

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.В.ДВ.04.01

(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Обработка естественного языка

(наименование дисциплины)

по направлению подготовки

09.03.03 Прикладная информатика

направленность (профиль)

Автоматизация бизнес-процессов и проектирование ИТ-решений

Форма обучения: заочная

Год набора: 2025

Общая трудоемкость: 4 ЗЕ

Распределение часов дисциплины по семестрам

Семестр	10	Итого
Форма контроля	зачет	
Вид занятий		
Лекции	4	4
Лабораторные		
Практические		
Руководство: курсовые работы (проекты) / РГР		
Промежуточная аттестация	0,25	0,25
Контактная работа	4,25	4,25
Самостоятельная работа	136	136
Контроль	3,75	3,75
Итого	144	144

Рабочую программу составил(и)

старший преподаватель института цифровых технологий Тренина Марина Анатольевна

(должность, ученое звание, степень, И.О. Фамилия)

Рецензирование рабочей программы дисциплины:



Отсутствует



Рецензент

(должность, ученое звание, степень, И.О. Фамилия)

Рабочая программа дисциплины составлена на основании ФГОС ВО и учебного плана направления подготовки

09.03.03 Прикладная информатика

Срок действия рабочей программы дисциплины до «31» августа 2031 г.

УТВЕРЖДЕНО

На заседании института цифровых технологий

(протокол заседания № 1 от «05» сентября 2025 г.).

1. Цель и задачи изучения дисциплины (учебного курса)

Целью изучения данного курса является освоение базовых моделей обработки текстов на естественном языке.

2. Место дисциплины (учебного курса) в структуре ОПОП ВО

Дисциплины, учебные курсы, на освоении которых базируется данная дисциплина – «Основы дискретной математики и логики», «Основы программирования», «Объектно-ориентированное программирование», и ряд других базовых дисциплин.

Дисциплины, учебные курсы, для которых необходимы знания, умения, навыки, приобретаемые в результате изучения данной дисциплины – «Выполнение и защита выпускной квалификационной работы».

3. Планируемые результаты обучения

Формируемые и контролируемые компетенции (код и наименование)	Индикаторы достижения компетенций (код и наименование)	Планируемые результаты обучения
ПК-10 Способен осуществлять моделирование решений по реализации программного обеспечения и управлению БД	ПК-10.1 Знает технологии моделирования решений по реализации программного обеспечения и управлению БД	Знать: технологии моделирования по реализации программного обеспечения и управления БД Уметь: выбирать технологии моделирования для реализации программного обеспечения и управлению БД Владеть: навыками моделирования программного обеспечения и управления БД
	ПК-10.2 Умеет использовать знания по моделированию решений реализации программного обеспечения и управления БД	Знать: инструменты и методы управления БД. Уметь: использовать системы управления базами данными для управления БД Владеть: навыками использовать знания по моделированию решений реализации программного обеспечения и управления БД
	ПК-10.3 Владеет навыками моделирования решений по реализации программного обеспечения и управлению БД	Знать: современные инструментальные средства моделирования программного обеспечения Уметь: применять технологии моделирования программного обеспечения Владеть: навыками моделирования решений по реализации программного обеспечения и управлению БД

4. Структура и содержание дисциплины

	Вид учебной работы	Наименование тем занятий (учебной работы)	Семестр	Объем, ч.	Баллы	Интерактив, ч.	Формы текущего контроля (наименование оценочного средства)
Модуль 1. Начальные этапы обработки слов и текста .	Лек 1	Знакомство с технологией NLP	10	2	2		Промежуточный тест
	Сам	Инструменты для работы с NLP	10	15	2		Промежуточный тест
	Сам	Составление словаря: токенизация слов	10	2	2		Промежуточный тест
					4		Отчет по практической работе №1
	Сам	Арифметика слов: векторы TF-IDF	10	2	2		Промежуточный тест
	Сам	Семантический анализ	10	2	2		Промежуточный тест
					4		Отчет по практической работе №2
Модуль 2. Применение нейронных сетей для обработки естественного языка	Лек 2	Умозаключения на основе векторов слов (Word2vec)	10	2	2		Промежуточный тест
	Сам	Умозаключения на основе векторов слов (Word2vec)	10	2	2		Промежуточный тест
	Сам	Сверточные нейронные сети	10	2	2		Промежуточный тест
					4		Отчет по практической работе № 3
	Сам	Нейронные сети с обратной связью: рекуррентные нейронные сети	10	2	2		Промежуточный тест
	Сам	Эффективное сохранение информации с помощью сетей с долгой краткосрочной памятью	10	2	2		Промежуточный тест
	Сам	Модели sequence-to-sequence и механизм внимания	10	2	2		Промежуточный тест
					4		Отчет по практической работе №4
Модуль 3. Реальные задачи обработки естественного языка	Сам	Выделение информации: выделение поименованных сущностей и формирование ответов на вопросы	10	1	2		Промежуточный тест
	Сам	Выделение информации	10	2	2		Промежуточный тест

