



Проектирование систем обработки больших данных (Big Data).oi(dor_БАК)

- 1 Говоря о таком признаке больших данных, как ..., следует отметить, что накопленная база данных представляет собой гигантский объем информации, для которого обработка и хранение традиционными способами являются трудоемкими процессами; такой объем нуждается в новых подходах и в более усовершенствованных инструментах
- 2 Установите соответствие признаков больших данных и их характеристик:
- 3 ... данные – это текстовые данные с неустойчивым форматом, которые для обработки инструментами требуют больших временных затрат на преобразование
- 4 ... данные – это данные, у которых нет строго зафиксированного формата
- 5 Данные текстовых файлов с определенными паттернами для их обработки называются ...
- 6 Согласно принципу трех V английским словом «...» обозначают скорость, реакцию на текущую информацию за время, ограниченное приложением
- 7 Комплексный набор методов, подходов и инструментов обработки структурированных и неструктурированных данных колоссальных объемов – это ... данные
- 8 Термин Big Data был предложен Клиффордом Линчем в журнале Nature 3 сентября ... года по аналогии с терминами «Большая нефть», «Большая руда»
- 9 Установите правильную последовательность этапов анализа данных:
- 10 ESFRI – это краткое обозначение ...
- 11 Подборка тематически связанных данных, расположенных на одном рабочем листе, – это ...
- 12 На рисунке ниже приведены имена ... (столбцов) списка





- 13) Чтобы переименовать рабочий лист электронной таблицы, сначала нужно щелчком правой кнопки мыши вызвать контекстное меню на ярлыке листа и выбрать в нем команду «...»
- 14) Установите правильную последовательность действий, если необходимо создать собственный формат данных (лс) и применить его для поля электронной таблицы «Мощность двигателя» (см. рисунок ниже):
- 15) Операция над данными, с помощью которой можно переупорядочить строки в списке в соответствии с содержанием конкретных столбцов, – это ...
- 16) Карта с нанесенной на нее информацией в виде цвета или с помощью других способов – это ...
- 17) ... – это макрокоманда, содержащая последовательность действий, записанных пользователем
- 18) ... – это создание большого количества различных визуальных представлений одних и тех же данных
- 19) ... таблица – это специальный тип таблицы, суммирующий информацию из конкретных полей списка или базы данных; допускает после построения настройку на различные разрезы данных
- 20) Установите правильную последовательность шагов по созданию макроса, с помощью которого в поле «Цвет» определяются неповторяющиеся значения (эти значения выводятся в ячейке перечислением через запятую):
- 21) ... данные представляют собой высокоорганизованную, фактическую и точную информацию и обычно представлены в форме букв и цифр, которые хорошо вписываются в строки и столбцы таблиц
- 22) ... данные не имеют заранее определенной структуры и представлены во всем разнообразии форм (от изображений и текстовых файлов, таких как документы PDF, до видео- и аудиофайлов)
- 23) Неструктурированные данные также называются ... в том смысле, что они имеют субъективный и интерпретирующий характер
- 24) ... данных – это хранилище или система, предназначенная для хранения огромных объемов данных в естественном / необработанном формате





- (25) Для работы с реляционными базами данных и хранилищами используется специальный язык программирования – язык структурированных запросов ..., который был разработан IBM в 1970-х гг.
- (26) Установите соответствие понятий и их определений:
- (27) Полный набор объектов, связанных с поставленной перед изучением проблемой, – это ...
- (28) Установите правильный порядок процедуры анализа текстовой информации, осуществляемой аналитиком:
- (29) В числе наиболее часто используемых систем управления реляционными базами данных (СУДБ), инструментов и технологий для обработки данных можно выделить ... – бесплатную СУБД с открытым исходным кодом, которая поддерживает запросы как SQL, так и JSON, а также наиболее широко используемые языки программирования
- (30) ... вывод – это утверждение о том, что представляют собой законы, лежащие в основе изучаемой генеральной совокупности
- (31) ... данных – это процесс, в котором мы раскрываем и дополняем данные, добавляя к ним связанную информацию из стороннего источника
- (32) ... – это научное направление в области искусственного интеллекта, в частности робототехники, и связанные с ним технологии получения изображений объектов реального мира, их обработки и использования, полученных данных для решения разного рода прикладных задач без участия (полного или частичного) человека
- (33) ... – это один из основных процессов в управлении хранилищами данных
- (34) Установите хронологическую последовательность исторических прорывов в машинном зрении:
- (35) Установите соответствие дат и описаний исторических прорывов в машинном зрении:
- (36) Компонентом системы машинного зрения является ... для синхронизации частей обнаружения (часто оптический или магнитный), который служит для захвата и обработки изображений





