

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Волков В.В.

Должность: Ректор

Дата подписания: 18.08.2025 13:54:37

Уникальный программный ключ:

ed68fd4b85b778e0f0b1bfea5dbc56cf4148f1229917e799a70e51517ff6d591

УТВЕРЖДАЮ:

Ректор

В.В. Волков

« 26 » Февраль 2025 г.

Протокол УС № 1

от 26.02.2025 г.



**Рабочая программа дисциплины
Обработка естественного языка**

образовательная программа
направление подготовки
09.04.03 Прикладная информатика

направленность (профиль)
«Прикладной анализ данных и искусственный интеллект»
программа подготовки – магистратура

язык обучения – русский
форма обучения - очная

квалификация (степень) выпускника
Магистр

Санкт-Петербург

Автор:

Левшун Д.С., к.т.н., доцент, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рецензент:

Котельников Евгений Вячеславович, д. тех. н., профессор, Школа вычислительных социальных наук, АНООВО «ЕУСПб»

Рабочая программа дисциплины **«Обработка естественного языка»**, входящей в образовательную программу уровня магистратуры **«Прикладной анализ данных и искусственный интеллект»**, утверждена на заседании Совета Школы вычислительных социальных наук.

Протокол заседания № 4 от 25.02.2025 года.

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ **«Обработка естественного языка»**

Дисциплина «Обработка естественного языка» является дисциплиной обязательной части основной профессиональной образовательной программы высшего образования «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика.

Дисциплина «Обработка естественного языка» посвящена изучению современных методов и инструментов для анализа текста. Слушатели знакомятся с передовыми подходами, включая нейронные сети и трансформеры, лежащие в основе таких технологий, как ChatGPT и приобретают практические навыки работы с библиотеками и фреймворками для NLP. Особое внимание уделяется оценке качества моделей, анализу социальных сетей и разработке интеллектуальных чат-ботов.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточный контроль в форме зачета с оценкой.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы, 108 часов.

Содержание

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ	5
3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	6
4. ОБЪЕМ ДИСЦИПЛИНЫ	6
5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ	6
5.1 Содержание дисциплины	7
5.2 Структура дисциплины	7
6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	8
6.1 Общие положения	8
6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины	8
6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине	8
6.4 Перечень литературы для самостоятельной работы обучающегося	9
6.5 Перечень учебно-методического обеспечения для самостоятельной работы	9
7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ	9
7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации	9
7.2 Контрольные задания для текущей аттестации	10
7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации	11
7.4 Типовые задания к промежуточной аттестации	13
7.5 Средства оценки индикаторов достижения компетенций	15
8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА	16
8.1. Основная литература	16
8.2 Дополнительная литература	16
9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	16
9.1 Программное обеспечение	16
9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:	17
9.3 Лицензионные электронные ресурсы библиотеки Университета	17
9.4 Электронная информационно-образовательная среда Университета	17
10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА	18
ПРИЛОЖЕНИЕ 1	19

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины «Обработка естественного языка» является знакомство с основными методами и приложениями автоматической обработки естественного языка, получение практических навыков работы с инструментами NLP.

Задачи освоения дисциплины «Обработка естественного языка» включают:

1. знакомство с основными методами и подходами к NLP,
2. понимание принципов оценки качества методов NLP,
3. овладение практическими навыками обработки больших коллекций текстов.

2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В результате изучения учебной дисциплины обучающийся должен овладеть следующими компетенциями: общепрофессиональными (ОПК). Планируемые результаты формирования компетенций и индикаторы их достижения в результате освоения дисциплины представлены в Таблице 1.

Планируемые результаты освоения дисциплины, соотнесенные с индикаторами достижения компетенций обучающихся

Таблица 1

Код и наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (знать, уметь, владеть)
ОПК-2 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ИД.ОПК-2.1. Разрабатывает и осуществляет отладку работоспособности оригинальных алгоритмов и программных средств для решения профессиональных задач ИД.ОПК-2.2. Использует современные интеллектуальные технологии для разработки оригинальных алгоритмов и программных средств для решения профессиональных задач ИД.ОПК-2.3. Интегрирует разработанные программные модули и компоненты и верифицирует выпуски программного продукта ИД.ОПК-2.4. Разрабатывает требования и проектирует программное обеспечение для решения профессиональных задач	Знать: методологию разработки оригинальных алгоритмов и программных средств, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач З (ОПК-2) Уметь: самостоятельно разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач У (ОПК-2) Владеть: навыками разработки и интеграции программных модулей и компонент для решения профессиональных задач В (ОПК-2)

В результате освоения дисциплины магистрант должен:

Знать:

- ключевые этапы и методы обработки естественного языка,
- принципы работы языковых моделей, включая классические и нейронные подходы,
- архитектуру трансформеров и их применение в современных NLP-системах,
- методы оценки качества NLP-моделей,
- способы представления лингвистической информации в цифровом виде,
- современные библиотеки и фреймворки для NLP (spaCy, HuggingFace Transformers, NLTK и др.).

Уметь:

- применять методы предварительной обработки текста,
- использовать алгоритмы машинного обучения и глубокого обучения для задач NLP,

- проводить анализ текстовых данных и извлекать из них информацию,
- строить и обучать нейронные языковые модели,
- разрабатывать базовые интеллектуальные чат-боты,
- применять методы оценки качества моделей и интерпретировать результаты.

Владеть:

- навыками построения и обучения NLP-моделей на практике,
- инструментами для работы с текстовыми корпусами и аннотированными данными,
- методами визуализации и анализа текстов,
- технологиями создания приложений на основе NLP (например, чат-ботов, систем классификации и извлечения информации),
- подходами к решению прикладных задач обработки естественного языка в различных предметных областях.

3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Обработка естественного языка» является обязательной дисциплиной Блока 1 «Дисциплины (модули)» обязательной части образовательной программы «Прикладной анализ данных и искусственный интеллект». Курс читается в шестом модуле, форма промежуточной аттестации – зачет с оценкой.

Для успешного освоения данной дисциплины требуются знания, полученные в рамках прохождения обучения на уровне бакалавриата/ специалитета.

