

**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

Цифровые технологии компаративного исследования текста

(наименование дисциплины/модуля)

Рекомендована МСЧН для направления подготовки/специальности:

45.04.01. Филология

(код и наименование направления подготовки/специальности)

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

Цифровые инновации в филологии (магистратура)

(наименование (профиль/специализация) ОП ВО)

2022 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Цифровые технологии компаративного исследования текста» ознакомление слушателей с цифровыми методами обработки естественного языка.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Цифровые технологии компаративного исследования 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)
ОПК-4	Способен использовать цифровые технологии и методы в профессиональной деятельности в области филологии для изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации.	ОПК-4.1. Использует цифровые технологии и методы в профессиональной деятельности в области филологии для изучения и моделирования объектов профессиональной деятельности.
		ОПК-4.2. Использует цифровые технологии и методы в профессиональной деятельности для анализа данных.
		ОПК-4.3. Применяет цифровые технологии и методы в профессиональной деятельности для представления информации.

В результате освоения курса обучающийся должен:

Знать:

- фундаментальные понятия и идеи в области компьютерной обработки текстов;
- основные этапы обработки естественно-языковых текстов, основные задачи и подходы к их решению;
- направления исследований и основные проблемы, возникающие при обработке текстов.

Уметь:

- решать задачи из области обработки текстов;
- проектировать системы для анализа отдельных текстовых документов и коллекций текстовых документов;
- применять методы статистического анализа и машинного обучения для решения прикладных задач.

Владеть:

- культурой разработки и реализации системного программного обеспечения современных компьютеров;
- приемами и алгоритмами самостоятельной работы с современными технологиями и программными инструментами для обработки текстов;
- технологиями решения задач обработки естественно-языковых текстов с использованием современных подходов, лингвистических методов и моделей машинного обучения;

- методами и технологиями решения прикладных задач с помощью современных библиотек и языков программирования.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Цифровые технологии компаративного исследования текста» относится к части, формируемой участниками образовательных отношений, блока Б1 ОП ВО.

В рамках ОП ВО обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Цифровые технологии компаративного исследования 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*
ОПК-4	Способен использовать цифровые технологии и методы в профессиональной деятельности в области филологии для изучения и моделирования объектов профессиональной деятельности, анализа данных, представления информации.		

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

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Цифровые технологии компаративного исследования текста» составляет 3 зачетные единицы (108 ак.часов).

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

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)			
		1	2	3	4
<i>Контактная работа, ак.ч.</i>	26			26	
в том числе:					
Лекции (ЛК)	13			13	
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)	13			13	
<i>Самостоятельная работа обучающихся, ак.ч.</i>	64			64	
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	18			18	
Общая трудоемкость дисциплины	ак.ч.	108		108	
	зач.ед.	3		3	

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

Не предусмотрена.

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)			
		1	2	3	4
Контактная работа, ак.ч.					
в том числе:					
Лекции (ЛК)					
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)					
Самостоятельная работа обучающихся, ак.ч.					
Контроль (экзамен/зачет с оценкой), ак.ч.					
Общая трудоемкость дисциплины	ак.ч.				
	зач.ед.				

* - заполняется в случае реализации программы в заочной форме

Таблица 4.3. Виды учебной работы по периодам освоения ОП ВО для **ЗАОЧНОЙ** формы обучения*

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)			
		3	4	5	6
Контактная работа, ак.ч.	22			22	
в том числе:					
Лекции (ЛК)	10			10	
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)	12			12	
Самостоятельная работа обучающихся, ак.ч.	77			77	
Контроль (зачет с оценкой), ак.ч.	9			9	
Общая трудоемкость дисциплины	ак.ч.			108	
	зач.ед.			3	

* - 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. Задачи обработки естественного языка	Тема 1.1. Категоризация текстов, выделение тем, машинный перевод, ответы на вопросы, анализ тональности текста, автоматическое реферирование, использование машинного обучения в обработке естественного языка.	ЛК
	Тема 1.2. Предобработка данных и их инструменты.	ЛК
Раздел 2. Векторные представления слов	Тема 2.1. Построение векторных представлений документов. Методы Term Frequency и Inverted Document Frequency. Метод Word2Vec. Целевая функция и архитектура нейронной сети в алгоритме Word2vec. Понятие контекста. Архитектуры CBOW и skip-gram. Метод Negative Sampling.	ЛК
	Тема 2.3. Применение Word2Vec.	СЗ
Раздел 3.	Тема 3.1. Общие принципы функционирования	ЛК

Рекуррентные нейронные сети		рекуррентных нейронных сетей. Архитектура LSTM-сетей. Двухнаправленные и многослойные LSTM-сети.	
		Тема 3.2. Применение рекуррентных сетей для построения языковой модели.	СЗ
Раздел Классификация текста	4.	Тема 4.1. Постановка задачи классификации текста. Классификация текста с помощью рекуррентных нейронных сетей (LSTM-сетей). Классификация текста с помощью свёрточных нейронных сетей.	ЛК
		Тема 4.2. Метод классификации текста FastText.	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	Доступ к сети Интернет. Язык программирования Python и программа Anaconda

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
	персональными компьютерами (в количестве 12 шт.), доской (экраном) и техническими средствами мультимедиа презентаций.	
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 Microsoft Azure, доступ к интернету

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

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1 Хиценко В.П., Основы программирования : учебное пособие / Хиценко В.П. Новосибирск: Издательство Новосибирского государственного технического университета, 2015. Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785778227064.html>

2. Manning C.D., Schütze H. Foundations of Statistical Natural Language Processing. MIT Press, 1999.

Дополнительная литература:

1. Батура Т.В. Математическая лингвистика и автоматическая обработка текстов: учебное пособие / Батура Т.В. Новосибирск : РИЦ НГУ, 2016. Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785443705484.html>

2. Рассел С., Норвиг П. Искусственный интеллект: современный подход. М.: ООО «И.Д. Вильямс», 2016.

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/>
- <http://www.opennet.org> ,
- <http://www.javascript.ru>
- www.phpclub.ru,
- www.w3c.org,
- <http://apache.org>
- Российская ассоциация цифровых гуманитарных наук (dhrussia.ru)

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

1. Курс лекций по дисциплине «Цифровые технологии компаративного исследования текста».

2. Презентационные материалы по дисциплине «Цифровые технологии компаративного исследования 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 «Цифровые технологии компаративного исследования 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

Резаиан 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.