

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 25.05.2026 09:45:33
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

ТЕКСТОВАЯ АНАЛИТИКА

(наименование дисциплины/модуля)

Рекомендована МССН 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

(наименование (профиль/специализация) ОП ВО)

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Текстовая аналитика» входит в программу магистратуры «Цифровые инновации в филологии» по направлению 45.04.01 «Филология» и изучается во 2 семестре 1 курса. Дисциплину реализует Кафедра прикладной информатики и интеллектуальных систем в гуманитарной сфере. Дисциплина состоит из 4 разделов и 13 тем и направлена на изучение методов и моделей классификации и кластеризации текстов.

Целью освоения дисциплины является овладение методами и моделями классификации и кластеризации текстовой информации, изучение соответствующих метрик, алгоритмов и процедур работы с текстовыми данными, а также знакомство с практикой применения анализа тональности в гуманитарных исследованиях.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Текстовая аналитика» 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-7	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать 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; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	УК-7.1 Способен найти источники информации и данные, воспринимать, анализировать, запоминать и передавать 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 профессиональных задач; УК-7.2 Оценивает информацию, ее достоверность для решения профессиональных задач; УК-7.3 Строит логические умозаключения на основании информации и поступающих данных для решения задач в профессиональной сфере;
ПК-2	Способен применять системы, использующие средства баз данных и лингвистического обеспечения, использовать алгоритмы и программы интеллектуального и лингвистического анализа данных	ПК-2.1 Знает теоретические основы разработки баз данных и систему требований, предъявляемых к лингвистическому обеспечению; ПК-2.2 Применяет современные системы управления базами данных для практической работы по созданию, модернизации и использованию баз данных в гуманитарных областях; ПК-2.3 Использует лингвистическое обеспечение информационных систем для решения профессиональных задач;

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Текстовая аналитика» относится к обязательной части 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 «Текстовая аналитика».

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

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
------	--------------------------	---	--

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-7	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать 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; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	Программирование на языке Python; Программное обеспечение для гуманитарной сферы; Математические методы в гуманитарной сфере; Табличный редактор: продвинутый Excel; Информационные базы данных;	Анализ социальных сетей**; Мониторинг медиапространства**; Научно-исследовательская работа;
ПК-2	Способен применять системы, использующие средства баз данных и лингвистического обеспечения, использовать алгоритмы и программы интеллектуального и лингвистического анализа данных	Введение в специальность;	

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Текстовая аналитика» 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)
			2
Контактная работа, ак.ч	15		15
Лекции (ЛК)	0		0
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	15		15
Самостоятельная работа обучающихся, ак.ч.	48		48
Контроль (экзамен/зачет с оценкой), ак.ч.	9		9
Общая трудоемкость дисциплины ак.ч.	ак.ч.	72	72
	зач.ед.	2	2

Общая трудоемкость дисциплины «Текстовая аналитика» составляет «2» зачетные единицы.

