

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



подписано электронно-цифровой подписью

Программа дисциплины

Анализ интернет-данных

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и): доцент, к.н. Пинягина О.В. (кафедра анализа данных и технологий программирования, отделение фундаментальной информатики и информационных технологий), Olga.Piniaguina@kpfu.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ПК-4	Способен строить математические модели и анализировать данные, обосновывать и выбирать решение в прикладных задачах

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

- основные методы Data mining, пригодные для работы с Интернет-данными,

Должен уметь:

- работать в среде R с пакетами Data mining,

Должен владеть:

- применять на практике знания, полученные при изучении курса, для извлечения, трансформации и интеллектуального анализа Интернет-данных.

Должен продемонстрировать способность и готовность:

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.08 Дисциплины (модули)" основной профессиональной образовательной программы 01.04.02 "Прикладная математика и информатика (Анализ данных и его приложения)" и относится к части ОПОП ВО, формируемой участниками образовательных отношений.

Осваивается на 2 курсе в 3 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 2 зачетных(ые) единиц(ы) на 72 часа(ов).

Контактная работа - 32 часа(ов), в том числе лекции - 16 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 16 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 40 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет в 3 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)						Самостоятельная работа
			Лекции, всего	Лекции в эл. форме	Практические занятия, всего	Практические в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	
N	Разделы дисциплины / модуля	Семестр	Лекции, всего	Лекции в эл. форме	Практические занятия, всего	Практические в эл. форме	Лабораторные работы, всего	Лабораторные в эл. форме	Самостоятельная работа
1.	Тема 1. Data mining и Web mining	3	2	0	0	0	0	0	0
2.	Тема 2. Основные Интернет-технологии для Web mining	3	2	0	0	0	4	0	10
3.	Тема 3. Технология R в Web Mining	3	2	0	0	0	4	0	10
4.	Тема 4. Text mining	3	2	0	0	0	4	0	10
5.	Тема 5. Социальные сети. Social mining	3	2	0	0	0	0	0	0
6.	Тема 6. Рекомендательные системы	3	2	0	0	0	0	0	0
7.	Тема 7. Интеллектуальные агенты	3	2	0	0	0	4	0	10
4.2	Содержание дисциплины (модуля)								
8.	Тема 8. Принципы функционирования поисковых систем	3	2	0	0	0	0	0	0
Интеллектуальный анализ данных и интеллектуальный анализ Интернет-данных.									
Задачи Web mining. Этапы Web mining: постановка задачи, получение данных, предварительная обработка, выбор методов, моделирование, оценка результатов и анализ модели.									40

Web structure mining, Web usage mining, Web content mining.

Тема 2. Основные Интернет-технологии для Web mining

Основные Интернет-технологии для Web mining. Клиент-серверная архитектура. Протокол HTTP - основной протокол уровня приложений. Структура запроса клиента. Структура ответа сервера. HTTP-заголовки. XML - расширяемый язык запросов. Язык навигации внутри XML-документа (XPath). Формат JSON (Java Script Object Notation).

Тема 3. Технология R в Web Mining

Технология R в Web Mining. Установка и настройка R и среды RStudio. Пакеты для связи с базами данных: RODBC, пакеты для MySQL, Postgres. Пакет XML. Пакет Jsonlite. Пакет RCurl. Пакет rvest. Web-scraping структурированных данных. Сравнение возможностей навигации в HTML-документе с помощью XPath и с помощью селекторов. Примеры приложений.

Тема 4. Text mining

Text mining. Пакет tm, основные объекты: корпус слов, матрицы терминов-документов и документов-терминов. Web-scraping текстов. Задачи классификации документов. Задачи кластеризации документов. Аннотирование документов. Анализ тональности текстов. Построение облака слов. Примеры приложений.

Тема 5. Социальные сети. Social mining

Задачи интеллектуального анализа данных на основе информации из социальных сетей. Понятие социальной сети и её анализа. социальная сеть как большая система. Виртуальная социальная сеть. Задачи Social Mining - анализ информационных потоков, персонификация предложений, поиск аномалий, компьютерных ботов и мошенников.

Тема 6. Рекомендательные системы

Рекомендательные системы. Задачи рекомендательных систем. Критерии оценки качества РС - точность, покрытие, скорость обучения, степень новизны. Системы на основе сходства товаров. Системы на основе сходства поведения потребителей. Алгоритм коллаборативной фильтрации. Задача поиска ассоциативных правил.

Тема 7. Интеллектуальные агенты

Интеллектуальные агенты. Свойства интеллектуальных агентов: автономность, социальное поведение, реактивность, активность, базовые знания, убеждения, цели, желания, обязательства, намерения. Открытые системы: расширяемость/масштабируемость, мобильность/переносимость, интероперабельность, дружелюбность. Виды агентов: рефлексивный агент; агент, действующий на основе цели; агент, действующий на основе полезности.

Тема 8. Принципы функционирования поисковых систем

Основные характеристики поисковых систем: полнота, точность, актуальность, скорость поиска, наглядность. Принципы функционирования поисковых систем. Поисковый индекс. Поисковый сервер. Поисковый кластер. Поисковый робот. Ранжирование документов. Глобальные поисковые системы. Структура информационно-поисковых запросов. Стратегия поиска информации в Интернете. Язык поисковых запросов. Специализированные поисковые системы. Расширение запросов.

