

Курс

**СИЛЛАБУС**  
**Весенний семестр 2024-2025 учебного года**  
**Образовательная программа «7М06101 – Вычислительная лингвистика»**

ID и наименование дисциплины	Самостоятельная работа обучающегося (СРМ)	Кол-во кредитов			Общее кол-во кредитов	Самостоятельная работа обучающегося под руководством преподавателя (СРМП)
		Лекции (Л)	Практ. занятия (ПЗ)	Лаб. занятия (ЛЗ)		
102269 Интеллектуальный анализ данных	4	1,7	3,3	0	5	8

**АКАДЕМИЧЕСКАЯ ИНФОРМАЦИЯ О ДИСЦИПЛИНЕ**

Формат обучения	Цикл, компонент	Типы лекций	Типы практических занятий	Форма и платформа итогового контроля
<i>Офлайн</i>	БД, КВ	Проблемно-ориентированный	Изучение концепций обработки естественных языков с помощью моделей машинного обучения	Устный оффлайн
<b>Лектор - (ы)</b>	Карюкин Владислав Игоревич			
<b>e-mail:</b>	<a href="mailto:vladislav.karyukin@gmail.com">vladislav.karyukin@gmail.com</a> <a href="mailto:vladislav.karyukin@kaznu.kz">vladislav.karyukin@kaznu.kz</a>			
<b>Телефон:</b>	+77019405992			
<b>Ассистент- (ы)</b>	Карюкин Владислав Игоревич			
<b>e-mail:</b>	<a href="mailto:vladislav.karyukin@gmail.com">vladislav.karyukin@gmail.com</a> <a href="mailto:vladislav.karyukin@kaznu.kz">vladislav.karyukin@kaznu.kz</a>			
<b>Телефон:</b>	+77019405992			

**АКАДЕМИЧЕСКАЯ ПРЕЗЕНТАЦИЯ ДИСЦИПЛИНЫ**

Цель дисциплины	Ожидаемые результаты обучения (РО)*	Индикаторы достижения РО (ИД)
Этот курс направлен на изучение концепций интеллектуального анализа данных, базовых принципов и задач NLP, включая обучение с учителем, обучение без учителя, глубокое обучение, включающее архитектуру сверточных нейронных сетей, рекуррентных нейронных сетей, трансформеров и больших языковых моделей, таких как BERT, GPT-3 и т.д.	1. (когнитивный) Теоретические и методологические концепции интеллектуального анализа данных	1.1 Понимает базовые и расширенные программы по анализу данных 1.2 Анализирует особенности методов извлечения данных 1.3 Применяет методы разработки приложений по обработке данных
	2. (функциональный) Работа с библиотеками NumPy, Pandas и Matplotlib	2.1 Использует знания установки библиотек NumPy, Pandas и Matplotlib 2.2 Применяет данные библиотеки для работы с моделями машинного обучения 2.3 Формирует навыки работы с данными библиотеками при создании приложений
	3.(функциональный) Разрабатывать программы парсинга текстовых данных из интернет-источников	3.1 Разрабатывает скрипты для парсинга текстов 3.2 Создает полнофункциональную программу парсинга текстов 3.3 Разрабатывает скрипты сохранения полученных данных в текстовые файлы
	4. (системный) Создание моделей машинного и глубокого обучения для классификации данных	4.1 Создает модели машинного обучения по обработке данных 4.2 Разрабатывает приложения, использующие модели машинного обучения 4.3 Применяет модели машинного обучения для классификации данных
	5. (системный) Создавать веб-приложения, использующие модели машинного обучения и нейронные сети	5.1 Создает основной каркас веб-приложения на HTML, CSS и JavaScript 5.2 Создает подключение моделей машинного обучения для определения

