

**ГОУ ВПО Российско-Армянский (Славянский)
университет**

Утверждено
Директор Института Математики и Информатики
Арамян Р.Г.



«21» *марта* 2025, протокол № 9.1

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ

Наименование дисциплины: NLP2

Автор: Варданян Эдгар Араевич

Направление подготовки: 01.04.02 Прикладная математика и информатика

Наименование образовательной программы: Искусственный интеллект и машинное обучение (Artificial Intelligence and Machine Learning)

1. АННОТАЦИЯ

1.1. Краткое описание содержания данной дисциплины

Целями освоения дисциплины NLP 2 (Natural Language Processing) являются знакомство с основными проблемами в области компьютерной лингвистики, базовыми алгоритмами, математическими методами моделирования языковых феноменов, основными инструментами и технологиями в области автоматической обработки естественного языка, умение представлять в алгоритмическом виде процессы анализа и синтеза текста.

1.2. Трудоемкость в академических кредитах и часах, формы итогового контроля

Трудоемкость курса составляет 4 академических кредита, что эквивалентно 144 часам (1 академический кредит равен 36 часам). Итоговый контроль будет проведен в форме экзамена.

1.3. Взаимосвязь дисциплины с другими дисциплинами учебного плана специальности

При изучении дисциплины «NLP» используются понятия и методы Python, Big Data, Mathematica

1.4. Результаты освоения программы дисциплины:

Код компетенции	Наименование компетенции	Код индикатора достижения компетенций	Наименование индикатора достижений компетенций
УК-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий	УК-1.1	Знать теории и методы критического анализа и системного подхода
		УК-1.2	Уметь применять методы поиска, анализа и синтеза информации для решения конкретных задач
		УК-1.3	Владеть способностью выработки стратегических решений на основе комплексного анализа

			ситуации и прогнозирования исходов
УК-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1	Знать основы социального взаимодействия и теории командной работы
		УК-3.2	Уметь эффективно коммуницировать и сотрудничать в команде
		УК-3.3	Владеть навыками стратегического планирования и лидерства, способными максимизировать потенциал команды для достижения общих целей
УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1	Знать ключевые коммуникативные технологии и инструменты, используемые для международного и межкультурного общения
		УК-4.2	Уметь эффективно использовать технологии для общения и сотрудничества в многоязычной и мультикультурной среде
		УК-4.3	Владеть мастерством межкультурной коммуникации, умением адаптировать сообщения для различных культурных и профессиональных контекстов
ПК-2	Способностью разрабатывать и	ПК-2.1	Знать теоретические основы и

	анализировать концептуальные и теоретические модели решаемых научных проблем и задач		концептуальные модели в научных исследованиях
		ПК-2.2	Уметь анализировать и разрабатывать концептуальные и теоретические модели для проектной и производственной деятельности
		ПК-2.3	Владеть глубоким пониманием теоретических подходов и их практического применения в создании новых исследовательских проектов
ПК-12	Способностью к взаимодействию в рамках международных проектов и сетевых сообществ в области прикладной математики и информационных технологий	ПК-12.1	Знать принципы международного сотрудничества в научных и проектных деятельности
		ПК-12.2	Уметь взаимодействовать в международных проектах и сетевых сообществах
		ПК-12.3	Владеть умениями эффективного межкультурного общения и сотрудничества

2. УЧЕБНАЯ ПРОГРАММА

2.1. Цели и задачи дисциплины

Цели дисциплины:

- Освоение теоретических основ:** Предоставить студентам глубокие знания в области теории и методов обработки естественного языка.
- Развитие практических навыков:** Развить у студентов навыки применения алгоритмов и инструментов NLP для анализа и интерпретации текстовых данных.
- Подготовка к инновационной деятельности:** Подготовить студентов к использованию NLP в различных профессиональных областях, включая IT, маркетинг, социальные науки и другие.

