

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»

ФАКУЛЬТЕТ БИЗНЕС–ИНФОРМАТИКИ И УПРАВЛЕНИЯ КОМПЛЕКСНЫМИ
СИСТЕМАМИ

КАФЕДРА ЭКОНОМИКИ И МЕНЕДЖМЕНТА В ПРОМЫШЛЕННОСТИ

ОДОБРЕНО УМС ФБИУКС

Протокол № 24/08

от 22.08.2024 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ТЕКСТОВ

Направление подготовки
(специальность)

[1] 38.04.05 Бизнес-информатика

Семестр	Трудоемкость, кред.	Общий объем курса, час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	В форме практической подготовки/ В	СРС, час.	КСР, час.	Форма(ы) контроля, экз./зач./КР/КП
3	2	72	12	0	20		40	0	3
Итого	2	72	12	0	20	20	40	0	

АННОТАЦИЯ

В рамках данной дисциплины рассматриваются основные понятия, задачи и современные методы и технологии в области интеллектуального анализа текстов, также называемого обработкой текстов на естественном языке (Natural Language Procession).

Рассматриваются основы теории языка. Рассматриваются задачи в области определения тональности текстов, машинного перевода, автоматического аннотирования и генерации текстов, и др. В качестве методов рассматриваются как методы классического машинного обучения, так и методы глубокого обучения, с использованием нейросетевых моделей. В качестве основного инструмента работы с текстами используется язык Python.

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ УЧЕБНОЙ ДИСЦИПЛИНЫ

Целью изучения дисциплины «Интеллектуальный анализ текстов» являются формирование у студентов теоретико-методологических знаний и закрепление профессиональных навыков в области решения прикладных задач в различных сферах государственной, корпоративной и общественной деятельности на основе методов и технологий обработки текстов на естественном языке.

2. МЕСТО УЧЕБНОЙ ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Учебная дисциплина «Интеллектуальный анализ текстов» относится к дисциплинам базовой части обще-профессионального модуля.

Содержание программы «Интеллектуальный анализ текстов» представляет собой развитие полученных ранее знаний в области анализа данных в рамках учебной дисциплины «Математические и инструментальные методы машинного обучения».

3. ФОРМИРУЕМЫЕ КОМПЕТЕНЦИИ И ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Универсальные и(или) общепрофессиональные компетенции:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции
--------------------------------	--

Профессиональные компетенции в соответствии с задачами и объектами (областями знаний) профессиональной деятельности:

Задача профессиональной деятельности (ЗПД)	Объект или область знания	Код и наименование профессиональной компетенции; Основание (профессиональный стандарт-ПС, анализ опыта)	Код и наименование индикатора достижения профессиональной компетенции
аналитический			

Исследование и разработка методов и инструментов создания и развития цифровых предприятий, цифровых двойников и сквозных цифровых технологий	Методы и инструменты создания и развития цифровых предприятий и цифровых двойников и сквозных цифровых технологий	ПК-2.3 [1] - Способен применять методы машинного обучения и интеллектуального анализа информации для развития цифровых предприятий, цифровых двойников и сквозных цифровых технологий <i>Основание:</i> Профессиональный стандарт: 24.078, Анализ опыта: По согласованию с Заказчиками образовательной программы Трудовая функция: «Выполнение деятельности по использованию методов машинного обучения и интеллектуального анализа информации для развития цифровых предприятий, цифровых двойников и сквозных цифровых технологий»	З-ПК-2.3[1] - Знать: Технологии, методы и инструментальные средства обработки больших данных; Основные методы интеллектуальной обработки данных (Data Mining), в том числе и больших данных; Математические и инструментальные методы машинного обучения; Методы интеллектуального анализа текстов; Основные положения цифровой экономики; Основные направления Стратегии внутренней цифровизации Госкорпорации «Росатом»; Основные положения дорожных карт сквозных цифровых технологий.; У-ПК-2.3[1] - Уметь: Использовать модели данных, адаптированных к технологиям больших данных; Пользоваться методами и инструментами получения, хранения, передачи, обработки больших данных; Применять методы интеллектуальной обработки данных (Data Mining), в том числе и больших данных, в зависимости от предметной области и специфики деятельности организации; Применять математические и
---	--	--	---