					4		Отчет по практической работе № 5
	Сам	Диалоговые системы	10	2	2		Промежуточный тест
	Сам	Масштабирование: оптимизация, распараллеливание и обработка по батчам	10	2	2		Промежуточный тест
					5		Отчет по практической работе №6
		Промежуточная аттестация	10	0,25	3		
	Конт	Подготовка к зачету	10	3,75	40		Итоговый тест
Итого:				144	100		

5. Образовательные технологии

В рамках изучения дисциплины предусмотрено использование дистанционных технологий.

6. Методические указания по освоению дисциплины

6.1. Рекомендации по подготовке к тестированию по темам курса

Тесты – это вопросы или задания, предусматривающие конкретный, краткий, четкий ответ на имеющиеся эталоны ответов.

При самостоятельной подготовке к тестированию студенту необходимо:

а) готовясь к тестированию, проработайте информационный материал по дисциплине. Проконсультируйтесь с преподавателем по вопросу выбора учебной литературы;

б) четко выясните все условия тестирования заранее. Вы должны знать, сколько тестов Вам будет предложено, сколько времени отводится на тестирование, какова система оценки результатов и т.д.;

в) приступая к работе с тестами, внимательно и до конца прочтите вопрос и предлагаемые варианты ответов. Выберите правильные (их может быть несколько). На отдельном листке ответов выпишите цифру вопроса и буквы, соответствующие правильным ответам;

г) в процессе решения желательно применять несколько подходов в решении задания. Это позволяет максимально гибко оперировать методами решения, находя каждый раз оптимальный вариант.

д) если Вы встретили чрезвычайно трудный для Вас вопрос, не тратьте много времени на него. Переходите к другим тестам. Вернитесь к трудному вопросу в конце.

е) обязательно оставьте время для проверки ответов, чтобы избежать механических ошибок.

Тестирование - позволяет оценить знание фактического материала, умение логически мыслить, способность к рефлексии и творчески подходить к решению поставленной задачи.

6.2. Рекомендации по выполнению практических заданий

Основные задачи выполняемой работы:

- 1) закрепление полученных ранее теоретических знаний;
- 2) выработка навыков самостоятельной работы;
- 3) выяснение подготовленности студента к будущей лабораторной работе.

Весь процесс написания работы можно условно разделить на следующие этапы:

- а) выбор темы и составление предварительного плана работы;
- б) сбор научной информации, изучение литературы;
- в) анализ составных частей проблемы, изложение темы;
- г) обработка материала в целом.

Подготовку выполнения работы следует начинать с повторения соответствующего раздела учебника, учебных пособий по данной теме. Приступать к выполнению работы без изучения основных положений и понятий науки, не следует, так как в этом случае студент, как правило, плохо ориентируется в материале, не может отграничить смежные вопросы и сосредоточить внимание на основных, первостепенных проблемах рассматриваемой темы.

7. Оценочные средства

7.1. Паспорт оценочных средств

Семестр	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
10	ПК-10	Отчеты по практическим работам 1-6

7.2. Типовые задания или иные материалы, необходимые для текущего контроля

7.2.1. Типовые тестовые материалы

(наименование оценочного средства)

Модуль 1. Начальные этапы обработки слов и текста

Тема 1.1 Знакомство с технологией NLP

1. Выберите задачи НЛП

диалоговые агенты / диалоговые системы

машинный перевод

дедукция и мышление

представление знаний и планирование

2. Назовите две проблемы НЛП, которые "в основном решены"

обнаружение спама

POS-маркировка

языковое моделирование

преобразование текста в речь

3. Какие из этих утверждений о Python верны?

