

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 03.06.2025 12:01:24
Уникальный программный ключ:
ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»**

Институт русского языка

(наименование основного учебного подразделения (ОУП)-разработчика ОП ВО)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

СОВРЕМЕННЫЕ ТЕХНОЛОГИИ ОБРАБОТКИ ТЕКСТОВ НА ЕСТЕСТВЕННЫХ ЯЗЫКАХ

(наименование дисциплины/модуля)

Рекомендована МССН для направления подготовки/специальности:

45.04.04 ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ В ГУМАНИТАРНОЙ СРЕДЕ

(код и наименование направления подготовки/специальности)

Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):

ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ И АНАЛИЗ ДАННЫХ В ГУМАНИТАРНОЙ СФЕРЕ

(наименование (профиль/специализация) ОП ВО)

2025 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Современные технологии обработки текстов на естественных языках» входит в программу магистратуры «Интеллектуальные технологии и анализ данных в гуманитарной сфере» по направлению 45.04.04 «Интеллектуальные системы в гуманитарной среде» и изучается в 3 семестре 2 курса. Дисциплину реализует Кафедра прикладной информатики и интеллектуальных систем в гуманитарной сфере. Дисциплина состоит из 3 разделов и 10 тем и направлена на изучение методов извлечения отношений, анализа тональности, аннотирования и кластеризации текстов, а также существующих программных реализаций этих методов.

Целью освоения дисциплины является формирование у студентов компетенций, необходимых для применения в профессиональной деятельности методов обработки текста на естественном языке, а также методов обработки слабоструктурированных данных и извлечения информации.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Современные технологии обработки текстов на естественных языках» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.1 Владеет принципами и методами критического анализа и оценки проблемных ситуаций на основе системного подхода; УК-1.2 Вырабатывает стратегию действий через постановку задач и определение алгоритма решения проблемных ситуаций;
ПК-1	Способен анализировать, формировать и согласовывать требования к интеллектуальным системам для задач гуманитарной сферы	ПК-1.1 Выявляет и формализует цели заинтересованных сторон, проблемы, решаемые построением интеллектуальной системы, и рамки автоматизации; ПК-1.2 Вырабатывает предложения по проектным решениям; ПК-1.3 Выявляет несоответствие требованиям заказчика к интеллектуальной системе с точки зрения архитектуры интеллектуальной системы;
ПК-3	Способен разрабатывать техническое задание на создание интеллектуальной системы	ПК-3.2 Разрабатывает разделы технического задания на создание интеллектуальной системы;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Современные технологии обработки текстов на естественных языках» относится к части, формируемой участниками образовательных отношений блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Современные технологии обработки текстов на естественных языках».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-1	Способен осуществлять поиск, критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	Теория систем и системный анализ; Проблемы современной философии; Введение в специальность. Цифровая гуманитаристика; Эксплуатационная практика (учебная);	Научно-исследовательская работа; Преддипломная практика;
ПК-1	Способен анализировать, формировать и согласовывать требования к интеллектуальным системам для задач гуманитарной сферы	Методология проектирования интеллектуальных систем; Техническая документация в ИТ-проектах; Математические методы в гуманитарных исследованиях;	Преддипломная практика; Научно-исследовательская работа; <i>Информационная безопасность интеллектуальных систем**;</i> <i>Интеллектуальные технологии машинного перевода**;</i> <i>Компьютерная лингвистика**;</i> <i>Инструменты разработки и запуска бизнес-проекта**;</i> <i>Создание технологического бизнеса**;</i> <i>Технологическая и эксплуатационная безопасность программного обеспечения**;</i>
ПК-3	Способен разрабатывать техническое задание на создание интеллектуальной системы	Методология проектирования интеллектуальных систем; Техническая документация в ИТ-проектах; Математические методы в гуманитарных исследованиях; Архитектура интеллектуальных систем;	Научно-исследовательская работа; Преддипломная практика; <i>Технологическая и эксплуатационная безопасность программного обеспечения**;</i> <i>Информационная безопасность интеллектуальных систем**;</i> <i>Инструменты разработки и запуска бизнес-проекта**;</i> <i>Создание технологического бизнеса**;</i> <i>Интеллектуальные технологии машинного перевода**;</i> <i>Компьютерная лингвистика**;</i>

* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Современные технологии обработки текстов на естественных языках» составляет «2» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			3
<i>Контактная работа, ак.ч.</i>	34		34
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	17		17
<i>Самостоятельная работа обучающихся, ак.ч.</i>	29		29
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	9		9
Общая трудоемкость дисциплины	ак.ч.	72	72
	зач.ед.	2	2