			инструментальные методы машинного обучения в зависимости от поставленной задачи; Применять интеллектуального анализа текстов в зависимости от поставленной задачи.; В-ПК-2.3[1] - Владеть: Разрабатывать модели данных, адаптированных к технологиям больших данных; Разрабатывать предложения по развитию и совершенствованию системы получения, хранения, передачи, обработки больших данных; Использовать математические и инструментальные методы машинного обучения и интеллектуального анализа информации в целях развития цифровых предприятий, цифровых двойников и сквозных цифровых технологий.
научно-исследовательский			
Исследование, разработка и внедрение новых моделей, методов и средств в области экономики, управления и ИКТ для совершенствования архитектуры предприятия	Архитектура предприятия (бизнес-архитектура, архитектура информации, архитектура приложений, инфраструктура)	ПК-1 [1] - Способен проводить исследования и поиск новых моделей и методов в области экономики, управления и ИКТ для совершенствования архитектуры предприятия и выявления инновации; <i>Основание:</i> Профессиональный стандарт: 06.012	З-ПК-1[1] - Знать: современные ИТ, широкий кругозор в области ИТ, понимание соотношения целей и путей реализации стратегии развития ИТ; предметная функциональная область применения ИТ; принципы инновационной деятельности ; У-ПК-1[1] - Уметь:

			определять возможности использования инноваций ИТ в стратегическом управлении; интегрировать ИТ в деятельность организации; В-ПК-1[1] - Владеть навыками: формирование целей, приоритетов и ограничений формирования ИТ в создание и реализацию инновационной стратегии и изменение их по мере изменения внешних условий и внутренних потребностей; организация работы персонала и выделение ресурсов для формирования вклада ИТ в создание и реализацию инновационной стратегии; контроль формирования вклада ИТ в создание и реализацию инновационной стратегии; анализ формирования вклада ИТ в создание и реализацию инновационной стратегии, целей, приоритетов и ограничений процесса и выполнение управленческих действий по результатам анализа
--	--	--	---

4. СТРУКТУРА И СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Разделы учебной дисциплины, их объем, сроки изучения и формы контроля:

№ п.п	Наименование раздела учебной дисциплины	Недели	Лекции/ Практи. (семинары) / Лабораторные работы, час.	Обязат. текущий контроль (форма*, неделя)	Максимальный балл за раздел**	Аттестация раздела (форма*, неделя)	Индикаторы освоения компетенции
	<i>3 Семестр</i>						
1	Первый раздел	1-8	6/0/10	Т-7 (20), ЛР-8 (20)	40	КИ-8	3-ПК-1, 3-ПК-2.3, 3-ПК-2.3
2	Второй раздел	9-16	6/0/10	Т-15 (20), ЛР-16 (20)	40	КИ-16	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-2.3, У-ПК-2.3, В-ПК-2.3, 3-ПК-2.3, У-ПК-2.3, В-ПК-2.3
	<i>Итого за 3 Семестр</i>		12/0/20		80		
	Контрольные мероприятия за 3 Семестр				20	3	3-ПК-1, У-ПК-1, В-ПК-1, 3-ПК-2.3, У-ПК-2.3, В-ПК-2.3, 3-ПК-2.3, У-ПК-2.3, В-ПК-2.3

* – сокращенное наименование формы контроля

** – сумма максимальных баллов должна быть равна 100 за семестр, включая зачет и (или) экзамен

Сокращение наименований форм текущего контроля и аттестации разделов:

Обозначение	Полное наименование
Т	Тестирование
ЛР	Лабораторная работа
КИ	Контроль по итогам
З	Зачет