		угроз 5.3 Создает полную конфигурацию работы веб-приложения
<b>Пререквизиты</b>	Технологии программирования для NLP, Методы естественного языка, Языковые ресурсы	Машинного обучения в обработке
<b>Постреквизиты</b>	Технология сентимент анализа, Глубокое обучение	
<b>Учебные ресурсы</b>	<p><b>Литература:</b> основная, дополнительная.</p> <ol style="list-style-type: none"> <li>1. Python for Everybody: Exploring Data in Python 3 by Dr. Charles Russell Severance, Sue Blumenberg, Elliott Hauser, Aimee Andrion, 2016.</li> <li>2. Natural Language Processing with Python and spaCy: A Practical Introduction, Yuli Vasiliev, 2021.</li> <li>3. Machine Learning and Deep Learning in Natural Language Processing, Anitha S. Pillai, Roberto Tedesco, 2023.</li> <li>4. Natural Language Processing: A Machine Learning Perspective Yue Zhang, Zhiyang Teng, 2021.</li> <li>5. Natural Language Processing Projects: Build Next-Generation NLP Applications Using AI Techniques, Akshay Kulkarni, Adarsha Shivananda, Anoosh Kulkarni, 2021.</li> <li>6. Learning Scientific Programming with Python, Christian Hill, 2021</li> <li>7. Deep Learning for Natural Language Processing: Creating Neural Networks with Python. Palash Goyal, Sumit Pandey, Karan Jain, 2018</li> </ol> <p><b>Исследовательская инфраструктура</b></p> <ol style="list-style-type: none"> <li>1. Бизнес-инкубатор №12</li> <li>2. Лабораторная аудитория 517</li> <li>3. Лабораторная аудитория 323</li> <li>4. Лабораторная аудитория 514</li> </ol> <p><b>Профессиональные научные базы данных</b></p> <ol style="list-style-type: none"> <li>1. Scopus – <a href="https://www.scopus.com">https://www.scopus.com</a></li> <li>2. Elsevier – <a href="https://www.elsevier.com/">https://www.elsevier.com/</a></li> <li>3. Researchgate – <a href="https://www.researchgate.net/">https://www.researchgate.net/</a></li> </ol> <p><b>Интернет-ресурсы:</b></p> <ol style="list-style-type: none"> <li>1. Python Exercises, Practice, Solution – <a href="https://www.w3resource.com/python-exercises/">https://www.w3resource.com/python-exercises/</a></li> <li>2. Сайт программирования – <a href="https://metanit.com/python/tutorial/1.1.php">https://metanit.com/python/tutorial/1.1.php</a></li> <li>3. Бесплатный курс по Python для начинающих – <a href="https://codebasics.com/ru/languages/python">https://codebasics.com/ru/languages/python</a></li> <li>4. Python. Введение в программирование – <a href="https://younglinux.info/python/course">https://younglinux.info/python/course</a></li> <li>5. Самоучитель Python – <a href="https://pythonworld.ru/samouchitel-python">https://pythonworld.ru/samouchitel-python</a></li> <li>6. Python Exercises, Practice, Solution – <a href="https://www.w3resource.com/python-exercises/">https://www.w3resource.com/python-exercises/</a></li> <li>7. Natural Language Toolkit – <a href="https://www.nltk.org/">https://www.nltk.org/</a></li> <li>8. Tensorflow – <a href="https://www.tensorflow.org/?hl=ru">https://www.tensorflow.org/?hl=ru</a></li> <li>9. Machine learning mastery – <a href="https://machinelearningmastery.com/start-here/">https://machinelearningmastery.com/start-here/</a></li> </ol> <p><b>Программное обеспечение:</b></p> <ol style="list-style-type: none"> <li>1. Python IDE</li> <li>2. Anaconda Navigator Python</li> <li>3. Microsoft Visual Studio</li> <li>4. SQL Lite</li> <li>5. Microsoft SQL Server</li> <li>6. Microsoft Office Word</li> <li>7. WinRAR</li> <li>8. WordPad</li> <li>9. Power Point</li> <li>10. Adobe Reader</li> <li>11. Paint.</li> </ol>	