- 37) Такой компонент распознавания текста, как ..., включает в себя определение ориентации определенного объекта относительно камеры
- 38) Такой метод обработки изображений, как ..., используется для поиска и (или) подсчета деталей, его цель заключается в упрощении и/или изменении представления изображения, чтобы его было проще и легче анализировать
- 39) Одной из задач машинного зрения является ..., например: определение трехмерного движения камеры; слежение, то есть следование за перемещениями объекта (машины, человека)
- 40) Такой метод обработки изображения, как ..., заключается в преобразовании изображения в серых тонах в белые и черные пиксели; значения каждого пикселя условно кодируются как 0 и 1
- 41) Английское название хранилищ данных масштаба предприятия, где данные хранятся в неструктурированном / необработанном виде, – Data ...
- 42) Эффективным подходом к формированию поискового словаря в условиях слабого знакомства с предметной областью исследований является ...
- 43) В результате построения частотно-рангового распределения длин слов в данном тексте выбирается специфичный для данного языка диапазон ... слов – именно в этом диапазоне будут содержаться те слова, которые наилучшим образом отражают тематику и содержание текста
- 44) Термин на английском языке, обозначающий процесс анализа данных на основе определенных образцов и поиск соответствия им в больших наборах данных, – Data ...
- 45) Среди разработчиков программного обеспечения, реализующих передовые поисковые лингвистические технологии, следует выделить ... – разработки этой фирмы «Следопыт», «Классификатор» и «Аннотатор» позволяют осуществлять оперативный поиск документов, их индексирование, классификацию и автоматическое аннотирование
- 46) Программный продукт TextAnalyst научно-производственного инновационного центра «...», использует для решения задач распознавания слов нейросетевые технологии и, в основном, предназначен для решения задачи автоматического реферирования документов





- 47) ... – английское сокращенное обозначение специализированных комплектов разработчика, которые позволяют при необходимости создавать собственное программное обеспечение, адаптированное к задачам, решаемым той или иной организацией
- 48) Сервисы ... обучения в облаках помогают использовать неструктурированные данные на практике: распознавать лица, объекты, номера автомобилей, анализировать интонации речи
- 49) Установите соответствие элементов специфики работы со неструктурированными данными и их содержания:
- 50) Установите соответствие основания классификации и соответствующих примеров корпусов текстов:
- 51) Установите соответствие английского обозначения признака больших данных и характеристики данного признака:
- 52) Установите соответствие признаков больших данных и их характеристик:
- 53) Текстовые данные с неустойчивым форматом, которые для обработки инструментами требуют больших временных затрат на преобразование, – это ... данные
- 54) Данные, у которых нет строго зафиксированного формата, называются ...
- 55) ... данные – это данные текстовых файлов с определенными паттернами для их обработки
- 56) Согласно принципу трех V английским словом «...» обозначают разнообразие, способность обработки множества типов, источников и форматов данных
- 57) Комплексный набор методов, подходов и инструментов обработки структурированных и неструктурированных данных колоссальных объемов –это ...
- 58) ... впервые использовал термин «большие данные» в журнале Nature по аналогии с терминами «большая нефть», «большая руда»
- 59) Установите правильную последовательность этапов анализа данных:
- 60) Совокупность методов и средств извлечения их организованных данных для принятия решений – это ... данных
- 61) Подборка тематически связанных данных, расположенных на одном рабочем листе, – это ...





- 62 На рисунке ниже приведены ... списка
- 63 Чтобы переименовать рабочий лист электронной таблицы, сначала нужно ... кнопки мыши вызвать контекстное меню на ярлыке листа и выбрать в нем команду «Переименовать»
- 64 Установите правильную последовательность действий, если необходимо создать собственный формат данных (лс) и применить его для поля электронной таблицы «Мощность двигателя» (см. рисунок ниже):
- 65 Операция над данными, с помощью которой можно переупорядочить строки в списке в соответствии с содержанием конкретных столбцов, – это ...
- 66 Карта с нанесенной на нее информацией в виде цвета или с помощью других способов – это ...
- 67 Макрокоманда, содержащая последовательность действий, записанных пользователем, – это ...
- 68 Создание большого количества различных визуальных представлений одних и тех же данных – это ...
- 69 ... таблица – это специальный тип таблицы, суммирующий информацию из конкретных полей списка или базы данных; допускает после построения настройку на различные разрезы данных
- 70 Установите правильную последовательность шагов по созданию макроса, с помощью которого в поле «Цвет» определяются неповторяющиеся значения (эти значения выводятся в ячейке перечислением через запятую):
- 71 ... данные представляют собой высокоорганизованную, фактическую и точную информацию и обычно представлены в форме букв и цифр, которые хорошо вписываются в строки и столбцы таблиц
- 72 ... данные обычно существуют в таблицах, подобных файлам Excel и электронным таблицам Google Docs
- 73 Структурированные данные часто называют ..., что означает, что их объективный и заранее определенный характер позволяет нам легко подсчитывать, измерять и выражать данные в числах
- 74 ... – это утверждение о том, что представляют собой законы, лежащие в основе изучаемой генеральной совокупности





