

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Ястребов Олег Александрович
Должность: Ректор
Дата подписания: 15.05.2026 12:35:05
Уникальный программный ключ:
ca953a01204891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования
«Российский университет дружбы народов имени Патриса Лумумбы»
Филологический факультет**

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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА

(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):

ЯЗЫК И КУЛЬТУРА: ТЕОРИЯ И ПРАКТИКА

(наименование (профиль/специализация) ОП ВО)

2026 г.

1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Обработка естественного языка» входит в программу магистратуры «Язык и культура: теория и практика» по направлению 45.04.01 «Филология» и изучается в 3 семестре 2 курса. Дисциплину реализует Кафедра русского языка и методики его преподавания. Дисциплина состоит из 4 разделов и 26 тем и направлена на изучение полного цикла анализа текста: от установки среды и работы с текстовыми корпусами до методов предобработки текста, морфологического анализа, построения языковых моделей.

Целью освоения дисциплины является формирование у студентов базовых навыков обработки и анализа текстов на естественном языке с использованием современных библиотек и инструментов; умение выполнять основные этапы обработки текста: предобработку, извлечение лингвистических признаков, векторизацию текста, а также применять простые модели классификации и анализа тональности.

2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Обработка естественного языка» 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 Эффективно ищет и использует информацию, применяя цифровые средства и алгоритмы работы с данными из различных источников; УК-7.2 Использует предварительно проверенные на достоверность данные и информацию для построения умозаключений;
ОПК-2	Способен использовать в профессиональной, в том числе педагогической, деятельности знания современной научной парадигмы в области филологии и динамики ее развития, системы методологических принципов и методических приемов филологического исследования;	ОПК-2.1 Корректно применяет различные методы научно-исследовательской работы;
ПК-2	Владеет навыками самостоятельного исследования системы языка в синхроническом и диахроническом аспектах; изучения устной, письменной и	ПК-2.1 Умеет решать исследовательские задачи в области синхронического анализа системы языка, речевых произведений и актов коммуникации;

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
	виртуальной коммуникации с изложением аргументированных выводов	

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Обработка естественного языка» 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 «Обработка естественного языка».

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 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.	Information Technologies in Language Studies; Multimodal Communication; Language of Art; Academic Research Databases; Research Work;	<i>Methods of Linguistic Analysis**;</i> <i>Sourcing and Referencing**;</i> Scientific and Research Training;
ОПК-2	Способен использовать в профессиональной, в том числе педагогической, деятельности знания современной научной парадигмы в области филологии и динамики ее развития, системы методологических принципов и методических приемов филологического исследования;	Research Work; General Linguistics; Theory of Intercultural Communication; Planning a Research Project; Theory and History of Linguistics; Multimodal Communication; Philology in the System of Modern Humanities;	<i>Scientific and Research Training;</i>
ПК-2	Владеет навыками самостоятельного исследования системы языка в синхроническом и диахроническом аспектах;	Research Work; Theory of Intercultural Communication; <i>Linguistic Pragmatics**;</i> <i>Theory of Politeness**;</i>	<i>Scientific and Research Training;</i>

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	изучения устной, письменной и виртуальной коммуникации с изложением аргументированных выводов	General Linguistics; Multimodal Communication;	

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

** - элективные дисциплины /практики

4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Обработка естественного языка» составляет «3» 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)
			3
<i>Контактная работа, ак.ч.</i>	34		34
Лекции (ЛК)	17		17
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	17		17
<i>Самостоятельная работа обучающихся, ак.ч.</i>	65		65
<i>Контроль (экзамен/зачет с оценкой), ак.ч.</i>	9		9
Общая трудоемкость дисциплины	ак.ч.	108	108
	зач.ед.	3	3

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	Natural language processing: tools	1.1	Introduction to NLP	Definition of natural language processing, major applications, basic terms, examples of NLP tasks in real- world products	ЛК
		1.2	NLP Environment Setup	Installation of Python, Jupyter Notebook, NLTK, spaCy, pandas, scikit-learn; first text processing script	СЗ
		1.3	Language Data and Corpora	Types of text data, corpora, annotation, structured and unstructured text, common file formats	ЛК
		1.4	Working with Text Data	Loading texts from files, reading CSV and TXT data, corpus inspection with pandas	СЗ
		1.5	Python Tools for NLP	Overview of libraries for beginner- level NLP, role of tokenizers, taggers, vectorizers, classifiers	ЛК
		1.6	Exploring NLP Libraries	Basic operations in NLTK and spaCy: tokenization, sentence splitting, linguistic pipeline overview	СЗ
Раздел 2	Text Representation and Vectorization	2.1	Text Representation I: Bag-of-Words and TF-IDF	Document-term matrix, frequency- based representation, sparsity, weighting schemes	ЛК
		2.2	Vectorization Practice	Creating Bag-of-Words and TF-IDF matrices with scikit-learn	СЗ
		2.3	Text Representation II: Word Embeddings	Distributed representation of words, semantic similarity, introduction to Word2Vec and GloVe	ЛК
		2.4	Embedding Exploration	Loading pretrained embeddings, measuring word similarity, simple semantic analogy tasks	СЗ
Раздел 3	Classification, Evaluation, and Sentiment Analysis	3.1	Text Classification	Supervised learning for text, training data, class labels, common algorithms	ЛК
		3.2	First Text Classifier	Building a Naive Bayes classifier for spam detection or sentiment classification	СЗ
		3.3	Evaluation of NLP Models	Train/test split, validation, accuracy, precision, recall, F1-score, confusion matrix	ЛК
		3.4	Evaluation Metrics Lab	Evaluating a classifier and interpreting common error patterns	СЗ
		3.5	Sentiment Analysis	Polarity, subjectivity, lexicon-based and machine-learning approaches	ЛК
		3.6	Sentiment Analysis Mini-Lab	Classifying review texts or social media posts by sentiment	СЗ
Раздел 4	Advanced NLP, Neural Models, and Ethics	4.1	Named Entity Recognition	Names, locations, organizations, dates; role of NER in information extraction	ЛК
		4.2	NER with spaCy	Running pretrained NER models, checking entity spans, modifying	СЗ

