

**КАЗАНСКИЙ КООПЕРАТИВНЫЙ ИНСТИТУТ (ФИЛИАЛ)
АВТОНОМНОЙ НЕКОММЕРЧЕСКОЙ ОБРАЗОВАТЕЛЬНОЙ
ОРГАНИЗАЦИИ ВЫСШЕГО ОБРАЗОВАНИЯ
ЦЕНТРОСОЮЗА РОССИЙСКОЙ ФЕДЕРАЦИИ
«РОССИЙСКИЙ УНИВЕРСИТЕТ КООПЕРАЦИИ»**

ОЦЕНОЧНЫЕ МАТЕРИАЛЫ

**БЕЗОПАСНЫЕ МЕТОДЫ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ
СЕТЕВЫХ ИСТОЧНИКОВ**

Специальность: 10.05.01 Компьютерная безопасность

Специализация: «Разработка защищенного программного обеспечения»

Формы обучения: очная

Объем дисциплины:

в зачетных единицах: 4 з.е.

в академических часах: 144 ак.ч.

Форма промежуточной аттестации: экзамен

Оценочные материалы по дисциплине Безопасные методы извлечения информации из сетевых источников

обсуждены и рекомендованы к утверждению решением кафедры естественных дисциплин, сервиса и туризма от 25 марта 2025 г., протокол № 6

одобрены Научно-методическим советом Казанского кооперативного института (филиала) от 2 апреля 2025 г., протокол № 3

СОДЕРЖАНИЕ

1. Паспорт оценочных материалов по дисциплине
2. Оценочные материалы по дисциплине:
 - 2.1. Оценочные материалы для проведения текущего контроля.
 - 2.2. Оценочные материалы для проведения промежуточной аттестации.

Обновление оценочных материалов дисциплины

Целью оценочных материалов по дисциплине (модулю) (далее –ОМ) является комплексная оценка результатов обучения по дисциплине Безопасные методы извлечения информации из сетевых источников и установление уровня сформированности компетенций обучающегося.

ОМ предназначены для проведения текущего контроля успеваемости и промежуточной аттестации.

ОМ включают оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации.

Оценочные материалы разработаны в соответствии с рабочей программой дисциплины Безопасные методы извлечения информации из сетевых источников.

1. Паспорт оценочных материалов по дисциплине «Безопасные методы извлечения информации из сетевых источников»

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения	Контролируемые разделы, темы дисциплины	Наименование оценочного средства ¹	
				Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6
ОПК-10. Способен анализировать тенденции развития методов и средств криптографической защиты информации, использовать средства криптографической защиты информации при решении задач профессиональной деятельности	ОПК-10.1. Применяет в профессиональной деятельности знания об основных алгоритмах, используемых в криптографии, а также методы оценивания сложности таких алгоритмов	Знать: Основные этапы научного исследования и их содержание Уметь: Разрабатывать план научного исследования с учетом целей и задач. Владеть: Навыками работы с научными базами данных и информационными ресурсами.	Тема 1. Принципы работы поисковых движков. Тема 2. Методологии безопасного сбора данных из сетевых источников	Реферат (Тема 1-2) Тест (тема 1-2)	Тест
	ОПК-10.2. Демонстрирует при решении задач профессиональной деятельности умение использовать основные преобразования симметричной и асимметричной	Знать: Основные принципы управления проектами и их применение в науке. Уметь: Составлять графики выполнения проектов и определять ключевые этапы. Владеть: Компетенциями в ведении документации и отчетности по проектам.	Тема 3. Типы информационных систем. Оценка эффективности и безопасности работы поисковых систем. Тема 4. Модели и технологии децентрализованного поиска	Реферат (Тема 3-4) Тест (Тема 3-4)	Тест

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения	Контролируемые разделы, темы дисциплины	Наименование оценочного средства ¹	
				Текущий контроль	Промежуточная аттестация
1	2	3	4	5	6
	криптографии и анализировать их криптостойкость к возможным компьютерным атакам				
	ОПК-10.3. Демонстрирует умения разработки криптографически х протоколов и исследования их стойкости к различным атакам в составе средств защиты компьютерных систем	Знать: Основы командной работы и принципы эффективного взаимодействия. Уметь: Работать в команде, учитывая мнения и идеи других участников. Владеть Компетенциями в использовании цифровых инструментов для совместной работы (например, Zoom, Slack, Google Docs).	Тема 5. Способы обработки и хранения больших данных в сетевых источниках	Реферат (Тема 5) Тест (Тема 5)	Тест

2. Оценочные материалы по дисциплине «Безопасные методы извлечения информации из сетевых источников»

2.1 Оценочные материалы для проведения текущего контроля успеваемости

Тематика рефератов

Тема 1 Принципы работы поисковых движков.

