

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ВЛАДИМИРА ДАЛЯ»

Экономический институт
Кафедра экономической кибернетики и прикладной статистики
(наименование кафедры)

УТВЕРЖДАЮ
Директор

- 6 -

« 28 » февраля 20 года



ФОНД ОЦЕНОЧНЫХ СРЕДСТВ по учебной дисциплине (практике)

Методы поиска структурированной и неструктурированной информации

(наименование учебной дисциплины, практике)

38.03.05 Бизнес-информатика

(код и наименование направления подготовки (специальности))

Экономическая аналитика и бизнес-статистика

(назначение профиля подготовки (специальности, магистерской программы): при отсутствии ставится прочерк)

Разработчик (разработчики):

Разработчик (разработчики):

доцент

Спорняк С.А.

старший преподаватель

Лофиченко А.А.

ФОС рассмотрен и одобрен на заседании кафедры экономической кибернетики и прикладной статистики от « 15 » 02 2025 г., протокол № 15

Заведующий кафедрой экономической кибернетики и прикладной статистики

Велигуря А.В.

(подпись)

Луганск 2025 г.

Комплект оценочных материалов по дисциплине
«Методы поиска структурированной и неструктурированной информации»

Задания закрытого типа

Задания закрытого типа на выбор правильного ответа

1. Выберите один правильный ответ.

Информация – это:

А) сведения об объектах и явлениях окружающей среды, их параметрах, свойствах и состоянии, которые уменьшают имеющуюся о них степень неопределенности или неполноты знаний;

Б) организованный социально-экономический и научно-технический процесс;

В) отыскание рациональных решений в любой сфере;

Г) процесс, использующий совокупность средств и методов сбора, обработки и передачи данных.

Правильный ответ: А.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

2. Выберите один правильный ответ.

Какая модель имеет структуру в виде дерева и выражает вертикальные связи подчинения нижнего уровня высшему: относительное изменение какого-либо явления во времени;

А) сетевая;

Б) иерархическая;

В) реляционная.

Правильный ответ: Б.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

3. Выберите один правильный ответ.

Данные превращаются в информацию в том случае, если появляется возможность:

А) использовать их для уменьшения неопределенности о чем-либо;

Б) использовать их для увеличения неопределенности о чем-либо;

В) использовать их для вычислений.

Правильный ответ: А.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

4. Выберите один правильный ответ.

Для решения хорошо структурированных задач, по которым имеются необходимые входные данные и известны алгоритмы и другие стандартные процедуры их обработки предназначена:

- А) информационная технология обработки данных;
- Б) информационная технология анализа и регулирования;
- В) информационная технология автоматизированного офиса.

Правильный ответ: А

Компетенции (индикаторы): ПК-5 (ПК-5.1).

5. Выберите один правильный ответ.

Содержит ли какую-либо информацию таблица, в которой нет ни одной записи?

- А) пустая таблица не содержит никакой информации;
- Б) пустая таблица содержит информацию о структуре базы данных; +
- В) пустая таблица содержит информацию о будущих записях;
- Г) таблица без записей существовать не может.

Правильный ответ: Б.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

Задания закрытого типа на установление соответсвия

1. Установите соответствие между поисковыми системами и ее описанием. Каждому элементу левого столбца соответствует только один элемент правого столбца.

- | | |
|------------------------|--|
| 1) CRAWLER. | А) (система выдачи результатов): извлекает результаты поиска из базы данных |
| 2) SPIDER | Б) (паук). браузероподобная программа, которая скачивает веб-страницы |
| 3) Indexer | В) (индексатор): «слепая» программа, которая анализирует веб-страницы, скаченные пауками. |
| 4) Thedatabase | Г) (база данных): хранилище скаченных и обработанных страниц. |
| 5) Searchengineresults | Д) «путешествующий» паук, который автоматически идет по всем ссылкам, найденным на странице. |

Правильный ответ: 1-Д; 2-Б; 3-В; 4-Г; 5-А.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

2. Установите соответствие между коэффициентами связи и их применением. Каждому элементу левого столбца соответствует только один элемент правого столбца.

