

Документ подписан простой электронной подписью Информация о владельце: ФИО: Гаскаев Сергей Валерьевич Должность: Ректор	МИНОБРАЗОВАНИЯ И НАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)	
Дата подписания: 03.07.2024 14:09:00 Уникальный программный ключ: 0919240801978145010064019307888871255	Рабочая программа дисциплины "Информационный поиск, анализ и предобработка данных" по направлению подготовки (специальности) 01.04.02 "Прикладная математика и информатика" направленности (профилю) Технологии и методы искусственного интеллекта в фундаментальных и прикладных исследованиях ФГБОУ ВО «ЧелГУ»	стр. 1

**Рабочая программа дисциплины (модуля)\***  
**Информационный поиск, анализ и предобработка данных**

Направление подготовки (специальность)

01.04.02 Прикладная математика и информатика

Направленность (профиль)

Технологии и методы искусственного интеллекта в фундаментальных и прикладных исследованиях

Присваиваемая квалификация (степень)

магистр

Форма обучения

очная

Год(ы) набора 2024

\*Рабочая программа дисциплины (модуля) адаптирована для инклюзивного обучения инвалидов и лиц с ограниченными возможностями здоровья

Челябинск 2024 г.



## Содержание

1. Цели освоения дисциплины
2. Место дисциплины в структуре ОПОП
3. Компетенции обучающегося, формируемые в результате освоения дисциплины (модуля)
4. Объем дисциплины (модуля)
5. Структура и содержание дисциплины (модуля)
6. Фонд оценочных средств
  - 6.1. Перечень видов оценочных средств
  - 6.2. Типовые контрольные задания и иные материалы для текущей аттестации
  - 6.3. Типовые контрольные вопросы и задания для промежуточной аттестации
  - 6.4. Критерии оценивания
7. Учебно-методическое и информационное обеспечение дисциплины (модуля)
  - 7.1. Рекомендуемая литература
  - 7.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"
  - 7.3. Перечень информационных технологий
8. Материально-техническое обеспечение дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Специальные условия освоения дисциплины обучающимися с инвалидностью и ограниченными возможностями здоровья



### 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью дисциплины является изучение методологии поиска и подготовки данных для анализа и обработки с использованием искусственного интеллекта и машинного обучения. Задачами дисциплины являются изучение современных технологий поисковых систем, методов анализа и предварительной обработки данных, а также приобретение навыков программной реализации изучаемых методов с использованием различных программных сред.

Краткое содержание дисциплины

В курсе рассматриваются основные понятия информационного поиска, технологии поисковых систем, этапы подготовки данных для машинного обучения в рамках методологии CRISP-DM, методы анализа и предварительной обработки временных рядов.

Результаты обучения по данной дисциплине направлены на достижение индикаторов соответствующих компетенций:

### 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Цикл (раздел) ОПОП: Б1.В.01

#### 2.1 Требования к предварительной подготовке обучающегося:

Знание основ языка программирования Python

#### 2.2 Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:

Компьютерное зрение

Интеллектуальный анализ текстов

### 3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

**ПК-4: Способен выбирать и применять методы инженерии знаний для создания систем, основанных на знаниях**

#### Знать:

Для достижения ПК-4.1: методологические подходы к выбору и разработке методов получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов и применения соответствующих инструментальных средств

#### Уметь:

Для достижения ПК-4.2: выбирать и применять методы и средства получения знаний инженером по знаниям от экспертов; извлечения знаний из данных и текстов

**ПК-6: Способен разрабатывать и применять методы и алгоритмы машинного обучения для решения задач искусственного интеллекта**

#### Знать:

Для достижения ПК-6.1: унифицированные и обновляемые методологии описания, сбора и разметки данных, а также механизмы контроля за соблюдением указанных методологий

#### Владеть:

Для достижения ПК-6.3: навыком разработки унифицированных и обновляемых методологий описания, сбора и разметки данных, а также механизмов контроля за соблюдением указанных методологий

**ПК-12: Способен разрабатывать и применять алгоритмы анализа данных для решения прикладных задач**

#### Знать:

Для достижения ПК-12.1: основные принципы сбора, хранения и предобработки данных

#### Уметь:

Для достижения ПК-12.2: выбирать методы и средства для анализа данных, оценивать возможности и ограничения используемых методов, осуществлять дискретизацию непрерывных данных с учётом решаемой задачи