1. Всемирная паутина.
2. Развитие интернет в XXI веке.
3. Организационная структура Интернета.
4. Схема адресации в сети Интернет.
5. Модель BOW TIE.
6. Понятия и различия WEB 2.0- WEB 4.0.
7. Структура WEB, Deep WEB, его возможности и характеристики.
8. Инструменты и технологии работы в невидимом WEB.;

Тема 2 Методологии безопасного сбора данных из сетевых источников.

1. Технологии извлечения знаний из WEB - WEB-mining.
2. Задачи и этапы извлечения знаний из WEB.
3. Направления WEB-mining: Извлечение Web-контента (Web Content Mining);
4. Извлечение Web-структур (Web Structure Mining);
5. Исследование использования Web-ресурсов (WebUsage Mining)
6. Понятие data scraping или «срезание данных с поверхности».
7. Классификация способов извлечения информации из WEB-источников.
8. Модели информационного поиска.
9. Булева модель, векторная модель, вероятностная модель, гибридная модель.
10. Математические особенности обработки информации разными моделями.
11. Сферы их применения.

Тема 3. Типы информационных систем. Оценка эффективности и безопасности работы поисковых систем.

1. Понятие поисковой системы.
2. Методы лингвистического анализа.
3. Методы ранжирования поисковой выдачи.
4. Виды поисковых роботов.
5. Порядок индексации сайтов.
6. Алгоритмы работы поисковых систем Яндекс и Google.
7. Выбор ключевых слов для продвижения сайта.
8. Типы запросов по частотности.
9. Типы запросов по степени конверсии.
10. Создание семантического ядра.
11. Выбор ключевых страниц сайта.
12. Распределение семантического ядра.
13. Выбор основной стратегии поискового продвижения сайта.

Тема 4. Модели и технологии децентрализованного поиска.

1. Организация P2P.
2. Принцип работы Bittorrent.
3. Примеры и способы работы децентрализованных поисковых систем.

Тема 5. Способы обработки и хранения больших данных в сетевых источниках

1. NoSQL хранилища.
2. Требования к хранилищам данных, OLTP и OLAP системы.
3. Нереляционные базы данных.
4. Принципы работы и примеры использования Map-Reduce.

Критерии оценки выполнения рефератов по учебной дисциплине

Оценка «отлично» выставляется студенту, если тема реферата раскрыта в полной мере и все требования по написанию реферата соблюдены.

Оценка «хорошо» выставляется студенту, если недостаточно раскрыта тема реферата, но все требования по написанию реферата соблюдены.

Оценка «удовлетворительно» выставляется студенту, если недостаточно раскрыта тема реферата, не все требования по написанию реферата соблюдены, присутствуют ошибки и неточности в работе.

Оценка «неудовлетворительно» выставляется студенту, если отсутствует реферат.

Рекомендации по написанию реферата:

Реферат — письменная работа, выполняемая обучающимся в течение определенного срока.

Реферат — краткое точное изложение сущности какого-либо вопроса, темы на основе одной или нескольких книг, монографий или других первоисточников. Реферат должен содержать основные фактические сведения и выводы по рассматриваемому вопросу.

Реферат отвечает на вопрос — что содержится в данной публикации (публикациях). Реферат — не механический пересказ работы, а изложение ее сущности.

Кроме реферирования прочитанной литературы, от обучающегося требуется аргументированное изложение собственных мыслей по рассматриваемому вопросу. Тему реферата предлагает преподаватель или сам обучающийся, в последнем случае она должна быть согласована с преподавателем.

В реферате нужны развернутые аргументы, рассуждения, сравнения. Материал подается не столько в развитии, сколько в форме констатации или описания.

Содержание реферируемого произведения излагается объективно от имени автора. Если в первичном документе главная мысль сформулирована недостаточно четко, в реферате она должна быть конкретизирована и выделена.

Структура реферата:

1. Титульный лист.
2. После титульного листа на отдельной странице следует оглавление (план, содержание), в котором указаны названия всех разделов (пунктов плана) реферата и номера страниц, указывающие начало этих разделов в тексте реферата.
3. После оглавления следует введение. Объем введения составляет 1,5-2 страницы.
4. Основная часть реферата может иметь одну или несколько глав, состоящих из 2-3 параграфов (подпунктов, разделов) и предполагает осмысленное и логичное изложение главных положений и идей, содержащихся в изученной литературе. В тексте обязательны ссылки на первоисточники. В том случае если цитируется или используется чья-либо неординарная мысль, идея, вывод, приводится какой-либо цифрой материал, таблицу — обязательно сделайте ссылку на того автора у кого вы взяли данный материал.
5. Заключение содержит главные выводы, и итоги из текста основной части, в нем отмечается, как выполнены задачи и достигнуты ли цели, сформулированные во введении.
6. Приложение может включать графики, таблицы, расчеты.
7. Библиография (список литературы) здесь указывается реально использованная для

написания реферата литература. Список составляется согласно правилам библиографического описания.