- 75) Для работы с реляционными базами данных и хранилищами используется специальный язык программирования – язык структурированных запросов ..., который был разработан IBM в 1970-х гг.
- 76) Установите соответствие понятий и их определений:
- 77) ... – это полный набор объектов, связанных с поставленной перед изучением проблемой
- 78) Установите правильный порядок процедуры анализа текстовой информации, осуществляемой аналитиком:
- 79) Установите соответствие примеров и характеристик систем управления реляционными базами данных (СУДБ), инструментов и технологий для обработки данных:
- 80) ... данных – это процесс, посредством которого извлекается необходимая для обработки информация, впоследствии предоставляющая решение поставленной перед ним задачи/проблемы
- 81) ... данных являются соцсети, приложения, веб-визоры, камеры наблюдения, видеорегистраторы, датчики и другие системы, где генерируются разнородные данные
- 82) ... включает данные, которые относятся к нескольким однородным объектам, при этом каждый объект описан в виде набора характеристик; обычно его можно представить в виде таблицы в несколько строк и столбцов
- 83) ... – это сокращенное английское обозначение одного из основных процессов в управлении хранилищами данных, которое можно перевести как «извлечение, преобразование, загрузка / извлечение, загрузка, преобразование»
- 84) Установите соответствие дат и описаний исторических прорывов в машинном зрении:
- 85) Установите хронологическую последовательность исторических прорывов в машинном зрении:
- 86) Компонентом системы машинного зрения является датчик для ... частей обнаружения (часто оптический или магнитный датчик), который служит для захвата и обработки изображений
- 87) Установите соответствие компонентов распознавания текста и их характеристик:





- 88) Такой метод обработки изображений, как ..., используется для поиска и (или) подсчета деталей, его цель заключается в упрощении и/или изменении представления изображения, чтобы его было проще и легче анализировать
- 89) Такой метод обработки изображения, как ..., заключается в преобразовании изображения в серых тонах в белые и черные пиксели; значения каждого пикселя условно кодируются как 0 и 1
- 90) Одной из задач машинного зрения является ..., например: определение трехмерного движения камеры; слежение, то есть следование за перемещениями объекта (машины, человека)
- 91) Английское название хранилищ данных масштаба предприятия, где данные хранятся в неструктурированном / необработанном виде, – Data ...
- 92) В числе наиболее доступных источников информации на сегодня – ... (укажите 2 варианта ответа)
- 93) После того как первая версия поискового словаря создана, дальнейшие операции ...
- 94) Услуга эвристического поиска, основанного на использовании систем искусственного интеллекта, формирующих расширенный запрос на основе применения специализированных отраслевых тезаурусов и семантических сетей, в Интернете на сегодня ...
- 95) Среди разработчиков программного обеспечения, реализующих передовые поисковые лингвистические технологии, следует выделить ЗАО «...» – такие разработки этой фирмы, как «Следопыт», «Классификатор» и «Аннотатор», позволяют осуществлять оперативный поиск документов, их индексирование, классификацию и автоматическое аннотирование
- 96) Программный продукт TextAnalyst научно-производственного инновационного центра ..., использует для решения задач распознавания слов нейросетевые технологии и, в основном, предназначен для решения задачи автоматического реферирования документов
- 97) ... – английское сокращенное обозначение специализированных комплектов разработчика позволяющих при необходимости создавать собственное программное обеспечение, адаптированное к задачам, решаемым той или иной организацией