<p><b>Академическая политика дисциплины</b></p>	<p>Академическая политика дисциплины определяется <u>Академической политикой и Политикой академической честности КазНУ имени аль-Фараби</u>. Документы доступны на главной странице ИС Univer.</p> <p><b>Интеграция науки и образования.</b> Научно-исследовательская работа студентов, магистрантов и докторантов – это углубление учебного процесса. Она организуется непосредственно на кафедрах, в лабораториях, научных и проектных подразделениях университета, в студенческих научно-технических объединениях. Самостоятельная работа обучающихся на всех уровнях образования направлена на развитие исследовательских навыков и компетенций на основе получения нового знания с применением современных научно-исследовательских и информационных технологий. Преподаватель исследовательского университета интегрирует результаты научной деятельности в тематику лекций и семинарских (практических) занятий, лабораторных занятий и в задания СРОП, СРО, которые отражаются в силлабусе и отвечают за актуальность тематик учебных занятий и заданий.</p> <p><b>Посещаемость.</b> Дедлайн каждого задания указан в календаре (графике) реализации содержания дисциплины. Несоблюдение дедлайнов приводит к потере баллов.</p> <p><b>Академическая честность.</b> Практические/лабораторные занятия, СРО развивают у обучающегося самостоятельность, критическое мышление, креативность. Недопустимы плагиат, подлог, использование шпаргалок, списывание на всех этапах выполнения заданий. Соблюдение академической честности в период теоретического обучения и на экзаменах помимо основных политик регламентируют <u>«Правила проведения итогового контроля»</u>, <u>«Инструкции для проведения итогового контроля осеннего/весеннего семестра текущего учебного года»</u>, <u>«Положение о проверке текстовых документов обучающихся на наличие заимствований»</u>. Документы доступны на главной странице ИС Univer.</p> <p><b>Основные принципы инклюзивного образования.</b> Образовательная среда университета задумана как безопасное место, где всегда присутствуют поддержка и равное отношение со стороны преподавателя ко всем обучающимся и обучающимся друг к другу независимо от гендерной, расовой/ этнической принадлежности, религиозных убеждений, социально-экономического статуса, физического здоровья студента и др. Все люди нуждаются в поддержке и дружбе ровесников и сокурсников. Для всех студентов достижение прогресса скорее в том, что они могут делать, чем в том, что не могут. Разнообразие усиливает все стороны жизни. Все обучающиеся, особенно с ограниченными возможностями, могут получать консультативную помощь по телефону/ e-mail <a href="mailto:vladislav.karyukin@gmail.com">vladislav.karyukin@gmail.com</a> / +77019405992 либо посредством видеосвязи <span style="float: right;">MS Teams</span> <a href="https://teams.microsoft.com/l/channel/19%3AICQ8sYa8ULqib_TBCwW8DDnlu17ttBZYyxF2e_Ix9R.A1%40thread.tacv2/General?groupId=f27cb6af-83e5-4326-8379-0eead760ef80&amp;tenantId=">https://teams.microsoft.com/l/channel/19%3AICQ8sYa8ULqib_TBCwW8DDnlu17ttBZYyxF2e_Ix9R.A1%40thread.tacv2/General?groupId=f27cb6af-83e5-4326-8379-0eead760ef80&amp;tenantId=</a></p>
---	--

### ИНФОРМАЦИЯ О ПРЕПОДАВАНИИ, ОБУЧЕНИИ И ОЦЕНИВАНИИ

Балльно-рейтинговая буквенная система оценки учета учебных достижений			Методы оценивания															
Оценка	Цифровой эквивалент баллов	Баллы, % содержание	Оценка по традиционной системе															
A	4,0	95-100	Отлично	<p><b>Критериальное оценивание</b> – процесс соотнесения реально достигнутых результатов обучения с ожидаемыми результатами обучения на основе четко выработанных критериев. Основано на формативном и суммативном оценивании.</p> <p><b>Формативное оценивание</b> – вид оценивания, который проводится в ходе повседневной учебной деятельности. Является текущим показателем успеваемости. Обеспечивает оперативную взаимосвязь между обучающимся и преподавателем. Позволяет определить возможности обучающегося, выявить трудности, помочь в достижении наилучших результатов, своевременно корректировать преподавателю образовательный процесс. Оценивается выполнение заданий, активность работы в аудитории во время лекций, семинаров, практических занятий (дискуссии, викторины, дебаты, круглые столы, лабораторные работы и т. д.). Оцениваются приобретенные знания и компетенции.</p> <p><b>Суммативное оценивание</b> – вид оценивания, который проводится по завершению изучения раздела в соответствии с программой дисциплины. Проводится 3-4 раза за семестр при выполнении СРО. Это оценивание освоения ожидаемых результатов обучения в соотнесенности с дескрипторами. Позволяет определять и фиксировать уровень освоения дисциплины за определенный период. Оцениваются результаты обучения.</p>														
A-	3,67	90-94																
B+	3,33	85-89	Хорошо															
B	3,0	80-84																
B-	2,67	75-79																
C+	2,33	70-74																
C	2,0	65-69																
C-	1,67	60-64																
D+	1,33	55-59	Удовлетворительно															
D	1,0	50-54																
FX	0,5	25-49		Неудовлетворительно														
F	0	0-24																
				<table border="1" style="width: 100%;"> <thead> <tr> <th>Формативное и суммативное оценивание</th> <th>Баллы % содержание</th> </tr> </thead> <tbody> <tr> <td>Активность на лекциях</td> <td>0</td> </tr> <tr> <td>Работа на практических занятиях</td> <td>25</td> </tr> <tr> <td>Самостоятельная работа</td> <td>25</td> </tr> <tr> <td>Проектная и творческая деятельность</td> <td>10</td> </tr> <tr> <td>Итоговый контроль (экзамен)</td> <td>40</td> </tr> <tr> <td><b>ИТОГО</b></td> <td><b>100</b></td> </tr> </tbody> </table>	Формативное и суммативное оценивание	Баллы % содержание	Активность на лекциях	0	Работа на практических занятиях	25	Самостоятельная работа	25	Проектная и творческая деятельность	10	Итоговый контроль (экзамен)	40	<b>ИТОГО</b>	<b>100</b>
Формативное и суммативное оценивание	Баллы % содержание																	
Активность на лекциях	0																	
Работа на практических занятиях	25																	
Самостоятельная работа	25																	
Проектная и творческая деятельность	10																	
Итоговый контроль (экзамен)	40																	
<b>ИТОГО</b>	<b>100</b>																	