Знания, умения и навыки, полученные при освоении данной дисциплины, применяются магистрантами в процессе прохождения Б2.О.01(У) Технологической (проектно-технологической) практики и выполнения выпускной квалификационной работы.

4. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 3 (три) зачетных единицы, 108 часов.

Таблица 2
Объем дисциплины

Типы учебных занятий и самостоятельная работа	форма	Объем дисциплины									
		Всего		Модуль							
		1	2	3	4	5	6	7	8	9	10
Контактная работа обучающихся с преподавателем в соответствии с УП:		28	-	-	-	-	28	-	-	-	-
Лекции (Л)		14	-	-	-	-	14	-	-	-	-
Лабораторные занятия (ЛЗ)		14	-	-	-	-	14	-	-	-	-
Самостоятельная работа (СР)		80	-	-	-	-	80	-	-	-	-
Промежуточная аттестация	Зачет с оценкой	-	-	-	-	-	Зачет с оценкой	-	-	-	-
	час.	-	-	-	-	-		-	-	-	-
Общая трудоемкость дисциплины (час./з.е.)		108/3	-	-	-	-	108/3	-	-	-	-

5. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Содержание дисциплины соотносится с планируемыми результатами обучения по дисциплине: через задачи, формируемые компетенции и их компоненты (знания, умения, навыки – далее ЗУВ) по средствам индикаторов достижения компетенций в соответствии с Таблицей 3.

5.1 Содержание дисциплины

Таблица 3

Содержание дисциплины

№ п/п	Наименование тем (разделов)	Содержание тем (разделов)	Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Коды ЗУВ (в соотв. с Таблицей 1)
1	Лингвистическая предобработка текста	токенизация, нормализация, стемминг, лемматизация, удаление стоп-слов, POS-теггинг, морфологический анализ, синтаксический разбор, построение словарей и частотных моделей	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	З (ОПК-2) У (ОПК-2) В (ОПК-2)
2	Машинное обучение в NLP	наивный байесовский классификатор, логистическая регрессия, SVM, деревья решений, методы векторизации текста (TF-IDF, Bag-of-Words), word embeddings (Word2Vec, GloVe), трансформеры, обучение с учителем и без учителя, fine-tuning предобученных моделей, оценка качества моделей (accuracy, precision, recall, F1, BLEU, ROUGE и пр.)	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	З (ОПК-2) У (ОПК-2) В (ОПК-2)

5.2 Структура дисциплины

Таблица 4

Структура дисциплины

№ п/п	Наименование тем (разделов)	Объем дисциплины, час.			Форма текущего контроля успеваемости*, промежуточной аттестации	
		Всего	Контактная работа обучающихся с преподавателем по типам учебных занятий в соответствии с УП	СР		
			Л			
<i>Очная форма обучения</i>						
Тема 1	Лингвистическая предобработка текста	54	7	7	40	KР
Тема 2	Машинное обучение в NLP	54	7	7	40	KР
Промежуточная аттестация		-	-	-	-	Зачет с оценкой
Всего:		180/5	14	14	80	

*Примечание: формы текущего контроля успеваемости: контрольная работа (КР).

6. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

6.1 Общие положения

Знания и навыки, полученные в результате лекций и семинарских занятий, закрепляются и развиваются в результате повторения материала, усвоенного в аудитории, путем чтения текстов и исследовательской литературы (из списков основной и дополнительной литературы) и их анализа.

Самостоятельная работа является важнейшей частью процесса высшего образования. Ее следует осознанно организовать, выделив для этого необходимое время и соответственным образом организовав рабочее пространство. Важнейшим элементом самостоятельной работы является проработка материалов прошедших занятий (анализ конспектов, чтение рекомендованной литературы) и подготовка к следующим лекциям/семинарским занятиям. Литературу, рекомендованную в программе курса, следует, по возможности, читать в течение всего семестра, концентрируясь на обусловленных программой курса темах.

Существенную часть самостоятельной работы магистранта представляет самостоятельное изучение вспомогательных учебно-методических изданий, лекционных конспектов, интернет-ресурсов и пр. Подготовка к семинарским занятиям, контрольному тесту также является важной формой работы магистранта. Самостоятельная работа может вестись как индивидуально, так и при содействии преподавателя.

6.2 Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины

Тема 1. Лингвистическая предобработка текста:

1.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 20 часов.

1.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 20 часов. Итого: 40 часов.

Тема 2. Машинное обучение в NLP:

2.1. Изучение вопросов, представленных в списке тем лекций. Повторение изученного на предыдущих лекциях материала при подготовке к последующим лекциям – 20 часов.

2.2. Подготовка к лабораторным занятиям по предложенным темам, самостоятельное изучение рекомендованной литературы, повторение материала лекций – 20 часов. Итого: 40 часов.

6.3 Перечень основных вопросов по изучаемым темам для самостоятельной работы обучающихся по дисциплине

Вопросы для самостоятельной подготовки по темам дисциплины:

1. Современные подходы к исправлению опечаток в текстах: правила, статистика, нейросети.
2. Методы и инструменты распознавания звучащей речи.
3. Морфологический анализ: словари, правила, алгоритмы, современные библиотеки.
4. Архитектура и принципы работы нейронных языковых моделей.
5. Подходы к анализу тональности текста: лексические, статистические и нейронные.

6. Синтаксический анализ: методы, формализмы, инструменты, оценка качества.
7. Извлечение информации из текста с применением методов машинного обучения.
8. Архитектуры нейронных сетей для автоматического реферирования текста.
9. Дистрибутивные семантические модели: Skip-Gram, CBOW, FastText.
10. Методы оценки качества машинного перевода: BLEU, METEOR, TER.
11. Применение трансформеров в задачах NLP: BERT, GPT, T5.
12. Разработка чат-ботов: архитектура, диалоговые модели, инструменты.
13. Этические аспекты и проблемы предвзятости в NLP-моделях.
14. Использование корпусной лингвистики в задачах анализа текстов.
15. Проблемы многозначности и контекстной интерпретации слов в NLP.

