

**МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ**
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Воронежский государственный технический университет»

УТВЕРЖДАЮ
Декан факультета экономики менеджмента и
информационных технологий
С.А.Баркалов
«30» августа 2017 г.

РАБОЧАЯ ПРОГРАММА
дисциплины
«Технологии обработки информации»

Направление подготовки 09.03.02 ИНФОРМАЦИОННЫЕ СИСТЕМЫ И
ТЕХНОЛОГИИ

Профиль Информационные системы и технологии строительстве

Квалификация выпускника Бакалавр

Нормативный период обучения 4 года

Форма обучения очная

Год начала подготовки 2017

Автор программы _____ /Курипта О.В./

Заведующий кафедрой
Информационных технологий
и автоматизированного
проектирования в
строительстве _____ /Смольянинов А.В./

Руководитель ОПОП _____ /Курипта О.В./

Воронеж 2017

1. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

1.1. Цели дисциплины

Изучение принципов организации и оперирования большими объемами данных с применением современных информационных средств и технологий.

1.2. Задачи освоения дисциплины

- изучение архитектуры и возможностей построения хранилищ данных;
- рассмотрение инструментов создания и работы с хранилищами данных;
- получение практических навыков работы с аналитическими инструментами Microsoft;
- изучение принципов работы с большими объемами данных на основе NoSQL подхода.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Технологии обработки информации» относится к дисциплинам базовой части блока Б1.

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Процесс изучения дисциплины «Технологии обработки информации» направлен на формирование следующих компетенций:

ОПК-5 - способностью использовать современные компьютерные технологии поиска информации для решения поставленной задачи, критического анализа этой информации и обоснование принятых идей и подходов к решению

ПК-24 - способностью обосновывать правильность выбранной модели, сопоставляя результаты экспериментальных данных и полученных решений

Компетенция	Результаты обучения, характеризующие сформированность компетенции
ОПК-5	знать особенности обработки больших объемов данных; типы основных источников производственных данных; категории ошибок, возникающих при интеграции данных в единое хранилище; способы реализации нереляционных моделей данных.
	уметь определять параметры ETL процесса преобразования информации; выбирать способ реализации нереляционной базы данных, удовлетворяющий задачам связанной информационной системы.
	владеть навыками реализации NoSQL баз данных; инструментами извлечения, преобразования и загрузки данных в хранилища данных.
ПК-24	знать принципы построения хранилищ данных и этапы работы с ними; многомерную модель данных и способы работы с ней.
	уметь проектировать многомерные кубы данных; организовывать доступ к хранилищам данных; выполнять

	очистку данных в процессе загрузки в хранилище.
	владеть навыками проектирования моделей хранилищ данных на основе различных топологий; средствами представления результатов анализа.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины «Технологии обработки информации» составляет 5 з.е.

Распределение трудоемкости дисциплины по видам занятий
очная форма обучения

Виды учебной работы	Всего часов	Семестры
		4
Аудиторные занятия (всего)	54	54
В том числе:		
Лекции	18	18
Лабораторные работы (ЛР)	36	36
Самостоятельная работа	90	90
Часы на контроль	36	36
Виды промежуточной аттестации - экзамен	+	+
Общая трудоемкость:		
академические часы	180	180
зач.ед.	5	5

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

5.1 Содержание разделов дисциплины и распределение трудоемкости по видам занятий

очная форма обучения

№ п/п	Наименование темы	Содержание раздела	Лекц	Лаб. зан.	СРС	Всего, час
1	Хранилища данных	Введение. Информация, данные, знания. Технологии обработки информации. Эволюция информационных систем. Общие сведения об анализе данных. Хранилища и витрины данных. Варианты реализации хранилищ данных. Компоненты хранилища данных.	6	8	20	34
2	Оперативный анализ данных (OLAP)	Многомерные кубы, основные определения. Действия, выполняемые над многомерными данными. Таблицы фактов и измерений. Топология хранилищ данных «звезда» и «снежинка». Серверные и клиентские средства OLAP. Типы реализаций OLAP.	6	12	26	44
3	Работа с данными в хранилище	Источники данных, типы и способы работы с ними. Извлечение, преобразование и загрузка данных (ETL). Очистка данных. Обогащение данных.	4	8	18	38
4	Нереляционные модели данных (NoSQL)	Понятие больших данных, подходы к работе с ними. Нереляционные СУБД, особенности NoSQL подхода, преимущества и недостатки. Способы реализации NoSQL хранилищ: ключ-значение, документоориентированные, колоночные, основанные на графах.	2	8	18	28
Итого			18	36	90	144