Этапы работы над рефератом.

Работу над рефератом можно условно подразделить на три этапа:

1. Подготовительный этап, включающий изучение предмета исследования;
2. Изложение результатов изучения в виде связного текста;
3. Устное сообщение по теме реферата.

1. Подготовительный этап работы.

Формулировка темы. Подготовительная работа над рефератом начинается с формулировки темы. Тема в концентрированном виде выражает содержание будущего текста, фиксируя как предмет исследования, так и его ожидаемый результат. Для того чтобы работа над рефератом была успешной, необходимо, чтобы тема заключала в себе проблему, скрытый вопрос (даже если наука уже давно дала ответ на этот вопрос, студент, только знакомящийся с соответствующей областью знаний, будет вынужден искать ответ заново, что даст толчок к развитию проблемного, исследовательского мышления).

Поиск источников. Грамотно сформулированная тема зафиксировала предмет изучения; задача обучающегося — найти информацию, относящуюся к данному предмету и разрешить поставленную проблему. Выполнение этой задачи начинается с поиска источников. На этом этапе необходимо вспомнить, как работать с энциклопедиями и энциклопедическими словарями.

Работа с источниками. Работу с источниками надо начинать с ознакомительного чтения, т. е. просмотреть текст, выделяя его структурные единицы. При ознакомительном чтении закладками отмечаются те страницы, которые требуют более внимательного изучения. В зависимости от результатов ознакомительного чтения выбирается дальнейший способ работы с источником. Если для разрешения поставленной задачи требуется изучение некоторых фрагментов текста, то используется метод выборочного чтения. Если в книге нет подробного оглавления, следует обратить внимание на предметные и именные указатели.

Избранные фрагменты или весь текст (если он целиком имеет отношение к теме) требуют вдумчивого, неторопливого чтения с «мысленной проработкой» материала. Такое чтение предполагает выделение: 1) главного в тексте; 2) основных аргументов; 3) выводов. Особое внимание следует обратить на то, вытекает тезис из аргументов или нет. Необходимо также проанализировать, какие из утверждений автора носят проблематичный, гипотетический характер и уловить скрытые вопросы.

Понятно, что умение таким образом работать с текстом приходит далеко не сразу. Наилучший способ научиться выделять главное в тексте, улавливать проблематичный характер утверждений, давать оценку авторской позиции — это сравнительное чтение, в ходе которого студент знакомится с различными мнениями по одному и тому же вопросу, сравнивает весомость и доказательность аргументов сторон и делает вывод о наибольшей убедительности той или иной позиции.

Создание конспектов для написания реферата.

Подготовительный этап работы завершается созданием конспектов, фиксирующих основные тезисы и аргументы. Здесь важно вспомнить, что конспекты пишутся на одной стороне листа, с полями и достаточным для исправления и ремарок межстрочным расстоянием (эти правила соблюдаются для удобства редактирования). Если в конспектах приводятся цитаты, то непременно должно быть дано указание на источник (автор, название, выходные данные, № страницы).

По завершении предварительного этапа можно переходить непосредственно к созданию текста реферата.

2. Создание текста.

Общие требования к тексту.

Текст реферата должен подчиняться определенным требованиям: он должен

раскрывать тему, обладать связностью и цельностью.

Раскрытие темы предполагает, что в тексте реферата излагается относящийся к теме материал и предлагаются пути решения содержащейся в теме проблемы; связность текста предполагает смысловую соотносительность отдельных компонентов, а цельность - смысловую законченность текста.

С точки зрения связности все тексты делятся на тексты-констатации и тексты-рассуждения. Тексты-констатации содержат результаты ознакомления с предметом и фиксируют устойчивые и несомненные суждения. В текстах-рассуждениях одни мысли извлекаются из других, некоторые ставятся под сомнение, дается им оценка, выдвигаются различные предположения.

План реферата.

Универсальный план реферата – введение, основной текст и заключение.

Требования к введению.

Во введении аргументируется актуальность исследования, - т. е. выявляется практическое и теоретическое значение данного исследования. Далее констатируется, что сделано в данной области предшественниками; перечисляются положения, которые должны быть обоснованы. Введение может также содержать обзор источников или экспериментальных данных, уточнение исходных понятий и терминов, сведения о методах исследования. Во введении обязательно формулируются цель и задачи реферата.