6.4 Перечень литературы для самостоятельной работы обучающегося:

1. Богданов Е. П. Интеллектуальный анализ данных: практикум для магистрантов направления 09.04.03 «Прикладная информатика» профиль подготовки «Информационные системы и технологии корпоративного управления» / Е. П. Богданов. Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. 112 с. Текст: электронный. URL: <https://znanium.com/catalog/product/1087885> . Режим доступа: по подписке.
2. Бутусова А. С. Машинный и автоматизированный перевод: учебное пособие/ А. С. Бутусова, Ю. В. Бец; Южный федеральный университет. Ростов-на-Дону; Таганрог: Издательство Южного федерального университета, 2021. 106 с. ISBN 978-5-9275-3982-6. Текст: электронный. URL: <https://znanium.com/catalog/product/2057597> . Режим доступа: по подписке.
3. Тапскотт Д. Викиномика: как массовое сотрудничество изменяет все / Д. Тапскотт, Э. Д. Уильямс. Москва: Интеллектуальная Литература, 2020. 456 с. ISBN 978-5-6042878-7-3. Текст: электронный. URL: <https://znanium.com/catalog/product/1220225> . Режим доступа: по подписке.

6.5 Перечень учебно-методического обеспечения для самостоятельной работы

Для обеспечения самостоятельной работы магистрантов по дисциплине «Обработка естественного языка» разработано учебно-методическое обеспечение в составе:

1. Контрольные задания для подготовки к процедурам текущего контроля (п. 7.2 Рабочей программы).
2. Типовые задания для подготовки к промежуточной аттестации (п. 7.4 Рабочей программы).
3. Рекомендуемые основная, дополнительная литература, Интернет-ресурсы и справочные системы (п. 8, 9 Рабочей программы).
4. Рабочая программа дисциплины размещена в электронной информационно-образовательной среде Университета на электронном учебно-методическом ресурсе АНООВО «ЕУСПб» — образовательном портале LMS Sakai — Sakai@EU.

7. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

7.1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому лабораторному занятию, выполнение контрольных работ, активное слушание на лекциях. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 5

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Лингвистическая предобработка текста	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	З (ОПК-2) У (ОПК-2) В (ОПК-2)	Контрольная работа 1	зачтено/ не зачтено
Машинное обучение в NLP	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	З (ОПК-2) У (ОПК-2) В (ОПК-2)	Контрольная работа 2	зачтено/ не зачтено

Таблица 6

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Контрольная работа	магистрант выполняет задания контрольной работы частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение заданий контрольной работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

7.2 Контрольные задания для текущей аттестации

Примерные задания для контрольных работ

Тема 1. Лингвистическая предобработка текста

1. Преобразуйте текст в список токенов, объясните, какие сложности могут возникнуть.
2. Выполните стемминг и лемматизацию, сравните результаты.
3. Объясните различия между POS-теггингом и лемматизацией.
4. Разработайте регулярное выражение для извлечения e-mail адресов.
5. Постройте частотный словарь для коллекции текстов.
6. Объясните роль морфологического анализа в NLP.
7. Проведите синтаксический разбор простого предложения.
8. Напишите функцию для удаления стоп-слов.
9. Объясните, почему нормализация важна при анализе текста.
10. Сравните различные методы токенизации.
11. Какие ошибки могут возникнуть при автоматическом разборе текста?
12. Постройте биграммы из заданного текста.

13. Объясните понятие "лемма" на примере.
14. Сравните NLTK и spaCy для лингвистической предобработки.
15. Оцените роль частеречной разметки в анализе текста.

Тема 2. Машинное обучение в NLP

1. Объясните принцип работы наивного байесовского классификатора.
2. Постройте классификатор для анализа тональности текста.
3. Примените TF-IDF в задаче классификации новостей.
4. Сравните методы Bag-of-Words и Word2Vec.
5. Объясните, как работает логистическая регрессия в NLP.
6. Постройте confusion matrix и интерпретируйте её.
7. Обучите модель SVM на наборе текстов и оцените её качество.
8. Объясните разницу между обучением с учителем и без учителя в NLP.
9. Используйте предобученную модель BERT для классификации.
10. Сравните GloVe и FastText по качеству эмбеддингов.
11. Напишите код для fine-tuning модели DistilBERT.
12. Примените метод кластеризации для группировки текстов.
13. Объясните метрику BLEU и приведите пример ее расчета.
14. Проведите оценку модели по метрике F1-score.
15. Постройте пайплайн для задачи извлечения именованных сущностей.

7.3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет с оценкой в форме тестирования.

Перед зачетом с оценкой проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 20 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте - 1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балла; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Задания открытого типа

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,
 K – количество осваиваемых в рамках дисциплины компетенций,
 k_n – максимально возможное количество баллов за вопросы по компетенции,
 x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Таблица 7
Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Коды ЗУВ (в соот. с Таблицей 1)	Критерии оценивания	Оценка
Зачет с оценкой/ Тест	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	3 (ОПК-2) У (ОПК-2) В (ОПК-2)	81-100% правильных ответов	Зачтено, отлично
				61-80% правильных ответов	Зачтено, хорошо
				41-60% правильных ответов	Зачтено, удовлетворительно
				0-40% правильных ответов	Не зачтено, неудовлетворительно

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 7а.

Таблица 7а
Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, отлично», «зачтено, хорошо», «зачтено, удовлетворительно», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценках «не зачтено, неудовлетворительно», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

7.4 Типовые задания к промежуточной аттестации

ОПК-2 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

Комбинированные задания

Базовый уровень сложности

Задание 1

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Из представленных вариантов наиболее точным определением автоматического перевода является

Варианты ответа:

- 1) перевод, полностью выполненный человеком с помощью словаря
- 2) перевод, выполненный машиной без участия человека
- 3) перевод, при котором человек участвует в каждом этапе
- 4) перевод, сделанный с использованием онлайн-словаря
- 5) перевод, сделанный человеком с последующей проверкой компьютером

Поле для ответа:

--	--	--

Обоснование _____

Задание 2

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Из предложенных вариантов наилучшим образом отражает суть автоматизированного перевода следующий вариант

Варианты ответа:

- 1) перевод, при котором человек только набирает текст
- 2) перевод, при котором машина выполняет все этапы без вмешательства человека
- 3) перевод, при котором человек помогает машине на отдельных этапах
- 4) перевод, сделанный полностью вручную
- 5) перевод, при котором используется только словарь

Поле для ответа:

--	--	--

Обоснование _____

Задание 3

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Из приведённых вариантов термин «машинные помощники переводчика» лучше всего описывает

Варианты ответа:

- 1) автоматическую систему перевода без участия человека
- 2) систему, заменяющую переводчика полностью
- 3) инструменты, облегчающие работу переводчика
- 4) перевод, сделанный вручную и отправленный по электронной почте
- 5) программу, которая сама выбирает язык перевода

Поле для ответа:

--	--	--

Обоснование _____

Задание 4

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Из предложенных определений наилучшим образом описывает цель машинного перевода в его начальной стадии

Варианты ответа:

- 1) перевод, не нуждающийся в человеке
- 2) перевод, равный по качеству переводу опытного переводчика
- 3) перевод, выполняемый только вручную
- 4) перевод, основанный на устной речи
- 5) перевод, ориентированный на художественные тексты

Поле для ответа:

--	--	--

Обоснование _____

Задание 5

Инструкция: Выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа

Вопрос: Из предложенных вариантов наиболее точным примером автоматизированной системы перевода является

Варианты ответа:

- 1) программа, переводящая текст без возможности вмешательства
- 2) словарь с переводом слов
- 3) переводческая система, в которую человек вносит правки до и после перевода
- 4) перевод, выполненный вручную с помощью бумажного справочника
- 5) система, выполняющая только синтаксический анализ

Поле для ответа:

--	--	--

Обоснование _____

Задания открытого типа

Повышенный уровень сложности

Задание 1

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Назовите не менее двух причин, по которым полностью автоматический машинный перевод не может обеспечить высокое качество перевода. Приведите примеры возможных трудностей, с которыми сталкиваются такие системы.

Поле для ответа: _____

Задание 2

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Назовите два отличия между автоматическим и автоматизированным переводом. Объясните, в чём заключается роль человека в каждом из этих типов перевода.

Поле для ответа: _____

Задание 3

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Укажите не менее двух функций, которые выполняют машинные помощники переводчика. Объясните, как они облегчают работу профессионала.

Поле для ответа: _____

Задание 4

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Назовите не менее двух факторов, повлиявших на развитие машинного перевода в 20 веке. Объясните, почему именно эти факторы были значимыми.

Поле для ответа: _____

Задание 5

Инструкция: Прочитайте задание и запишите развернутый обоснованный ответ

Вопрос: Опишите устройство переводческой системы, предложенной П. П. Смирновым-Троянским. Укажите не менее двух этапов в работе этой системы и их особенности.

Поле для ответа: _____

7.5 Средства оценки индикаторов достижения компетенций

Таблица 8

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Средства оценки (в соотв. с Таблицами 5, 7)
ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	Контрольная работа, тест

Таблица 9

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соотв. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Контрольная работа	Магистрант в ходе подготовки и выполнения контрольной работы показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Разрабатывать и осуществлять отладку работоспособности оригинальных алгоритмов и программных средств для решения профессиональных задач, используя современные интеллектуальные технологии, интегрировать разработанные программные модули и компоненты и верифицировать выпуски программного продукта, разрабатывать требования и проектировать программное обеспечение для решения профессиональных задач

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Тест	<p>Магистрант в ходе подготовки и выполнения тестов показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности:</p> <p>— Разрабатывать и осуществлять отладку работоспособности оригинальных алгоритмов и программных средств для решения профессиональных задач, используя современные интеллектуальные технологии, интегрировать разработанные программные модули и компоненты и верифицировать выпуски программного продукта, разрабатывать требования и проектировать программное обеспечение для решения профессиональных задач</p>

8. ОСНОВНАЯ И ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

8.1. Основная литература

1. Бутусова А. С. Машинный и автоматизированный перевод: учебное пособие/ А. С. Бутусова, Ю. В. Бец; Южный федеральный университет. Ростов-на-Дону; Таганрог: Издательство Южного федерального университета, 2021. 106 с. ISBN 978-5-9275-3982-6. Текст: электронный. URL: <https://znanium.com/catalog/product/2057597> . Режим доступа: по подписке..

8.2 Дополнительная литература

1. Богданов Е. П. Интеллектуальный анализ данных: практикум для магистрантов направления 09.04.03 «Прикладная информатика» профиль подготовки «Информационные системы и технологии корпоративного управления» / Е. П. Богданов. Волгоград : ФГБОУ ВО Волгоградский ГАУ, 2019. 112 с. Текст: электронный. URL: <https://znanium.com/catalog/product/1087885> . Режим доступа: по подписке.

9. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ, ИСПОЛЬЗУЕМЫЕ ПРИ ОСУЩЕСТВЛЕНИИ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

9.1 Программное обеспечение

При осуществлении образовательного процесса магистрантами и профессорско-преподавательским составом используется следующее лицензионное программное обеспечение:

1. ABBYY FineReader 11 Corporate Edition
2. ABBYY Lingvo x5
3. Adobe Acrobat Professional 11.0 MLP AOO License RU
4. Adobe CS5.5 Design Standart Win IE EDU CLP
5. Adobe Acrobat Reader – бесплатно
6. Git (версия 2.40 и выше)
7. Google Chrome
8. Mozilla – бесплатно
9. MS Office (OVS Office Platform)
10. Opera – бесплатно
11. OS Microsoft Windows (OVS OS Platform)
12. VLC – бесплатно
13. Яндекс.Браузер (Yandex Browser) – бесплатно
14. Anaconda - бесплатно

9.2 Перечень информационно-справочных систем и профессиональных баз данных информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

Информационно-справочные системы

1. Гарант.Ру. Информационно-правовой портал: <http://www.garant.ru>
2. Информационная система «Единое окно доступа к образовательным ресурсам»: <http://window.edu.ru/>
3. Открытое образование. Ассоциация «Национальная платформа открытого образования»: <http://npoed.ru>
4. Официальная Россия. Сервер органов государственной власти Российской Федерации: <http://www.gov.ru>
5. Официальный интернет-портал правовой информации. Государственная система правовой информации: <http://pravo.gov.ru>
6. Правовой сайт КонсультантПлюс: <http://www.consultant.ru/sys>
7. Российское образование. Федеральный портал: <http://www.edu.ru>