Неделя	Название темы	Кол-во часов	Макс. балл
<b>МОДУЛЬ 1 Основы операции работы с текстовыми данными</b>			
1	Л 1. Основные задачи и понятия интеллектуального анализа данных	1	0
	СЗ 1. Применение Data Mining в различных областях	2	6
2	Л 2. Поиск источников и методы сбора данных	1	0
	СЗ 2. Создание программы обработки текстовых данных	2	6
	СРМП 1. Консультации по выполнению СРМ1 на тему «Реализация проекта с базовыми операциями обработки текстов»		
3	Л 3. Выполнение операции стемминга с текстовыми данными	1	0
	СЗ 3. Создание программы стемминга текстовых данных	2	6
	СРМП 2. Прием СРМ 1		25
4	Л 4. Выполнение операции лемматизации с текстовыми данными	1	0
	СЗ 4. Создание программы лемматизации текстовых данных	2	6
5	Л 5. Выполнение операции векторизации текстовых данных	1	0
	СЗ 5. Создание программы векторизации текстовых данных	2	7
<b>МОДУЛЬ 2 Обработка текстовых данных моделями машинного обучения</b>			
6	Л 6. Подготовка текстовых данных для классификации моделями машинного обучения	1	0
	СЗ 6. Создание программы подготовки текстовых данных для классификации моделями машинного обучения	2	7
	СРМП 3. Консультация по выполнению СРМ 2 на тему «Создание программы классификации текстовых данных»		
7	Л 7. Классификация данных моделями машинного обучения	1	0
	СЗ 7. Создание программы классификации текстов моделями Наивного Байеса, Логистической регрессии, Машины опорных векторов, Деревя решений, Случайного леса и т.д.	2	7
	СРМП 4. Прием СРМ 2		30
<b>Рубежный контроль 1</b>			<b>100</b>
8	Л 8. Классификация текстовых данных нейронными сетями	1	0
	СЗ 8. Создание программы классификации текстов моделями Deep neural network, Convolutional neural network и Long short term memory neural network	2	5
	СРМП 5. Консультация по выполнению СРМ 3 на тему «Разработка программы анализа тональности текстов с помощью BERT»		
9	Л 9. Большие языковые модели BERT, GPT	1	0
	СЗ 9. Создание программы обработки текстовых данных моделью BERT	2	5
10	Л 10. Анализ и обработка текстов с помощью запросов ChatGPT	1	0
	СЗ 10. Создание программы обработки текстов с API ChatGPT	2	5
	СРМП 6. Прием СРМ 3		25
<b>МОДУЛЬ 3 Работа с парсингом текстовых данных</b>			
11	Л 11. Основные методы анализа текстовых документов в HTML формате	1	0
	СЗ 11. Создание программы парсинга текстов библиотекой BeautifulSoup	2	5
	СРМП 7. Консультация по выполнению СРМ 4 на тему «Создание приложения веб-краулера»		
12	Л12. Работа с парсингом HTML страниц с помощью BeautifulSoup и Scrapy	1	0
	СЗ 12. Создание программы парсинга текста библиотекой Scrapy	2	5
13	Л 13. Работа с инструментами для работы с большими данными: Hadoop, Spark	1	0
	СЗ 13. Создание методов визуализации многомерных данных и построения дашбордов	2	5
	СРМП 8. Прием СРМ 4		30
14	Л 14. Добавление функций поиска данных в социальных сетях	1	0
	СЗ 14. Разработка многофункционального веб-краулера	2	5
15	Л 15. Основные этапы создания веб-приложения с моделями машинного обучения	1	0
	СЗ 15. Разработка веб-приложения на Django	2	10
<b>Рубежный контроль 2</b>			<b>100</b>
<b>Итоговый контроль (экзамен)</b>			<b>100</b>
<b>ИТОГО за дисциплину</b>			<b>100</b>