5.2 Перечень лабораторных работ

1. Работа со сводными таблицами в Microsoft Excel;
2. Проектирование структуры хранилища данных;
3. Создание базы данных, являющейся источником для хранилища;
4. Создание хранилища данных средствами Microsoft SQL Server Analysis Services;
5. Доступ к хранилищу данных средствами Microsoft Excel;
6. Создание отчетов средствами Microsoft SQL Server Reporting Services;
7. Разработка собственных приложений, осуществляющих доступ к хранилищу данных;
8. Проектирование и реализация нереляционной базы данных.

6. ПРИМЕРНАЯ ТЕМАТИКА КУРСОВЫХ ПРОЕКТОВ (РАБОТ) И КОНТРОЛЬНЫХ РАБОТ

В соответствии с учебным планом освоение дисциплины не предусматривает выполнение курсового проекта (работы) или контрольной работы.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

7.1.1 Этап текущего контроля

Результаты текущего контроля знаний и межсессионной аттестации оцениваются по следующей системе:

«аттестован»;

«не аттестован».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Аттестован	Не аттестован
ОПК-5	знать особенности обработки больших объемов данных; типы основных источников производственных данных; категории ошибок, возникающих при интеграции данных в единое хранилище; способы реализации нереляционных моделей данных.	Активное участие в устном опросе Выполнение лабораторного практикума	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь определять параметры ETL процесса преобразования информации; выбирать способ реализации нереляционной базы данных, удовлетворяющий	Выполнение лабораторного практикума	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

	задачам связанной информационной системы.			
	владеть навыками реализации NoSQL баз данных; инструментами извлечения, преобразования и загрузки данных в хранилища данных.	Выполнение лабораторного практикума	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
ПК-24	знать принципы построения хранилищ данных и этапы работы с ними; многомерную модель данных и способы работы с ней.	Активное участие в устном опросе Выполнение лабораторного практикума	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	уметь проектировать многомерные кубы данных; организовывать доступ к хранилищам данных; выполнять очистку данных в процессе загрузки в хранилище.	Выполнение лабораторного практикума	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах
	владеть навыками проектирования моделей хранилищ данных на основе различных топологий; средствами представления результатов анализа.	Выполнение лабораторного практикума	Выполнение работ в срок, предусмотренный в рабочих программах	Невыполнение работ в срок, предусмотренный в рабочих программах

7.1.2 Этап промежуточного контроля знаний

Результаты промежуточного контроля знаний оцениваются в 4 семестре для очной формы обучения по четырехбалльной системе:

«отлично»;

«хорошо»;

«удовлетворительно»;

«неудовлетворительно».

Компетенция	Результаты обучения, характеризующие сформированность компетенции	Критерии оценивания	Отлично	Хорошо	Удовл.	Неудовл.
ОПК-5	знать особенности обработки больших объемов данных; типы основных источников производственных данных; категории ошибок, возникающих при интеграции данных в единое хранилище; способы реализации нереляционных	Опрос	Получен развернутый ответ на все вопросы билета и дополнительные вопросы	Получен ответ на все вопросы билета и большую часть дополнительных вопросов	Получен ответ на один вопросов билета и большую часть дополнительных вопросов	Отсутствие ответов на основные вопросы билета и дополнительные вопросы