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы

Номер раздела	Наименование раздела дисциплины	Содержание раздела (темы)		Вид учебной работы*
Раздел 1	Языковые модели. Основы обработки текстов на естественном языке словарными подходами и методами, основанными на правилах.	1.1	Языковые модели. Анализ текста в парадигме когнитивных исследований. Анализ текста в парадигмах автоматического понимания текста.	ЛК, СЗ
		1.2	Коммуникативная и информационная (смысловая) структуры текста. Избыточность. Морфологический анализ и синтез. Словарный морфологический анализ и синтез.	ЛК, СЗ
		1.3	Компрессия текста. Лемматизация. Приведение текста в нормальную форму. Классификация текста. Оценка качества классификации.	ЛК, СЗ
		1.4	Создание словарей. Методы разработки правил.	ЛК, СЗ
Раздел 2	Обработка текстов на естественном языке методами машинного обучения.	2.1	Наивный Байесовский классификатор. Теорема Байеса. Функции ошибки и регуляризация. KL-расстояние и перекрестная энтропия.	ЛК, СЗ
		2.2	Градиентный спуск: основы. Граф вычислений и дифференцирование на нем. Логистическая регрессия.	ЛК, СЗ
		2.3	Метод опорных векторов. Деревья решений. Автоматическое извлечение признаков из текста. Подбор оптимальных признаков.	ЛК, СЗ
Раздел 3	Прикладные задачи обработки текстов на естественном языке.	3.1	Извлечение информации из текстов, выделение именованных сущностей. Анализ тональности текстов. Задача поиска отношений.	ЛК, СЗ
		3.2	Разрешение анафоры и кореференции. Машинный перевод. Синтаксические парсеры русского и английского языка.	ЛК, СЗ
		3.3	Вопросно-ответные системы. Диалоговые системы и чат-боты. Онтологии и тезаурусы. Тематическое моделирование. Кластеризация текстов.	ЛК, СЗ

* - заполняется только по **ОЧНОЙ** форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная		Технические средства мультимедиа презентаций (компьютер/ноутбук, экран, проектор), доска
Семинарская		Проектор и/или большой экран/монитор, компьютер, ОС MS Windows 10, MS Office 2016 или MS Office 2019

		или MS Office 365, портал Microsoft Azure, доступ к интернету
Для самостоятельной работы		Проектор и/или большой экран/монитор, компьютер, ОС MS Windows 10, MS Office 2016 или MS Office 2019 или MS Office 365, портал Microsoft Azure, доступ к интернету

* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Гольдберг, Й. Нейросетевые методы в обработке естественного языка : руководство / Й. Гольдберг ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 282 с. — ISBN 978-5-97060-754-1. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131704> (дата обращения: 24.11.2023). — Режим доступа: для авториз. пользователей.

2. Бессмертный, И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511999> (дата обращения: 24.11.2023).

Дополнительная литература:

1. Советов, Б. Я. Информационные технологии : учебник для вузов / Б. Я. Советов, В. В. Цехановский. — 7-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2023. — 327 с. — (Высшее образование). — ISBN 978-5-534-00048-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/510751> (дата обращения: 24.11.2023).

2. Обработка естественного языка с использованием языка программирования Python : учебное пособие : в 2 частях / составитель А. Б. Мантусов. — Элиста : КГУ, 2022. — Часть 1 — 2022. — 56 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/360923> (дата обращения: 24.11.2023). — Режим доступа: для авториз. пользователей.

3. Алексеев, Д. С. Технологии интеллектуального анализа данных / Д. С. Алексеев, О. В. Щекочихин. — 2-е изд., стер. — Санкт-Петербург : Лань, 2024. — 176 с. — ISBN 978-5-507-48763-9. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/362915> (дата обращения: 24.11.2023). — Режим доступа: для авториз. пользователей.

4. Рабчевский, А. Н. Синтетические данные и развитие нейросетевых технологий : учебное пособие для вузов / А. Н. Рабчевский. — Москва : Издательство Юрайт, 2023. — 187 с. — (Высшее образование). — ISBN 978-5-534-17716-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/533606> (дата обращения: 24.11.2023).

Ресурсы информационно-телекоммуникационной сети «Интернет»:

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН

<http://lib.rudn.ru/MegaPro/Web>

- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
- ЭБС Юрайт <http://www.biblio-online.ru>
- ЭБС «Консультант студента» www.studentlibrary.ru
- ЭБС «Троицкий мост»
- ЭБС «Лань» <http://e.lanbook.com/>
- ТУИС РУДН <https://esystem.rudn.ru/>

2. Базы данных и поисковые системы

- электронный фонд правовой и нормативно-технической документации
<http://docs.cntd.ru/>

- поисковая система Яндекс <https://www.yandex.ru/>
- поисковая система Google <https://www.google.ru/>
- реферативная база данных SCOPUS

<http://www.elsevierscience.ru/products/scopus/>

Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля:*

1. Курс лекций по дисциплине «Современные технологии обработки текстов на естественных языках».

2. Материалы для семинарских занятий и самостоятельной работы.

* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

РАЗРАБОТЧИКИ:

Старший преподаватель

Должность, БУП

Подпись

Мухин Илья Николаевич

Фамилия И.О.

Профессор, д.т.н.

Должность, БУП

Подпись

Царегородцев Анатолий
Валерьевич

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой, доцент,
к.т.н.

Должность БУП

Подпись

Софронова Елена
Анатольевна

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Доцент, к.т.н.

Должность, БУП

Подпись

Страшнов Станислав
Викторович

Фамилия И.О.