- | | |
|---------------------------------|--|
| 1). Коэффициент контингенции | А) Используется для измерения связи между двумя номинальными переменными, которые могут иметь более двух значений. |
| 2). Коэффициент Чупрова-Крамера | Пример: Связь между образованием (высшее/среднее/начальное) и занятостью (работающий/безработный/учащийся). |
| 3). Коэффициент ассоциации | |

- 4) Коэффициент корреляции
- Б) Является частным случаем коэффициента Чупрова-Крамера и используется для измерения связи между двумя номинальными переменными, каждая из которых имеет два значения.
 Пример: Связь между полом (мужской/женский) и участием в спортивных мероприятиях (участвует/не участвует).
- В) Используется для измерения связи между двумя номинальными переменными, имеющими два значения (дихотомические переменные).
 Пример: Связь между полом (мужской/женский) и курением (курящий/некурящий).
- Г) Используется для измерения силы и направления линейной связи между двумя количественными переменными.
 Пример: Связь между возрастом и доходом.
- Д) Используется для измерения изменений в структуре какого-либо показателя или процесса.
 Пример: оценка изменений в доле различных отраслей промышленности в валовом внутреннем продукте (ВВП) страны.

Правильный ответ: 1- В; 2-А; 3-Б; 4-Г.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

3. Установите соответствие между алгоритмами и их определениями.
Каждому элементу левого столбца соответствует только один элемент правого столбца.

- | | |
|--|--|
| 1) Алгоритм Apriori | А) это метод кластеризации, который разделяет данные на k кластеров. Он пытается минимизировать внутрикластерное рассеивание, перемещая центры кластеров до тех пор, пока они не стабилизируются. |
| 2) Алгоритм K-means | Б) это метод кластеризации, который последовательно объединяет объекты в кластеры на основе их близости. |
| 3) Алгоритм C-means (Fuzzy C-means) | В) это разновидность алгоритма K-means, но с возможностью частичной принадлежности объектов к нескольким кластерам. Каждый объект имеет степень принадлежности к каждому кластеру, выражаемую в диапазоне от 0 до 1. |
| 4) Иерархический алгоритм | Г) используется для поиска ассоциативных правил в больших наборах данных. Он основан на принципе подъема (lift) и ищет наборы элементов, которые часто встречаются вместе в транзакциях. |

Правильный ответ: 1- Г; 2-А; 3-В; 4-Б.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

4. Установите соответствие между методами масштабирования и их определения (назначения). Каждому элементу левого столбца соответствует только один элемент правого столбца.

- | | |
|-----------------------------------|--|
| 1) Горизонтальное масштабирование | A) увеличение мощности одного узла путем добавления большего количества ресурсов (процессоры, оперативная память, хранилище). |
| 2) Вертикальное масштабирование | B) способ увеличения вычислительной мощности системы путем добавления дополнительных узлов (серверов, компьютеров) в существующую инфраструктуру. |
| | B) симметричный механизм репликации, в котором каждый узел сети имеет равные права на чтение и запись данных. Все узлы взаимодействуют друг с другом, и изменения, вносимые на любом узле, распространяются на все остальные узлы. |

Правильный ответ: 1- Б; 2-А.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

5. Установите соответствие между методами репликации и их определения (назначения). Каждому элементу левого столбца соответствует только один элемент правого столбца.

- | | |
|----------------------------|---|
| 1) Master-slave репликация | A) это симметричный механизм репликации, в котором каждый узел сети имеет равные права на чтение и запись данных. Все узлы взаимодействуют друг с другом, и изменения, вносимые на любом узле, распространяются на все остальные узлы. Такой подход обеспечивает высокую отказоустойчивость и гибкость, но требует сложного механизма согласования изменений. |
| 2) Peer-to-peer репликация | B) увеличение мощности одного узла путем добавления большего количества ресурсов (процессоры, оперативная память, хранилище). |
| | B) это модель асимметричного реплицирования данных, в которой один главный узел (master) управляет записью данных, а подчиненные узлы (slaves) используются для чтения данных. Изменения, внесенные на master-узле, автоматически распространяются на slaves. Такая модель помогает повысить доступность данных и снизить нагрузку на основной сервер. |

Правильный ответ: 1- В; 2-А.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

Задания закрытого типа на установление правильной последовательности

1. Установите правильную последовательность этапов выполнения вертикального масштабирования. Запишите правильную последовательность букв слева направо:

- А) Анализ текущих потребностей и ограничений
- Б) Выбор подходящего оборудования
- В) Резервное копирование данных
- Г) Установка нового оборудования
- Д) Настройка и конфигурирование
- Е) Тестирование и проверка
- Ж) Мониторинг и оптимизация
- З) Документирование изменений

Правильный ответ: А, Б, В, Г, Д, Е, Ж, З.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

2. Установите правильную последовательность этапов выполнения горизонтального масштабирования. Запишите правильную последовательность букв слева направо:

- А) Мониторинг и оптимизация
- Б) Выбор архитектуры и платформы
- В) Разделение нагрузки
- Г) Добавление новых узлов
- Д) Синхронизация данных
- Е) Тестирование и проверка
- Ж) Анализ текущих потребностей и ограничений

Правильный ответ: Ж, Б, В, Г, Д, Е, А.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

3. Установите правильную последовательность этапов проведения алгоритма Apriori. Запишите правильную последовательность букв слева направо.

