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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

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

2025

Рабочая программа по дисциплине «Безопасные методы извлечения информации из сетевых источников» по специальности 10.05.01 Компьютерная безопасность, специализации «Разработка защищенного программного обеспечения», составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования – специалитет по специальности 10.05.01 Компьютерная безопасность, утвержденного приказом Министерства науки и высшего образования Российской Федерации от «26» ноября 2020 г. № 1459, профессионального стандарта 06.032 «Специалист по безопасности компьютерных систем и сетей», утвержденного приказом Министерства труда и социальной защиты Российской Федерации от 14.09.2022 г. № 533н, профессионального стандарта 06.033 «Специалист по защите информации в автоматизированных системах», утвержденного приказом Министерства труда и социальной защиты Российской Федерации от 14.09.2022 г. №525н.

Рабочая программа:

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

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

СОДЕРЖАНИЕ

1. Цель и задачи освоения дисциплины

Целью изучения дисциплины «Безопасные методы извлечения информации из сетевых источников» является обучение передовым методам, моделям, средствам и технологиям поиска и компьютерной обработки информации с учетом аспектов безопасности. Углубленное изучение принципов работы поисковых движков, алгоритмов работы с текстовой, графической и иными видами информации.

Задачи:

Изучить историю и тенденции развития информационно-поисковых систем, работы крупных ученых, участвовавших в их разработке. Научиться безопасным принципам обмена данными в глобальной сети Интернет; основным методам функционирования информационно- поисковых систем; методам программирования поисковых систем, как на стороне сервера, так и на стороне клиента, знать методы передачи конфиденциальной информации по каналам связи, методы установления подлинности передаваемых сообщений и хранимой информации (документов, баз данных);научиться использовать современные инструментальные средства разработки поисковых систем и оценке их качества и безопасности. комбинаторного анализа.

2. Место дисциплины в структуре образовательной программы

Дисциплина «Комбинаторный анализ» относится к части, формируемой участниками образовательных отношений Блока 1 «Дисциплины (модули)» основной образовательной программы – программы специалитета по специальности 10.05.01 Компьютерная безопасность специализация «Разработка защищенного программного обеспечения».

Код компетенции	Предшествующие дисциплины (модули), практики	Изучаемые в текущем семестре дисциплины (модули), практики	Последующие дисциплины (модули), практики
ОПК-10	Основы военной подготовки Криптографические протоколы	**Безопасные методы извлечения информации из сетевых источников** Методы и средства криптографической защиты информации	

3. Перечень планируемых результатов обучения по дисциплине

Изучение дисциплины направлено на формирование у обучающихся компетенций.

Формируемые компетенции (код и наименование компетенции)	Индикаторы достижения компетенций	Планируемые результаты обучения
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..
	ОПК-10.2. Демонстрирует при решении задач профессиональной деятельности умение использовать основные преобразования симметричной и асимметричной криптографии и анализировать их криптостойкость к возможным компьютерным атакам	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.
	ОПК-10.3. Демонстрирует умения разработки криптографических 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)..

4. Объем дисциплины и виды учебной работы

Объем дисциплины и виды учебной работы в академических часах с выделением объема контактной работы обучающихся с преподавателем и самостоятельной работы обучающихся

очная форма обучения

Вид учебной деятельности		ак.часов	
		Всего	По семестрам
			10
1. Контактная работа обучающихся с преподавателем:		69	69
Аудиторные занятия, часов всего, в том числе:		68	68
• занятия лекционного типа		34	34
• занятия семинарского типа:		34	34
практические занятия		-	-
лабораторные занятия		34	34
в том числе занятия в форме практической подготовки		-	-
Консультации		0,5	0,5
Контактные часы на аттестацию в период экзаменационных сессий		0,5	0,5
2. Самостоятельная работа студентов всего, в том числе		39	39
- подготовка курсовой работы		-	-
- проработка учебного (теоретического) материала		15	15
- выполнение домашних заданий по практическим занятиям		15	15
- подготовка к экзамену		9	9
Промежуточная аттестация: экзамен		36	36
ИТОГО:	часов	144	144
общая трудоемкость	зач. ед.	4	4

5. Содержание дисциплины, структурированное по темам (разделам) с указанием количества академических часов и видов учебных занятий

5.1. Содержание дисциплины

Тема 1 Принципы работы поисковых движков.

Всемирная паутина. Развитие интернет в XXI веке. Организационная структура Интернета. Схема адресации в сети Интернет. МодельBOW TIE. Понятия и различия WEB 2.0- WEB 4.0. Структура WEB, Deep WEB, его возможности и характеристики. Инструменты и технологии работы в невидимом WEB.;

Тема 2 Методологии безопасного сбора данных из сетевых источников.