КАЛЕНДАРНЫЙ ПЛАН

Недели	Темы занятий / Содержание	Лек., час.	Пр./сем., час.	Лаб., час.
	<i>3 Семестр</i>	12	0	20
1-8	Первый раздел	6	0	10
1	Обработка текстов на естественном языке.	Всего аудиторных часов		

	Применение методов машинного обучения Основные понятия и задачи интеллектуального анализа текстов (Natural Language Procession). Морфология. Синтаксис. Семантика. Прагматика. Дискурс. Токенизация. Векторизация. Модель Мешок слов. Метрика Tf-idf. N-граммы. Регулярные выражения. Библиотека re. Классификаторы в NLP. Метрики качества. Точность. Полнота. F-мера. Разметка частей речи (Part-of-Speech). Библиотека nltk. Морфологический анализ. Стемминг. Лемматизация. Библиотека rymorphy. Оценка тональности. Библиотека sklearn. Тематическое моделирование. Латентный семантический анализ. Нейросети прямого распространения. Распределенные модели слов. Модель CBOW. word2vec и doc2vec.	6	0	10
		Онлайн		
		0	0	0
9-16	Второй раздел	6	0	10
	Обработка текстов на естественном языке. Применение методов глубокого обучения Сверточные нейронные сети. Обучение с первого раза (One-Shot Learning). Энкодеры и декодеры. Библиотека Keras. Автоматическое аннотирование (суммаризация текста). Метрика ROUGE (Recall-Oriented Understudy for Gisting Evaluation). Модели seq2seq. Генерация текстов. Рекуррентные нейросети. Long-short Term Memory (LSTM). Определение смысла слова. Генерация изображений по текстовым описаниям. Машинный перевод. Механизм внимания. Вопрос-ответ. Трансформеры. Модель BERT. Библиотека DeepPavlov. Современные архитектуры. Обзор нейросетей ChatGpt, Midjourney.	Всего аудиторных часов		
		6	0	10
		Онлайн		
		0	0	0

Сокращенные наименования онлайн опций:

Обозначение	Полное наименование
ЭК	Электронный курс
ПМ	Полнотекстовый материал
ПЛ	Полнотекстовые лекции
ВМ	Видео-материалы
АМ	Аудио-материалы
Прз	Презентации
Т	Тесты
ЭСМ	Электронные справочные материалы
ИС	Интерактивный сайт

ТЕМЫ ЛАБОРАТОРНЫХ РАБОТ

Недели	Темы занятий / Содержание
	<i>3 Семестр</i>
	Обработка текстов на естественном языке. Применение методов машинного обучения Лабораторная работа №1. Освоение библиотеки sklearn. Векторизация корпуса текстов. Визуализация полученных векторов. Кластеризация текстов. Лабораторная работа №2. Освоение библиотеки re. Использование регулярных

	выражений для подготовки текстов. Использование <code>rumorphy</code> для определения части речи и лемматизации текстов. Лабораторная работа №3. Освоение библиотеки <code>nlTK</code>. Тематическое моделирование. LSA. Лабораторная работа №4. Освоение библиотеки <code>gensim</code>. Модель CBOW. Нахождение близких текстов. Лабораторная работа №5. Освоение библиотеки <code>spacy</code>. Распознавание именованных сущностей.
	Обработка текстов на естественном языке. Применение методов глубокого обучения Лабораторная работа №6. Освоение библиотеки Keras. Автоматическое аннотирование текстов. Лабораторная работа №7. Генерация текстов. Рекуррентные сети. LSTM. Лабораторная работа №8. Машинный перевод. <code>seq2seq</code>. Лабораторная работа №9. Генерация изображений по текстовым описаниям. Лабораторная работа №10. Освоение библиотеки DeepPavlov. Вопрос-ответ

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

При реализации программы дисциплины используются различные интерактивные образовательные технологии – занятия проводятся в виде анализа деловых ситуаций с визуализацией с использованием технических средств обучения.