## РУБРИКАТОР СУММАТИВНОГО ОЦЕНИВАНИЯ

### КРИТЕРИИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

#### СРМ 1. Реализация проекта с базовыми операциями обработки текстов (25% от 100% РК1)

Критерий	«Отлично» 21-25%	«Хорошо» 11-20%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Знание и понимание основных концепций обработки текстовых данных	Понимание степени соответствия, актуальности найденных данных. Знание и понимание элементов операций обработки текстовых данных	Понимание степени соответствия, актуальности найденных данных. Знание больше части операций обработки текстовых данных	Ограниченное понимание степени соответствия, актуальности элементов и операций обработки текстовых данных	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности данных. Отсутствие знания элементов и операций обработки текстовых данных
Навыки написания программного кода обработки текстовых данных	Четкое и ясное представление программного кода, отсутствие синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и корректность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать вза содержанием. Много ошибок в тексте

**СРМ 2. Создание программы классификации текстовых данных (30% от 100% РК1)**

Критерий	«Отлично» 26-30%	«Хорошо» 16-25%	«Удовлетворительно» 6-15%	«Неудовлетворительно» 0-5%
Работа с моделями машинного обучения классификации текстовых данных	Понимание степени соответствия, актуальности работы с данными в приложении. Знание операций и понимание всех основных операций классификации машинного обучения текстовых данных моделями машинного обучения	Понимание степени соответствия, актуальности работы с данными. Знание операций создания моделей машинного обучения	Ограниченное понимание актуальности операций по созданию моделей машинного обучения	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности работы с базами данных. Отсутствие знаний операций создания моделей машинного обучения
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие ошибок в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и ясность. В основном отсутствуют ошибки.	В письме есть некоторые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать в тексте

**СРМ 3. Разработка программы анализа тональности текстов с помощью BERT (25% от 100% РК2)**

Критерий	«Отлично» 16-25%	«Хорошо» 11-15%	«Удовлетворительно» 6-10%	«Неудовлетворительно» 0-5%
Работа с моделями машинного обучения классификации текстовых данных с помощью BERT	Понимание степени соответствия, актуальности работы с данными в приложении. Знание операций классификации текстовых данных с помощью BERT	Понимание степени соответствия, актуальности найденных данных. Знание операций создания моделей BERT	Ограниченное понимание актуальности операций по созданию моделей BERT	Поверхностное понимание/ отсутствие понимания степени соответствия, актуальности и достоверности работы с базами данных. Отсутствие знаний операций создания моделей BERT
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие ошибок в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических синтаксических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и ясность. В основном отсутствуют ошибки.	В письме есть некоторые ошибки, и ясность нуждается в улучшении.	Написанное неясно, трудно следовать в тексте

СРМ 4. Создание приложения веб-краулера (30% от 100% РК2)

Критерий	«Отлично» 26-30%	«Хорошо» 16-25%	«Удовлетворительно» 6-15%	«Неудовлетворительно» 0-5%
Знание и понимание библиотек создания веб-краулера	Понимание степени соответствия, актуальности работы с веб-краулером	Понимание степени соответствия, актуальности работы с веб-краулером	Ограниченное понимание работы с веб-краулером	Поверхностное понимание/ отсутствие понимания основных операций работы с веб-краулером
Навыки написания программного кода	Четкое и ясное представление программного кода, отсутствие в коде синтаксических ошибок	В программном коде имеются небольшие логические ошибки	Большое количество логических ошибок в программном коде, что делают его практически неработоспособным	Отсутствие программного кода или наличие нескольких строк кода
Написание отчета	Письмо демонстрирует ясность, лаконичность и правильность.	Письмо демонстрирует ясность, лаконичность и ясность, лаконичность и ясность. В основном отсутствуют ошибки.	В письме есть некоторые ключевые ошибки, и ясность нуждается в уточнении.	Написанное неясно, трудно следовать вза содержанию. Много ошибок в тексте

И.о. декана

Турар О.Н.

Председатель Академического комитета

по качеству преподавания и обучения



Адилжанова С.А.

Заведующий кафедрой

Шормакова А.Н.

Лектор

Карюкин В.И.