

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

Институт русского языка

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Инструменты искусственного интеллекта для анализа обработки текста

(наименование дисциплины/модуля)

Рекомендована МСЧН для направления подготовки/специальности:

45.03.01. Филология

(код и наименование направления подготовки/специальности)

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

Прикладная цифровая филология

(наименование (профиль/специализация) ОП ВО)

2022 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Инструменты искусственного интеллекта для анализа обработки текста» является знакомство с технологиями автоматической обработки естественного языка с целью обучения применению таких технологий в различных сферах филологической и лингвистической деятельности. В рамках курса предусматривается ознакомление учащихся с основными областями и задачами применения технологий искусственного интеллекта, а также с современными программными средствами для решения базовых филологических и лингвистических прикладных задач. Результатом занятий должно стать приобретение студентами навыков работы с приложениями для обработки естественного языка, системами автоматического перевода, анализа тональности, а также получение представлений о возможном применении данных средств и ресурсов при проведении исследований.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Инструменты искусственного интеллекта для анализа обработки текста» 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-6.3.	Находит и использует источники получения дополнительной информации для повышения уровня общих и профессиональных знаний с учетом условий, средств, личностных возможностей, этапов карьерного роста, временной перспективы развития деятельности и требований рынка труда	Знать: алгоритмы поиска дополнительной информации в целях повышения собственного уровня знаний; Уметь: осуществлять поиск дополнительной информации для повышения уровня общих и профессиональных знаний; Владеть: готовностью находить и использовать источники получения дополнительной информации для повышения уровня общих и профессиональных знаний
УК-12.1	Ищет нужные источники информации и данные, воспринимает, анализирует, запоминает и передает информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач	Знать инструменты основных цифровых экосистем для получения, обработки и анализа информации Знать особенности различных расширений и форматов хранения данных, принципы работы различных поисковых сервисов, методы организации и обработки данных, информации и цифрового контента, современные программные средства, используемые для анализа, обработки и представления информации, инструменты структурирования и оформления информации в профессиональной деятельности Уметь формулировать потребность в информации, искать данные в цифровой среде с использованием

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
		<p>фильтров и ключевых слов, создавать и менять собственные стратегии поиска информации; модифицировать информацию для повышения качества контента, интегрировать его в единую совокупность знаний для создания нового контента; использовать современные программные средства для анализа и обработки информации; выбирать оптимальный способ, формат и место хранения информации и данных с помощью цифровых инструментов, а также методы из защиты; создавать резервные копии данных на различных носителях; представлять информацию с помощью цифровых инструментов; оценивать практическую значимость результатов поиска, выделять в профессиональной информации главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями. Владеть цифровыми технологиями для получения, хранения, обработки информации с использованием для этого программных средств; языком международного общения и терминологией цифровой среды для эффективного освоения и использования цифровых устройств и программного обеспечения</p>
ОПК-2.2	Обладает навыками филологического анализа и интерпретации языкового материала, различных типов текстов	<p>Знать: методы и приемы проведения филологического анализа; стилевые, жанровые особенности различных типов текстов. Уметь: проводить филологический анализ языкового материала и различного типа текстов, идентифицировать, анализировать и интерпретировать филологические проблемы, выявленные в процессе проведения анализа. Владеть: навыком филологического анализа и интерпретации языкового материала, различных типов текстов.</p>
ОПК-3.2.	Соотносит знания в области теории литературы с конкретным литературным материалом и дает историко-литературную интерпретацию прочитанного и определяет жанровую специфику фольклорного и литературного явления	<p>Знать: основные положения, понятия и объект исследования теории литературы и основные положения учения о литературном произведении, его составе, структуре и строе. Уметь: применять полученные знания в области теории литературы при работе с конкретным литературным материалом. Владеть: способностью соотносить знания в области теории литературы с конкретным литературным материалом в целях решения задач профессиональной деятельности</p>
ОПК-3.3	Корректно осуществляет библиографический поиск в традиционных и электронных источниках и составляет корректное описание на основе полученных результатов	<p>Знать: алгоритмы и приемы формирования библиографического поиска и его осуществления в традиционных и электронных источниках; методы обработки и структурирования информации; формы и виды анализа полученных результатов. Уметь: составлять корректное описание библиографической информации, полученной в соответствии с личным и профессиональным целеполаганием из традиционных и электронных источников. Владеть: способностью составлять корректное описание результатов на основании</p>