	моделей данных.					
	уметь определять параметры ETL процесса преобразования информации; выбирать способ реализации нереляционной базы данных, удовлетворяющий задачам связанной информационной системы.	Выполнение лабораторного практикума	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	владеть навыками реализации NoSQL баз данных; инструментами извлечения, преобразования и загрузки данных в хранилища данных.	Выполнение лабораторного практикума	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
ПК-24	знать принципы построения хранилищ данных и этапы работы с ними; многомерную модель данных и способы работы с ней.	Опрос	Получен развернутый ответ на все вопросы билета и дополнительные вопросы	Получен ответ на все вопросы билета и большую часть дополнительных вопросов	Получен ответ на один вопросов билета и большую часть дополнительных вопросов	Отсутствие ответов на основные вопросы билета и дополнительные вопросы
	уметь проектировать многомерные кубы данных; организовывать доступ к хранилищам данных; выполнять очистку данных в процессе загрузки в хранилище.	Выполнение лабораторного практикума	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены
	владеть навыками проектирования моделей хранилищ данных на основе различных топологий; средствами представления результатов анализа.	Выполнение лабораторного практикума	Задачи решены в полном объеме и получены верные ответы	Продемонстрирован верный ход решения всех, но не получен верный ответ во всех задачах	Продемонстрирован верный ход решения в большинстве задач	Задачи не решены

7.2 Примерный перечень оценочных средств (типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности)

7.2.1 Примерный перечень заданий для подготовки к тестированию

1. Упорядочивание системы профессиональных знаний – это

- А классификация
 - Б кодирование
 - В тезариус
 - Г дифференциация
2. Мнемонические коды
- А формируются из символов класса
 - Б используются для неупорядоченных категорий
 - В имеют иерархические связи
 - Г состоят из несвязанных частей
3. Применительно к системам классификации ось – это
- А родительская связь
 - Б признак разделения
 - В список условий
 - Г выделенный код
4. Автоматизация документооборота обеспечивают
- А АРМ работника
 - Б мониторные системы
 - В экспертные системы
 - Г информационно-справочные системы
5. Для обмена данными между медицинскими инструментальными комплексами используется
- А HL7
 - Б DICOM
 - В MeSH
 - Г USB
6. Использование XML позволяет
- А объединять различные ИС
 - Б упростить передачу информации
 - В сделать информацию доступной всем пользователям сети
 - Г использовать различные стандарты обмена информацией
7. Техническим налогом телемониторинга является
- А удаленное администрирование
 - Б резервное копирование
 - В шифрование информации
 - Г управление настройками сервера через Internet
8. Телеметрия – это
- А передача информации с прибора посредством средств коммуникации
 - Б обмен информацией между пользователями по телефону
 - В дистанционное управление прибором
 - Г удаленный мониторинг параметров
9. Знание – это
- А информация, представленная в удобном для обработки виде.
 - В проверенный практикой результат познания действительности, ее верное отражение в сознании человека.
 - Г сведения, сообщения, являющиеся объектами хранения, преобразования, передачи и помогающие решать поставленную задачу.

10. Реквизит – это

А минимальная составная единица информации, сохраняющая информативность.

В простейшая структурная единица информации, неделимая на смысловом уровне, отражающая количественную или качественную характеристику сущностей (объекта, процесса и т.п.) предметной области

Г составная единица информации, которая представлена на бумажном носителе и имеет самостоятельное значение.

11 Система кодирования - это

А совокупность правил и результат распределения заданного множества объектов на подмножества в соответствии с признаками сходства или различия.

Б упорядочивание объектов в соответствии со значениями признаков.

В совокупность правил образования кода.

12. Абстрактные модели - это

А макеты.

Б моделирование на реальном объекте.

В экспериментирование на реальном объекте.

Г математические, мысленные модели.

13. Математической моделью является

А модель автомобиля.

Б сборник правил дорожного движения.

В формула закона всемирного тяготения.

Г номенклатура списка товаров на складе.

14. Информационные системы – это

А компьютерные сети.

Б хранилище информации.

В системы, управляющие работой компьютера.

Г системы хранения, обработки и передачи информации в специально организованной форме.

15. Экспертные системы по своей сути – это

А операционные системы.

Б системы программирования.

В системы искусственного интеллекта.

Г авторские системы.

16. ГИС (геоинформационные системы) – это

А информационные системы в предметной области – география.

Б системы, содержащие топологические базы данных на электронных картах.

В электронные географические карты.

Г глобальные фонды и архивы географических данных.