Задачи дисциплины:

- Изучение основных компонентов NLP:** Освоение методов морфологического, синтаксического и семантического анализа.
- Разработка и применение NLP-моделей:** Обучение студентов созданию моделей для задач классификации текстов, машинного перевода, извлечения информации, и автоматического реферирования.
- Анализ естественно-языковых данных:** Развитие умений в обработке и анализе больших объемов текстовых данных, включая предварительную обработку, векторизацию и визуализацию.
- Исследование новых подходов в NLP:** Поощрение самостоятельного изучения современных исследований и технологий в области NLP, а также их применение в проектной и научной деятельности.

2.2. Трудоемкость дисциплины и виды учебной работы

Виды учебной работы	Всего, в акад. часах	Распределение по семестрам					
		сем	сем	3 сем	сем.	сем	сем.
1	2	3	4	5	6	7	8
1.Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	32			3			
1.1. Аудиторные занятия, в т. ч.:	32			32			
1.1.1. Лекции							
1.1.2. Практические занятия, в т. ч.	32			32			
Итоговый контроль				Экз аме н			

2.3. Содержание дисциплины

2.3.1. Тематический план и трудоемкость аудиторных занятий (модули, разделы дисциплины и виды занятий) по рабочему учебному плану

Разделы и темы дисциплины	Всего ак. часов	Лекции, ак. часов	Практ. занятия, ак. часов	Семина- ры, ак. часов	Лабор, ак. часов	Другие виды занятий, ак. часов
1	3=4+5+6 +7+8	4	5	6	7	8
Модуль 1.	32		32			
Введение в НЛП	1		1			
Раздел 1. Нейронные сети для НЛП	5		5			
Тема 1.1. Простой набор слов и подходы N-грамм	2		2			
Тема 1.2. Рекуррентные нейронные сети для последовательностей.	1		1			

Тема 1.3. Эффективная реализация	1		1			
Раздел 2. Стробированные RNN и генерация естественного языка	9		9			
Тема 2.1. Исчезновение градиента. Линейная / длинная память. GRU и LSTM	3		3			
Тема 2.2. Архитектура Encoder-Decoder. Архитектура Seq2Seq	3		3			
Тема 2.3. Машинный перевод	3		3			
Раздел 3. Обсуждение проблем, приложений и общих подходов НЛП	10		10			
Тема 3.1. SQUAD: сети указателей, сети памяти	2		2			
Тема 3.2. Вывод на естественном языке	2		2			
Тема 3.3. Системы диалога, чат-боты: проблема контекста	3		3			
Тема 3.4. Машинный перевод. Извлечение информации. Извлечение отношений.	3		3			
Раздел 4. Трансформаторы и контекстные представления слов.	7		7			
Тема 4.1. Трансформеры. Многоголовое самовнимание. Неавторегрессивная тренировка. Позиционные вложения.	4		4			
Тема 4.2 «Момент ImageNet». Предварительная подготовка языковой модели. Моделирование маскированного языка, задача Cloze.	3		3			
ИТОГО	32		32			

2.3.2. Краткое содержание разделов дисциплины в виде тематического плана

Тема 1. Введение в НЛП

- **Раздел 1. Нейронные сети для НЛП**
 - **Тема 1.1. Простой набор слов и подходы N-грамм**
 - Изучение базовых методов обработки текста, включая модели на основе набора слов и N-грамм, их применение для анализа текста.
 - **Тема 1.2. Рекуррентные нейронные сети для последовательностей**
 - Обзор рекуррентных нейронных сетей (RNN) и их использования для работы с последовательностями текста.