Объем введения - в среднем около 10% от общего объема реферата.

Основная часть реферата.

Основная часть реферата раскрывает содержание темы. Она наиболее значительна по объему, наиболее значима и ответственна. В ней обосновываются основные тезисы реферата, приводятся развернутые аргументы, предполагаются гипотезы, касающиеся существа обсуждаемого вопроса. Важно проследить, чтобы основная часть не имела форму монолога. Аргументируя собственную позицию, можно и должно анализировать, и оценивать позиции различных исследователей, с чем-то соглашаться, чему-то возражать, кого-то опровергать. Текст основной части делится на главы, параграфы, пункты. План основной части может быть составлен с использованием различных методов группировки материала: классификации (эмпирические исследования), типологии (теоретические исследования), периодизации (исторические исследования).

Заключение.

Заключение — последняя часть научного текста. В ней краткой и сжатой форме излагаются полученные результаты, представляющие собой ответ на главный вопрос исследования. Здесь же могут намечаться и дальнейшие перспективы развития темы. Небольшое по объему сообщение также не может обойтись без заключительной части - пусть это будут две-три фразы. Но в них должен подводиться итог проделанной работы.

Список использованной литературы.

Реферат любого уровня сложности обязательно сопровождается списком используемой литературы. Названия книг в списке располагают по алфавиту с указанием выходных данных использованных книг.

Требования, предъявляемые к оформлению реферата

Объемы рефератов – от 10-18 машинописных страниц. Работа выполняется на одной стороне листа стандартного формата. По обеим сторонам листа оставляются поля размером 35 мм. слева и 15 мм. справа, рекомендуется шрифт 12-14, интервал - 1,5. Все листы реферата должны быть пронумерованы. Каждый вопрос в тексте должен иметь заголовок в точном соответствии с наименованием в плане-оглавлении. При написании и оформлении реферата следует избегать типичных ошибок, например, таких:

- поверхностное изложение основных теоретических вопросов выбранной темы, когда автор не понимает, какие проблемы в тексте являются главными, а какие второстепенными,

- в некоторых случаях проблемы, рассматриваемые в разделах, не раскрывают

основных аспектов выбранной для реферата темы,

– дословное переписывание книг, статей, заимствования рефератов из интернета и т.

д.

При проверке реферата преподавателем оцениваются:

1. Знания и умения на уровне требований стандарта конкретной дисциплины: знание фактического материала, усвоение общих представлений, понятий, идей.

2. Характеристика реализации цели и задач исследования (новизна и актуальность поставленных в реферате проблем, правильность формулирования цели, определения задач исследования, правильность выбора методов решения задач и реализации цели; соответствие выводов решаемым задачам, поставленной цели, убедительность выводов).

3. Степень обоснованности аргументов и обобщений (полнота, глубина, всесторонность раскрытия темы, логичность и последовательность изложения материала, корректность аргументации и системы доказательств, характер и достоверность примеров, иллюстративного материала, широта кругозора автора, наличие знаний интегрированного характера, способность к обобщению).

4. Качество и ценность полученных результатов (степень завершенности реферативного исследования, спорность или однозначность выводов).

5. Использование литературных источников.

6. Культура письменного изложения материала.

7. Культура оформления материалов работы.

Комплект типовых лабораторных заданий

Тема 1 Вводное занятие. Определение комбинаторики и её значимость.

Изучение основных понятий: перестановки, сочетания и размещения.

Примеры задач на подсчет различных комбинаций

Тема 2.

Классификация интернет-сервисов.

Составление безопасных запросов по теме ВКР, выполнение поиска в открытых и закрытых сетевых источниках, сравнение эффективности поиска с помощью различных инструментов.

Тема 3. Системы управления контентом..

Обсуждение преимуществ и недостатков различных CMS, особенностей разработки WEB-ресурсов с их помощью..

Тема 4. Технологии извлечения знаний из WEB – WEB

Mining. Понятие data scraping или «срезание данных с поверхности». Используя любой из приведенных либо найденных вами способов извлечения информации с web страниц, разработать программу по сбору статей на тему ВКР методами Web- scrappinga и продемонстрировать результат ее работы. Продемонстрировать применение безопасных приемов разработки программы.

Тема 5. Модели информационного поиска

Модели информационного поиска. Устройство и принцип работы поисковых систем. Определение и анализ характеристик выбранной поисковой системы: Google, Yandex, Rambler, Yahoo, Bing, AltaVista.

Тема 6. Разработать классификатор собранных статей..

Определение линейных однородных рекуррентных соотношений.