Профессиональные базы данных информационно-телекоммуникационной сети «Интернет»:

1. ЕНИП — Электронная библиотека «Научное наследие России»: <http://e-heritage.ru/>
2. Интелрос. Интеллектуальная Россия: <http://www.intelros.ru/>
3. Национальная электронная библиотека НЭБ: <http://www.rusneb.ru>
4. Президентская библиотека: <http://www.prlib.ru>
5. Российская государственная библиотека: [http://www.rsl.ru/](http://www.rsl.ru)
6. Российская национальная библиотека: <http://www.nlr.ru/poisk/>

9.3 Лицензионные электронные ресурсы библиотеки Университета

Профессиональные базы данных:

Полный перечень доступных обучающимся профессиональных баз данных представлен на официальном сайте Университета <https://eusp.org/library/electronic-resources>, включая следующие базы данных:

1. **eLIBRARY.RU** — Российский информационно-аналитический портал в области науки, технологии, медицины и образования, содержащий рефераты и полные тексты научных статей и публикаций, научометрическая база данных: <http://elibrary.ru>;
2. Электронные журналы по подписке (текущие номера научных зарубежных журналов)

Электронные библиотечные системы:

1. **Znanium.com** — Электронная библиотечная система (ЭБС) — <http://znanium.com/>;
2. Университетская библиотека онлайн — Электронная библиотечная система (ЭБС) — <http://biblioclub.ru/>

9.4 Электронная информационно-образовательная среда Университета

Образовательный процесс по дисциплине поддерживается средствами электронной информационно-образовательной среды Университета, которая включает в себя электронный учебно-методический ресурс АНООВО «ЕУСПб» — образовательный портал LMS Sakai — Sakai@EU, лицензионные электронные ресурсы библиотеки Университета, официальный сайт Университета (Европейский университет в Санкт-Петербурге

[<https://eusp.org/>]), локальную сеть и корпоративную электронную почту Университета, и обеспечивает:

- доступ к учебным планам, рабочим программам дисциплин (модулей), практик и к изданиям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах;
- фиксацию хода образовательного процесса, результатов промежуточной аттестации и результатов освоения основной образовательной программы;
- формирование электронного портфолио обучающегося, в том числе сохранение работ обучающегося, рецензий и оценок за эти работы со стороны любых участников образовательного процесса;
- взаимодействие между участниками образовательного процесса, в том числе синхронное и (или) асинхронное взаимодействие посредством сети «Интернет» (электронной почты и т.д.).

Каждый обучающийся в течение всего периода обучения обеспечен индивидуальным неограниченным доступом к электронным ресурсам библиотеки Университета, содержащей издания учебной, учебно-методической и иной литературы по изучаемой дисциплине.

10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКАЯ БАЗА, НЕОБХОДИМАЯ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА

В ходе реализации образовательного процесса используются специализированные многофункциональные аудитории для проведения занятий лекционного типа, занятий семинарского типа (практических занятий, лабораторных работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории.

Проведение занятий лекционного типа обеспечивается демонстрационным оборудованием.

Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Для лиц с ограниченными возможностями здоровья и инвалидов предоставляется возможность присутствия в аудитории вместе с ними ассистента (помощника). Для слабовидящих предоставляется возможность увеличения текста на экране ПК. Для самостоятельной работы лиц с ограниченными возможностями здоровья в помещении для самостоятельной работы организовано одно место (ПК) с возможностями бесконтактного ввода информации и управления компьютером (специализированное лицензионное программное обеспечение – Camera Mouse, веб камера). Библиотека университета предоставляет удаленный доступ к электронным ресурсам библиотеки Университета с возможностями для слабовидящих увеличения текста на экране ПК. Лица с ограниченными возможностями здоровья могут при необходимости воспользоваться имеющимся в университете креслом-коляской. В учебном корпусе имеется адаптированный лифт. На первом этаже оборудован специализированный туалет. У входа в здание университета для инвалидов оборудована специальная кнопка, входная среда обеспечена информационной доской о режиме работы университета, выполненной рельефно-точечным тактильным шрифтом (азбука Брайля).

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ
«Обработка естественного языка»

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

1 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Информация о содержании и процедуре текущего контроля успеваемости, методике оценивания знаний, умений и навыков обучающегося в ходе текущего контроля доводятся научно-педагогическими работниками Университета до сведения обучающегося на первом занятии по данной дисциплине.

Текущий контроль предусматривает подготовку магистрантов к каждому лабораторному занятию, выполнение контрольных работ, активное слушание на лекциях. Магистрант должен присутствовать на семинарских занятиях, отвечать на поставленные вопросы, показывая, что прочитал разбираемую литературу, представлять содержательные реплики по обсуждаемым вопросам.

Текущий контроль проводится в форме оценивания выполненных контрольных работ, демонстрирующих степень знакомства с дополнительной литературой.

Таблица 1

Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе текущей аттестации

Наименование тем (разделов)	Коды компетенций	Индикаторы компетенций	Коды ЗУВ (в соот. с Таблицей 1)	Формы текущего контроля успеваемости	Результаты текущего контроля
Лингвистическая предобработка текста	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	З (ОПК-2) У (ОПК-2) В (ОПК-2)	Контрольная работа 1	зачтено/ не зачтено
Машинное обучение в NLP	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	З (ОПК-2) У (ОПК-2) В (ОПК-2)	Контрольная работа 2	зачтено/ не зачтено

Таблица 2

Критерии оценивания

Формы текущего контроля успеваемости	Критерии оценивания
Контрольная работа	магистрант выполняет задания контрольной работы частично или с существенными недочетами (некорректно сформулирован исследовательский вопрос, не определены основные агенты, некорректно выбраны методы исследования, требования к содержанию, структуре, логике, аргументации, оформлению не выполнены) – не зачтено, полное и правильное выполнение заданий контрольной работы в соответствии с требованиями к содержанию, структуре, логике, аргументации, оформлению с возможным небольшим количеством погрешностей (например, плохо выдержанная структура текста, недостаточная аргументация отдельных тезисов) – зачтено

2 Контрольные задания для текущей аттестации

Задания для контрольных работ

Тема 1. Лингвистическая предобработка текста

1. Преобразуйте текст в список токенов, объясните, какие сложности могут возникнуть.
2. Выполните стемминг и лемматизацию, сравните результаты.
3. Объясните различия между POS-теггингом и лемматизацией.

