

МИНОБРНАУКИ РОССИИ
ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Рабочая программа дисциплины (модуля)
ТЕХНОЛОГИИ СБОРА И ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ДАННЫХ

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Искусственный интеллект и машинное обучение в управлении и принятии решений

Год набора на ОПОП
2020

Форма обучения
очная

Владивосток 2023

Рабочая программа дисциплины (модуля) «Технологии сбора и предварительной обработки данных» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика (утв. приказом Минобрнауки России от 19.09.2017г. №916) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Назаров Д.А., кандидат технических наук, доцент, Научно-образовательный центр "Искусственный интеллект"

Утверждена на заседании кафедры информационных технологий и систем от 31.05.2023 , протокол №

9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчик)

Кийкова Е.В.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575633692
Номер транзакции	0000000000BBE2A7
Владелец	Кийкова Е.В.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью освоения дисциплины являются формирование у студентов магистратуры компетенции в области автоматизации сбора и обработки информации.

Основные задачи изучения дисциплины:

- формирование углублённых знаний об основных методах использования современных информационных технологий для сбора и обработки данных;
 - формирование практических навыков в области:
1. создания автоматизированных информационных систем сбора и обработки информации;
 2. поиска, анализа и оценки источников информации для обеспечения возможности их сбора в автоматическом режиме;
 3. выбор наиболее подходящих инструментов для создания информационных систем сбора и обработки информации.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине		
			Код результата	Формулировка результата	
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-3 : Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем	РД1	Знание	методологию системного подхода
			РД11	Умение	проводить анализ структуры информационных ресурсов
			РД4	Знание	концептуальные основы методологии функциональной декомпозиции программных систем
			РД5	Умение	выполнить анализ требований и создание алгоритмов функционирования программного обеспечения
		ОПК-3.2к : Формулирует обоснованные выводы и рекомендации по предлагаемым техническим и программным решениям по профилю деятельности	РД10	Знание	теоретических проблем прикладной информатики, в том числе семантической обработки информации, развития представлений об оценке качества информации в информационных системах
			РД3	Навык	владение методами анализа и синтеза
			РД5	Умение	выполнить анализ требований и создание алгоритмов функционирования программного обеспечения
			РД8	Умение	применять на практике перспективных методик исследования прикладных и информационных процессов
	ОПК-4 : Способен применять на практике новые научные принципы и методы исследований	ОПК-4.1к : Изучает теоретико-методологические основы инновационных научных принципов исследований	РД1	Знание	методологию системного подхода
			РД9	Навык	навыками логико-методологического анализа научного исследования
		ОПК-4.2к : Применяет на практике новые инструментальные средства научных исследований при решении поставленных задач	РД12	Навык	сбора и обработки информации из открытых информационных ресурсов
			РД8	Умение	применять на практике перспективных методик исследования прикладных и информационных процессов
	ОПК-6 : Способен исследовать современные проблемы и методы	ОПК-6.1к : Исследует технологии в сфере прикладной информатики,	РД12	Навык	сбора и обработки информации из открытых информационных ресурсов
			РД7	Знание	основных понятия методов интеллектуальной обработки данных и организации баз знаний

	прикладной информатики и развития информационного общества	влияющие на развитие информационного общества	РД9	Навык	навыками логико-методологического анализа научного исследования
	УК-1 : Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.2к : Принимает конкретные решения для повышения эффективности процедур анализа проблем, принятия решений и разработки стратегий	РД1	Знание	методологию системного подхода
			РД2	Умение	решать задачи, требующие навыков абстрактного мышления
			РД3	Навык	владение методами анализа и синтеза

2 Место дисциплины (модуля) в структуре ОПОП

Данная дисциплина относится к обязательной части направления подготовки.