Тема 7. Технологии хранения данных в WEB.
Организация безопасных хранилищ.

Критерии оценивания выполнения практического задания:

Оценка «отлично» выставляется, если: работа выполнена в полном объеме с соблюдением необходимой последовательности действий; работа проведена в условиях, обеспечивающих получение правильных результатов и выводов; соблюдены правила техники безопасности; в ответе правильно и аккуратно выполняет все записи, таблицы, рисунки, чертежи, графики, вычисления; правильно выполняет анализ ошибок.

Оценка «хорошо» выставляется, если: работа выполнена правильно с учетом 1-2 мелких погрешностей или 2-3 недочетов, исправленных самостоятельно по требованию преподавателя.

Оценка «удовлетворительно» выставляется, если: работа выполнена правильно не менее чем наполовину, допущены 1-2 погрешности или одна грубая ошибка.

Оценка «неудовлетворительно» выставляется, если: допущены две (и более) грубые ошибки в ходе работы, которые обучающийся не может исправить даже по требованию преподавателя или работа не выполнена полностью.

Типовые тесты

Тема 1. Принципы работы поисковых движков.

Тест 1

Основной задачей индексации сайта поисковым роботом является:

- A. Добавление сайта в черный список
- B. Удаление ненужных страниц сайта
- C. Сбор и структурирование данных о сайте для последующего поиска
- D. Улучшение дизайна веб-сайта

Ответ: C

Тест 2

Что такое ранжирование в поисковых системах?

- A. Автоматическое удаление нерелевантных ссылок
- B. Определение приоритета страницы в выдаче по запросу пользователя
- C. Загрузка картинок на сайт
- D. Оптимизация кода HTML

Ответ: B

Тест 3

Что такое PageRank?

- A. Сервис проверки грамматики текстов
- B. Один из факторов ранжирования Google, определяющий важность страницы на основе ссылок
- C. Сайт-конструктор для начинающих пользователей
- D. Метод сжатия фотографий для оптимизации загрузки сайта

Ответ: B

Тест 4

Что означают термины SERP и сниппет?

- A. SERP — список рекламных объявлений, сниппет — ошибка сервера
- B. SERP — страница результатов поиска, сниппет — фрагмент описания страницы в результатах поиска
- C. SERP — сокращение от Social Engine Rating Pages, сниппет — логотип сайта
- D. SERP — серверное программное обеспечение, сниппет — служебная программа для анализа трафика

Ответ: B

Тест 5

Что такое робот поисковой системы (краулер)?

- A. Скрипт для автоматического заполнения форм регистрации
- B. Программа, индексирующая контент в Интернете и собирающая информацию для поисковика
- C. Интернет-магазин товаров для дома
- D. Рекламная кампания, нацеленная на повышение позиций сайта в поисковиках

Ответ: B

Тема 2. Методологии безопасного сбора данных из сетевых источников

Тест 1

Что означает соблюдение этических норм при сборе данных из открытых источников Интернета?

- A. Игнорирование авторских прав владельцев ресурсов
- B. Сохранение анонимности сборщика данных
- C. Четкое следование законодательству и правилам пользования ресурсами
- D. Открытость информации обо всех собранных данных третьим лицам

Ответ: C

Тест 2

Зачем важно учитывать политику конфиденциальности ресурса при сборе данных?

- A. Чтобы избежать блокировки IP-адреса или аккаунта
- B. Для предотвращения нарушений закона о защите личной информации
- C. Для лучшего понимания структуры сайта
- D. Для автоматической идентификации владельца ресурса

Ответ: B

Тест 3

Что из перечисленного не является безопасной практикой при сборе данных?

- A. Использование прокси-серверов для скрытия реального местоположения
- B. Предоставление ложной информации о цели сбора данных
- C. Периодическое обновление скриптов парсинга для соответствия изменениям сайта
- D. Ограничение частоты запросов к ресурсам

Ответ: B

Тест 4

Что такое бот-лимитирование (rate limiting) при сборе данных?

- A. Ограничение скорости сбора данных с целью снижения нагрузки на сервер
- B. Техника очистки данных от дубликатов
- C. Автоматическое распознавание капчей для входа на сайт
- D. Искусственный интеллект, улучшающий качество извлечённых данных

Ответ: A

Тест 5

Какая из техник предотвращает обнаружение парсера сайтом-жертвой?

- A. Отправка постоянных запросов без паузы
- B. Работа только ночью, когда нагрузка на сайт снижена
- C. Передача реальных заголовков браузера и уникального User-Agent
- D. Использование огромного количества прокси-серверов одновременно

Ответ: C

Тема 3. Типы информационных систем. Оценка эффективности и безопасности работы поисковых систем.