#### Владеть:

Для достижения ПК-12.3: навыком сбора первичной информации, организации и хранения данных для конкретного



исследования, применения методов предобработки данных

**В результате освоения дисциплины обучающийся должен**

<b>3.1</b>	<b>Знать:</b>
3.1.1	основные понятия информационного поиска, технологии поисковых систем, этапы подготовки данных для машинного обучения в рамках методологии CRISP-DM, методы анализа и предварительной обработки временных рядов
<b>3.2</b>	<b>Уметь:</b>
3.2.1	выбирать методы и средства для анализа данных, оценивать возможности и ограничения используемых методов
<b>3.3</b>	<b>Владеть:</b>
3.3.1	сбора первичной информации, организации и хранения данных для конкретного исследования, применения методов предобработки данных

**4. ОБЪЕМ ДИСЦИПЛИНЫ (МОДУЛЯ)**

Общая трудоемкость	<b>3 ЗЕТ</b>
Часов по учебному плану : 108 в том числе : аудиторные занятия : 32 самостоятельная работа : 69,75 : контактная работа: 38,25 ИКР: 6,25	Виды контроля в семестрах:  зачеты 1

**5. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)**

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Квнс	Часов	Литература
	<b>Раздел 1. Информационный поиск</b>			
1.1	Основные понятия информационного поиска /Лек/	1	2	Л1.1 Л1.2Л2.1
1.2	Современные технологии поисковых систем /Лек/	1	2	Л1.1 Л1.2Л2.1
1.3	Вычисление показателей ссылочной авторитетности /Лаб/	1	4	Л1.1 Л1.2Л2.1
1.4	Подготовка к лабораторным работам /Ср/	1	32	Л1.1 Л1.2Л2.1
	<b>Раздел 2. Анализ и предварительная обработка данных</b>			
2.1	Понятие датасета для машинного обучения /Лек/	1	2	Л1.1 Л1.2Л2.1
2.2	Подготовка данных в рамках методологии CRISP-DM /Лек/	1	2	Л1.1 Л1.2Л2.1
2.3	Восстановление пропусков в данных /Лек/	1	2	Л1.1 Л1.2Л2.1
2.4	Отбор информативных признаков /Лек/	1	2	Л1.1 Л1.2Л2.1
2.5	Анализ и предварительная обработка временных рядов /Лек/	1	4	Л1.1 Л1.2Л2.1
2.6	Восстановление пропусков в данных /Лаб/	1	4	Л1.1 Л1.2Л2.1
2.7	Отбор информативных признаков /Лаб/	1	4	Л1.1 Л1.2Л2.1
2.8	Анализ временных рядов /Лаб/	1	4	Л1.1 Л1.2Л2.1
2.9	Доработка лабораторных работ /Ср/	1	16	Л1.1 Л1.2Л2.1
	<b>Раздел 3. Подготовка к зачету</b>			
3.1	Подготовка к зачету /Ср/	1	21,75	Л1.1 Л1.2Л2.1
	<b>Раздел 4. Иная контактная работа</b>			
4.1	Индивидуальные консультации, текущий контроль /ИКР/	1	6,25	

**6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

**6.1. Перечень видов оценочных средств**

Лабораторная работа №1-4.



Собеседование.

### 6.2. Типовые контрольные задания и иные материалы для текущей аттестации

Лабораторная работа 1

1. Привести описание двух методов вычисления показателя PageRank.
2. Написать программу для вычисления значений PageRank по матрице смежности страниц.
3. Сравнить результаты для нескольких исходных данных.

Лабораторная работа 2

1. Привести описание двух методов восстановления пропусков в данных.
2. Написать программу для обработки пропусков выбранными методами.
3. Подготовить несколько наборов данных с пропусками.
4. Сравнить результаты обработки данных разными методами.

Лабораторная работа 3

1. Привести описание двух методов отбора информативных признаков.
2. Написать программу для анализа набора данных выбранными методами.
3. Подготовить набор данных, включив в него заведомо избыточные атрибуты.
4. Сравнить результаты отбора признаков разными методами.