- **Тема 1.3. Эффективная реализация**
 - Рассмотрение техник и стратегий для эффективной реализации и оптимизации NLP-моделей.
- **Раздел 2. Стробированные RNN и генерация естественного языка**
 - **Тема 2.1. Исчезновение градиента. Линейная / длинная память. GRU и LSTM**
 - Обсуждение проблем исчезновения градиента и архитектур LSTM и GRU для решения этих проблем.
 - **Тема 2.2. Архитектура Encoder-Decoder. Архитектура Seq2Seq**
 - Изучение моделей Encoder-Decoder и Seq2Seq для задач, таких как машинный перевод.
 - **Тема 2.3. Машинный перевод**
 - Подробный разбор механизмов и подходов к автоматическому переводу текстов на разные языки.
- **Раздел 3. Обсуждение проблем, приложений и общих подходов НЛП**
 - **Тема 3.1. SQUAD: сети указателей, сети памяти**
 - Рассмотрение продвинутых технологий для задач, связанных с пониманием и ответами на вопросы.
 - **Тема 3.2. Вывод на естественном языке**
 - Исследование методов генерации текста и их применение.
 - **Тема 3.3. Системы диалога, чат-боты: проблема контекста**
 - Анализ разработки и реализации систем диалога и чат-ботов, обсуждение проблемы контекста.
 - **Тема 3.4. Машинный перевод. Извлечение информации. Извлечение отношений**
 - Обзор применений НЛП в различных областях, включая извлечение информации и анализ отношений.
- **Раздел 4. Трансформаторы и контекстные представления слов**
 - **Тема 4.1. Трансформеры. Многоголовое самовнимание. Неавторегрессивная тренировка. Позиционные вложения**
 - Изучение архитектуры трансформеров, механизмов внимания и их влияния на развитие NLP.
 - **Тема 4.2 «Момент ImageNet». Предварительная подготовка языковой модели. Моделирование маскированного языка, задача Cloze**
 - Обсуждение эры "Момент ImageNet" в NLP, методов предварительной подготовки моделей и задач маскированного языка.

2.3.3. Краткое содержание семинарских/практических занятий/лабораторного практикума

Лабораторные работы: Студенты выполняют задачи по программированию, применяя NLP-библиотеки и инструменты (например, NLTK, spaCy, TensorFlow) для обработки текстовых данных, создания и тестирования моделей.

Проект (при наличии)								
Другие формы (при наличии)								
Веса результирующих оценок текущих контролей в итоговых оценках промежуточных контролей						0,3		
Веса оценок промежуточных контролей в итоговых оценках промежуточных контролей						0,7		
Вес итоговой оценки 1-го промежуточного контроля в результирующей оценке промежуточных контролей								
Вес итоговой оценки 2-го промежуточного контроля в результирующей оценке промежуточных контролей							1	
Вес результирующей оценки промежуточных контролей в результирующей оценке итогового контроля								0,4
Вес итогового контроля (Экзамен/зачет) в результирующей оценке итогового контроля								0,6
	$\Sigma = 1$							

3. Теоретический блок (указываются материалы, необходимые для освоения учебной программы дисциплины)

3.1. Материалы по теоретической части курса

3.1.1. Учебники

- Stanford CS224N Natural Language Processing course,
- Speech and Language Processing textbook by Dan Jurafsky,
- LxMLS Lisbon Machine Learning / NLP Summer School course,
- Oxford Deep NLP course

4. Фонды оценочных средств (указываются материалы, необходимые для проверки уровня знаний в соответствии с содержанием учебной программы дисциплины).

4.1. Планы практических и семинарских занятий

Практические занятия:

- Практические занятия включают выполнение заданий на программирование для решения задач NLP, таких как текстовая классификация, машинный перевод, и извлечение информации. Задания разрабатываются так, чтобы проверять способность студентов применять теоретические знания на практике.

Семинарские занятия:

- Семинары организованы для обсуждения ключевых тем и проблемных вопросов дисциплины. Включают анализ научных статей, кейс-стади и презентации проектов. Студентам предлагаются вопросы и задания для самостоятельной работы, направленные на углубление понимания материала и развитие критического мышления.

Материалы для оценки:

- Контрольные работы, тесты для проверки знаний по основным темам курса.
- Проектные задания и лабораторные работы, оценивающие умение студентов применять изученные техники в различных условиях и контекстах.
- Портфолио с выполненными заданиями и проектами, позволяющее оценить комплексные умения студентов в области NLP.

5. Методический блок

5.1. Методика преподавания

5.1.1. Методические рекомендации для студентов по подготовке к семинарским, практическим или лабораторным занятиям, по организации самостоятельной работы студентов при изучении конкретной дисциплины.

- **Изучение материала:** Перед посещением лекций или семинаров студентам рекомендуется заранее ознакомиться с предстоящими темами через учебники, научные статьи, и онлайн-ресурсы.
- **Активное участие:** На семинарах и практических занятиях важно активно участвовать в обсуждениях, задавать вопросы и делиться своими идеями и решениями с другими.