ПЗ 6

**Вопросы**

1

Как классифицируется IT-анализ по режиму и темпу?

2

В чем состоят задачи и содержание OLAP-анализа?

3

Что означают требования к OLAP-системе под рубрикой

«быстрый»?

4

Что означают требования к OLAP-системе под рубрикой

«анализ»?

5

Что означают требования к OLAP-системе под рубрикой

«разделяемой»?

6

Что означают требования к OLAP-системе под рубрикой

«многомерной»?

7

Что означают требования к OLAP-системе под рубрикой

«информации»?

8

Какие типы OLAP-систем ВЫ знаете, в чем их различие?

9

В чем сущность MOLAP-системы?

10

В чем сущность ROLAP-системы?

11

В чем сущность HOLAP-системы?

12

Какие функции выполняют процедуры сечение и поворот?

13

Какие функции выполняют процедуры свертка и развертка?

14

Какие функции выполняют процедуры проекция и

построение трендов?

15

Для каких целей предназначена подсистема

интеллектуального анализа данных?

16

Какие направления и методы поддерживают средства

подсистемы интеллектуального анализа ИАС?

17

Раскройте содержание понятия «знания».

18

Каким образом классифицируются виды знаний?

19

В чем специфика методов интеллектуального анализа?

20

Какие методы интеллектуального анализа Вы знаете?

21

В чем состоит содержание методов нечеткой логики, системы

рассуждений на основе аналогичных случаев?

22

В чем состоит содержание методов нейронных сетей и

генетических алгоритмов?

23

В чем заключаются методы ассоциаций, кластеризации и

классификации?

24

Раскройте сущность методов эволюционного

программирования и алгоритмов ограниченного перебора.

25

Назовите процессы, явления, закономерности, при

исследовании которых используются методы интеллектуального анализа.

26

В каких предметных областях и для каких целей

используются методы интеллектуального анализа?

**I. Ответьте на вопросы:**

1) IT-анализ классифицируется по режиму и темпу следующим образом:

* По режиму:

1. регулярный анализ - проводится на постоянной основе для контроля за работоспособностью системы или определенных ее компонентов;
2. ситуативный анализ - проводится в ответ на нештатную ситуацию, связанную с неправильной работой системы или с нарушением безопасности;
3. проектный анализ - проводится в рамках проекта для уточнения требований и определения оптимального решения задачи.

* По темпу:

1. оперативный анализ - проводится в режиме реального времени для быстрого выявления и устранения проблем;
2. стратегический анализ - проводится для дальнейшего планирования и разработки стратегии развития IT-системы;
3. тактический анализ - проводится в целях оптимизации текущих процессов и улучшения работоспособности системы.
4. OLAP-анализ (анализ многомерных данных) – это процесс анализа данных, хранящихся в многомерной структуре (кубе данных). Он позволяет пользователям анализировать данные, чтобы получать информацию о различных аспектах бизнес-процессов и принимать на их основании решения.

Основными задачами и содержанием OLAP-анализа являются:

1. Определение ключевых производственных показателей (KPI) – это ключевые показатели, используемые для измерения эффективности бизнес-процессов. В рамках OLAP-анализа, KPI могут быть построены с использованием данных из различных источников и отображены в многомерной форме.
2. Анализ трендов и паттернов – OLAP-анализ позволяет отображать данные в различных формах, таких как графики и диаграммы, что позволяет быстро определить тренды и паттерны. На основе этой информации можно разрабатывать стратегии и планы для улучшения бизнес-процессов.
3. Выявление связей и зависимостей – OLAP-анализ может использоваться для поиска связей и зависимостей между различными показателями. Эта информация может помочь бизнесу лучше понимать его производственные процессы и разрабатывать более эффективные стратегии.
4. Сокращение времени анализа данных – OLAP-анализ позволяет быстро анализировать и обрабатывать данные в режиме реального времени, что сокращает время необходимое для принятия важных бизнес-решений.
5. Систематизация информации – OLAP-анализ помогает систематизировать большие объёмы данных и лучше понимать бизнес-среду. Это позволяет улучшить принятие решений, разрабатывать стратегии и повышать эффективность бизнес-процессов.
6. Требования к OLAP-системе "быстрый" означают, что она должна быстро выполнять запросы и обладать мгновенным доступом к данным, обрабатывать большие объёмы данных и иметь высокую производительность благодаря оптимизированным алгоритмам обработки запросов и определенным аппаратным характеристикам.
7. Требования к OLAP-системе под рубрикой "анализ" означают, что она должна обладать возможностью многомерного анализа данных, позволяя быстро и эффективно анализировать, интерпретировать и визуализировать данные, чтобы выявить связи и паттерны, выяснить причины и определить возможности улучшения бизнес-процессов. OLAP-система также должна иметь возможность генерировать отчёты и диаграммы, чтобы предоставлять видимость и прозрачность в бизнес-процессы, и быть легко доступной и понятной для пользователей без особых навыков IT.
8. Требования к OLAP-системе "разделяемая" означают, что она должна обеспечивать возможность доступа к данным нескольким пользователям одновременно и поддерживать многопользовательскую конкурентность. OLAP-система должна иметь функциональность управления доступом к данным, где каждый пользователь получает доступ только к тем данным, которые ему нужны, и работает только с той частью данных, которая ему разрешена. OLAP-система должна также поддерживать возможность совместной работы и обмена данными между пользователями, чтобы облегчить процесс принятия бизнес-решений, основанных на общих датах и выводах.
9. Требования к OLAP-системе под рубрикой "многомерная" означают, что она должна поддерживать анализ данных в нескольких измерениях и масштабах одновременно, чтобы обеспечивать глубокое и всестороннее исследование данных. OLAP-система должна поддерживать множество измерений (например, продукт, регион, дата) и осуществлять многомерный анализ по пересечению этих измерений. Также система должна иметь возможность создавать сводные таблицы, диаграммы и графики, позволяющие сравнивать и анализировать данные в разных измерениях и контекстах. Кроме того, OLAP-система должна обладать высокой скоростью работы, эффективностью и надёжностью, чтобы оптимизировать процесс принятия бизнес-решений.
10. Требования к OLAP-системе под рубрикой "информация" означают, что она должна предоставлять высококачественную и точную информацию о бизнес-процессах компании. OLAP-система должна иметь возможность обрабатывать и хранить данные из различных источников, агрегировать их и создавать централизованное хранилище данных для дальнейшего анализа. OLAP-система также должна обеспечивать достоверность данных и защиту информации от несанкционированного доступа. Наконец, система должна предоставлять разные виды отчётов и диаграмм, чтобы сделать информацию более доступной и понятной для разных пользователей в организации.
11. Мультидименсиональные OLAP-системы (MOLAP) - используют ячеистые кубы для хранения данных и быстрого анализа.

Реляционные OLAP-системы (ROLAP) - используют реляционные базы данных для хранения данных.

Гибридные OLAP-системы (HOLAP) - комбинация MOLAP и ROLAP, использующая кубы для хранения наиболее часто используемых данных, а для остальных данных - реляционные базы.

OLAP на основе облачных вычислений - данные хранятся и обрабатываются в облачных хранилищах и вычислительных мощностях, что обеспечивает масштабируемость и гибкость системы.

Различия между ними заключаются в способах хранения данных, взаимодействии пользователей с данными (например, наличие интерактивных панелей и отчётов) и возможностях обработки и анализа данных.

1. Мультидименсиональная OLAP-система (MOLAP) является типом OLAP-системы, основанной на ячеистых кубах, которые позволяют быстро хранить и анализировать многомерную информацию. Данные в MOLAP-системах обычно представлены в виде огромных кубов, позволяющих быстро выполнять операции агрегации, фильтрации и преобразований данных по разным измерениям (например, по времени, продуктам, регионам и пр.). Пользователи могут взаимодействовать с данными через интуитивно понятные интерактивные интерфейсы, получая быстрые и точные ответы на запросы. MOLAP-системы часто используются в бизнес-аналитике для прогнозирования и планирования бизнес-операций.
2. Реляционная OLAP-система (ROLAP) – это тип OLAP-систем, который использует реляционные базы данных для хранения больших объемов информации и быстрого анализа данных. Иными словами, данные хранятся в виде таблиц со строками и столбцами, как в обычных базах данных.
3. Гибридная OLAP-система (HOLAP) - это тип OLAP-системы, который является комбинацией мультидименсиональных (MOLAP) и реляционных (ROLAP) OLAP-технологий. HOLAP-система использует кубы данных для хранения и быстрого анализа наиболее часто используемых данных, а для менее часто используемых данных, используется реляционная база данных. Это позволяет найти баланс между быстротой доступа к данным и гибкостью анализа на больших объемах информации. HOLAP-система используется для анализа данных на зависимости между большим количеством измерений (например, по времени, продуктам, регионам и пр.), и для получения полного представления о бизнес-процессах.
4. Процедура сечения - позволяет отобразить данные на основе выбранных условий. Например, можно выбрать данные только для определенных месяцев, продуктов, географических регионов и т.д. Это позволяет более точно сфокусироваться на конкретной информации и вывести ее на экран.