Лабораторная работа 4

1. Привести описание алгоритма спектрального сингулярного анализа.
2. Написать программу, которая выполняет декомпозицию временного ряда на аддитивные составляющие и выводит их основные характеристики (частотный спектр, корреляционную функцию, индекс фрактальности).
3. Подготовить временной ряд, который содержит тренд и сезонные колебания.
4. Привести результаты анализа временного ряда.

### 6.3. Типовые контрольные вопросы и задания для промежуточной аттестации

Собеседование

- Вопрос 1. Основные модули интеллектуальной поисковой системы и их назначение.  
Вопрос 2. Типы атрибутов в датасете. Примеры и особенности операций с атрибутами.

### 6.4. Критерии оценивания

Студент представляет преподавателю отчет о выполнении лабораторной работы, подготовленный в соответствии с правилами оформления. На защите студент отвечает на вопросы по отчету и содержанию соответствующей темы курса. Лабораторная работа считается выполненной в срок, если студент представил отчет на проверку не позднее срока, установленного преподавателем. Общий балл за контрольное мероприятие складывается из следующих показателей.

1. Соответствие заданию:
  - все пункты задания выполнены правильно – 2 балла;
  - некоторые пункты задания выполнены с незначительными недостатками – 1 балл;
  - некоторые пункты задания не выполнены или выполнены с существенными ошибками – 0 баллов.
2. Оценка программного кода:
  - отсутствуют замечания к программе – 2 балла;
  - незначительные недостатки в программе – 1 балл;
  - существенные ошибки в программе – 0 баллов.
3. Ответы на вопросы:
  - студент правильно ответил на все вопросы – 2 балла;
  - студент ответил на вопросы с незначительными недостатками – 1 балл;
  - студент затруднялся отвечать на вопросы или допустил существенные ошибки – 0 баллов.
4. Оформление отчета:
  - отсутствуют замечания к оформлению отчета – 2 балла;
  - незначительные замечания к оформлению отчета – 1 балл;
  - оформление отчета существенно не соответствует правилам – 0 баллов.
5. Соблюдение сроков выполнения:
  - лабораторная работа выполнена в срок – 2 балла;
  - лабораторная работа выполнена в течение недели после окончания срока – 1 балл;
  - лабораторная работа не выполнена в срок без уважительной причины – 0 баллов.

Макс. балл: 10.

Вес одного КМ: 0,25.



**Собеседование.**

На собеседовании студенту предлагается ответить на два вопроса. Во время собеседования преподаватель также может задать дополнительные вопросы, если студент не полностью раскрывает содержание основного вопроса. Ответ на каждый вопрос оценивается следующим образом:

- студент полностью и без ошибок раскрыл содержание вопроса – 5 баллов;
- студент полностью раскрыл содержание вопроса, но допустил незначительные ошибки в ответе – 4 балла;
- студент представил неполный ответ на вопрос, но правильно ответил на дополнительные вопросы – 3 балла;
- студент представил неполный ответ на вопрос и не ответил на дополнительные вопросы – 2 балла;
- студент не представил ответ на вопрос, но правильно ответил на дополнительные вопросы – 1 балл;
- студент не представил ответ на вопрос и не ответил на дополнительные вопросы – 0 баллов

Макс. балл: 10.

Вес: 1.

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

### 7.1. Рекомендуемая литература

#### 7.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л1.1	Теофили Т.	Глубокое обучение для поисковых систем ( <a href="https://e.lanbook.com/book/140574">https://e.lanbook.com/book/140574</a> )	Москва : ДМК Пресс, 2020	ЭБС
Л1.2	Келлехер Д., Тирни Б.	Наука о данных: базовый курс: учебное пособие ( <a href="https://biblioclub.ru/index.php?page=book&amp;id=598235">https://biblioclub.ru/index.php?page=book&amp;id=598235</a> )	Москва : Альпина Паблишер, 2020	ЭБС

#### 7.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год	Ресурс
Л2.1	Маккинни У.	Python и анализ данных ( <a href="https://e.lanbook.com/book/131721">https://e.lanbook.com/book/131721</a> )	Москва : ДМК Пресс, 2020	ЭБС

### 7.3 Перечень информационных технологий

#### 7.3.1 Программное обеспечение

LMS Moodle

Mathcad Prime (Лицензия Математический факультет)

Python

Visual Studio

LibreOffice

#### 7.3.2 Профессиональные базы данных и информационно-справочные системы

1. Интернет университет информационных технологий [Электронный ресурс]. – Электрон. дан. – URL: <http://www.intuit.ru/>. – Текст : электронный.