Тест 1

Что такое корпоративная информационная система (КИС)?

- A. Информационная система, применяемая в глобальном масштабе
- B. Система, созданная для автоматизации бизнес-процессов предприятия
- C. Общедоступная информационно-справочная система
- D. Комплексная инфраструктура общегородских служб

Ответ: B

Тест 2

Критерий, который оценивает эффективность поисковой системы по отношению к качеству выдачи релевантных результатов:

- A. Скорость индексации сайтов
- B. Объем хранимой информации
- C. Релевантность результатов поиска
- D. Безопасность передачи данных

Ответ: C

Тест 3

Какой тип информационной системы предназначен для оперативного мониторинга и анализа производственных процессов?

- A. ERP (Enterprise Resource Planning)
- B. CRM (Customer Relationship Management)
- C. SCADA (Supervisory Control And Data Acquisition)
- D. MIS (Management Information System)

Ответ: C

Тест 4

Что показывает коэффициент эффективности (precision) поисковой системы?

- A. Долю правильных результатов среди выданных
- B. Время отклика на запросы пользователей
- C. Общий объем базы данных поисковой системы
- D. Индекс цитируемости страниц

Ответ: A

Тест 5

Показатель recall (охват) поисковой системы отражает:

- A. Способность выдавать точные результаты поиска
- B. Среднюю продолжительность сессии пользователя
- C. Долю найденных релевантных документов среди существующих релевантных документов
- D. Надежность и безопасность передаваемых данных

Ответ: C

Тема 4. Модели и технологии децентрализованного поиска

Тест 1

Что такое децентрализованный поиск?

- A. Поисковая система, управляемая централизованным сервером
- B. Архитектура поиска, где узлы сети совместно обрабатывают запросы и распределяют нагрузку
- C. Технология, позволяющая искать файлы только на одном компьютере
- D. Поиск, осуществляемый только в пределах конкретной локальной сети

Ответ: B

Тест 2

Что такое пиринговая сеть (peer-to-peer network)?

- A. Сеть, где каждый узел может выступать как клиент и сервер одновременно
- B. Классическая архитектура клиент-сервер
- C. Сеть с центральным сервером, обслуживающим клиентов
- D. Сеть, объединяющая мобильные устройства через Bluetooth

Ответ: A

Тест 3

Какая технология используется для уменьшения времени поиска в децентрализованных сетях?

- A. Централизация данных на единственном сервере
- B. Фрагментация и репликация данных по множеству узлов
- C. Запрещение файлообмена между пользователями
- D. Использование исключительно защищенных соединений VPN

Ответ: B

Тест 4

Какие риски существуют при использовании децентрализованной поисковой инфраструктуры?

- A. Низкая скорость передачи данных
- B. Сложность обеспечения безопасности и приватности данных
- C. Недостаточно высокая емкость каналов связи
- D. Проблемы с интеграцией с мобильными платформами

Ответ: B

Тест 5

Что такое хештаблица в контексте децентрализованных сетей?

- A. Таблица для хранения криптографических ключей
- B. Список адресов серверов сети
- C. Данные, хранимые на центральном сервере
- D. Структура данных, сопоставляющая ключ и значение для быстрого поиска информации в сети

Ответ: D

Тема 5.Способы обработки и хранения больших данных в сетевых источниках

Тест 1

Какой метод подходит для быстрого анализа больших объемов данных?

- A. SQL-запросы к традиционным базам данных
- B. Big Data-решения вроде Apache Hadoop и Spark
- C. Локальное хранение данных на рабочем столе
- D. Ручная обработка данных Excel-таблицах

Ответ: B

Тест 2

Что такое NoSQL базы данных?

- A. Традиционные реляционные базы данных
- B. Базы данных, оптимизированные для маленьких объемов данных
- C. Базы данных, предназначенные для обработки больших объемов неструктурированных данных
- D. Система резервного копирования данных

Ответ: C

Тест 3

Какой инструмент используется для хранения огромных объемов неструктурированной информации?

- A. MySQL
- B. PostgreSQL
- C. MongoDB
- D. Microsoft Access

Ответ: C

Тест 4

Какая технология позволяет обрабатывать огромные объемы данных параллельно на многих серверах?

- A. MapReduce
- B. REST API
- C. ODBC
- D. SSH

Ответ: A

Тест 5

Какой тип хранилища данных используется для временной агрегации и первичной обработки больших потоков данных?