4. Разработайте регулярное выражение для извлечения e-mail адресов.
5. Постройте частотный словарь для коллекции текстов.
6. Объясните роль морфологического анализа в NLP.
7. Проведите синтаксический разбор простого предложения.
8. Напишите функцию для удаления стоп-слов.
9. Объясните, почему нормализация важна при анализе текста.
10. Сравните различные методы токенизации.
11. Какие ошибки могут возникнуть при автоматическом разборе текста?
12. Постройте биграммы из заданного текста.
13. Объясните понятие "лемма" на примере.
14. Сравните NLTK и spaCy для лингвистической предобработки.
15. Оцените роль частеречной разметки в анализе текста.

Тема 2. Машинное обучение в NLP

1. Объясните принцип работы наивного байесовского классификатора.
2. Постройте классификатор для анализа тональности текста.
3. Примените TF-IDF в задаче классификации новостей.
4. Сравните методы Bag-of-Words и Word2Vec.
5. Объясните, как работает логистическая регрессия в NLP.
6. Постройте confusion matrix и интерпретируйте её.
7. Обучите модель SVM на наборе текстов и оцените её качество.
8. Объясните разницу между обучением с учителем и без учителя в NLP.
9. Используйте предобученную модель BERT для классификации.
10. Сравните GloVe и FastText по качеству эмбеддингов.
11. Напишите код для fine-tuning модели DistilBERT.
12. Примените метод кластеризации для группировки текстов.
13. Объясните метрику BLEU и приведите пример ее расчета.
14. Проведите оценку модели по метрике F1-score.
15. Постройте пайплайн для задачи извлечения именованных сущностей.

3 Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации – зачет с оценкой в форме тестирования.

Перед зачетом с оценкой проводится консультация, на которой преподаватель отвечает на вопросы магистрантов.

В результате промежуточного контроля знаний студенты получают оценку по дисциплине.

Тест включает 20 вопросов по всем компетенциям дисциплины, 10 из них вопросы закрытого типа, 10 – открытого типа, все вопросы разного уровня сложности.

Тест оценивается в баллах в соответствии со следующими критериями:

Задания закрытого типа

Базовый уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте -1 балл; ответ отличен от эталонного - 0 баллов.

Повышенный уровень сложности: задание считается выполненным верно, если ответ полностью совпадает с эталоном ответа: каждый символ в ответе стоит на своём месте, лишние символы в ответе отсутствуют - 2 балл; если на любой одной позиции ответа записан не тот символ, который представлен в эталоне ответа - 1 балл; во всех других случаях выставляется 0 баллов

Задания открытого типа

Высокий уровень сложности: магистрант демонстрирует умение применять знания в нестандартной ситуации, решать нетиповые задачи, приводит корректные обоснования и доказательства, ответ полный, в ответе отсутствуют фактические ошибки, изложение связное, структура прозрачная, логика изложения прослеживается - 3 балла; ответ значительно отличается от эталонного, имеются фактические ошибки, искажающие его смысл или ответ сформулирован неверно или не сформулирован - 0 баллов.

Итоговый балл за тест рассчитывается по формуле:

$$F = \frac{100}{K} * \left(\frac{x_1}{k_1} + \frac{x_2}{k_2} + \dots + \frac{x_n}{k_n} \right),$$

где F – итоговое количество баллов за тест,
 K – количество осваиваемых в рамках дисциплины компетенций,
 k_n – максимально возможное количество баллов за вопросы по компетенции,
 x_n – количество баллов, набранное магистрантом, за правильные ответы на вопросы по соответствующей компетенции.

Таблица 3
Показатели, критерии и оценивание компетенций и индикаторов их достижения в процессе промежуточной аттестации

Форма промежуточной аттестации/вид промежуточной аттестации	Коды компетенций	Индикаторы компетенций (в соотв. с Таблицей 1)	Коды ЗУВ (в соотв. с Таблицей 1)	Критерии оценивания	Оценка
Зачет с оценкой/ Тест	ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	3 (ОПК-2) У (ОПК-2) В (ОПК-2)	81-100% правильных ответов	Зачтено, отлично
				61-80% правильных ответов	Зачтено, хорошо
				41-60% правильных ответов	Зачтено, удовлетворительно
				0-40% правильных ответов	Не зачтено, неудовлетворительно

Результаты сдачи промежуточной аттестации по направлениям подготовки уровня магистратуры оцениваются по стобалльной системе оценки в соответствии с Положением о формах, периодичности и порядке организации и проведения текущего контроля успеваемости и промежуточной аттестации обучающихся в АНООВО «ЕУСПб» следующим образом согласно таблице 3а.

Таблица 3а
Система оценки знаний обучающихся

Пятибалльная (стандартная) система	Стобалльная система оценки	Бинарная система оценки
5 (отлично)	100-81	зачтено
4 (хорошо)	80-61	
3 (удовлетворительно)	60-41	
2 (неудовлетворительно)	40 и менее	не зачтено

Результаты промежуточного контроля по дисциплине, выраженные в оценках «зачтено, отлично», «зачтено, хорошо», «зачтено, удовлетворительно», показывают уровень сформированности у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

Результаты промежуточного контроля по дисциплине, выраженные в оценках «не зачленено, неудовлетворительно», показывают несформированность у обучающегося компетенций по дисциплине в соответствии с картами компетенций образовательной программы «Прикладной анализ данных и искусственный интеллект» по направлению подготовки 09.04.03 Прикладная информатика (уровень магистратуры).

4 Задания к промежуточной аттестации

ОПК-2 Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач

Задания закрытого типа

Базовый уровень сложности

1. Прочтайте задание, выберите правильный ответ.

Что такое токенизация в обработке естественного языка?

- 1) Объединение слов в предложения
- 2) Перевод текста с одного языка на другой
- 3) Разделение текста на части, например, слова или предложения
- 4) Удаление всех гласных из текста

Поле для ответа:

2. Прочтайте задание, выберите правильный ответ.

Что такое стемминг?