Самостоятельная работа студентов подразумевает под собой проработку теоретического материала с использованием рекомендуемой литературы для выполнения практических заданий и подготовке к итоговой форме контроля, а также интерактивные формы обучения в виде выполнения тестов с помощью электронных учебных элементов для системы электронного обучения ИНФОМИФИСТ.

6. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств по дисциплине обеспечивает проверку освоения планируемых результатов обучения (компетенций и их индикаторов) посредством мероприятий текущего, рубежного и промежуточного контроля по дисциплине.

Связь между формируемыми компетенциями и формами контроля их освоения представлена в следующей таблице:

Компетенция	Индикаторы освоения	Аттестационное мероприятие (КП 1)
ПК-1	З-ПК-1	З, КИ-8, КИ-16, Т-7, ЛР-8, Т-15, ЛР-16
	У-ПК-1	З, КИ-16, Т-15, ЛР-16
	В-ПК-1	З, КИ-16, Т-15, ЛР-16
ПК-2.3	З-ПК-2.3	З, КИ-8, КИ-16, Т-7, ЛР-8, Т-15, ЛР-16
	У-ПК-2.3	З, КИ-16, Т-15, ЛР-16
	В-ПК-2.3	З, КИ-16, Т-15, ЛР-16

Шкалы оценки образовательных достижений

Шкала каждого контрольного мероприятия лежит в пределах от 0 до установленного максимального балла включительно. Итоговая аттестация по дисциплине оценивается по 100-балльной шкале и представляет собой сумму баллов, заработанных студентом при выполнении заданий в рамках текущего и промежуточного контроля.

Итоговая оценка выставляется в соответствии со следующей шкалой:

Сумма баллов	Оценка по 4-х балльной шкале	Оценка ECTS	Требования к уровню освоению учебной дисциплины
90-100	5 – «отлично»	A	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, использует в ответе материал монографической литературы.
85-89	4 – «хорошо»	B	Оценка «хорошо» выставляется студенту, если он твёрдо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос.
75-84		C	
70-74		D	
65-69	3 – «удовлетворительно»	E	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала.
60-64			
Ниже 60	2 – «неудовлетворительно»	F	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

ОСНОВНАЯ ЛИТЕРАТУРА:

1. ЭИ П 37 Машинное обучение : учебное пособие для вузов, Платонов А. В., Москва: Юрайт, 2022
2. ЭИ М 77 Машинное обучение с участием человека : , Монарх Р., Москва: ДМК Пресс, 2022

3. ЭИ Б 53 Системы искусственного интеллекта : учебное пособие для вузов, Бессмертный И. А., Москва: Юрайт, 2022
4. ЭИ О-48 Управление результативностью маркетинга : учебник и практикум для вузов, Ойнер О. К., Москва: Юрайт, 2022

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА:

1. 33 П90 Введение в технологический маркетинг при использовании атомной энергии : , Путилов А.В., Москва: Руда и металлы, 2005
2. 005 И66 Инновационный менеджмент. Концепции, многоуровневые стратегии и механизмы инновационного развития : учебное пособие для вузов, , Москва: Дело, 2007
3. ЭИ С 38 Маркетинг в предпринимательской деятельности : , Земляк С. В., Синяева И. М., Синяев В. В., Москва: Дашков и К, 2015
4. 33 Д79 Маркетинговые коммуникации : учебник для вузов, Дубровин И.А., Москва: Дашков и Ко, 2012
5. 005 У67 Управление инновационными проектами : учебное пособие, , Москва: ИНФРА-М, 2012
6. ЭИ Б 93 Управление маркетингом: учебник : , Жираткова А.Н., Казаков А.А., Бутова Т.Г., Moscow: Проспект, 2016

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ:

Специальное программное обеспечение не требуется

LMS И ИНТЕРНЕТ-РЕСУРСЫ:

<https://online.mephi.ru/>

<http://library.mephi.ru/>

8. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Специальное материально-техническое обеспечение не требуется

9. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ СТУДЕНТОВ

При реализации программы дисциплины «Интеллектуальный анализ текстов» используются различные интерактивные образовательные технологии с использованием технических средств обучения.