- А) Генерация кандидатных наборов
- Б) Сбор данных и предварительная обработка
- В) Подсчет поддержки
- Г) Фильтрация по минимальной поддержке
- Д) Генерация ассоциативных правил
- Е) Оценка уверенности правил
- Ж) Фильтрация по минимальной уверенности
- З) Анализ и интерпретация результатов

Представление результатов группировки в табличной форме.

Правильный ответ: Б, А, В, Г, Д, Е.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

4. Установите правильную последовательность этапов проведения алгоритма *Apriori*. Запишите правильную последовательность букв слева направо.

- А) Настройка подчинённых узлов (Slaves)
 - Б) Выбор главного узла (Master)
 - В) Инициализация репликации
 - Г) Постоянная синхронизация данных
 - Д) Управление конфликтами
 - Е) Мониторинг и диагностика
 - Ж) Резервное копирование и восстановление
- 3) Представление результатов группировки в табличной форме
- И) Обслуживание и поддержка

Правильный ответ: Б, А, В, Г, Д, Ж, И, З.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

Задания открытого типа

Задания открытого типа на дополнение

1. Напишите пропущенное слово.

В общем виде _____ данные можно классифицировать на две категории:

- генерируемые людьми неструктурированные данные, к которым относятся различные виды создаваемого людьми контента: текстовые документы, электронные письма, посты в соцсетях, изображения, видео и так далее;

- машиногенерируемые неструктурированные данные создаются устройствами и датчиками, это файлы журналов, данные GPS, результаты работы Internet of Things (IoT) и другая телеметрическая информация.

Правильный ответ: неструктурированные.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

2. Напишите пропущенное слово.

Какие свойства должна иметь система для хранения неструктурированных данных?

Веб-страницы могут быть статичными или создаваться «на лету» в ответ на действия пользователя, и в таком случае они, вероятно, будут содержать информацию из множества источников. В обоих случаях программа может читать веб-страницу и извлекать ее части. Это называется _____.

Правильный ответ: веб-скрейпинг (web scraping).
Компетенции (индикаторы): ПК-5 (ПК-5.1).

3. Напишите пропущенное слово.

Основные принципы определения _____ следующие: Количество слов запроса в текстовом содержимом документа (т.е. в html-коде). Тэги, в которых эти слова располагаются. Местоположение искомых слов в документе. Удельный вес слов, относительно которых определяется релевантность, в общем количестве слов документа. Время - как долго страница находится в базе поискового сервера. Индекс цитируемости - как много ссылок на данную страницу ведет с других страниц, зарегистрированных в базе поисковика.

Правильный ответ: релевантности.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

4. Вставьте пропущенное слово.

С помощью алгоритма _____ можно найти ассоциативное правило?

Правильный ответ: *a priori*.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

Задания открытого типа с кратким свободным ответом

1. Вставьте пропущенное словосочетание.

Выделение таблицы или группы таблиц на отдельный сервер это _____.

Правильный ответ: Вертикальный шардинг.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

2. Вставьте пропущенное словосочетание.

Технология машинного обучения, когда нет ответов и требуется искать зависимости между объектами, называется _____.

Правильный ответ: Обучение без учителя.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

3. Напишите пропущенное словосочетание.

_____ обеспечивают гибкий и экономный способ управления неструктуризованными данными и их хранения с высокой надёжностью и доступностью. Они могут хранить большие объёмы сырых данных в их нативном формате, позволяя организациям выполнять аналитику big data и в то же время предоставляя возможность преобразования и интеграции данных с различными инструментами и платформами.

Правильный ответ: Озёра данных.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

Задания открытого типа с развернутым ответом

1. Решите задачу, связанную с методами поиска информации,

Компания хочет выпустить новый продукт — умные часы. Она уже провела небольшое исследование и узнала, что потенциальные клиенты заинтересованы в функциях мониторинга здоровья и уведомлениях. Однако компания хочет уточнить, какие именно функции мониторинга здоровья наиболее важны для потребителей, и понять, готовы ли они платить за премиум-версию с дополнительными функциями.

Привести расширенное решение.

Время выполнения – 30 мин.

Ожидаемый результат:

1. Постановка цели поиска информации

Прежде чем начать поиск информации, необходимо чётко определить цель. В данном случае целью является сбор дополнительной информации о предпочтениях потребителей в отношении функций умных часов и их готовности платить за премиальные версии.

2. Определение источников информации

Существует несколько типов источников информации, которые могут быть использованы:

Первичные источники: данные, собранные непосредственно компанией (анкетирование, интервью, фокус-группы).