17. Представление знаний в виде инструкций, позволяющих рационально организовать получение результата

А технология

Б информация

В автоматизация

Г управление

18. Результатом применения информационной технологии является

А информационный продукт

Б обработка и передача данных

В выработка первичной информации

Г сбор данных

19. Если объект ищется по произвольным фрагментам, позволяющим отличить объект от остальных, то это

А автоассоциация

Б гетероассоциация

В кластеризация

Г обучение с поощрением

20. Какое характерное отличие искусственных нейронных сетей от других разделов искусственного интеллекта

А адаптивность

Б нелинейность

В параллелизм

Г обобщаемость

21. Поставьте в соответствие задачу и ее описание

автоассоциация	выходной образ выбирается по произвольным фрагментам, позволяющим отличить объект от остальных
кластеризация	выбирается указание на группу похожих объектов
регрессия	каждому входному образу ставится в соответствие выходной
классификация	из группы известных выходных образов выбирается один в зависимости от поступившего входного образа

22. Укажите класс задач наиболее эффективно решаемых сетями встречного распространения

А сжатие с потерями

Б краткосрочное прогнозирование

В распознавание образов

Г аппроксимация функций

23. Какая операция над гиперкубом формирует изменение расположения измерений, представленных в отчете или на отображаемой странице?

1. консолидация;

2. срез;

3. вращение;

24. Какая операция над гиперкубом определяет переход от детального представления данных к агрегированному?

1. консолидация;

2. срез;

3. вращение;

25. Какое понятие определяет следующее высказывание «Множественная перспектива, состоящая из нескольких независимых измерений, вдоль которых могут быть проанализированы определенные совокупности данных»?

1. реляционная модель данных;
2. многомерное представление данных;
3. хранилище данных;

26. Технологию оперативной аналитической обработки данных, использующую методы и средства для сбора, хранения и анализа многомерных данных в целях поддержки процессов принятия решений, называют кратко?

1. OLAP технологией;
2. OLTP технологией;
3. технологией Data Mining;

27. Что означает термин HOLAP, в технологии OLAP?

1. для реализации многомерной модели используют реляционные БД;
2. для реализации многомерной модели используют многомерные БД;
3. для реализации многомерной модели используют и многомерные и реляционные БД;

28. К какому типу задач Data Mining, относится задача в которой необходимо определить независимые группы и их характеристики во всем множестве анализируемых данных?

1. задача классификации;
2. задача регрессии;
3. задача кластеризации;

29. К какому типу задач Data Mining, относится задача в которой необходимо определить зависимости между объектами или событиями?

1. задача распознавания образов;
2. задача поиска ассоциативных правил;
3. задача нормализации;

7.2.2 Примерный перечень заданий для решения стандартных задач

№1. Исходные данные: текстовый файл data1.txt, хранящийся на сетевом диске. Содержит даты в специальном формате (год/месяц/день) и числовые значения (дробная часть от целой отделяется точкой).

Требуется вычислить производные величины – среднее количество дней между парами дат, выделить даты, попадающие на 1 и 3 месяц от текущей даты, сформировать выборку только целых чисел, добавить случайное значение и сформировать новую даты, перевести отсчет из дней в часы. Из полученных результатов сформировать файл данных data2.txt.

Для полученных величин подготовить статистическое описание выборки – среднее, 95% доверительный интервал, минимаксное значение, медиана, мода.

Для вновь полученных категоризованных переменных вычислить частоту встречаемости каждого значения.

№ 2. Исходные данные: текстовый файл data2.txt, хранящийся на сетевом диске и содержащий значения количественных параметров.

Требуется исключить аномальные значения и построить функции распределения переменных (гистограммы) с наложением кривой нормального и логнормального распределения. Из полученных результатов сформировать файл данных data3.txt.

№ 3. Исходные данные: текстовый файл data4.txt, хранящийся на сетевом диске и содержащий выборку из 30 пар наблюдений

Требуется выполнить сплайновую интерполяцию и получить 300 промежуточных значений пропущенных наблюдений.

Требуется выполнить аппроксимацию данных с использованием экспоненциальной, линейной и полиномиальной функции и получить по 5 значений за пределами наблюдений.