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
		информации из традиционных и электронных библиографических источников
ОПК-4.2.	Интерпретирует тексты разных типов и жанров на основе существующих методик	Знать: основные методики, приемы и алгоритмы интерпретации текстов разных типов и жанров. Уметь: выбирать корректную и эффективную методику для интерпретации текстов различных типов и жанров в соответствии с личным и профессиональным целеполаганием. Владеть: навыком интерпретирования текстов различных типов и жанров на основе существующих методик в соответствии с личным и профессиональным целеполаганием
ОПК-7.1.	Знает основные цифровые технологии, принципы их работы и методы для изучения и моделирования объектов профессиональной, в том числе педагогической, деятельности, анализа данных, представления информации и пр.	Знать основные требования к оформлению документации, зафиксированные в стандартах; принципы анализа обработки и представления информации
ОПК-7.2.	Применяет современные цифровые методы и технологии анализа, обработки и представления информации в профессиональной, в том числе педагогической, сфере деятельности	Уметь применять современные методы обработки различных видов информации с применением цифровых технологий; применять современное программное обеспечение для анализа данных; применять современное программное обеспечение для обработки данных; применять современное программное обеспечение для визуализации данных; оценивать функциональные возможности программных систем для работы с данными и осваивать технологию работы с ними Владеть навыками использования баз данных и систем электронного и дистанционного образования; цифровыми средствами анализа, обработки и визуализации информации
ПК-8.2.	Применяет основные законы естественнонаучных дисциплин в профессиональной в том числе педагогической, деятельности, использует методы математическо-статистического анализа, логики и моделирования, теоретического и экспериментального исследования в гуманитарной сфере, в том числе во взаимосвязи с основными филологическими законами и методами	Знать основы математического анализа, логики, математического моделирования Уметь использовать математические методы для построения моделей, статистические методы для подготовки данных, выявления закономерностей, проверки гипотез и принятия решений и применять методы машинного обучения (в базовом виде, без доработки и модификации) для решения практических задач анализа данных Владеть методами математического анализа; статистическими методами
ПК-9.1.	Решает стандартные задачи в профессиональной, в том числе педагогической, деятельности с использованием специализированных цифровых инструментов и технологий с учетом требований информационной безопасности	Знать правила, нормы поведения в процессе использования цифровых технологии и коммуникации в цифровых средах. Знать правила информационной безопасности Знать специализированные цифровые инструменты и технологии для профессиональной деятельности Уметь применять цифровые инструменты и технологии в профессиональной деятельности, включая совместную работу и для совместного производства ресурсов и знаний Уметь использовать цифровые инструменты для создания знаний и инноваций

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
		Уметь управлять информацией с применением специализированных цифровых инструментов и технологий
ПК-9.2.	Использует цифровые технологии и высокоуровневые языки программирования общего назначения для разработки алгоритмов и программ в области интеллектуального анализа данных	Знать основы современных систем управления базами данных, теорию баз данных, основы программирования Уметь использовать цифровые технологии и элементарные языки программирования для разработки алгоритмов и программ в области интеллектуального анализа данных Владеть навыками программирования на высокоуровневых языках программирования общего назначения
ПК-10.2	Применяет, комбинирует и адаптирует существующие информационно-коммуникационные технологии для решения задач в профессиональной, в том числе педагогической, деятельности с учетом требований информационной безопасности	Знать основные технические средства обучения, электронные ресурсы, необходимые для организации различных видов деятельности обучающихся; психолого-педагогические основы и методику применения технических средств обучения, ИКТ, электронных образовательных и информационных ресурсов, дистанционных образовательных технологий и электронного обучения. Уметь: создавать информационные образовательные ресурсы, осуществлять электронное обучение; использовать дистанционные образовательные технологии. Владеть педагогически обоснованными формами, методами, средствами и приемами организации деятельности учащихся с использованием электронных образовательных и информационных ресурсов; навыками создания и редактирования цифровых ресурсов и цифрового контента, в том числе образовательного, в различных форматах

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Инструменты искусственного интеллекта для анализа обработки текста» относится к обязательной части (вариативная компонента) блока Б1.О.02.16.

В рамках ОП ВО обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Инструменты искусственного интеллекта для анализа обработки текста».