Python использует промежуточный код

Python имеет обширную библиотеку стандартных модулей

язык Python применяется для быстрой разработки приложений

программы на Python транслируются в машинные коды, которые затем исполняются

Тема 1.2 Составление словаря: токенизация слов

4. N-граммы определяются как комбинация N ключевых слов вместе. Сколько биграмм можно сгенерировать из данного предложения: "Analytics Vidhya is a great source for studying data science"

7

8

9

10

5. В языке Python для обработки текста используется библиотека:

NLTK

NLP

SKLEAN

TKN

6. Этот токенизатор языка Python разбивает исходный текст на подстроки, используя регулярное выражение, передаваемое ему в качестве параметра

RegexpTokenizer

WhitespaceTokenizer

BlanklineTokenizer

WordPunctTokenizer

Тема 1.3 Арифметика слов: векторы TF-IDF

7. В корпусе из N документов случайным образом выбирается один документ. Документ содержит в общей сложности T терминов, а термин "данные" встречается K раз.

Каково правильное значение для произведения TF и IDF, если термин "данные" встречается примерно в одной трети всех документов?

$KT * \log(3)$

$K * \log(3) / T$

$T * \log(3) / K$

$\log(3) / KT$

8. Что делает эта формула? $TF(dog, documentA) = 3 / 30 = 0,1$

учитывает длину документа

считает количество строк

считает количество символов в документе

учитывает количество точек в документе

9. Что делает функция `copy.copy`? `vec = COPY.COPY(zero_vector)`

Копирует

Вставляет

создает независимую копию

переиспользует ссылку на местоположение в памяти исходного объекта

Тема 1.4 Семантический анализ

10. Корпуса более похожи между собой при метрике стремящейся к?

Минимуму

Нулю

Максимуму

Бесконечности

11. В основе латентно-семантического анализа лежит

сингулярное разложение

обращение матрицы

понижение размерности

регулярное выражение

12. С помощью какого метода можно производить семантический анализ

для разложения и преобразования векторов TF-IDF и BOW

SVD

SVW

SDD

GVD

Модуль 2. Применение нейронных сетей для обработки естественного языка

Тема 2.1 Умозаключения на основе векторов слов (Word2vec)

13. Если произвести группировку объектов по какому-то признаку, то их можно считать...?

Образом

Системой

Группой

Структурой

14. Что такое t-SNE?

Фоторедактор

Баллистическое упорядочивание объектов

метод визуализации, способный визуализировать многомерные данные

метод визуализации, неспособный визуализировать многомерные данные

15. CBOW где работает лучше?

в семантических задачах
в синтаксических задачах
одинаково во всех задачах
не может работать в этих задачах

Тема 2.2 Сверточные нейронные сети

16. Что является основной моделирующей единицей в персептронах?

Аксон
Дендрит
Нейрон
Синапс

17. Последовательность ситуаций с указанием, к какому классу они относятся, называется?

обучающая последовательность
последовательность слов
последовательность обучения
последовательность текстов

18. Какими понятиями характеризуется способность к обучению?

коэффициентом влияния правила на обучение
степенью значимости решающего правила
надежность получения решающего правила с заданным качеством
качеством полученного решающего правила

Тема 2.3 Нейронные сети с обратной связью: рекуррентные нейронные сети

19. Токены представлены в виде векторов слов, каждый длиной 300 элементов. Сколько у каждого нейрона будет весов?

300
200
5
1000

20. Сколько единиц компонентов смещения прибавляется к входным элементам?

1
0
7
9

21. Какое значение определяется выбранной моделью word2vec и так же может быть любым числом?