№4 Исходные данные: текстовые файлы data1.txt, data 3, полученные в ходе предыдущих работ.

Требуется сформировать базу данных. Структура измерений и процессов должна соответствовать данным в текстовых файлах. Результаты сохранить как хранилище данных data5.

Для заданных показателей сформировать 3-мерную выборку и провести OLAP-анализ.

№ 5. Исходные данные: хранилище данных, созданное в ходе работы 4.

Из хранилища согласно индивидуальному заданию (указаны два показателя, и правило формирования третьего) требуется построить таблицы кросс-табуляции.

Требуется выполнить линейный регрессионный анализ, если условия его применения соблюдаются в противном случае подобрать и выполнить трансформацию выборочных данных или рассмотреть возможность использования полиномиальной регрессии для поиска зависимости между заданными показателями.

№ 6. Исходные данные: хранилище данных, созданное в ходе работы 4.

Из хранилища согласно индивидуальному заданию сформировать выборку из семи показателей, указанных в индивидуальном задании.

Требуется выполнить кластерный анализ с целью разбиения выборки на 3 части. Исследовать все возможные варианты разделения на кластеры, результаты сравнить по расстоянию между кластерами, однородности элементов кластера, статистическим характеристикам.

7.2.3 Примерный перечень заданий для решения прикладных задач

№1 Постановка задачи

Реализовать, используя механизм общей памяти, передачу данных между двумя потоками. Первый поток проводит вычислительную работу, оперируя

данными, считываемых из ранее заготовленного файла. Результат вычислений для каждого цикла вычислений появляется асинхронно. Время появления результата на каждом цикле вычислений зависит от множества различных факторов. При появлении результата вычислений первый поток должен передавать его второму потоку для дальнейших вычислений. Исходный файл – содержит~10000 строк чисел типа float. Первый поток считывает их последовательно по~1000 и производит над ними вычислительные операции (например нахождение среднеквадратичного). Результат каждого вычислительного цикла передается второму потоку. Второй поток также производит над ним мат. операции и конечный результат пишет в файл.

Требования реализации:

Использование системных объектов синхронизации обязательно.

Все записи считывания и записи в файл дублировать на консоль (форму).

Моменты переключения между потоками сообщать на консоль (форму).

№2 Разработать структуру хранилища данных (ХД) для следующего примера: имеется история продаж различных товаров по дням в нескольких торговых объектах. Товары объединены в группы. Все данные представлены в 4 таблицах: Товарные группы, Товары, Отделы, Продажи. Их фрагменты приведены ниже (таблицы 1-4).

Таблица 1 – Товарные группы

Код группы	Наименование группы
33	Иммуномодуляторы
48	Общетонизирующие средства и адаптогены
50	Местные анестетики
108	Микро- и макроэлементы
198	Витамины и витаминоподобные средства
223	Желчегонные средства и препараты желчи
247	Антисептики и дезинфицирующие средства

Таблица 2 – Товары (фрагмент)

Код товара	Наименование товара	Код группы
774	Альмагель	1
810	Иммунорм	33
824	Ревит	198
898	Настойка пустырника	48
...		

Таблица 3 – Отделы

Код отдела	Наименование отдела
1	Аптека 1
2	Аптека 2
3	Аптека 3

Таблица 4 – Продажи (фрагмент)

Дата	Код отдела	Код товара	Час покупки	Количество	Сумма
01.01.2006	1	31052	13	1	56,5
01.01.2006	1	36259	16	1	72,48
01.01.2006	1	40315	15	1	15,84
01.01.2006	3	810	14	1	163,50
01.01.2006	2	40315	15	3	47,52
...	...	...	...	...	...

Реализовать многоагентную систему для предоставления товаров или услуг клиентам. Многоагентная система должна содержать несколько видов агентов. Например, агенты, предоставляющие товар или услуги за определенную стоимость; агенты, покупающие указанные товары или услуги; агенты-посредники между агентом-продавцом и агентом — конечным покупателем. Каждый агент получает начальную информацию от пользователя. Общение пользователя и агента происходит при помощи графического пользовательского интерфейса. Один пользователь вводит описания товаров и услуг, а другой пользователь вводит информацию о том, какой товар или услугу необходимо приобрести.