Таблица 4.2. Виды учебной работы по периодам освоения образовательной программы высшего образования для заочной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			4
Контактная работа, ак.ч	4		4
Лекции (ЛК)	0		0
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	4		4
Самостоятельная работа обучающихся, ак.ч.	64		64
Контроль (экзамен/зачет с оценкой), ак.ч.	4		4
Общая трудоемкость дисциплины ак.ч.	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	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Анализ текстов на естественном языке	1.1	Введение в обработку естественного языка	Введение. Область применения систем анализа естественного языка (информационно-поисковые системы, вопросно-ответные системы, системы автоматизированного машинного перевода, системы извлечения содержательной информации из текстов). Методы Data mining. Область применения обработки текстовой информации. Основные понятия	СЗ
		1.2	Методы семантического анализа	Методы семантического анализа текстов на естественном языке. Подходы к выделению и представлению смысла текста.	СЗ
Раздел 2	Сбор, предобработка, хранение и представление текстовых данных	2.1	Технологии сбора текстового контента	Методы сбора пользовательского текстового контента из сети Интернет. Разработка адаптивных парсеров (веб-скраперов).	СЗ
		2.2	Основы предобработки текста	Этапы очистки, нормализации и лемматизации текстов. Удаление стоп-слов с использованием словарей.	СЗ
		2.3	Хранение текстовых данных	Хранение текстового контента в реляционных (PostgreSQL, SQLite3) и нереляционных (NoSQL, MongoDB) базах данных.	СЗ
Раздел 3	Классификация и кластеризация текстов	3.1	Основы обработки естественного языка (NLP)	Введение в NLP и его роль в анализе текстов. Основные задачи: токенизация, лемматизация.	СЗ
		3.2	Классификация текстов	Методы представления текстов (векторизация): Bag-of-Words, TF-IDF, Word2Vec, ruBERT. Алгоритмы машинного обучения для классификации: наивный Байес, метод опорных векторов.	СЗ
		3.3	Кластеризация текстов	Понятие и применение кластеризации текстов. Алгоритмы кластеризации: k-means, k-means++, C-means, Affinity Propagation, DBSCAN, BIRCH. Метрики оценки качества кластеризации: accuracy, precision, recall, f-score.	СЗ
		3.4	Лексико-семантические модели для классификации текстов	Использование лексико-семантических моделей в задачах классификации текстовых данных.	СЗ
Раздел 4	Анализ тональности текстов	4.1	Применение семантических моделей для анализа тональности	Использование семантических моделей при анализе тональности текстового контента.	СЗ
		4.2	Работа с эмоциональной лексикой и оценка качества модели	Использование словарей эмоциональной окраски. Построение матрицы ошибок и анализ ошибок модели.	СЗ
		4.3	Модели машинного обучения для анализа тональности	Модели для анализа тональности: наивный Байес, метод опорных векторов, рекуррентные нейронные сети (LSTM, GRU), сверточные нейронные сети. Применение предобученных моделей (ruBERT, GPT).	СЗ
		4.4	Практическое применение анализа тональности	Мониторинг и анализ тональности сообщений в социальных сетях. Оценка настроений пользователей.	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	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
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.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

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

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература:

1. Воронов Э.С. «Методы и алгоритмы анализа естественно-языковых сообщений» Технологии Microsoft в теории и практике программирования
2. Толпегин П.В. «Информационные технологии анализа русских естественных языковых текстов» – Российская академия наук вычислительный центр им. А.А. Дородницына, 2023
3. Дюк В., Самойленко А. «Data Mining: учебный курс (+CD)». — СПб.: Изд. Питер, 2023
4. Лукашевич, Н. В., Сорокин, А. А. Компьютерная лингвистика и автоматическая обработка текста. / Н. В. Лукашевич, А. А. Сорокин. — Москва: МАКС-Пресс, 2025. — 608 с. — ISBN 978-5-317-07437-1. Текст : электронный // Сайт Кафедры алгоритмических языков ВМК МГУ. — URL: https://al.cs.msu.ru/books/Inv_saa_cl-aot (дата обращения: 15.03.2026)
5. Маккинли У. Python и анализ данных. — Перевод с английского. — М.: ДМК Пресс, 2023. — 482 с. — ISBN 978-5-9706-0315-4

Дополнительная литература:

1. Хайдуков Д. С. Применение кластерного анализа в государственном управлении// Философия математики: актуальные проблемы. — М.: МАКС Пресс, 2009. — 287 с.
2. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze An Introduction to Information Retrieval Draft. Online edition. Cambridge University Press. - 2009. - 544 p.

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 <https://mega.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 «Знаниум» <https://znanium.ru/>
2. Базы данных и поисковые системы
 - Sage <https://journals.sagepub.com/>
 - Springer Nature Link <https://link.springer.com/>

- Wiley Journal Database <https://onlinelibrary.wiley.com/>

- Научно-метрическая база данных Lens.org <https://www.lens.org>

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

1. Курс лекций по дисциплине «Текстовая аналитика».

2. Презентационные материалы по дисциплине «Текстовая аналитика».

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

РАЗРАБОТЧИКИ

Старший преподаватель

Должность

РУКОВОДИТЕЛЬ БУП

Заведующий кафедрой, доцент, к.т.н.

Должность

РУКОВОДИТЕЛЬ ОП ВО

Доцент кафедры русского языка №5

Должность

Студеникина К.А.

Фамилия И.О

Софронова Е.А.

Фамилия И.О

Бирюкова Ю.Н.

Фамилия И.О