- 1) Преобразование чисел в текст
- 2) Перевод текста в верхний регистр
- 3) Приведение слова к его корню
- 4) Разделение текста на абзацы

Поле для ответа:

3. Прочтайте задание, выберите правильный ответ.

Какую задачу решает морфологический разбор?

- 1) Определение темпа речи
- 2) Определение части речи и формы слова
- 3) Перевод текста на другой язык
- 4) Анализ эмоциональной окраски текста

Поле для ответа:

4. Прочтайте задание, выберите правильный ответ.

Что такое частотный словарь в NLP?

- 1) Список редких слов
- 2) Таблица с переводами слов
- 3) Перечень слов с указанием частоты их употребления
- 4) Список синонимов слов

Поле для ответа:

5. Прочтите задание, выберите правильный ответ.

Что такое стоп-слова в обработке текста?

Самые важные слова в тексте

Слова, которые подчеркиваются

Часто употребляемые слова, не несущие смысловой нагрузки

Слова, содержащие ошибки

Поле для ответа:

6. Прочтите задание, выберите правильный ответ.

Какую задачу решает лемматизация?

- 1) Добавление новых слов в текст
- 2) Перевод текста в аудио
- 3) Преобразование слова к его нормальной форме
- 4) Удаление всех прилагательных из текста

Поле для ответа:

7. Прочтите задание, выберите правильный ответ.

Какая из перечисленных задач относится к обработке естественного языка?

- 1) Обработка изображений
- 2) Сегментация клиентов
- 3) Синтаксический разбор текста
- 4) Предсказание движения цены акций

Поле для ответа:

8. Прочтите задание, выберите правильный ответ.

Какой подход к представлению текста использует подсчёт количества вхождений слов?

- 1) One-hot encoding
- 2) TF-IDF
- 3) Transformer
- 4) word2vec

Поле для ответа:

9. Прочтите задание, выберите правильные ответы.

Какие задачи входят в область обработки естественного языка (NLP)?

- 1) Распознавание речи
- 2) Кластеризация изображений
- 3) Машинный перевод
- 4) Генерация числовых таблиц

Поле для ответа:

10. Прочтите задание, выберите правильные ответы.

Какие этапы входят в лингвистический анализ текста?

- 1) Морфологический анализ
- 2) Сжатие изображений
- 3) Синтаксический анализ
- 4) Предсказание временного ряда

Поле для ответа:

11. Прочтайте задание, выберите правильные ответы.

Что включает в себя этап предварительной обработки текста?

- 1) Удаление пунктуации
- 2) Объединение видеофайлов
- 3) Приведение слов к начальной форме
- 4) Генерация случайных чисел

Поле для ответа:

12. Прочтайте задание, выберите правильные ответы.

Какие операции выполняются при морфологическом анализе текста?

- 1) Определение частей речи
- 2) Определение синтаксических зависимостей
- 3) Нормализация слов
- 4) Построение деревьев решений

Поле для ответа:

13. Прочтайте задание, выберите правильные ответы.

Какие утверждения верны для синтаксического анализа?

- 1) Выявляет структуру предложения
- 2) Используется для кодирования изображений
- 3) Находит связи между словами
- 4) Работает только с аудиоданными

Поле для ответа:

14. Прочтайте задание, выберите правильные ответы.

Какие характеристики справедливы для модели представления текста Bag of Words?

- 1) Не учитывает порядок слов
- 2) Использует плотные вектора
- 3) Учитывает частотность слов
- 4) Использует контекст слов

Поле для ответа:

15. Прочтайте задание, выберите правильные ответы.

Что верно для модели TF-IDF?

- 1) Учитывает важность слова в документе и корпусе
- 2) Игнорирует частотность слов

- 3) Использует веса, а не бинарные значения
- 4) Основана на последовательности пикселей

Поле для ответа:

16. Прочтайте задание, выберите правильные ответы.

Какие понятия связаны с большими языковыми моделями?

- 1) Токенизация
- 2) Сжатие звуковых сигналов
- 3) Prompt engineering
- 4) Построение гистограмм

Поле для ответа:

Повышенный уровень сложности

17. Прочтайте задание и установите соответствие.

Соотнесите задачи обработки естественного языка (слева) с их примерами (справа).

- A) Классификация текста
- B) Машинный перевод
- B) Извлечение информации
- Г) Распознавание именованных сущностей

- 1) Перевод с английского на русский
- 2) Определение категории текста
- 3) Поиск фактов в тексте
- 4) Выделение имен, локаций, организаций

Поле для ответа:

18. Прочтайте задание и установите соответствие.

Соотнесите этапы лингвистического анализа (слева) с их описаниями (справа).

- A) Морфологический анализ
- Б) Синтаксический анализ
- В) Семантический анализ

- 1) Определение структуры предложения
- 2) Определение частей речи и форм слов
- 3) Анализ смысловых связей между словами

Поле для ответа:

19. Прочтайте задание и установите соответствие.

Соотнесите этапы предварительной обработки текста (слева) с их функциями (справа).

- A) Токенизация

- Б) Лемматизация
- В) Удаление стоп-слов

- 1) Разделение текста на слова или фразы
- 2) Приведение слов к начальной форме
- 3) Исключение часто встречающихся, но малозначимых слов

Поле для ответа:

20. Прочтайте задание и установите соответствие.

Соотнесите модели представления текста (слева) с их характеристиками (справа).

- А) One-hot encoding
- Б) Частотная модель Bag of Words
- В) TF-IDF
- Г) word2vec

- 1) Учитывает частотность и редкость слов в корпусе
- 2) Представляет слово единичным вектором без учета контекста
- 3) Учитывает только количество вхождений слов
- 4) Использует распределённые представления слов

Поле для ответа:

21. Прочтайте задание и установите соответствие.

Соотнесите типы векторных моделей (слева) с их описаниями (справа).

- А) One-hot encoding
- Б) Frequency model
- В) TF-IDF

- 1) Отображает количество вхождений слов в документ
- 2) Отражает важность слова с учетом его встречаемости в корпусе
- 3) Представляет слова в виде векторов с одним значением "1"

Поле для ответа:

22. Прочтайте задание и установите соответствие.

Соотнесите модели представления текста (слева) с их описаниями (справа).