№ 4

Постановка задачи.

Составить программу простейшего автоматизированного англо — немецко-русского словаря, данные словаря хранятся в файле. Программа должна выполнять следующие задачи:

- 1 Добавление новых слов и редактирование имеющихся слов в словаре.
- 2 По запросу пользователя словарем выдавать перевод слова с любого языка на два других.
- 3 Производить поиск нужного слова (по полному и частичному совпадению).
- 4 Сортировку слов.

№5

Постановка задачи.

А) Существует файл СКЛАД, в котором хранится информация об ассортименте продукции на складе:

- наименование продукции;
- цена единицы продукции;
- количество единиц наличия;
- сумма.

Б) Имеется сеть магазинов, которые получают продукцию со склада. Информация храниться в файле РАСХОД:

- потребитель;
- наименование продукции;
- количество единиц отпущено;
- сумма;
- дата.

Программа следит за движением продукции со склада, обеспечивает ввод, просмотр и удаление информации, каждому потребителю выдает счет-накладную.

№6

Написать программу, использующую стек для моделирования Т — образного сортировочного узла на железной дороге. Программа должна разделять на два направления состав, состоящий из вагонов двух типов (на каждое направление формируется состав из вагонов одного типа). Предусмотреть

возможность формирования состава из файла и с клавиатуры. Предусмотреть графическую интерпретацию.

№7

Создать приложение «Генеалогическое дерево». Паспортные данные членов некоторого родового клана; ссылки на детей (или на родителей). Поиск всех потомков или всех предков для указанного лица, возможность редактирования, добавления, удаления. Графическое изображение дерева.

№8

Написать программу, имитирующую перевод чисел в различные системы счисления и выполнение арифметических операций над ними. Разработать удобную систему подсказок пользователю и графический интерфейс.

№9

Дан текстовый файл большого размера произвольной структуры. Реализовать различные способы чтения файла: посимвольно, поблочно, построчно с использованием разных функций чтения: низкоуровневые, потоковые. Организовать поиск заданного слова в тексте при полном и частичном совпадении, посчитать количество встречаемости слова в тексте. Определить скорость работы каждого способа.

7.2.4 Примерный перечень вопросов для подготовки к зачету

Не предусмотрено учебным планом

7.2.5 Примерный перечень заданий для подготовки к экзамену

1. Понятие технологии обработки информации. Этапы технологии обработки.
2. Классификация источников информации.
3. Данные как вид информации. Формы данных. Типы данных.
4. Структурированные данные. Не структурированные и слабоструктурированные данные. Структуры данных.
5. Основные компоненты технологии обработки информации. Информационные процессы обработки информации.
6. Системы сбора и обработки информации. Виды систем сбора информации.
7. Общие визуализаторы: графики, диаграммы, гистограммы, статистика.
8. OLAP- анализ, манипуляции с OLAP- кубами. Матрицы классификации, диаграммы рассеяния.
9. Детализированные и агрегированные данные, метаданные.
10. Многомерное представление данных и многомерный куб MOLAP. Операции с многомерным кубом ROLAP, схемы «снежинка» и «звезда».
11. Машинное обучение и классы задач DataMining.
12. Классификация программного обеспечения для создания аналитических решений.
13. Группировка/разгруппировка данных, объединение данных, преобразование даты и времени.