2. Справочник «Информо» (<http://www.informio.ru/>) ИНФОРМИО : электронный справочник [обеспечение всех типов образовательных учреждений нормативными, методическими, научнопрактическими материалами]. – URL: <http://www.informio.ru/>. – Режим доступа: для зарегистрир. пользователей ЧелГУ. – Текст : электронный.

## 8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Для реализации дисциплины используются учебные аудитории для проведения занятий лекционного типа, лабораторных занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.

Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения: проектором, экраном, магнитно-маркерной доской, маркером; с возможностью демонстрации электронных презентаций при уровне освещения, достаточном для работы с конспектом.

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий (мультимедийные презентации по отдельным темам, рисунки, таблицы, схемы и т.д).

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с подключением к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.

## 9. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)



Рабочей программой дисциплины предусмотрена самостоятельная работа студентов. Самостоятельная работа проводится с целью углубления знаний по дисциплине и предусматривает:

- проработку теоретического материала по учебникам или конспекту лекций с обязательным разбором приведенных примеров;
- подготовку к занятиям;
- выполнение контрольной работы;
- подготовку доклада.

При планировании времени на самостоятельную работу студентам необходимо предусмотреть регулярное повторение пройденного материала. Теоретический материал, законспектированный на лекциях, необходимо дополнять сведениями из литературных источников, представленных в рабочей программе.

В случае применения при изучении дисциплины электронного обучения, дистанционных образовательных технологий общение обучающихся и преподавателя осуществляется в режиме реального или отложенного времени, при этом используются возможности системы дистанционного обучения Moodle и электронная почта.

Большую часть времени обучающиеся самостоятельно работают с учебно-методическими материалами. Студенты имеют возможность консультироваться с преподавателем по всем вопросам, возникающим в ходе самостоятельной работы, посредством электронной почты, сообщений системы дистанционного обучения Moodle.

Доступ обучающегося к учебным ресурсам в режиме отложенного времени, самостоятельной работы осуществляется через сеть Интернет в удобном для него месте, времени и темпе.

При обучении инвалидов и лиц с ограниченными возможностями здоровья электронное обучение, дистанционные образовательные технологии предусматривают возможность приема-передачи информации в доступных для них формах. Реализация дисциплины с применением электронного обучения, дистанционных образовательных технологий (далее – ЭО, ДОТ) осуществляется на основании «Положения о реализации основных и дополнительных образовательных программ с применением электронного обучения и дистанционных образовательных технологий в федеральном государственном бюджетном образовательном учреждении высшего образования «Челябинский государственный университет», «Положения о порядке зачета обучающимся по основным профессиональным образовательным программам высшего образования в ФГБОУ ВО «ЧелГУ» результатов освоения в организациях, осуществляющих образовательную деятельность, учебных предметов, курсов, дисциплин (модулей), практик, дополнительных образовательных программ» посредством электронной информационно-образовательной среды ФГБОУ ВО «ЧелГУ». В исключительных случаях (форс-мажор и т.п.) при реализации образовательной деятельности с применением ЭО, ДОТ могут применять компоненты, не входящие в перечень электронной информационно-образовательной среды.

## 10. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ ОБУЧАЮЩИМИСЯ С ИНВАЛИДНОСТЬЮ И ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Освоение дисциплины инвалидами и лицами с ограниченными возможностями здоровья осуществляется с использованием специальных технических средств и голо информационных технологий, предоставляемых Ресурсным учебно-методическим центром по обучению инвалидов и лиц с ограниченными возможностями здоровья ЧелГУ по запросу обучающегося.

1. Мобильные специальные технические средства для лиц с нарушениями зрения: портативный компьютер с вводом/выводом шрифтом Брайля с синтезатором речи «E1Braille-W14J G2»; ноутбуки с программной экранного доступа NVDA; электронные увеличители для удаленного просмотра; видеоувеличители портативные; тифлоплеер; цифровые диктофоны.