Процедура поворота - позволяет отобразить данные в соответствии с необходимыми условиями и измерениями. Например, можно выбрать данные о продажах за определенные месяца, в определенной географической зоне и лишь по некоторым товарам. Это позволяет визуализировать данные в разных контекстах и получить более широкое понимание бизнес-процессов и трендов в них.

1. Процедура свертки позволяет объединить данные на более высоком уровне агрегации. Например, можно объединить данные по году продаж, чтобы увидеть суммарные данные за год. Свертка позволяет получить общую статистику и тенденции по всей иерархии данных и найти основные факторы, влияющие на результаты.

Процедура развертки позволяет детализировать данные на более низкий уровень агрегации. Например, можно просмотреть данные о продажах по месяцам, чтобы лучше понимать динамику продаж и выявить сезонность. Развертка позволяет узнать дополнительную информацию об интересующих данных и найти причинно-следственные связи между различными факторами в бизнес-процессах.

1. Процедура проекции (projection) позволяет сравнить данные на основе разных временных периодов. Например, можно сравнить данные о продажах в текущем месяце с прошлогодним месяцем, чтобы оценить эффективность текущих продаж и найти пути для улучшения.

Процедура построения трендов (trend analysis) позволяет проанализировать данные на основе изменения тенденций во времени. Например, можно проанализировать данные о продажах за последние несколько месяцев, чтобы определить, есть ли рост или спад и дать прогноз на будущее. Построение трендов позволяет получить полную картину бизнес-процессов, определить главные факторы, влияющие на результаты и принять соответствующие меры.

1. Подсистема интеллектуального анализа данных (Data Mining) предназначена для автоматического поиска в больших объёмах данных скрытых, ранее неизвестных, но практически важных шаблонов и закономерностей. Data Mining используется для выявления неочевидных связей между данными, прогнозирования будущих событий, определения наиболее значимых факторов влияющих на результаты, построения рекомендательных систем и т.д. Целью использования этой подсистемы является получение нового знания и повышение эффективности принятия решений в различных сферах бизнеса, науки, медицины и других областей деятельности.
2. Кластерный анализ - метод, организующий данные в группы схожих объектов.

Классификация - метод, позволяющий определить к какому классу относится каждый объект на основе системы правил.

Ассоциативный анализ - метод, позволяющий искать связи между объектами и атрибутами, находить наборы связанных объектов (таких, как товары, которые часто покупают вместе), и строить рекомендательные системы.

Регрессионный анализ - метод, используемый для прогнозирования будущих значений на основе имеющихся данных.

Деревья решений - метод, позволяющий искать зависимости между объектами и атрибутами и предсказать возможные исходы.

Нейронные сети - метод, используемый для моделирования сложных зависимостей между данными и выявления скрытых паттернов.

Генетические алгоритмы - методы, используемые для оптимизации сложных систем.

1. Знания - это информация, которая была интерпретирована, понята и находится в контексте, что позволяет ей быть использованной в принятии решений, планировании действий, решении задач и выполнении других процессов.
2. Знания объективные или «фактические» - это информация о реальном мире, которая может обьективироваться.

Знания процедурные - это информация о том, как выполнять определенные действия и задачи.

Знания концептуальные - это информация о том, как объекты и явления связаны между собой и как они могут быть классифицированы и описаны.

Знания метапознания - это информация о том, как мы знаем то, что мы знаем, или как мы могли бы улучшить наши знания или подходы к решению проблем.

Знания эвристические - это информация, которая позволяет нам принимать решения в неопределенной ситуации, используя интуицию и опыт.