14. Оценка качества данных по происхождению, профайлинг данных. Набор инструментов предобработки данных в аналитическом приложении. Фильтрация данных.
15. Виды аномалий данных. Обнаружение аномальных значений специальными методами. Происхождение пропусков данных, восстановление пропущенных значений.
16. Основные внешние устройства хранения вычислительного комплекса, назначение и использование.
17. Обработка двумерных сигналов (изображений), обработка графической информации.
18. Детерминированная математическая модель описания дискретного изображения
19. Статистическая модель дискретного изображения (Гистограмма яркостей изображения).
20. Цифровая обработка цифровых изображений: двумерное дискретное преобразование Фурье, получение спектра изображения.
21. Задача ассоциации, кластеризации, классификация и восстановления зависимостей.
22. Системы обработки входящей текстовой информации, методы поиска текстовой информации. Качество информационно-поисковых систем.
23. Системы поддержки принятия решений
24. Хранилища данных
25. Физические и виртуальные хранилища данных
26. Витрины (киоски) данных
27. Проблемы построения хранилищ данных
28. Категории данных в ХД
29. Информационные потоки в ХД
30. Топологии ХД
31. Таблицы фактов
32. Таблицы измерений
33. Архитектура OLAP-систем
34. Клиентские OLAP-средства
35. Серверные OLAP-средства
36. ETL. Основные цели и задачи
37. Извлечение данных
38. Выбор источников данных
39. Очистка данных
40. Критерии оценки качества данных
41. Преобразование структуры данных
42. Агрегирование данных
43. Создание новых данных
44. Загрузка данных в ХД
45. Преимущества и недостатки ХД
46. NoSQL подход
47. Классификация NoSQL хранилищ

7.2.6. Методика выставления оценки при проведении промежуточной аттестации

Экзамен проводится по билетам, каждый из которых содержит два теоретических вопроса.

1. Оценка «Неудовлетворительно» ставится в случае, если студент не ответил на все основные вопросы билета или все дополнительные.
2. Оценка «Удовлетворительно» ставится в случае, если студент ответил только на один вопрос билета и большую часть дополнительных.
3. Оценка «Хорошо» ставится в случае, если студент ответил на все вопросы билета и большую часть дополнительных.
4. Оценка «Отлично» ставится в случае, если студент дал развернутый ответ на все вопросы билета и все дополнительные.

7.2.7 Паспорт оценочных материалов

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции	Наименование оценочного средства
1	Хранилища данных	ОПК-5, ПК-24	Опрос, защита лабораторных работ.
2	Оперативный анализ данных (OLAP)	ОПК-5, ПК-24	Опрос, защита лабораторных работ.
3	Работа с данными в хранилище	ОПК-5, ПК-24	Опрос, защита лабораторных работ.
4	Нереляционные модели данных (NoSQL)	ОПК-5, ПК-24	Опрос, защита лабораторных работ.

7.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности

Устный экзамен проводится на основании билета на бумажном носителе. На подготовку к ответу отводится 20 минут. Затем осуществляется проверка ответа экзаменатором и обсуждение дополнительных вопросов. После этого выставляется оценка согласно методике.

Защита лабораторных работ осуществляется согласно требованиям, описанным в методических материалах. Примерное время защиты на одного студента составляет 20 минут.

8 УЧЕБНО МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

8.1 Перечень учебной литературы, необходимой для освоения дисциплины

1. Нестеров С.А. Интеллектуальный анализ данных средствами MS SQL Server 2008 [Электронный ресурс]/ Нестеров С.А.— Электрон. текстовые данные.— Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016.— 303 с.— Режим доступа: <http://www.iprbookshop.ru/62813.html>.— ЭБС «IPRbooks»

2. Парфенов Ю.П. Постреляционные хранилища данных [Электронный ресурс]: учебное пособие/ Парфенов Ю.П.— Электрон. текстовые данные.—

Екатеринбург: Уральский федеральный университет, ЭБС АСВ, 2016.— 120 с.— Режим доступа: <http://www.iprbookshop.ru/68372.html>.— ЭБС «IPRbooks»

3. Полубояров В.В. Использование MS SQL Server Analysis Services 2008 для построения хранилищ данных [Электронный ресурс]/ Полубояров В.В.— Электрон. текстовые данные.— Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), 2016.— 663 с.— Режим доступа: <http://www.iprbookshop.ru/73682.html>.— ЭБС «IPRbooks»

4. Чубукова И.А. DataMining [Электронный ресурс]: учебное пособие/ Чубукова И.А.— Электрон.текстовые данные.— М.: БИНОМ. Лаборатория знаний, Интернет-Университет Информационных Технологий (ИНТУИТ), 2008.— 382 с.— Режим доступа: <http://www.iprbookshop.ru/15860>.