- A. Lakehouse
- B. Warehouse
- C. Data lake
- D. Database

Ответ: C

Критерии оценивания тестовых заданий для текущего контроля

Максимальное количество баллов за решение одного теста составляет 5 баллов, которые переводятся в академические оценки следующим образом:

- оценка «отлично» (5 баллов) выставляется при условии правильного ответа студента не менее чем на 85% контрольных заданий
- оценка «хорошо» (4 балла) выставляется при условии правильного ответа студента не менее чем на 70% контрольных заданий
- оценка «удовлетворительно» (3 балла) выставляется при условии правильного ответа студента не менее чем на 51% контрольных заданий
- оценка «неудовлетворительно» (1-2 балла) выставляется при условии правильного ответа студента менее чем на 50% контрольных заданий

2.2 Оценочные материалы для проведения промежуточной аттестации

Тест 1

Какая практика помогает предотвратить злоупотребление ресурсами при автоматизированном извлечении данных?

- A. Установка лимитов на частоту запросов
- B. Парсинг только в ночное время суток
- C. Ограничение объема извлекаемых данных
- D. Отправка только GET-запросов

Ответ: A

Тест 2

Что такое политика robots.txt?

- парсинг
- A. Документ, устанавливаемый владельцами сайтов для указания разрешений на парсинг
 - B. Веб-сервис для фильтрации вредоносных запросов
 - C. Протокол безопасности для обмена информацией
 - D. Файл конфигурации для оптимизации поисковой системы

Ответ: A

Тест 3

Какой тип атаки возможен при некорректном подходе к извлечению данных?

- A. DoS/DDoS
- B. SQL-инъекция
- C. Cross-site scripting (XSS)
- D. Phishing

Ответ: A

Тест 4

Какую технику применяют для сокрытия источника автоматических запросов?

- A. Прoxy-серверы
- B. Сжатие данных
- C. Кэширование данных
- D. Шифрование данных

Ответ: A

Тест 5

Какой метод помогает снизить риск блокировки вашего IP при массовом парсинге?

- A. Постоянные обращения без интервалов
- B. Медленный темп отправки запросов с промежутками
- C. Многоуровневая аутентификация
- D. Обход капчи вручную

Ответ: B

Тест 6

Что такое "Rate Limiting"?

- A. Лимит на число обращений к серверу за единицу времени
- B. Лимит на максимальный объем извлекаемых данных
- C. Лимит на глубину вложенности извлекаемого контента
- D. Лимит на сложность используемых запросов

Ответ: A

Тест 7

Что представляет собой captcha-код?

- A. Графический файл, помогающий визуальному оформлению сайта
- B. Дополнительный фильтр для замедления работы парсера
- C. Средство проверки принадлежности пользователя к человеку, а не программе
- D. Сообщение, отправляемое владельцу сайта при подозрительной активности

Ответ: C

Тест 8

Какая мера помогает минимизировать угрозу утечки конфиденциальных данных при парсинге?

- A. Передача данных открытым текстом
- B. Использование зашифрованных протоколов (HTTPS)
- C. Неправильное кэширование паролей
- D. Хранение всех данных в открытом доступе

Ответ: В

Тест 9

Как называется процесс сохранения копии содержимого веб-ресурса для дальнейшего анализа?

- A. Репликация
- B. Бэкап
- C. Веб-архивация
- D. Клонирование

Ответ: C

Тест 10

Что рекомендуется предпринять, если владельцы сайта запретили доступ к своему ресурсу для парсинга?

- A. Продолжить парсинг, игнорируя запреты
- B. Обратиться к владельцам ресурса с просьбой разрешить доступ
- C. Сделать временную копию сайта для изучения офлайн
- D. Попытаться взломать защиту сайта

Ответ: B

Тест 11

Какой принцип защищает вас от неправомерного доступа к личным данным при использовании публичных сервисов?

- A. Открытый доступ к любым данным
- B. Согласие пользователя на обработку данных
- C. Автономный доступ без уведомления
- D. Универсальность и доступность любых сведений

Ответ: B

Тест 12

Какая рекомендация обеспечит безопасность при разработке собственной системы парсинга?

- A. Не обращать внимания на инструкции сайта в файле robots.txt
- B. Использовать мощные серверы для массовой обработки данных
- C. Осуществлять парсинг постепенно, соблюдая рекомендации ресурса
- D. Оставлять личные данные открытыми для удобства разработки

Ответ: C

Тест 13

Какую стратегию целесообразно применить для борьбы с чрезмерной нагрузкой на серверы стороннего ресурса?

- A. Периодически отправлять большое количество запросов
- B. Установить разумные интервалы между запросами
- C. Использовать низкоскоростные каналы связи
- D. Работать только в часы пиковой нагрузки

Ответ: B

Тест 14

Что такое этичный парсинг?