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-6.3.	Находит и использует источники получения дополнительной информации	Программирование на Python	Преддипломная практика

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	для повышения уровня общих и профессиональных знаний с учетом условий, средств, личностных возможностей, этапов карьерного роста, временной перспективы развития деятельности и требований рынка труда		
УК-12.1.	Ищет нужные источники информации и данные, воспринимает, анализирует, запоминает и передает информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач		
ОПК-7.1.	Знает основные цифровые технологии, принципы их работы и методы для изучения и моделирования объектов профессиональной, в том числе педагогической, деятельности, анализа данных, представления информации и пр.		
ОПК-7.2.	Применяет современные цифровые методы и технологии анализа, обработки и представления информации в профессиональной, в том числе педагогической, сфере деятельности		
ПК-9.2	Использует цифровые технологии и высокоуровневые языки программирования общего назначения для разработки алгоритмов и программ в области интеллектуального анализа данных		
ПК 8.2.	Применяет основные законы естественнонаучных дисциплин в профессиональной в том числе педагогической, деятельности, использует методы математическо-статистического анализа, логики и моделирования, теоретического и экспериментального исследования в гуманитарной сфере, в том числе во взаимосвязи с основными филологическими законами и методами		
ПК-10.1	Знает современные цифровые технологии и средства управления информацией для применения в профессиональной, в том числе педагогической, деятельности		
ПК-10.2	Применяет, комбинирует и адаптирует существующие информационно-коммуникационные технологии для решения задач в профессиональной, в том числе педагогической, деятельности с учетом требований информационной безопасности		
УК-6.3.	Находит и использует источники получения дополнительной информации	Инструменты обработки и визуализации данных	

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	для повышения уровня общих и профессиональных знаний с учетом условий, средств, личностных возможностей, этапов карьерного роста, временной перспективы развития деятельности и требований рынка труда		
УК-12.1.	Ищет нужные источники информации и данные, воспринимает, анализирует, запоминает и передает информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач		
ОПК-7.2.	Применяет современные цифровые методы и технологии анализа, обработки и представления информации в профессиональной, в том числе педагогической, сфере деятельности		
ПК-9.1	Решает стандартные задачи в профессиональной, в том числе педагогической, деятельности с использованием специализированных цифровых инструментов и технологий с учетом требований информационной безопасности		
ПК-9.2.	Использует цифровые технологии и высокоуровневые языки программирования общего назначения для разработки алгоритмов и программ в области интеллектуального анализа данных		
ПК-8.2	Применяет основные законы естественнонаучных дисциплин в профессиональной в том числе педагогической, деятельности, использует методы математико-статистического анализа, логики и моделирования, теоретического и экспериментального исследования в гуманитарной сфере, в том числе во взаимосвязи с основными филологическими законами и методами		
ПК-10.2	Применяет, комбинирует и адаптирует существующие информационно-коммуникационные технологии для решения задач в профессиональной, в том числе педагогической, деятельности с учетом требований информационной безопасности		
УК-1.5	Анализирует и контекстно обрабатывает информацию для решения поставленных задач		
УК-6.3.	Находит и использует источники получения дополнительной информации для повышения уровня общих и профессиональных знаний с учетом	Введение в корпусную лингвистику	

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	условий, средств, личностных возможностей, этапов карьерного роста, временной перспективы развития деятельности и требований рынка труда		
УК-12.2.	Проводит оценку информации, ее достоверность, строит логические умозаключения на основании поступающих информации и данных		
ОПК-1.2.	Осуществляет первичный отбор и анализ языкового и (или) литературного материала		
ОПК-1.3.	Обладает навыками анализа и интерпретации филологических проблем в диахроническом и синхроническом аспектах		
ОПК-4.1.	Владеет методикой сбора и филологического анализа языковых и литературных фактов и текстов разной степени сложности		
ПК-4.1.	Владеет базовыми навыками поиска и извлечения графической и/или текстовой информации с учетом требований законодательства Российской Федерации в области интеллектуальной собственности и ограничений, касающихся использования информационных материалов в сети Интернет		
ПК-5.1	Владеет навыками контроля качества текстов		
ПК-5.2.	Владеет навыками работы со словарями и справочниками, электронными информационными ресурсами в избранной сфере профессиональной, в том числе педагогической, деятельности		

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

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Инструменты искусственного интеллекта для анализа обработки текста» составляет 5 зачетных единиц.