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

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

Понятие поисковой системы.Методы лингвистического анализа. Методы ранжирования поисковой выдачи. Виды поисковых роботов.

Порядок индексации сайтов. Алгоритмы работы поисковых систем Яндекс и Google. Выбор ключевых слов для продвижения сайта. Типы запросов по частотности. Типы запросов по степени конверсии. Создание семантического ядра. Выбор ключевых страниц сайта. Распределение семантического ядра. Выбор основной стратегии поискового продвижения сайта.

Тема 4. Модели и технологии децентрализованн ого поиска.

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

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

NoSQL хранилища. Требования к хранилищам данных, OLTP и OLAPсистемы. Нереляционные базы данных. Принципы работы и примеры использования Map-Reduce.

5.2. Разделы, темы дисциплины и виды занятий

очная форма обучения

№ п/п	Наименование раздела, темы дисциплины	Виды занятий, включая самостоятельную работу студентов (в часах)			Индикаторы достижения компетенций
		занятия лекционного типа	занятия семинарского типа / из них в форме практической подготовки (лабораторные работы)	самос-тоятельная работа	
1	Тема 1. Принципы работы поисковых движков.	6	6/6	7	ОПК-10 ОПК-10.1 ОПК-10.2 ОПК-10.3
2	Тема 2. Методологии безопасного сбора данных из сетевых источников	6	6/6	8	ОПК-10 ОПК-10.1 ОПК-10.2 ОПК-10.3
3	Тема 3. Типы информационных систем. Оценка эффективности и безопасности работы поисковых систем.	6	6/6	8	ОПК-10 ОПК-10.1 ОПК-10.2 ОПК-10.3
4	Тема 4. Модели и технологии децентрализованн ого поиска	6	6/6	8	ОПК-10 ОПК-10.1 ОПК-10.2 ОПК-10.3

№ п/п	Наименование раздела, темы дисциплины	Виды занятий, включая самостоятельную работу студентов (в часах)			Индикаторы достижения компетенций
		занятия лекционного типа	занятия семинарского типа / из них в форме практической подготовки (лабораторные работы)	самос-тоятельная работа	
5	Тема 5. Способы обработки и хранения больших данных в сетевых источниках	10	10/10	8	ОПК-10 ОПК-10.1 ОПК-10.2 ОПК-10.3
	Всего	**34**	**34/34**	**39**	

6. Практические занятия

Практические занятия не предусмотрены.

7. Лабораторные работы

№ п/п	Наименование раздела, темы дисциплины	Содержание лабораторных работ	Объем (час.)
			Очная форма обучения
1.	Тема 1 Вводное занятие	Определение комбинаторики и её значимость. – Изучение основных понятий: перестановки, сочетания и размещения. – Примеры задач на подсчет различных комбинаций	6
2.	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.	6
3.	Тема3. Системы управления контентом..	Обсуждение преимуществ и недостатков различных CMS, особенностей разработки WEB-ресурсов с их помощью..	6
4.	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.	6
5.	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.	10
	Всего		**34**

8. Примерная тематика курсовых работ (проектов)

Курсовые работы (проекты) не предусмотрены.

9. Самостоятельная работа студента

Самостоятельная работа студента при изучении дисциплины «Безопасные методы извлечения информации из сетевых источников» направлена на:

– освоение нормативных правовых актов и учебной литературы, необходимых для освоения дисциплины;

– самостоятельный поиск информации в Интернете и других источниках;

– выполнение домашних заданий по практическим занятиям;

– подготовку к экзамен.

Краткие рекомендации по выполнению самостоятельной работы:

Успешное усвоение дисциплины предполагает большой, упорный, серьезный, систематический труд студентов. Важнейшая его составная часть – выполнение разных видов самостоятельной работы.

1. **Составление тематического конспекта** на основе изученной основной и дополнительной учебной литературы. В тематическом конспекте за основу берется содержание темы, вопросы для обсуждения.

Этапы работы.

1.1 Конспектирование делается только после того, как прочитан или усвоен материал для конспектирования.

1.2. Необходимо мысленно или письменно составить план конспекта. По этому плану и будет строиться конспект далее.

1.3. Составление самого конспекта. Можно сказать, что конспект – это расширенные тезисы, дополненные рассуждениями и доказательствами, содержащимися в материалах для конспекта, а также собственными мыслями и положениями составителя конспекта.

Писать конспект рекомендуется четко и разборчиво. В конспекте можно выделять места текста в зависимости от их значимости. Для этого применяются различного размера буквы, подчеркивания, замечания на полях.

Перечень вопросов

1.Опишите структуру, пропорции, охарактеризуйте размеры и динамику WEB.

2.Понятие “Сильной связности» WEB-графа, типы его узлов

3.Какому функциональному закону подчиняются сети «тесного мира»?