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обуче- ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо- емкость	Объем контактной работы (час)						СРС	Форма аттес- тации
				(З.Е.)	Всего	Аудиторная			Внеауди- торная			
						лек.	прак.	лаб.	ПА	КСР		
09.04.03 Прикладная информатика	ОФО	М01.Б	2	4	33	8	24	0	1	0	111	Э

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы	Код ре-зультата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	
1	Информация в современном мире	РД1, РД6, РД7	2	0	0	10	не предусмотрен
2	Источники и способы хранения цифровой информации	РД4, РД5, РД6, РД7	2	0	0	12	не предусмотрен
3	Методы и средства сбора данных	РД1, РД6, РД8, РД9	2	0	0	16	не предусмотрен
4	Методы анализа данных	РД2, РД3, РД7, РД8, РД9	2	0	0	18	не предусмотрен
5	Работа с пропущенными данными	РД3, РД7, РД8	0	4	0	14	отчет о выполнении практического задания
6	Веб-технологии хранения, обработки и передачи информации	РД2, РД5, РД6, РД8	0	8	0	14	отчет о выполнении практического задания
7	API-интерфейсы предоставления информации	РД4, РД6, РД8, РД9	0	6	0	14	отчет о выполнении практического задания
8	Структурированные виды данных	РД5, РД7	0	6	0	13	отчет о выполнении практического задания
Итого по таблице			8	24	0	111	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Информация в современном мире.

Содержание темы: 1.1. Основные определения информации. Типы и виды носителей и накопителей информации. 1.2. Жизненный цикл информации. Преобразования информации в цифровой вид.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем литературы и использование интернет-ресурсов.

Тема 2 Источники и способы хранения цифровой информации.

Содержание темы: 2.1. Информационные ресурсы. 2.2. Базы данных. 2.3. Веб-ресурсы. 2.4. Распределенные информационные системы.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция.

Виды самостоятельной подготовки студентов по теме: изучение лекционного материала и литературы по данной теме.

Тема 3 Методы и средства сбора данных.

Содержание темы: 3.1. Особенности хранения информации на локальных и глобальных ресурсах. 3.2. Глобальные информационные ресурсы. 3.3. Веб-технологии хранения, обработки и передачи информации. 3.4. API-интерфейсы предоставления информации. 3.5. Структурированные виды данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем литературы и использование интернет-ресурсов.

Тема 4 Методы анализа данных.

Содержание темы: 4.1. Распределения вероятностей. 4.2. Статистика выводов. 4.3. Анализ табличных данных. 4.4. Основы регрессионного анализа. 4.5. Интеллектуальный анализ данных.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Лекция.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем литературы и использование интернет-ресурсов.

Тема 5 Работа с пропущенными данными.

Содержание темы: 1.Обнаружения пропущенных, недопустимых и бесконечных значений с помощью функции `is.na()` , `is.nan()` и `is.infinite()` программы R-Studio. 2.Изучение структуры пропусков в соотношении с другими переменными массива с помощью функции `md.pattern()` библиотеки `mice` программы R-Studio. 3. Функции для визуализации структуры пропущенных данных `aggr()`, `matrixplot()` и `scattMiss()` пакета `VIM` программы R-Studio. 4. Обнаружения строк в данных, которые не содержат пропущенных значений с помощью функции `complete.cases()`. Построчное удаление. 5.Простое восстановление данных с помощью пакета `Hmisc`. 6.Многомерное восстановление данных с помощью пакета `mice`. Моделирование набора массивов для пропущенных значений с помощью функции `mice()` . Применение необходимого статистического метода с помощью функции `with()` (например, линейная регрессия). Объединение результатов с помощью функции `pool()`. 7. Специализированные методы работы с пропущенными данными. Оценка методом максимального правдоподобия для многомерных нормальных данных с пропущенными значениями с помощью пакета `Mvnmle` .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Практическое занятие.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем литературы и использование интернет-ресурсов.

Тема 6 Веб-технологии хранения, обработки и передачи информации.

Содержание темы: 1.Анализ структуры HTML документа. 2. Программная реализация GET и POST-запросов. 3. Разбор DOM-дерева HTML документа. 4. Автоматизация сбора данных из открытого веб-ресурса.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Практическое занятие.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем литературы и использование интернет-ресурсов.

Тема 7 API-интерфейсы предоставления информации.

Содержание темы: 1.Структура и формат данных XML и JSON. 2. Генерация и разбор XML-документа. 3. Реализация сервера на основе RESTful API. 4. Реализация интерфейса взаимодействия с API-интерфейсом .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Практическое занятие.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем

литературы и использование интернет-ресурсов.

Тема 8 Структурированные виды данных.

Содержание темы: 1. Структура и формат документов в формате MS Office. 2. COM-объекты MS Office. 3. Автоматизация анализа файлов формата MS Word и Excel .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: Практическое занятие.