- A. Незаконное извлечение данных с сайтов конкурентов
- B. Извлечение данных без нарушения условий использования ресурса
- C. Использование краулинга для сбора личных данных пользователей
- D. Продажа собранных данных третьим сторонам

Ответ: B

Тест 15

Как правильно поступать, если хотите собрать публичные данные с сайта, не нарушая законодательство?

- A. Игнорировать правовые нормы и политику ресурса
- B. Обязательно запрашивать разрешение у администрации сайта
- C. Следовать инструкциям файла robots.txt и условиям использования ресурса
- D. Копировать контент с использованием встроенных инструментов браузера

Ответ: C

Тест 16

Что такое CAPTCHA и зачем она используется?

- A. Капча предназначена для устранения ошибок в тексте
- B. Это контрольный пункт для подтверждения того, что запрос отправил человек, а не робот
- C. Используется для шифрования передаваемых данных
- D. Помогает ускорить загрузку страниц сайта

Ответ: B

Тест 17

Какой важный аспект следует учитывать при настройке тайминга запросов к веб-сайтам?

- A. Важно устанавливать слишком короткие интервалы между запросами
- B. Следует избегать строгих временных рамок между запросами
- C. Необходимо периодически проверять доступность ресурса
- D. Лучше увеличить интервал между запросами, чтобы не перегружать сервер

Ответ: D

Тест 18

Какой подход рекомендуется для отслеживания изменений на сайтах без риска блокировки?

- A. Часто обращаться к сайту с множеством запросов
- B. Использовать специальный сервис уведомлений о изменениях
- C. Регулярно загружать полную версию сайта целиком
- D. Скачивать страницу вручную каждые полчаса

Ответ: B

Тест 19

Какой документ регламентирует разрешённые и запрещённые операции по парсингу

данных на конкретном сайте?

- A. License Agreement
- B. Terms of Service
- C. Privacy Policy
- D. Robots Exclusion Protocol (robots.txt)

Ответ: D

Тест 20

Что такое геоблокировка и как она может повлиять на парсинг?

- A. Геоблокировка — ограничение доступа к ресурсу в зависимости от географического местоположения посетителя
- B. Такое ограничение невозможно обойти средствами прокси или VPN
- C. Ограничение распространяется только на российские IP-адреса
- D. Доступ ограничен только при большом количестве запросов

Ответ: A

Тест 21

Какой безопасный метод рекомендуется для длительного архивирования извлечённых данных?

- A. Хранение данных на бесплатных облачных сервисах
- B. Регулярное создание бэкапов на удалённом сервере
- C. Загрузка данных непосредственно на личный компьютер
- D. Использование флешек для резервного копирования

Ответ: B

Тест 22

Какая методика полезна для предупреждения блокировки при регулярном парсинге популярных сайтов?

- A. Генерировать фальшивые cookie-файлы
- B. Менять IP-адрес и user-agent для каждого запроса
- C. Постоянно изменять название программы-парсера
- D. Чаще менять браузеры для парсинга

Ответ: B

Тест 23

Какой протокол предпочтителен для передачи данных при извлечении информации с веб-ресурсов?

- A. HTTP
- B. HTTPS
- C. FTP
- D. Telnet

Ответ: B

Тест 24

Какая техника применяется для анализа динамики цен на товары онлайн-магазинах?

- A. Price scraping (парсинг цен)
- B. Web crawling (сканирование сети)

C. Data mining (добыча данных) D. SEO-аналитика

Ответ: А

Тест 25

Что рекомендуют специалисты по безопасности для обработки больших объемов данных?

- A. Передавать данные открыто без шифрования
- B. Использовать надежные механизмы шифрования и авторизации
- C. Проводить манипуляции с данными прямо на стороне клиента
- D. Хранить данные только на собственных устройствах

Ответ: В

Критерии оценки для проведения зачета по дисциплине (тестирование)

Шкала оценивания результатов промежуточной аттестации и критерии выставления оценок.

Оценка	Уровень сформированности компетенции	Критерии
«Отлично» («зачтено»)	Высокий	Выполнено более 80% заданий предложенного теста
«Хорошо» («зачтено»)	Хороший	Выполнено 60-80% заданий предложенного теста
«Удовлетворительно» («зачтено»)	Достаточный	Выполнено 35-60% заданий предложенного теста
«Неудовлетворительно» («не зачтено»)	Недостаточный	Выполнено менее 35% заданий предложенного теста

Обновление оценочных материалов дисциплины

обсуждено и рекомендовано к утверждению решением кафедры
наименование кафедры _____
от ____ ____ 20__ г., протокол № ____

одобрено Научно-методическим советом _____ от ____ ____ 20__ г.,
протокол № ____