Вторичные источники: опубликованная информация (исследования, отчёты, статьи, базы данных).

3. Методы поиска информации

Для решения поставленной задачи можно использовать следующие методы:

Анкетирование: разработка анкет для опроса потенциальных клиентов с целью получения информации о их предпочтениях, привычках и потребностях.

Интервью: проведение личных или телефонных интервью с представителями целевой аудитории для глубокого понимания их мнений и поведения.

Фокус-группы: организация дискуссий с небольшой группой представителей целевой аудитории для обсуждения различных аспектов продукта и рынка.

Анализ вторичных источников: изучение уже имеющихся исследований, отчётов и статей, касающихся целевого сегмента рынка.

Онлайн-исследования: использование социальных сетей и специализированных платформ для сбора данных о поведении потребителей.

4. Разработка инструмента для сбора данных

Для каждого метода поиска информации необходимо разработать соответствующие инструменты:

Анкеты: вопросы должны быть чёткими и конкретными, направленными на получение нужной информации.

Вопросы для интервью: должны быть открыты и стимулировать обсуждение.

План проведения фокус-групп: определение тем для обсуждения и подготовка ведущего.

Список вторичных источников: составление перечня публикаций и отчётов, которые необходимо изучить.

Онлайн-опросы: создание опросов и размещение их на платформах социальных сетей или специализированных сайтах.

5. Сбор данных

После разработки инструментов начинается процесс сбора данных. Это может включать рассылку анкет, проведение интервью, организацию фокус-групп и изучение вторичных источников.

6. Анализ данных

Собранные данные анализируются для выявления закономерностей и подтверждения гипотез. Для этого могут быть использованы статистические методы, анализ текста и контент-анализ.

7. Интерпретация результатов

Полученные результаты интерпретируются в контексте целей компании. На основе анализа делается вывод о том, насколько верны первоначальные гипотезы и какие дополнительные выводы можно сделать о целевой аудитории.

8. Формулирование рекомендаций

На основе интерпретации результатов формулируются рекомендации для компании. Они могут включать изменения в продукте, маркетинговую стратегию или дальнейшие исследования.

Ответ: разработана пошаговая методика поиска и анализа информации о функциях мониторинга здоровья потребителей и о готовности оплачивания премиум-версии на основании которой формулируются рекомендации для компаний.

Критерии оценивания:

– разработан пошаговый алгоритм поиска и анализа информации о функциях мониторинга здоровья потребителей и о готовности оплачивания премиум-версии в которую входит: определение цели, методов, источников, инструментов сбора информации; непосредственно сбор и анализ данных; интерпретация результатов и формулирование рекомендаций.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

2. Решите задачу с на определение релевантности полученных данных.

Система поиска возвращает 100 документов в ответ на запрос пользователя, и после ручной проверки оказалось, что 80 из них действительно релевантны запросу.

Чему равен коэффициент релевантности?

Привести расширенное решение.

Время выполнения – 10 мин.

Ожидаемый результат:

Формула для определения коэффициента релевантности:

$$R = Nr/Nt$$

где R – коэффициент релевантности,

Nr – количество релевантных документов,

Nt – общее количество возвращенных документов.

$$R = 100/80 = 0,8$$

Ответ: Коэффициент релевантности равен 0,8, что означает, что 80% возвращенных документов соответствуют запросу пользователя.

Критерии оценивания:

– приведение формулы и расчет коэффициента релевантности.

Компетенции (индикаторы): ПК-5 (ПК-5.1).

Экспертное заключение

Представленный фонд оценочных средств (далее – ФОС) по дисциплине «Методы поиска структурированной и неструктурированной информации» соответствует требованиям ФГОС ВО.

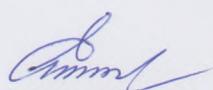
Предлагаемые формы и средства текущего и промежуточного контроля адекватны целям и задачам реализации основной профессиональной образовательной программы по направлению подготовки 38.03.05 Бизнес-информатика.

Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины представлены в полном объеме.

Виды оценочных средств, включенные в представленный фонд, отвечают основным принципам формирования ФОС.

Разработанный и представленный для экспертизы фонд оценочных средств рекомендуется к использованию в процессе подготовки обучающихся по указанному направлению.

Председатель учебно-методической комиссии экономического института



Шаповалова Е.Н.

Лист изменений и дополнений

| № п/п | Виды дополнений и изменений | Дата и номер протокола заседания кафедры (кафедр), на котором были рассмотрены и одобрены изменения и дополнения | Подпись (с расшифровкой) заведующего кафедрой (заведующих кафедрами) |
|------------------|--|---|---|
| | | | |
| | | | |
| | | | |
| | | | |