4.Закономерности и ограничения модели Bow Tie.

5.Понятие безопасного поиска. Приема безопасного поиска в сетевых источниках.

6.Deep WEB. Какие ресурсы его составляют. Какими средствами его можно исследовать.

7.Понятия Web Mining и Web Analytics. Этапы аналитики в соответствии со

стандартом CRISP-DM.
8.Задачи Data Mining. Направления Data Mining.
9.Понятие и задачи Web Content Mining.
10.Перечислите и охарактеризуйте средства WEB scraping.
11.Методы Text Mining в приложении к специфике WWW.
12.Методологии Web Graph Mining для подхода Web Structure Mining.
13.Основные задачи Web Usage Mining, средства их решения, назначение кластерного анализа в контексте Web Usage Mining.
14.Классификация способов извлечения информации из WEB-источников.
15.Задачи Web-scrapping, механизм его работы. Разновидность методов Web-scrapping.
16.Этапы работы поисковой системы. Компоненты поискового движка.
17.Как работают алгоритмы индексирования. Необходимость ранжирования и задачи машинного обучения в приложении к информационному поиску.
18.Охарактеризуйте модели информационного поиска.
19.Изложите подробно принцип булевой модели информационного поиска (ИП), возможные средства оптимизации запроса.
20.Суть векторной и вероятностной моделей ИП, их достоинства и недостатки.
21.Назовите и кратко охарактеризуйте этапы нормализации текста перед индексацией.
22.Перечислите и дайте краткую характеристику методов лингвистического анализа.
23.Безопасные способы хранения словарей. Способы нечеткого поиска.
24.Алгоритм составления электронной подписи.
25.Алгоритм проверки электронной подписи.
26.Функции хеширования.
27.Стандарты цифровой подписи.
28.Способы применения криптопротоколов при навигации в сетевых источниках
29.Технология Map-Reduce, механизмы работы, примеры использования. Как обеспечивается отказоустойчивость Map-Reduce.
30.Технология Hadoop. MapReduce в Hadoop. Структура программы в Hadoop.
31.Безопасные хранилища Больших данных. Примеры распределенных хранилищ.
32.NoSQL, типы NoSQL баз данных. Теорема CAP.
33.Понятия OLAP и OLTP, безопасное использование хранилищ.

10. Оценивание результатов обучения и уровня сформированности компетенций

Оценивание результатов обучения по дисциплине и уровня сформированности компетенций (части компетенции) осуществляется в рамках текущего контроля успеваемости и промежуточной аттестации в

соответствии с оценочными материалами.

11. Учебно-методическое обеспечение дисциплины

Основная литература:

1 1. Система формирования знаний в среде Интернет / В. И. Аверченков, А. В. Заболеева- Зотова, Ю. М. Казаков [и др.]. – 4-е изд., стер. – Москва : ФЛИНТА, 2021. – 181 с. : ил., схем., табл. – Режим доступа: по подписке. –URL: 4TUhttps://biblioclub.ru/index.php?page=book&id=93354U4T (дата обращения: 01.06.2024). – Библиогр. в кн. – ISBN 978-5-9765-1266-5. – Текст : электронный.

2. Симанков, В. С. Методы и алгоритмы поиска информации в Интернете=Search methods and algorithms for information retrieval on the Internet / В. С. Симанков, Д. М. Толкачев. – Москва : Библио-Глобус, 2017. – 332 с. : граф., схем., ил. – Режим доступа: по подписке.– URL: 4Thttps://biblioclub.ru/index.php?page=book&id=4990774T (дата обращения: 01.06.2024).– Библиогр. в кн. – ISBN 978-5-9500501-8-3. – DOI 10.18334/9785950050183. – Текст :электронный.

3. Гисин, В. Б. Криптография и распределенные реестры : учебное пособие : [16+] / В. Б. Гисин ; Финансовый университет при Правительстве Российской Федерации. – Москва : Прометей, 2022.– 186 с. : табл., схем. – Режим доступа: по подписке. URL: 4Thttps://biblioclub.ru/index.php?page=book&id=7009414T (дата обращения: 01.03.2025). – Библиогр. в кн. – ISBN 978-5-00172-257-1. – Текст : электронный.

4. Загинайлов, Ю. Н. Теория информационной безопасности и методология защиты информации [Электронный ресурс]: учебное пособие/ Ю. Н. Загинайлов. - Москва; Берлин: Директ-Медиа, 2015. - 253 с.: ил. - Режим доступа: http://biblioclub.ru/index.php?page=book&id=276557

Для освоения дисциплины инвалидами и лицами с ограниченными возможностями здоровья имеются издания в электронном виде в электронно-библиотечных системах «Лань» и «Юрайт».