Таблица 4.1. Виды учебной работы по периодам освоения ОП ВО для **ОЧНОЙ** формы обучения

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)			
		8			
<i>Контактная работа, ак.ч.</i>	48	48			
Лекции (ЛК)	24	24			
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)	24	24			
<i>Самостоятельная работа обучающихся, ак.ч.</i>	105	105			
<i>Контроль (экзамен), ак.ч.</i>	27	27			
Общая трудоемкость дисциплины	ак.ч.	180	180		
	зач.ед.	5	5		

Таблица 4.2. Виды учебной работы по периодам освоения ОП ВО для ЗАОЧНОЙ формы обучения*

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)			
		10			
Контактная работа, ак.ч.	26	26			
Лекции (ЛК)	10	10			
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)	16	16			
Самостоятельная работа обучающихся, ак.ч.	136	136			
Контроль (экзамен/зачет с оценкой), ак.ч.	18	18			
Общая трудоемкость дисциплины	ак.ч.	180	180		
	зач.ед.	5	5		

* - заполняется в случае реализации программы в заочной форме

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

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

Наименование раздела дисциплины	Содержание раздела (темы)	Вид учебной работы*
Раздел 1. графематический анализ	Тема 1.1. Изучение принципов обработки естественного языка, задачи обработки естественного языка, методы переобработки текстов, методы сходства текста	ЛК, СЗ
	Тема 1.2. Практические задания по нормализации, токенизации, закон Ципфа, мешок слов (BOW), TFIDF, N-грамм, методы сходства текста	СЗ
Раздел 2. Морфологический анализ	Тема 2.1 Изучение принципов лемматизации, Стемминга, автоматическая морфологическая разметка (POS tagging), морфологическая омонимия	ЛК, СЗ
	Тема 2.2. Практические задания по морфологическому анализу, система Диалинг–АОТ, система TreeTagger, система Pymorphy2, система Snowball, система MyStem	СЗ
Раздел 3. Визуализация данных	Тема 3.1. Изучение принципов визуализации текстовых данных	ЛК, СЗ
	Тема 3.2 Практические задания по визуализации	СЗ
Раздел 4. Практическая часть	Тема 4.1. Задания по автоматическому реферированию, анализ тональности, машинный перевод	СЗ

* - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – 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	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	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
Лаборатория	Аудитория для проведения лабораторных работ, индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и оборудованием.	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
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	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
Компьютерный класс	Компьютерный класс для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная персональными компьютерами (в количестве 12 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	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 Python (свободное лицензионное соглашение), Anaconda
Для самостоятельной работы обучающихся	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	Проектор и/или большой экран/монитор, компьютер, ОС MS Windows 10, MS Office 2016 или MS Office 2019 или MS Office 365, портал

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
		Microsoft Azure, доступ к интернету Python (свободное лицензионное соглашение), Anaconda

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

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

- 1- Автоматическая обработка текстов на естественном языке и анализ данных - Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С.
- 2- Прикладной анализ текстовых данных на Python – Бенджамин бенгфорт , Ребекка Билбор и Тони Охеда

Дополнительная литература:

- 1- NLP для новичков - Лейн, Ханке, Ховард.
- 2- Обработка естественного языка с TensorFlow - Ганегедара.
- 3- Обработка естественного языка - Юрий Васильев
- 4- Natural Language Processing with Python - Bird, Klein, Loper.

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 «Лань» <http://e.lanbook.com/>
- ЭБС «Троицкий мост»

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/>

- <https://habr.com/>
- Российская ассоциация цифровых гуманитарных наук (dhrussia.ru)

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

1. Курс лекций по дисциплине «Инструменты искусственного интеллекта для анализа обработки текста»

2. Презентационные материалы по дисциплине «Инструменты искусственного интеллекта для анализа обработки текста»

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

8. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ И БАЛЛЬНО-РЕЙТИНГОВАЯ СИСТЕМА ОЦЕНИВАНИЯ УРОВНЯ СФОРМИРОВАННОСТИ КОМПЕТЕНЦИЙ ПО ДИСЦИПЛИНЕ

Оценочные материалы и балльно-рейтинговая система* оценивания уровня сформированности компетенций (части компетенций) по итогам освоения дисциплины «Инструменты искусственного интеллекта для анализа обработки текста» представлены в Приложении к настоящей Рабочей программе дисциплины.

* - ОМ и БРС формируются на основании требований соответствующего локального нормативного акта РУДН.

РАЗРАБОТЧИКИ:

**Руководитель направления
развития Центра развития
цифровых технологий для
образовательных процессов
РУДН**

Должность, БУП



Подпись

Резаев Н.

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

**Директор
Института русского языка**

Наименование БУП



Подпись

Должикова А.В.

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

**Заведующий кафедрой
русского языка № 5
Института русского языка**

должность, БУП



Подпись

Куриленко В.Б.

Фамилия И.О.