Виды самостоятельной подготовки студентов по теме: изучение предлагаемой преподавателем литературы и использование интернет-ресурсов.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

Перечень и тематика самостоятельных работ студентов по дисциплинеПреимущества и недостатки ORM-подхода к манипулированию данными

1. Идентификаторы URI и URL.
2. Web 2.0 – анализ термина, история возникновения, описание основных признаков и критериев отнесения тех или иных сервисов к web 2.0.
3. HTML5. Различия между HTML и HTML5.
4. RSS/Atom – назначение, структурная организация. Возможности синдикации.
5. CURL – технология реализация html -запросов.
6. Структура и формат GET и POST -запросов формата html .
7. Основные отличия RESTful и REST API.
8. Конвертация данных из excel в xml -документ.
9. Использование элементов ActiveX в MS Office .

Самостоятельная работа студентов заключается в изучении предлагаемой литературы для усвоения и углубления полученных аудиторных знаний.

Для самостоятельной подготовки к занятиям рекомендуется использовать предложенную в РПД литературу и информационные ресурсы сети Интернет

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.
- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Практика ЭВМ - HTML / Д.В. Груздев .— Воронеж : Издательский дом ВГУ, 2017 .— 108 с. — 108 с. — URL: <https://lib.rucont.ru/efd/670081> (дата обращения: 18.07.2024)
2. Прикладной анализ данных / А.С. Цыбиков .— Улан-Удэ : Бурятский государственный

университет, 2017 .— 150 с. — ISBN 978-5-9793-1206-4 .— URL: <https://lib.rucont.ru/efd/704776> (дата обращения: 18.07.2024)

3. Шабашов В. Я. Организация доступа к данным из РНР приложений для различных СУБД : Учебники [Электронный ресурс] - Москва|Берлин : Директ-Медиа , 2019 - 121 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=499185

7.2 Дополнительная литература

1. Информационные технологии. HTML и XHTML : учеб. пособие / А.И. Костюк, С.М. Гушанский, М.Ю. Поленов, Б.В. Катаев; Южный федеральный ун-т .— Ростов-на-Дону : Изд-во ЮФУ, 2015 .— 131 с. : ил. — ISBN 978-5-9275-1329-1 .— URL: <https://lib.rucont.ru/efd/637103> (дата обращения: 18.07.2024)

2. П.В. Садчиков. Статистические приложения в Excel 2010 [Электронный ресурс] : Воронеж: Издательский дом Воронежского государственного университета , 2014 - 29 - Режим доступа: <https://lib.rucont.ru/efd/294539>

3. Стефанова, И.А. Изучение языка гипертекстовой разметки HTML5 : задания и метод. указания к лаб. работам по информатике для студентов телекоммуникационных и экономических направлений: 38.03.05 - Бизнес-информатика, 38.03.02 - Информационный менеджмент, 09.03.05 - Прикладная информатика / И.А. Стефанова .— Самара : Изд-во ПГУТИ, 2015 .— 55 с. — URL: <https://lib.rucont.ru/efd/319861> (дата обращения: 18.07.2024)

4. Тузовский А. Ф. ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА WEB-ПРИЛОЖЕНИЙ. Учебное пособие для СПО [Электронный ресурс] , 2019 - 218 - Режим доступа: <https://urait.ru/book/proektirovanie-i-razrabotka-web-prilozheniy-442423>

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Электронная библиотечная система «РУКОНТ» - Режим доступа: <https://lib.rucont.ru/>
2. Электронная библиотечная система «Университетская библиотека онлайн» - Режим доступа: <http://biblioclub.ru/>

3. Электронно-библиотечная система "РУКОНТ"

4. Электронно-библиотечная система издательства "Юрайт" - Режим доступа: <https://urait.ru/>

5. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: <http://oaji.net/>

6. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: <https://www.prlib.ru/>

7. Информационно-справочная система "Консультант Плюс" - Режим доступа: <http://www.consultant.ru/>

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Компьютеры
- Проектор

Программное обеспечение:

- Microsoft Office 2010 Standart
- Microsoft Windows 7 Russian
- RStudio
- Visual Studio
- СПС КонсультантПлюс: Версия Проф