5. Белов, В. М. Теория информации : Учебное пособие / Белов В. М. - Москва: Горячая линия - Телеком, 2012. - 143 с. - ISBN 978-5-9912-0237-4.URL: <http://www.iprbookshop.ru/12050>

6. Федин Ф.О. Анализ данных. Часть 1. Подготовка данных к анализу [Электронный ресурс]: учебное пособие/ Федин Ф.О., Федин Ф.Ф.— Электрон.текстовые данные.— М.: Московский городской педагогический университет, 2012.— 204 с.— Режим доступа: <http://www.iprbookshop.ru/26444>

7. Федин Ф.О. Анализ данных. Часть 2. Инструменты DataMining [Электронный ресурс]: учебное пособие/ Федин Ф.О., Федин Ф.Ф.— Электрон.текстовые данные.— М.: Московский городской педагогический университет, 2012. – 308 с.— Режим доступа: <http://www.iprbookshop.ru/26445>.

8.2 Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень лицензионного программного обеспечения, ресурсов информационно-телекоммуникационной сети «Интернет», современных профессиональных баз данных и информационных справочных систем:

- Система управления базами данных Microsoft SQL Server
- Система управления базами данных MongoDB
- Среда разработки приложений Visual Studio
- Табличный редактор Microsoft Excel
- специальное ПО –DeductorAcademic свободно распространяемая (для некоммерческого использования) версия 5.3(Разработчик: BaseGroupLabs, info@basegroup.ru.)

- Сайт компании BaseGroupLabs <http://basegroup.ru>, Документация по ПО DeductorAcademic. Презентации решений прикладных задач с использованием Deductor

- www.citforum.ru
- www.machinelearning.ru
- <http://archive.ics.uci.edu/ml/>
- <http://mloss.org/>

9 МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

Персональные компьютеры с установленной ОС и специальным программным обеспечением и доступом к сети Интернет.

Средства мониторинга – программа тестирования по модулям дисциплины с базами тестовых вопросов.

10. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

По дисциплине «Технологии обработки информации» читаются лекции, проводятся лабораторные работы.

Основой изучения дисциплины являются лекции, на которых излагаются наиболее существенные и трудные вопросы, а также вопросы, не нашедшие отражения в учебной литературе.

Лабораторные работы выполняются на лабораторном оборудовании в соответствии с методиками, приведенными в указаниях к выполнению работ.

Вид учебных занятий	Деятельность студента
Лекция	Написание конспекта лекций: кратко, схематично, последовательно фиксировать основные положения, выводы, формулировки, обобщения; помечать важные мысли, выделять ключевые слова, термины. Проверка терминов, понятий с помощью энциклопедий, словарей, справочников с выписыванием толкований в тетрадь. Обозначение вопросов, терминов, материала, которые вызывают трудности, поиск ответов в рекомендуемой литературе. Если самостоятельно не удастся разобраться в материале, необходимо сформулировать вопрос и задать преподавателю на лекции или на практическом занятии.
Лабораторная работа	Лабораторные работы позволяют научиться применять теоретические знания, полученные на лекции при решении конкретных задач. Чтобы наиболее рационально и полно использовать все возможности лабораторных для подготовки к ним необходимо: следует разобрать лекцию по соответствующей теме, ознакомиться с соответствующим разделом учебника, проработать дополнительную литературу и источники, решить задачи и выполнить другие письменные задания.
Самостоятельная работа	Самостоятельная работа студентов способствует глубокому усвоения учебного материала и развитию навыков самообразования. Самостоятельная работа предполагает следующие составляющие: - работа с текстами: учебниками, справочниками, дополнительной литературой, а также проработка конспектов лекций; - выполнение домашних заданий и расчетов; - работа над темами для самостоятельного изучения; - участие в работе студенческих научных конференций, олимпиад; - подготовка к промежуточной аттестации.
Подготовка к промежуточной аттестации	Готовиться к промежуточной аттестации следует систематически, в течение всего семестра. Интенсивная подготовка должна начаться не позднее, чем за месяц-полтора до промежуточной аттестации. Данные перед экзаменом, экзаменом три дня эффективнее всего использовать для повторения и систематизации материала.