Во время лекционных занятий по дисциплине студент должен уметь сконцентрировать внимание на рассматриваемых проблемах и включить в работу все виды памяти: словесную, образную и моторно-двигательную. Для этого ему необходимо конспектировать материал,

излагаемый преподавателем. Во время конспектирования в работу включается моторно-двигательная память, позволяющая эффективно усвоить лекционный материал. Весь иллюстративный материал, представляемый на лекции (на слайдах, на доске, в раздаточном материале) также должен быть зафиксирован в конспекте лекций. Каждому студенту необходимо помнить о том, что конспектирование лекции – это не диктант. Студент должен выделять главное и фиксировать основные моменты.

Важное место в учебном процессе занимают лабораторные работы, призванные закреплять полученные студентами теоретические знания. Перед лабораторной работой студенту необходимо восстановить в памяти теоретический материал по теме практического занятия. Для этого следует обратиться к первоисточникам, конспекту лекций, настоящим методическим указаниям. Каждое занятие начинается с повторения теоретического материала по соответствующей теме. Студенты должны уметь чётко ответить на вопросы, поставленные преподавателем. По характеру ответов преподаватель делает вывод о том, насколько тот или иной студент готов к выполнению различных заданий.

Самостоятельная работа студентов предусматривает: ознакомление с рекомендованной литературой и презентациями лекций, в том числе с использованием Интернет; повторение пройденного на лекциях материала; работу над электронными тестами; а также интерактивные формы обучения в виде выполнения тестов с помощью электронных учебных элементов для системы электронного обучения ИНФОМИФИСТ.

Оценочными средствами рубежного контроля является Контроль Итогов (КИ) - означающий выставление баллов на основании результатов Текущего контроля отдельно для первого раздела и второго раздела. Текущий контроль проводится на основе проверки теста, выполнения лабораторных работ/

В конце освоения дисциплины студент сдает зачет.

10. УЧЕБНО-МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ДЛЯ ПРЕПОДАВАТЕЛЕЙ

Целью изучения дисциплины «Интеллектуальный анализ текстов» являются формирование у будущих магистров теоретико-методологических знаний и закрепление профессиональных навыков в области решения прикладных задач в различных сферах государственной, корпоративной и общественной деятельности на основе учета закономерностей становления и развития цифровой экономики, общих свойств информации и особенностей информационных процессов.

Интерактивные занятия со студентами проводятся с использованием презентаций, которые публикуются в системе ИНФОМИФИСТ.

Видеозапись каждого проведенного занятия выкладывается на портал home.merphi.ru.

Для наиболее эффективного изучения предусмотрена самостоятельная проработка студентами отдельных тем, освоение которых проверяется при защите курсовой работы. Целесообразно для увеличения времени проработки важных тем предусмотреть рассмотрение отдельных вопросов в форме дискуссий и диспутов. Кроме того, необходимо предусмотреть дополнительные консультации по сложным темам. Самостоятельная работа студентов включает в себя так же интерактивные формы обучения в виде выполнения тестов с помощью электронных учебных элементов для системы электронного обучения ИНФОМИФИСТ.

Оценочными средствами рубежного контроля является Контроль Итогов (КИ) - означающий выставление баллов на основании результатов Текущего контроля отдельно для первого раздела и второго раздела. Текущий контроль проводится с устного опроса и

подготовки реферата. Текущий контроль проводится на основе проверки тестов, выполнения лабораторных работ.

Итоговая аттестация проводится в форме зачёта.

Автор(ы):

Киреев Василий Сергеевич, к.т.н.