Maxlen
Epochs
batch_size
embedding_dims

Тема 2.4 Эффективное сохранение информации с помощью сетей с долгой краткосрочной памятью

22.

$$\begin{aligned}r_t &= \sigma(x_t W_{xr} + h_{t-1} W_{hr} + b_r), \\z_t &= \sigma(x_t W_{xr} + h_{t-1} W_{hr} + b_r), \\ \hat{h}_t &= g(x_t W_{xh} + (r_t \circ h_{t-1}) W_{hh} + b_h), \\h_t &= (1 - z_t) \circ h_{t-1} + z_t \circ \hat{h}_t,\end{aligned}$$

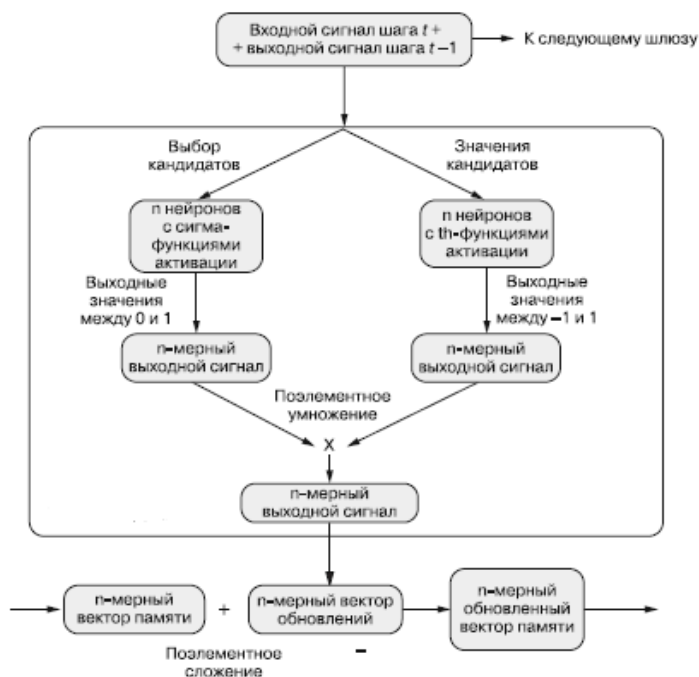
Основные уравнения для ...:
GRU

LSTM

NLP

NPL

23. Что показано на рисунке?



шлюз кандидатов

выходной шлюз

шлюз забывания

обновляющий/выходной шлюз

24. Что делает данный код?

```
>>> def test_len(data, maxlen):
...     total_len = truncated = exact = padded = 0
...     for sample in data:
...         total_len += len(sample)
...         if len(sample) > maxlen:
...             truncated += 1
...         elif len(sample) < maxlen:
...             padded += 1
...         else:
...             exact += 1
...     print('Padded: {}'.format(padded))
...     print('Equal: {}'.format(exact))
...     print('Truncated: {}'.format(truncated))
...     print('Avg length: {}'.format(total_len/len(data)))

>>> dataset = pre_process_data('./aclimdb/train')
>>> vectorized_data = tokenize_and_vectorize(dataset)
>>> test_len(vectorized_data, 400)
```

оптимизирует размер вектора идеи

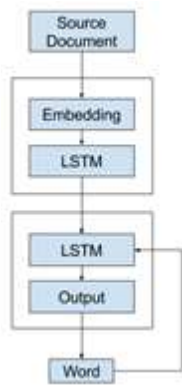
оптимизирует направленность вектора идеи

оптимизирует работу с вектором

оптимизирует вид вектора идеи

Тема 2.5 Модели sequence-to-sequence и механизм внимания

25. Где можно реализовать данную модель?

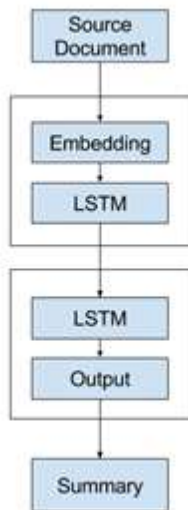


TensorFlow

Keras

Microsoft Cognitive Toolkit

26. Как называется данная модель?