- 98) ... – проект фонда Apache Software Foundation, платформа с открытым исходным кодом для обработки больших данных; благодаря распределенной файловой системе и распределенным вычислениям эта разработка позволяет обрабатывать и анализировать очень большие объемы данных.
- 99) Установите соответствие элементов специфики работы со неструктурированными данными и их содержания:
- 100) Установите соответствие основания классификации и соответствующих примеров корпусов текстов:
- 101) Рисунок ниже отражает так называемый «принцип трех V». Какие характеристики, согласно данному рисунку, играют решающую роль в больших данных?
- 102) Одна из характеристик больших данных определяет способность обработки множества типов, источников и форматов данных от сенсоров, умных устройств, социальных сетей. Также данная характеристика указывает на способность интегрировать все большее число источников, содержащих различные структурированные, полуструктурированные данные, извлекаемыми из веб-страниц, e-mail, документов и др. О какой характеристике больших данных идет речь?
- 103) Аналитика больших данных в различных областях человеческой деятельности имеет разные цели. В частности, одна из целей аналитики больших данных – уменьшение стоимости лечения различных заболеваний. К какой области относится данная цель?
- 104) С помощью такого вида операций над данными, как сортировка, можно переупорядочить строки в списке в соответствии с содержанием конкретных столбцов. Например, вам необходимо определить значения для критериев отбора: - Марка 1 – первая марка в списке, отсортированном по полю «владелец» в возрастающем (алфавитном) порядке. - Марка 2 – вторая марка в списке, отсортированном по полю «владелец» в возрастающем (алфавитном) порядке. Приведите технологию выполнения данного задания.
- 105) Существует такой вид операций над данными, как задание критериев – точное соответствие. Например, вам необходимо определить автомобили цвета «Цвет 1» (белый) и цвета «Цвет 2» (черный). Приведите технологию выполнения данного задания.
- 106) Существует такой вид операций над данными, как Задание критериев на основе сравнения. Например, вам необходимо определить автомобили марки «Марка1» (Hyundai). Приведите технологию выполнения данного задания.





- 107) Если мы применяем данные для аналитической обработки и используем так называемые конвейеры данных, конечным пунктом путешествия структурированных данных будут специальные хранилища данных. Чем больше объем данных, тем больше места требуется для их хранения. Например, картинка с высоким разрешением весит намного больше, чем текстовый файл. Следовательно, неструктурированные данные требуют больше места для хранения. Где обычно хранятся неструктурированные данные?
- 108) Существуют различные инструменты управления неструктурированными данными. Один из таких инструментов представляет собой систему управления базами данных, которая ориентирована на документы, не требующие жесткой схемы или структуры таблиц. Данный инструмент считается одним из классических примеров NoSQL и использует документы, подобные JSON. О каком инструменте управления неструктурированными данными говорится в данном примере?
- 109) Существуют различные инструменты управления неструктурированными данными. Один из таких инструментов представляет собой эффективную среду с открытым исходным кодом, которая используется для обработки больших объемов данных и их хранения на недорогих обычных серверах. Это не только мощный, но и гибкий инструмент, поскольку не требует наличия схемы или структуры для хранимых данных. Он помогает структурировать неструктурированные данные и затем экспортировать их в реляционные базы данных. О каком инструменте управления неструктурированными данными говорится в данном примере?
- 110) Изучите приведенный ниже рисунок. Реализация какой из задач машинного зрения здесь отражена?
- 111) Изучите приведенный ниже рисунок. Реализация какой из задач машинного зрения здесь отражена?
- 112) Реализация данной задачи машинного зрения основана на относительно простых и быстрых вычислениях. Иногда эта задача используется для нахождения небольших участков в анализируемом изображении, которые затем анализируются с помощью приемов, более требовательных к ресурсам, для получения правильной интерпретации. О какой задаче машинного зрения здесь идет речь?





- 113) Одной из задач компьютерной лингвистики является автоматическая классификация текстов, т. е. отнесение текста к той или иной области или ее подмножеству на основе некоторого алгоритма с некоторой вероятностью. Часть алгоритмов используют для этого только данные, полученные непосредственно из этого текста. Охарактеризуйте такие алгоритмы.
- 114) В комплексе инструментов автоматизированного анализа текстов реализованы инструменты анализа и исследования текстов на этапах морфологического, синтаксического анализа, с применением статистических методов, кроме того присутствует средство исследования полученных результатов на следующем – аналитическом – уровне. На основе инструментов комплекса созданы сервисы решения задач выделения ключевых слов, статистического анализа, классификации. Какие режимы имеет сервис классификации текстов?
- 115) Структурированные и неструктурированные данные обрабатывают и анализируют по-разному. Классические компьютерные алгоритмы бесполезны при работе с неупорядоченными данными. Поэтому существуют отдельные типы инструментов, созданные специально на этот случай. В частности, существует тип инструментов, который включает в себя классификаторы, нейронные сети, векторные машины – особые математические модели, которые могут работать практически с любыми типами данных. Они могут предсказывать, преобразовывать, распознавать и даже создавать новые данные. Назовите данный тип инструментов для работы с неструктурированными данными.

