

**ГОУ ВПО РОССИЙСКО-АРМЯНСКИЙ (СЛАВЯНСКИЙ)
УНИВЕРСИТЕТ**

Составлен в соответствии с
государственными требованиями к
минимуму содержания и уровню
подготовки выпускников по
направлению 01.04.02 Прикладная
математика и информатика
и Положением «ОБУМКД РАУ».

УТВЕРЖДАЮ:
Директор института
математики и информатики,
канд. физ.-мат. наук
Дарбинян Арман Араикович



“19” 07 2023 г.

Институт: Математики и Информатики

Кафедра: Математики и математического моделирования

Автор: Овсепян Нарек Грайрович

УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС

Дисциплина: Б1.О.06 NLP

Для магистерских программ:

**Магистерская программа: 01.04.02 Искусственный интеллект и
машинное обучение (Artificial Intelligence and Machine Learning)**

Направление: Прикладная математика и информатика
Название направления

ЕРЕВАН

1. Аннотация

Целями освоения дисциплины NLP (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.

2. Взаимосвязь с другими дисциплинами специальности

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

3. Требования к исходным уровням знаний и умений студентов

Дисциплина «NLP» базируется на знаниях курса Python, Big Data, Mathematica, программирования и теории алгоритмов для создания эффективных алгоритмов.

4. Объем дисциплины и виды учебной работы по рабочему учебному плану

Виды учебной работы	Всего часов	Количество часов по семестрам			
		1 сем.	2 сем.	3 сем.	4 сем.
1	2	3	4	5	6
1. Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	36		36		
1.1. Аудиторные занятия, в т. ч.:	36		36		
1.1.1. Лекции					
1.1.2. Практические занятия, в т. ч.	36		36		
2. Форма итогового контроля: Экзамен/Зачет			зачет		

6. Содержание дисциплины

6.1 Тематический план и трудоемкость аудиторных занятий (Модули, разделы дисциплины и виды занятий) по учебному плану

Разделы и темы дисциплины	Всего ак. часов	Лекции и, 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.	36		36			
Введение в NLP	1		1			
Раздел 1.	11		11			
Тема 1.1. Проблемы обработки естественного языка	1		1			
Тема 1.2. Обсуждение проблем и приложений НЛП	2		2			
Тема 1.3. Формальные языки, синтаксис и семантика	2		2			
Тема 1.4. Нормализация текста, токенизация	2		2			
Тема 1.5. Лемматизация, Поверхностные формы, POS-теги.	2		2			
Тема 1.6. Расстояние Левенштейна, Пространство	2		2			
Раздел 2. Простые модели и вложения слов	14		14			
Тема 2.1. N-граммное языковое моделирование	1		1			
Тема 2.2. TF-IDF, Пакет слов, наивные байесовские классификаторы	1		1			
Тема 2.3. Задача WordNet.	1		1			
Тема 2.4.Реализация Word2Vec, Glove	1		1			
Тема 2.5. Смысл сходства, арифметика встраивания слов	2		2			
Тема 2.6. Информация о подслове. Встраивание символов FastText и N-грамм.	2		2			
Тема 2.7. Информация о подслове. Модели подслов, кодирование Byte-Pair, модели WordPiece и Unigram.	2		2			
Тема 2.8. Структурированный	2		2			

ГОУ ВПО Российско-Армянский (Славянский) университет


языковой модели. Моделирование маскированного языка, задача Cloze.						
ИТОГО	36		36			

7. Рекомендуемая литература:

1. Stanford CS224N Natural Language Processing course,
2. Speech and Language Processing textbook by Dan Jurafsky,
3. LxMLS Lisbon Machine Learning / NLP Summer School course,
4. Oxford Deep NLP course

Учебная программа одобрена кафедрой Математики и математического моделирования

Зав. кафедрой: Дарбинян А.А.



(подпись)