Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 29.05.2025 15:40:53
Уникальный программный ключ:
ca953a01204891083f939673078ef1a989dae18a

**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)

2025 г.

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 изучение методов и моделей классификации и кластеризации текстов, включая метрики и алгоритмы кластеризации текстового контента данных и процессы работы с ними, ознакомление с примерами применения сентимент анализа в гуманитарной сфере.

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*
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 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;	<i>Критические исследования коммуникации**;</i> Научно-исследовательская работа;
ПК-2	Способен применять системы, использующие средства баз данных и лингвистического обеспечения, использовать алгоритмы и программы интеллектуального и лингвистического анализа данных.		<i>Риторика сетевой коммуникации**;</i>

* - 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.	30		30
Лекции (ЛК)	15		15
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	15		15
Самостоятельная работа обучающихся, ак.ч.	33		33
Контроль (экзамен/зачет с оценкой), ак.ч.	9		9
Общая трудоемкость дисциплины	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.	8		8
Лекции (ЛК)	4		4
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	4		4
Самостоятельная работа обучающихся, ак.ч.	55		55
Контроль (экзамен/зачет с оценкой), ак.ч.	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	Обработка естественно-языковых текстов	1.1	Введение. Область применения систем анализа естественно-языковых текстов (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	ЛК, СЗ
		1.2	Методы семантического анализа естественно-языковых текстов. Определение способов выделения и представления смысла	ЛК, СЗ
Раздел 2	Сбор, предобработка, хранение и представление текстовой информации	2.1	Технологии сбора текстового контента пользователей из сети интернет, разработка адаптивных парсеров (веб-скраперов)	ЛК, СЗ
		2.2	Основы предобработки текстового контента, этапы отчистки текста, нормализации, лемматизации, удаление стоп-слов (словари)	ЛК, СЗ
		2.3	Хранение текстового контента в реляционных БД (PostgreSQL, SQLite3) и NoSQL (MongoDB) базах данных	ЛК, СЗ
Раздел 3	Классификация и кластеризация текстов	3.1	Основы обработки естественного языка (Natural Language Processing, NLP): введение в NLP и его роль в анализе текстов, основные задачи NLP: токенизация, лемматизация	ЛК, СЗ
		3.2	Классификация текстов: Методы представления текстов (векторизации): Bag-of-Words, TF-IDF, word2vec, ruBERT. Алгоритмы машинного обучения для классификации текстов: наивный Байес, метод опорных векторов	ЛК, СЗ
		3.3	Кластеризация текстов: Концепция кластеризации текстов и ее применение, алгоритмы кластеризации текстов: k-means, k-means++, C-means, Affinity Propagation, DBDScan, BIRCH. Оценка качества результатов кластеризации (accuracy, precision, recall, f-measure)	ЛК, СЗ
		3.4	Лексико-семантические модели для классификации текстовых данных	ЛК, СЗ
Раздел 4	Сентимент анализ текстов	4.1	Использование семантических моделей для сентимент анализа текстового контента	ЛК, СЗ
		4.2	Работа с эмоциональной лексикой и словарями окраски, построение матрицы ошибок и анализ ошибок модели	ЛК, СЗ
		4.3	Модели машинного обучения для сентимент анализа: наивный Байес и метод опорных векторов, рекуррентные нейронные сети (LSTM, GRU) и сверточные нейронные сети, применение предобученных моделей, BERT (ruBERT) и GPT, для сентимент-анализа	ЛК, СЗ
		4.4	Практическое применение сентимент-анализа, анализ и мониторинг социальных сетей, оценка настроений пользователей	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	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. Чубукова И. А. «Data mining: учебное пособие». — М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2023
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.
3. Мандель И. Д. Кластерный анализ. — М.: Финансы и статистика, 1988.-176 с.

4. Жамбю М. Иерархический кластер-анализ и соответствия. — М.: Финансы и статистика, 1988. — 345 с.

5. Олдендерфер М. С., Блэшфилд Р. К. Кластерный анализ / Факторный, дискриминантный и кластерный анализ: пер. с англ.; Под. ред. И. С. Енюкова. — М.: Финансы и статистика, 1989—215 с.

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://znaniium.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/>

- Научометрическая 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:

Педагог ДО

Должность, БУП

Подпись

Калинин Владимир
Николаевич

Фамилия И.О.

Профессор, д.т.н.

Должность, БУП

Подпись

Жуков Дмитрий Олегович

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой, доцент,
к.т.н.

Должность БУП

Подпись

Софронова Елена
Анатольевна

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Доцент кафедры русского
языка №5

Должность, БУП

Подпись

Бирюкова Юлия
Николаевна

Фамилия И.О.