Дополнительная литература

1. Ромм, Я.Е. Детерминированный информационный поиск на основе сортировки с распараллеливанием базовых операций / Я.Е. Ромм, С.С. Белоконова. - Москва : Издательство Научный мир, 2014. - 197 с. [Электронныйресурс].URLhttp://biblioclub.ru/index.php?page=book&id=468725

2. Артемов, А.В. Мониторинг информации в интернете : учебно-методическоепособие / А.В. Артемов ; Межрегиональная Академия безопасности и выживания.- Орел : МАБИВ, 2014. - 160 с. [Электронный ресурс]. - URL:http://biblioclub.ru/index.php?page=book&id=428606

3. Скулкин, О. Шифровальщики : как реагировать на атаки с использованием программ- вымогателей : [16+] / О. Скулкин ; науч. ред. А.

Алексеев ; ред. К. Ахметов ; пер. с англ. А. Власюк. – Москва : Альпина Паблишер, 2023. – 205 с. : ил. – Режим доступа: по подписке. – URL: 4Thttps://biblioclub.ru/index.php?page=book&id=7078314T (дата обращения: 01.03.2025). – ISBN 978-5-206-00080-1. – Текст : электронный.

4. Лапонина, О. Р. Протоколы безопасного сетевого взаимодействия [Электронный ресурс]: курс/ О. Р. Лапонина. - 2-е изд., исправ. - Москва: ИНТУИТ, 2016. - 462 с. - Режим доступа: http://biblioclub.ru/index.php?page=book&id=429094

Электронные библиотечные системы (ЭБС) и электронные образовательные ресурсы

1. Электронно-библиотечная система ZNANIUM https:// www.znanium.com
2. Универсальная справочно-информационная полнотекстовая база данных периодических изданий «East View» (ИВИС) https://dlib.eastview.com/
3. Электронно-библиотечная система «BOOK.ru» www.book.ru
4. Электронно-библиотечная система «IPRbooks» http://www.iprbookshop.ru/
5. ЭБС «Национальный цифровой ресурс «Руконт» https://lib.rucont.ru/search
6. Образовательная платформа ЮРАЙТ http://www.urait.ru

Электронный каталог Университета – Автоматизированная информационная библиотечная система (АИБС «МегаПро»)

12. Перечень программного обеспечения, профессиональных баз данных и информационных справочных систем, необходимых для освоения дисциплины

Лицензионное программное обеспечение, в том числе отечественного производства:

– DsktpEdu ALNG LicSARk OLV F1 1Y Acdmc Ent (MS Windows) (подписка на ПО).

Свободно распространяемое программное обеспечение, в том числе отечественного производства:

– 7-Zip (свободный файловый архиватор с высокой степенью сжатия данных) (отечественное ПО).

– FastStone Image Viewer (программа для просмотра изображений в Microsoft Windows (лицензия бесплатного программного обеспечения)).

– Foxit Reader (Бесплатное прикладное программное обеспечение для просмотра электронных документов в стандарте PDF (лицензия бесплатного программного обеспечения)).

– Indigo (система программ для создания и проведения компьютерного

тестирования знаний).

– Google Crome, ЯндексБраузер (браузер).

– Adobe Acrobat Reader DC (• ПО для просмотра, печати и комментирования документов в формате PDF)

– Visual Studio Code (интегрированная среда разработки программного обеспечения)

Профессиональные базы данных не используются.

Информационные справочные системы:

– СПС КонсультантПлюс. Компьютерная справочная правовая система, широко используется учеными, студентами и преподавателями (подписка на ПО)

Каждый обучающийся в течение всего периода обучения обеспечивается индивидуальным неограниченным доступом к электронно-библиотечной

13. Материально–техническое обеспечение дисциплины

Образовательный процесс обеспечен учебными аудиториями для проведения учебных занятий, а именно для проведения занятий лекционного типа, лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещениями для самостоятельной работы студентов.

Учебная аудитория для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

Рабочее место преподавателя (стол, стул), рабочие места обучающихся, доска для написания мелом, кафедра.

Технические средства обучения: персональный компьютер-1, мультимедийное оборудование (проектор, проекционный экран) - 1

Учебная аудитория для проведения лабораторных занятий. Аудитория информационных технологий

Рабочее место преподавателя (стол, стул), рабочие места обучающихся:

Технические средства обучения:

персональный компьютер с подключением к сети «Интернет» - 15, сетевое оборудование.

Помещение для самостоятельной работы обучающихся, оснащенное компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду института.

Рабочие места обучающихся.

Технические средства обучения: персональный компьютер - 13.

Учебные аудитории соответствуют действующим противопожарным правилам и нормам, укомплектованы учебной мебелью.

Обновление рабочей программы дисциплины (модуля)

Наименование раздела рабочей программы, в который внесены изменения

(измененное содержание раздела)

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

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