

**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

Информационный поиск

(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 является формирование у обучающихся системного представления о методах автоматической обработки естественно-языкового текста на всех уровнях лингвистического анализа.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Информационный поиск» 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)
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	<i>Знать:</i> <ul style="list-style-type: none">– основные уровни обработки текста на естественном языке и существующие модели статистического, морфологического и синтаксического анализа текстов– типичные программные системы (приложения) в области автоматической обработки текста и их архитектурные особенности
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-10.1	Знает современные цифровые технологии и средства управления информацией для применения в профессиональной, в том числе педагогической, деятельности	<i>Уметь:</i> <ul style="list-style-type: none">– работать с автоматическими словарями (осуществлять запросы и извлекать информацию);

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
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	<ul style="list-style-type: none"> – охарактеризовать принципы морфологического, синтаксического анализа (на примере одной из систем) – осуществлять реферирование текста с использованием компьютерных средств.
		<p><i>Владеть:</i></p> <ul style="list-style-type: none"> – инструментами автоматической обработки естественного языка; – ориентированием в современных программных средствах обработки естественного языка.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Информационный поиск» относится к части, формируемой участниками образовательных отношений Б1.В.ДВ.10.01.

В рамках ОП ВО обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Информационный поиск».

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-1.3	Осуществляет поиск информации для решения поставленной задачи по различным типам запросов	Информационные системы анализа данных в гуманитарной сфере	Преддипломная практика
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		
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		

Шифр	Наименование компетенции	Предшествующие дисциплины/ модули, практики*	Последующие дисциплины/ модули, практики*
УК-12.2	Проводит оценку информации, ее достоверность, строит логические умозаключения на основании поступающих информации и данных		
ОПК-2.2	Обладает навыками филологического анализа и интерпретации языкового материала, различных типов текстов		
ОПК-3.3	Корректно осуществляет библиографический поиск в традиционных и электронных источниках и составляет корректное описание на основе полученных результатов		
ОПК-4.1	Владет методикой сбора и филологического анализа языковых и литературных фактов и текстов разной степени сложности		
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-9.1	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 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-10.2	Применяет, комбинирует и адаптирует существующие информационно-коммуникационные технологии для решения задач в профессиональной, в том числе педагогической, деятельности с учетом требований информационной безопасности		
ОПК-2.2	Обладает навыками филологического анализа и интерпретации языкового материала, различных типов текстов	Инструменты искусственного интеллекта для анализа и обработки текста	
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		

Шифр	Наименование компетенции	Предшествующие дисциплины/ модули, практики*	Последующие дисциплины/ модули, практики*
ОПК-3.3	Корректно осуществляет библиографический поиск в традиционных и электронных источниках и составляет корректное описание на основе полученных результатов		
ОПК-4.2.	Интерпретирует тексты разных типов и жанров на основе существующих методик		
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-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-9.2.	Использует цифровые технологии и высокоуровневые языки программирования общего назначения для разработки алгоритмов и программ в области интеллектуального анализа данных		
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		
ОПК-2.2	Обладает навыками филологического анализа и интерпретации языкового материала, различных типов текстов		
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		
ОПК-3.3	Корректно осуществляет библиографический поиск в традиционных и электронных источниках и составляет корректное описание на основе полученных результатов		
ОПК-4.2.	Интерпретирует тексты разных типов и жанров на основе существующих методик		
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		

Шифр	Наименование компетенции	Предшествующие дисциплины/ модули, практики*	Последующие дисциплины/ модули, практики*
	информации в профессиональной, в том числе педагогической, сфере деятельности		
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-9.2.	Использует цифровые технологии и высокоуровневые языки программирования общего назначения для разработки алгоритмов и программ в области интеллектуального анализа данных		
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 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.	Программирование на Python	
ОПК-7.2	Применяет современные цифровые методы и технологии анализа, обработки и представления информации в профессиональной, в том числе педагогической, сфере деятельности		
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		
ПК9.2	Использует цифровые технологии и высокоуровневые языки программирования общего		

Шифр	Наименование компетенции	Предшествующие дисциплины/ модули, практики*	Последующие дисциплины/ модули, практики*
	назначения для разработки алгоритмов и программ в области интеллектуального анализа данных		
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		

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

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Информационный поиск» 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)			
		7	8		
<i>Контактная работа, ак.ч.</i>	78	30	48		
Лекции (ЛК)	39	15	24		
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)	39	15	24		
<i>Самостоятельная работа обучающихся, ак.ч.</i>	66	33	33		
<i>Контроль (зачет с оценкой /экзамен), ак.ч.</i>	36	9	27		
Общая трудоемкость дисциплины	ак.ч.	180	72	108	
	зач.ед.	5	2	3	

