. Различные модели аналитической обработки данных: MOLAP, ROLAP, HOLAP.
 - 8.4. Методы реализации OLAP-систем.
9. Технология DataMining (часть 1):
 - 9.1. Основные понятия технологии.
 - 9.2. Специфика анализируемых данных: разреженные, отсутствующие и неточные данные.
 - 9.3. Разведочный анализ данных, анализ тенденций.
 - 9.4. Предсказательные и описательные модели.
 - 9.5. Преимущества использования технологии DataMining при выработке стратегических решений и примеры из реальных прикладных областей.
10. Технология DataMining (часть 2):
 - 10.1. Математические основы основных классов методов: таблицы и деревья решений, ассоциативные правила, иерархическая, агломеративная и дивизионная кластеризация, иерархические методы классификации, нейросетевые алгоритмы, генетические алгоритмы.
 - 10.2. Методы применения статистического моделирования.
 - 10.3. Кластеризация с использованием теории множеств.

11. Технология DataMining (часть 3):

- 11.1. Известные схемы и алгоритмы предварительной подготовки и преобразования данных: выбор атрибутов, дискретизация, многофакторный анализ.
- 11.2. Основные этапы поддержки принятия решений при использовании деревьев решений, ассоциативных правил и алгоритмов кластеризации.

Реферат предоставляется преподавателю в электронном виде. Преподаватель оценивает качество изложения материала, выдержано ли форматирование по заданным требованиям, наличие автоматического оглавления. Студенту задаются вопросы по теме реферата и по оформлению работы. После защиты из всех перечисленных выше факторов формируется оценка.

Вопросы к защите реферата

1. Актуальность выбранной темы.
2. Исследовательский компонент в работе.
3. Вывод по данным реферата.
4. Личный интерес к выбранной теме реферата.
5. Точка зрения по основным позициям реферата.

Обучающийся предоставляет преподавателю реферат в электронном виде и отвечает на вопросы, касающиеся содержания и оформления реферата.

Критерии оценки:

- оценка «зачтено» выставляется обучающемуся, если тема соответствует варианту, раскрыта, текст отформатирован, даны ответы на вопросы преподавателя;
- оценка «не зачтено» выставляется обучающемуся, если тема не соответствует варианту или полностью не раскрыта, отсутствует форматирование, обучающийся затрудняется ответить на вопросы преподавателя.