- А) TF-IDF
- Б) N-gram model
- В) word2vec
- Г) Bag of Words

- 1) Учитывает частотность слов и их уникальность в корпусе
- 2) Представляет текст как последовательности из N слов
- 3) Использует плотные векторы, обученные на контексте слов
- 4) Учитывает только количество вхождений слов без контекста

Поле для ответа:

23. Прочтайте задание и установите соответствие.

Соотнесите компоненты архитектуры Transformer (слева) с их функциями (справа).

- A) Механизм внимания
- Б) Многоголовое внимание
- В) Позиционные кодировки

- 1) Позволяет учитывать порядок слов
- 2) Усиливает способность модели обращать внимание на разные части входа
- 3) Вычисляет важность слов друг относительно друга

Поле для ответа:

24. Прочтайте задание и установите соответствие.

Соотнесите понятия, связанные с большими языковыми моделями (слева), с их определениями (справа).

- A) Transfer learning
- Б) Prompt engineering
- В) Токенизация
- Г) Галлюцинации

- 1) Разделение текста на минимальные единицы
- 2) Передача знаний от одной задачи к другой
- 3) Настройка запросов к языковой модели
- 4) Генерация правдоподобного, но вымышленного ответа

Поле для ответа:

25. Прочтайте задание и установите последовательность.

Расположите этапы лингвистического анализа текста в правильном порядке:

- 1) Морфологический анализ
- 2) Синтаксический анализ
- 3) Семантический анализ
- 4) Прагматический анализ

Поле для ответа:

26. Прочтайте задание и установите последовательность.

Расположите этапы морфологического анализа в правильном порядке:

- 1) Определение границ слов
- 2) Определение начальной формы слова
- 3) Определение части речи

4) Определение морфологических признаков (число, падеж и т. д.)

Поле для ответа:

27. Прочтайте задание и установите последовательность.

Расположите шаги работы частотной модели Bag of Words в правильном порядке:

- 1) Составление словаря корпуса
- 2) Подсчёт количества вхождений слов
- 3) Преобразование текста в вектор
- 4) Сбор текстов корпуса

Поле для ответа:

28. Прочтайте задание и установите последовательность.

Расположите шаги построения модели TF-IDF в правильном порядке:

- 1) Расчёт частоты слова в документе
- 2) Расчёт обратной частоты документа
- 3) Умножение TF на IDF
- 4) Формирование вектора документа

Поле для ответа:

29. Прочтайте задание и установите последовательность.

Расположите этапы работы модели N-gram в правильном порядке:

- 1) Разделение текста на токены
- 2) Формирование последовательностей по N токенов
- 3) Подсчёт частот появления N-грамм
- 4) Построение вероятностной модели

Поле для ответа:

30. Прочтайте задание и установите последовательность.

Расположите основные этапы архитектуры Transformer при обработке входного текста:

1. Преобразование слов в эмбеддинги
2. Добавление позиционных кодировок
3. Пропуск через блоки внимания
4. Получение выходного вектора

Поле для ответа:

31. Прочтайте задание и установите последовательность.

Расположите этапы работы модели seq2seq в правильном порядке:

1. Кодирование входной последовательности
2. Передача скрытого вектора
3. Генерация выходной последовательности
4. Обработка выходных данных

Поле для ответа:

32. Прочтайте задание и установите последовательность.

Расположите этапы применения большой языковой модели при генерации ответа в правильном порядке:

1. Ввод запроса (prompt)
2. Токенизация запроса
3. Генерация ответа моделью
4. Декодирование токенов в текст

Поле для ответа:

Задания открытого типа

Высокий уровень сложности

33. Прочтайте текст и напишите обоснованный ответ.

Что такое обработка естественного языка и для чего она применяется?

Поле для ответа:

34. Прочтайте текст и напишите обоснованный ответ.

Что происходит на этапе синтаксического анализа текста?

Поле для ответа:

35. Прочтайте текст и напишите обоснованный ответ.

В чём заключается задача морфологического анализа?

Поле для ответа:

36. Прочтайте текст и напишите обоснованный ответ.

Что такое токенизация и зачем она используется в NLP?

Поле для ответа:

37. Прочтайте текст и напишите обоснованный ответ.

Чем модель Bag of Words отличается от модели word2vec?

Поле для ответа:

38. Прочтите текст и напишите обоснованный ответ.

Для чего используется модель TF-IDF при анализе текста?

Поле для ответа:

39. Прочтите текст и напишите обоснованный ответ.

Что делает архитектура Transformer важной для языкового моделирования?

Поле для ответа:

40. Прочтите текст и напишите обоснованный ответ.

Что такое prompt engineering и зачем он нужен при работе с LLM?

Поле для ответа:

5 Средства оценки индикаторов достижения компетенций

Таблица 4

Средства оценки индикаторов достижения компетенций

Коды компетенций	Индикаторы компетенций (в соот. с Таблицей 1)	Средства оценки (в соот. с Таблицами 5, 7)
ОПК-2	ИД.ОПК-2.1. ИД.ОПК-2.2. ИД.ОПК-2.3. ИД.ОПК-2.4.	Контрольная работа, тест

Таблица 5

Описание средств оценки индикаторов достижения компетенций

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
Контрольная работа	Магистрант в ходе подготовки и выполнения контрольной работы показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Разрабатывать и осуществлять отладку работоспособности оригинальных алгоритмов и программных средств для решения профессиональных задач, используя современные интеллектуальные технологии, интегрировать разработанные программные модули и компоненты и верифицировать выпуски программного продукта, разрабатывать требования и проектировать программное обеспечение для решения профессиональных задач
Тест	Магистрант в ходе подготовки и выполнения тестов показывает наличие практической базы знаний в рамках дисциплины, необходимой для выполнения следующих действий в области профессиональной деятельности: — Разрабатывать и осуществлять отладку работоспособности оригинальных алгоритмов и программных средств для решения профессиональных задач, используя современные интеллектуальные технологии, интегрировать разработанные программные модули и компоненты и верифицировать выпуски программного

Средства оценки (в соот. С Таблицами 5, 7)	Рекомендованный план выполнения работы
	продукта, разрабатывать требования и проектировать программное обеспечение для решения профессиональных задач