Знания контекстуальные - это информация, которая зависит от ситуации и может быть полезной только в определенных условиях.

Знания экспертные - это информация, которая может быть использована для принятия решений в конкретной области и получена опытом и знаниями экспертов в этой области.

1. Методы интеллектуального анализа данных (ИАД) предназначены для обработки больших объемов данных и поиска скрытых зависимостей и трендов. Они отличаются от традиционных методов статистического анализа тем, что могут работать с неструктурированными и полуструктурированными данными.
2. Кластерный анализ, Анализ ассоциаций, Классификационный анализ, Регрессионный анализ, Машинное обучение, Нейронные сети, Генетические алгоритмы, Байесовские сети, Анализ текста и данных в интернете.
3. Методы нечёткой логики позволяют работать с нечёткими или неопределёнными данными, используя математические методы исчисления нечёткости. Это позволяет учитывать неопределенность и размытость в данных и принимать решения на основе степени затуманенности описываемого явления.

Системы рассуждений на основе аналогичных случаев используют накопленный опыт для решения новых проблем. Они анализируют ситуацию, выделяют сходства с известными ранее случаями, находят релевантный опыт и применяют его для решения текущей проблемы.

1. Нейронные сети являются структурами, моделирующими работу человеческого мозга. Они состоят из многих соединенных между собой нейронов, которые анализируют входные данные и передают сигналы далее по сети для принятия решений. Обучение нейронных сетей состоит в настройке весовых коэффициентов соединений между нейронами, чтобы минимизировать ошибки при прогнозировании или классификации.

Генетические алгоритмы - это эволюционные алгоритмы, имитирующие процесс естественного отбора. Они используются для нахождения оптимальных решений в сложных системах. Генетические алгоритмы включают в себя применение операторов скрещивания, мутации и отбора, которые помогают получить новые комбинации признаков, наиболее подходящие для решения конкретной задачи.

1. Методы ассоциаций позволяют находить связи между различными элементами исследуемых данных. Они выявляют поведенческие или пространственные закономерности в наборах данных, используя алгоритмы, основанные на измерении частоты совокупности различных элементов.

Методы кластеризации используются для группировки элементов данных по определенной логике или правилам. Они находят схожие объекты или события, группируют их в кластеры и создают модели для этих кластеров. Это позволяет более эффективно анализировать данные, выявлять тенденции и делать выводы.

Методы классификации используются для определения принадлежности объекта к одному из заранее определенных классов. Они используются для решения задачи распознавания образов и машинного обучения. Алгоритмы классификации используются для обучения систем ИАД на основе наборов данных, с помощью которых система может классифицировать новый объект в соответствии с известными шаблонами и образами.

1. Методы эволюционного программирования основаны на идеях эволюции в природе. Они используют методы генетических алгоритмов для создания программных решений, которые эволюционируют от поколения к поколению. Алгоритмы проводят отбор наилучших решений и комбинируют их для создания новых поколений программ.

Алгоритмы ограниченного перебора используются для поиска оптимального решения путем перебора всех возможных комбинаций элементов. Однако для сокращения времени поиска алгоритмы ограниченного перебора используют эффективные алгоритмы и структуры данных, чтобы уменьшить количество комбинаций, которые должны быть проверены.

1. Прогнозирование, классификация, кластеризация, ассоциативный анализ, анализ текстов, обработка изображений, анализ социальных сетей, отбор признаков, контроль качества, изучение поведения потребителей.
2. Финансы и банковское дело, здравоохранение, интернет-бизнес, промышленность, транспорт, экология, государственное управление. Цели включают улучшение производительности, сокращение расходов, предсказание и предотвращение аварий, улучшение качества продукции или определение наилучшего решения в сложных проблемах.

**II. Тесты**

1. Различают следующие виды функционирования информационно-аналитических систем по режиму и темпу:
   1. статический;
   2. статистический;
   3. динамический;
   4. выборочный.

Выберите правильные ответы.

Ответ: 1(статический) и 3 (динамический).

1. Поддержка принятия управленческих решений осуществляется в следующих режимах или базовых сферах:
   1. сфера детализированных данных;
   2. сфера агрегированных показателей;
   3. сфера взаимосвязей показателей;
   4. сфера закономерностей. Найдите неточный ответ.

Ответ: нет неточного ответа. Все 4 сферы являются базовыми.