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства науки и высшего образования Российской Федерации от 6 апреля 2021 года №245)

Письмо Министерства образования Российской Федерации №14-55-99бин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы. Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;
- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

The Comprehensive R Archive Network - <https://cran.gis-lab.info/index.html>

The R Project for Statistical Computing - <https://www.r-project.org/>

Анализ интернет-данных - <https://do.kpfu.ru/course/view.php?id=6226>

Наглядная статистика. Используем R! (электронный ресурс) - <https://cran.r-project.org/doc/contrib/Shipunov-rbook.pdf>

Пользовательский интерфейс для R - <https://www.rstudio.com/products/RStudio/>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	Желательно проводить лекционные занятия по курсу в аудитории, оборудованной мультимедийным проектором и экраном или компьютерном классе большой вместимости. По курсу разработаны презентации, включающие краткое описание каждой темы. Предполагается, что студенты используют их в качестве основы или пишут собственные конспекты лекций. Реализация данной дисциплины предполагает как очную, так и дистанционную форму обучения.
лабораторные работы	По курсу разработаны электронные материалы, которые объединены в электронный учебник https://do.kpfu.ru/course/view.php?id=6226 . В рамках практической части курса студенты выполняют индивидуальные задания. Используемое программное обеспечение: SQL server (или другая СУБД), среда R, программы для web-скрепинга. Реализация данной дисциплины предполагает как очную, так и дистанционную форму обучения.
самостоятельная работа	Поскольку времени, отводимого для компьютерных занятий, абсолютно недостаточно для выполнения заданий, студенты занимаются их выполнением также в рамках самостоятельной работы. Таким образом, аудиторные практические занятия предназначены, прежде всего, для того чтобы студенты задавали вопросы преподавателю и сдавали этапы задания. Остальная работа выполняется самостоятельно. Реализация данной дисциплины предполагает как очную, так и дистанционную форму обучения.
зачет	Подготовка к зачету способствует закреплению, углублению и обобщению знаний, получаемых, в процессе обучения, а также применению их к решению практических задач. Готовясь к зачету, студент ликвидирует имеющиеся пробелы в знаниях, углубляет, систематизирует и упорядочивает свои знания. На зачете студент демонстрирует то, что он приобрел в процессе обучения по конкретной учебной дисциплине. При подготовке к зачету необходимо ознакомиться со списком вопросов к зачету, повторно ознакомиться с лекционным материалом, систематизировать информацию по курсу. Особое внимание следует уделить разделам курса, изученным самостоятельно и вызывавшим наибольшие затруднения при теоретическом изучении и решении практических задач.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

Компьютерный класс.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи:
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;
- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 01.04.02 "Прикладная математика и информатика" и магистерской программе "Анализ данных и его приложения".

*Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.08 Анализ интернет-данных*

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Основная литература:

1. Баюк, О. А. Практикум по анализу данных на языках Python и R : учебное пособие / О. А. Баюк, М. Р. Исаева, М. О. Самсонкин. - Москва : Прометей, 2023. - 100 с. - ISBN 978-5-00172-356-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/2124862> (дата обращения: 10.01.2025). - Режим доступа: по подписке.
2. Кулаичев, А. П. Методы и средства комплексного статистического анализа данных : учебное пособие / А.П. Кулаичев. - 5-е изд., перераб. и доп. - Москва : ИНФРА-М, 2025. - 484 с. - (Высшее образование). - DOI 10.12737/25093. - ISBN 978-5-16-020053-8. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2155997> (дата обращения: 10.01.2025). - Режим доступа: по подписке.
3. Система формирования знаний в среде Интернет : монография / В. И. Аверченков, А. В. Заболеева-Зотова, Ю. М. Казаков [и др.]. - 4-е изд., стер. - ФЛИНТА : Флинта, 2016. - 181 с. - ISBN 978-5-9765-1266-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1843190> (дата обращения: 10.01.2025). - Режим доступа: по подписке.

Дополнительная литература:

1. Клячкин, В. Н. Статистические методы анализа данных : учебное пособие / В. Н. Клячкин, Ю. Е. Кувайскова, В. А. Алексеева. - Москва : Финансы и Статистика, 2021. - 240 с. - ISBN 978-5-00184-057-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1831431> (дата обращения: 10.01.2025). - Режим доступа: по подписке.
2. Мاستицкий, С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. - 2-е изд. - Москва : ДМК Пресс, 2023. - 497 с. Систем. требования: Adobe Reader XI либо Adobe Digital Editions 4.5 ; экран 10". - ISBN 978-5-89818-601-2. - Текст : электронный // ЭБС 'Консультант студента' : [сайт]. - URL : <https://www.studentlibrary.ru/book/ISBN9785898186012.html> (дата обращения: 10.01.2025). - Режим доступа : по подписке.
3. Дайитбеков, Д. М. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбеков. - 2-е изд., испр. и доп. - Москва : Вузовский учебник: ИНФРА-М, 2010. - 578 с. (Научная книга). ISBN 978-5-9558-0191-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/251791> (дата обращения: 10.01.2025). - Режим доступа: по подписке.

*Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.08 Анализ интернет-данных*

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 01.04.02 - Прикладная математика и информатика

Профиль подготовки: Анализ данных и его приложения

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2025

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.