Номер раздела	Наименование раздела дисциплины	Наименование темы	Содержание темы	Вид учебной работы*
			entity labels in examples	
		4.3 Topic Modeling	Introduction to document themes, latent topics, basic idea of LDA	ЛК
		4.4 Topic Modeling Practice	Running LDA on a small corpus and interpreting topic-word distributions	СЗ
		4.5 Neural NLP	Neural networks for language tasks, embeddings in neural models, RNN and LSTM overview	ЛК
		4.6 Neural Model Demo	Using a simple neural text classifier with pretrained tools or notebooks	СЗ
		4.7 Transformers and Pretrained Language Models	Attention mechanism, Transformer architecture, BERT, transfer learning for NLP	ЛК
		4.8 Transformer Practice	Using a pretrained transformer for text classification or question answering	СЗ
		4.9 Ethics, Bias, and Safety in NLP	Bias in datasets and models, privacy, fairness, responsible use of language technologies	ЛК
		4.10 Mini-Project and Error Analysis	Short project on a beginner NLP task, model testing, result interpretation, presentation of findings	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. Madnani, N. (2007). Getting Started on Natural Language Processing with Python.

XRDS:

Crossroads, The ACM Magazine for Students, 13(4), 5 DOI: 10.1145/1315325.1315330 Link: <https://doi.org/10.1145/1315325.1315330>

2. Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python.

O'Reilly

Media. DOI: 10.5555/1717171 Link: <https://dl.acm.org/doi/10.5555/1717171>

- Goldberg, Y. (2016). A Primer on Neural Network Models for Natural Language Processing. Journal of Artificial Intelligence Research, 57, 345–420. DOI: 10.1613/jair.4992 Link:

<https://doi.org/10.1613/jair.4992>

Дополнительная литература:

1. Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language

Processing Research. IEEE Computational Intelligence Magazine, 9(2), 48–57. DOI: 10.1109/MCI.2014.2307227 Link: <https://doi.org/10.1109/MCI.2014.2307227>

2. Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends in Deep

Learning

Based Natural Language Processing. IEEE Computational Intelligence Magazine, 13(3), 55–75. DOI: 10.1109/MCI.2018.2840738 Link: <https://doi.org/10.1109/MCI.2018.2840738>

- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation.

Journal of

Machine Learning Research, 3, 993–1022. DOI: 10.1162/jmlr.2003.3.4-5.993 Link: <https://doi.org/10.1162/jmlr.2003.3.4-5.993>

- Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for

Word

Representation. Proceedings of EMNLP 2014, 1532–1543. DOI: 10.3115/v1/D14-1162 Link: <https://doi.org/10.3115/v1/D14-1162>

- Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural

Networks. Advances in Neural Information Processing Systems 27. DOI: 10.5555/2969033.2969173 Link: <https://dl.acm.org/doi/10.5555/2969033.2969173>

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N.,

Kaiser, Ł., &

Polosukhin, I. (2017). Attention Is All You Need. DOI: 10.48550/arXiv.1706.03762 Link: <https://doi.org/10.48550/arXiv.1706.03762>

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training

of Deep

Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–4186. DOI: 10.18653/v1/N19-1423 Link: <https://doi.org/10.18653/v1/N19-1423>

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the

Dangers of

Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference

on Fairness, Accountability, and Transparency, 610–623. DOI: 10.1145/3442188.3445922 Link: <https://doi.org/10.1145/3442188.3445922>

- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language

processing: state of

the art, current trends and challenges. Multimedia Tools and Applications, 82(3), 3713–3744.

DOI:

10.1007/s11042-022-13428-4 Link: <https://doi.org/10.1007/s11042-022-13428-4>

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/>

- 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 «Обработка естественного языка».

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

РАЗРАБОТЧИК:

Доцент

Должность, БУП

Подпись

Гончарова Оксана
Владимировна

Фамилия И.О.

РУКОВОДИТЕЛЬ БУП:

Заведующий кафедрой

Должность БУП

Подпись

Шаклеин Виктор
Михайлович

Фамилия И.О.

РУКОВОДИТЕЛЬ ОП ВО:

Доцент

Должность, БУП

Подпись

Александрова Оксана
Ивановна

Фамилия И.О.