1. В …. Году Эдвард Кодд сформулировал … основных требований к средствам реализации оперативного OLAP-анализа.

Вставьте недостающие числа.

Ответ:В 1993 году Эдвард Кодд сформулировал 12 основных требований к средствам реализации оперативного OLAP-анализа.

1. В конце 90-х годов получил распространение свод требований к информационноаналитическим системам в виде «теста FASMI». Аббревиатура каких английских слов с русским переводом является названием теста? Ответ:Forms, Attributes, Selection, Metrics, Items
2. Оперативный анализ — это функция ИАС, обеспечивающая:

Дайте определение. Ответ: это функция ИАС, обеспечивающая быстрый и актуальный доступ к информации, которая необходима пользователю для принятия оперативных решений.

1. Извлечение информации, как правило, сопровождается обработкой ее по несложным алгоритмам, как-то:
   1. производится суммаризация;
   2. определение процентов от заданных величин;
   3. получение относительных показателей;
   4. выявление закономерностей;
   5. вычисление величин с заданными коэффициентами;
   6. определение взаимозависимостей;
   7. другие действия над данными на разных уровнях детализации. Определите не свойственные OLAP-анализу функции.

Ответ: 5. вычисление величин с заданными коэффициентами.

1. Извлечение необходимой информации для построения отчетов производится путем использования ряда процедур, в том числе:
   1. сечение или срез (slice and dice);
   2. поворот;
   3. свертка (drill up);
   4. развертка или раскрытие (roll up);
   5. проекция;
   6. создание кросс-таблиц;
   7. построение трендов.

Какая из процедур обеспечивает детализацию данных? Какая процедура противоположна 4-ой? Какая процедура производит манипуляции с координатами?

Ответ: Детализация данных обеспечивается процедурой "развертка" или "раскрытие". Противоположной процедурой является "свертка" или "drill up". Манипуляции с координатами производятся с помощью процедуры "поворот".

* 1. Различают три типа многомерных OLAP-систем:
     1. многомерный (Multidimensional) OLAP — MOLAP
     2. реляционный (Relation) OLAP — ROLAP
     3. смешанный или гибридный (Hibrid) OLAP — HOLAP.

Ответ: Более дешевая реализация OLAP-системы по типу ROLAP.

Скорость доступа к данным выше в у какого типа OLAP-систем? (MOLAP).

Наиболее целесообразно использование типа HOLAP.

* 1. Подсистема интеллектуального анализа данных (Data mining) предназначена для: автоматического извлечения скрытых, ранее неизвестных и полезных знаний из больших объемов данных.
  2. По определению Гавриловой Т.А. и Хорошевского В.Ф.[17], знания это: совокупность фактов, концепций, идей и общих принципов, которые лежат в основе деятельности человека и культуры в целом.
  3. Рассматриваются следующие виды знаний:
     1. фактические и стратегические;
     2. факты и эвристики;
     3. декларативные и процедурные;
     4. интенсиональные и экстенсиональные;
     5. глубинные и поверхностные;
     6. жесткие и мягкие.

Можно ли увязать фактические, жесткие знания, факты? Эвристики и мягкие знания имеют какую либо аналогию? Имеют ли какую либо схожесть декларативные и мягкие знания?

Ответ: Типы знаний несопоставимы между собой. Фактические, жесткие знания - это совокупность точных, измеримых фактов, в то время как эвристики и мягкие знания базируются на эмпирических наблюдениях, опыте и интуиции. Декларативные и мягкие знания также имеют существенные различия, так как первые базируются на заявлениях, а вторые - на эмоциях, чувствах и личном опыте.

* 1. Подсистема интеллектуального анализа ИАС поддерживает инструментальными средствами:
     1. математическую экономику;
     2. эконометрику;
     3. специфические методы Data mining.

Математическая экономика поддерживает методы многомерного статистического анализа? Используют ли специфические методы Data mining понятия моментов случайных величин? Поисковый метод «генетические алгоритмы» входит в арсенал средств Data mining?

Ответ: Подсистема интеллектуального анализа данных поддерживает инструментальными средствами специфические методы Data mining, которые могут включать в себя генетические алгоритмы, анализ ассоциаций и многое другое. Математическая экономика также используется для анализа данных, но методы многомерного статистического анализа не всегда являются подходящими.