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

Вид учебной работы	ВСЕГО, ак.ч.	Семестр(-ы)			
		10			
<i>Контактная работа, ак.ч.</i>	36	36			
Лекции (ЛК)	18	18			
Лабораторные работы (ЛР)					
Практические/семинарские занятия (СЗ)	18	18			
<i>Самостоятельная работа обучающихся, ак.ч.</i>	135	135			
<i>Контроль (зачет с оценкой), ак.ч.</i>	9	9			
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. Введение в информационный поиск.	Тема 1.1. Поисковая система: цель и задачи курса. История развития поисковых систем. Архитектура web-поиска. Булев поиск. Обратный индекс.	ЛК
	Тема 1.2. Применение инструментов информационного поиска по разбиению и сортировке, созданию обратного индекса, созданию матрицы термин-документа.	СЗ
Раздел 2. Информационный поиск по уровням языка.	Тема 2.1. Фонетика, фонология, морфология, синтаксис, семантика, прагматика. Историческая лингвистика. Лингвистическая типология. Социолингвистика. Диалектология. Лексикография.	ЛК
	Тема 2.2. Подходы к языку: рационалистический и эмпирический. Математическая лингвистика. Корпусная лингвистика	ЛК
	Тема 2.3. Использование библиотеки NLTK, применение закона Ципфа и формулы Мандельброта	СЗ
Раздел 3. Основы обработки текста.	Тема 3.1. Структурированные и неструктурированные данные. Критерии документа, кодировки. Уровни лингвистического анализа. Токены и термины. Детекция языка: графематический, N-граммный и лексический подходы. Нормализация. Проблемы токенизации.	ЛК
	Тема 3.2. Применение инструментов информационного поиска для выделения N-граммов, нормализации, токенизации, понижения регистра. Идентификация стоп-слов. Использование лемматизации. Инструменты для стемминга. Алгоритмы для снятия омонимии.	СЗ
Раздел 4. Использование инструментов	Тема 4.1. Определение коллокаций, их признаки. Частотность N-граммов. Фильтр по частям речи.	ЛК

информационного поиска для анализа коллокаций.	Отклонения, гистограммы отклонений. Поиск коллокаций.	
	Тема 4.2. Методы подсчёта вероятности: параметрический и непараметрический подходы, стандартные и биномиальные распределения, мультиномиальное и нормальное распределения, аппроксимирование.	ЛК
	Тема 4.3. Применение статистики для поиска коллокаций.	СЗ
Раздел 5. Использование инструментов информационного поиска для индексации.	Тема 5.1. Общая схема базы поиска. Назначение обратного индекса. Технические ограничения и дисковая подсистема. Состав обратного индекса и варианты его построения. Оптимизация пересечения блоков.	ЛК
	Тема 5.2. Применение инструментов информационного поиска для индексации.	СЗ
Раздел 6. Использование инструментов информационного поиска для текстового ранжирования.	Тема 6.1. Частотная матрица. Частота термина. Логарифмическое взвешивание. Документы как векторы. Методы оптимизации текстового ранжирования. Статические веса, общий вес. Классическое ранжирование.	ЛК
	Тема 6.2. Применение TF-IDF и метода «Мешок слов» для информационного поиска. Реализация информационного поиска с помощью бинарной вероятностной модели.	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,

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
		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, 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. Грудева, Е. В. Корпусная лингвистика [Электронный ресурс]: учеб. пособие / Е. В. Грудева. - 2-е изд., - М. : ФЛИНТА, 2012. - 165 с.
2. Потапова Р.К. Новые информационные технологии и лингвистика: учебное пособие для студентов вузов, обучающихся по специальности 021800 «Теоретическая и прикладная лингвистика» направления 620200 «Лингвистика и новые информационные технологии» / Р.К. Потапова ; Моск. гос. лингвист. ун-т . Изд. 5-е .
3. Маннинг К. Д., Рагхаван П., Шютце Х. Введение в информационный поиск. — Вильямс, 2011.
4. Федотова Е.Л., Федотов А.А. Информационные технологии в науке и образовании : учебное пособие / Е. Л. Федотова, А. А. Федотов . Москва : ФОРУМ : ИНФРА-М, 2011 . 334 с.
5. Щипицина, Л. Ю. Информационные технологии в лингвистике [Электронный ресурс] : учебное пособие / Л. Ю. Щипицина. М. : ФЛИНТА, 2013. 128 с.

Дополнительная литература:

1. Агеев М. С., Добров Б. В., Лукашевич Н. В. Автоматическая рубрикация текстов: методы и проблемы // Учёные записки Казанского государственного университета. Серия Физико-математические науки. — 2008. — Т. 150, № 4. — С. 25–40.
2. Айсина Р. М. Обзор средств визуализации тематических моделей коллекций текстовых документов // Машинное обучение и анализ данных (<http://jmlr.org>). — 2015. — Т. 1, № 11. — С. 1584–1618.
3. Воронцов К. В. Аддитивная регуляризация тематических моделей коллекций текстовых документов // Доклады РАН. — 2014. — Т. 456, № 3. — С. 268–271.
4. Воронцов К. В., Потапенко А. А. Регуляризация, робастность и разреженность вероятностных тематических моделей // Компьютерные исследования и моделирование. — 2012. — Т. 4, № 4. — С. 693–706.
5. Воронцов К. В., Потапенко А. А. Модификации EM-алгоритма для вероятностного тематического моделирования // Машинное обучение и анализ данных. — 2013. — Т. 1, № 6. — С. 657–686.
6. Лукашевич Н. В. Тезаурусы в задачах информационного поиска. — Издательство МГУ имени М. В. Ломоносова, 2011.
7. Павлов А. С., Добров Б. В. Метод обнаружения массово порожденных неестественных текстов на основе анализа тематической структуры // Вычислительные методы и программирование: новые вычислительные технологии. — 2011. — Т. 12. — С. 58–72

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/>

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

1. Курс лекций по дисциплине «Информационный поиск».

2. Презентационные материалы по дисциплине «Информационный поиск»

* - 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 «Информационный поиск» 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.