2. Мобильные специальные технические средства для лиц с нарушениями слуха: система свободного звукового поля со встроенной совместимостью с FM-устройствами; радиоклассы «Сонет-PCM» с передатчиком, заушным индуктором и индукционной петлей; система информационная для слабослышащих переносная «Исток» А2 со встроенным плеером – звуковым информатором; документ-камера; программируемые слуховые аппараты индивидуального пользования.

3. Ассистивные информационные технологии: программное обеспечение экранного доступа с синтезом речи NVDA; программы экранного увеличения; программы речевого синтеза для компьютеров и ноутбуков; программы речевого синтеза для мобильных устройств; экранная клавиатура; экранная лупа.

При необходимости для обучающихся с нарушениями зрения на рабочих местах для проведения практических или лабораторных занятий устанавливается специальное программное обеспечение (программа речевой навигации NVDA, речевые синтезаторы, экранные лупы).

В учебные аудитории обеспечивается беспрепятственный доступ для обучающихся инвалидов и обучающихся с ограниченными возможностями здоровья. В каждой аудитории, где обучаются инвалиды и лица с ограниченными возможностями здоровья, предусматривается соответствующее количество мест для обучающихся с учетом нарушений их здоровья.



Для освоения дисциплины инвалидам и лицам с ограниченными возможностями здоровья предоставляется доступ к печатным источникам, имеющимся в научной библиотеке ЧелГУ, с помощью специальных технических средств; доступ к электронным источникам, представленным в форме электронного документа в фонде научной библиотеки ЧелГУ или электронно-библиотечных системах, с помощью специальных технических и программных средств (рабочее место для незрячего пользователя с программным обеспечением экранного доступа с синтезом речи NVDA, рабочее место с компьютерным роллером и клавиатурой CleVu с большими кнопками и с разделяющей клавиши накладкой).

Учебно-методические материалы для обучающихся из числа инвалидов и лиц с ограниченными возможностями здоровья предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:

- в печатной форме увеличенным шрифтом,
- в форме электронного документа,
- в форме аудиофайла,
- в печатной форме шрифтом Брайля.

Для лиц с нарушениями слуха:

- в печатной форме,
- в форме электронного документа.

Для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме,
- в форме электронного документа,
- в форме аудиофайла.

Данный перечень может быть конкретизирован в зависимости от контингента обучающихся.

Для инвалидов и лиц с ограниченными возможностями здоровья освоение дисциплины может быть частично или полностью осуществлено с использованием дистанционных образовательных технологий (Moodle, Adobe Connect Pro и пр.).

В освоении дисциплины инвалидами и лицами с ограниченными возможностями здоровья используется индивидуальная работа. Под индивидуальной работой подразумевается две формы взаимодействия с преподавателем: индивидуальная учебная работа (консультации), т.е. дополнительное разъяснение учебного материала и углубленное изучение материала с теми обучающимися, которые в этом заинтересованы, и индивидуальная воспитательная работа. Индивидуальные консультации направлены на индивидуализацию обучения и установлению воспитательного контакта между преподавателем и обучающимся инвалидом или обучающимся с ограниченными возможностями здоровья.

При проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине обеспечивается выполнение следующих дополнительных требований в зависимости от индивидуальных особенностей, обучающихся:

- а) инструкция по порядку проведения процедуры оценивания предоставляется в доступной форме (устно, в письменной форме, в письменной форме шрифтом Брайля, устно с использованием услуг сурдопереводчика);
- б) доступная форма предоставления заданий оценочных средств (в печатной форме, в печатной форме увеличенным шрифтом, в печатной форме шрифтом Брайля, в форме электронного документа, задания зачитываются ассистентом, задания предоставляются с использованием сурдоперевода);
- в) доступная форма предоставления ответов на задания (письменно на бумаге, набор ответов на компьютере, письменно шрифтом Брайля, с использованием услуг ассистента, устно).

При проведении процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья предусматривается использование технических средств, необходимых им в связи с их индивидуальными особенностями. Эти средства могут быть предоставлены ЧелГУ или могут использоваться собственные технические средства. При необходимости инвалидам и лицам с ограниченными возможностями здоровья предоставляется дополнительное время для подготовки ответа на задания, процедура оценивания результатов обучения по дисциплине может проводиться в несколько этапов.

Проведение процедуры оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья допускается с использованием дистанционных образовательных технологий.