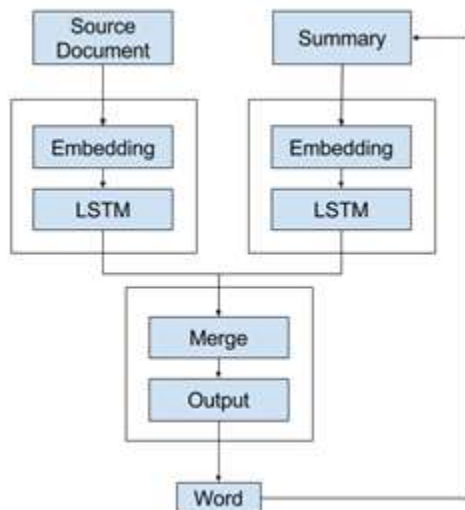
модель с одним выстрелом

рекурсивная модель А

рекурсивная модель В

рекуррентная модель

27. Как называется данная модель?



модель с одним выстрелом

рекурсивная модель А

рекурсивная модель В

рекуррентная модель

Модуль 3. Реальные задачи обработки естественного языка

Тема 3.1 Выделение информации: выделение поименованных сущностей и формирование ответов на вопросы

28. Какое из следующих регулярных выражений можно использовать для определения даты (дат), присутствующих в текстовом объекте: "The next meetup on data science will be held on 2017-09-21, previously it happened on 31/03, 2016"

$\backslash d\{4\}-\backslash d\{2\}-\backslash d\{2\}$

$(19|20)\backslash d\{2\}-(0[1-9]|1[0-2])-[0-2][1-9]$

$(19|20)\backslash d\{2\}-(0[1-9]|1[0-2])-([0-2][1-9]|3[0-1])$

29. Что это за формула $F = \frac{2PR}{P+R}$?

оценка систем выделения кол-во сущностей

оценка систем выделения второстепенных объектов

оценка систем выделения сущностей

оценка систем верно выделенных сущностей

30. Сколько существует признаков выделения именованных сущностей?

3

Тема 3.2 Диалоговые системы

31. Сколько этапов выделяют в процессе генерации текста?

32. Облегченные языки разметки:

Scribe

BBCode

GenCode

Markdown

33. Какие преимущества существуют у чат-ботов на базе NLP:

экономия затрат и времени

могут принимать решения

взаимодействие с людьми

немедленная помощь

Тема 3.3 Масштабирование: оптимизация, распараллеливание и обработка по батчам

34. При создании модели машинного обучения для текстовых данных вы создали матрицу терминов документа из входных данных 100 тыс. документов. Какие из следующих средств можно использовать для уменьшения размеров данных –

Скрытое распределение Дирихле

Нормализация ключевых слов

Скрытая семантическая индексация

Расстояние Левенштейна

35. Даны модели устройств: GRID K520, Tesla M60, Tesla P100, Tesla P6000. К какому виду устройству они относятся?

NLP

GPU

CPU

RAM

36. Какие существуют недостатки генератора языка Python:

генераторы могут выдавать только по два элемента за раз, так что количество содержащихся в нем элементов становится известно только в конце

генераторы могут выдавать только по одному элементу за раз, так что количество содержащихся в нем элементов становится известно только в конце

генераторы можно выполнять только один раз. Они являются одноразовыми и повторному использованию не подлежат

генераторы не возвращают никакой информации о размере массива тренировочных данных

Критерии оценки за пройденный тест по темам модуля:

- 6 баллов выставляется студенту, если он ответил правильно на 50-100 вопросов;
- 0 баллов выставляется студенту, если он ответил правильно на 0-49 вопросов.

7.2.1. Комплект отчетов по практическим работам

(наименование оценочного средства)

Задание 1.

1. Какие основные инструменты и приложения используются при обработке естественного языка?

2. История развития методов обработки естественного языка?

Указания к выполнению работы.

Написать реферат, который содержит ответы на вопросы задания. Реферат должен содержать не менее 10 страниц стандартного отформатированного текста.

Требования к отчету

Отчет по практическому заданию должен содержать:

1. Титульный лист.
2. Содержаний.
3. Основной текст.
4. Список используемой литературы.

Задание 2.

Лемматизировать текст (привести к словарной форме) и приписать леммам частеречные теги.

Указания к выполнению работы.

Для выполнения этого практического задания вы можете воспользоваться грамматическим словарём oDict или разметкой OpenCorpora и др.

На вход подаются предложения вида "токен1 токен2 ... токенN". В предложениях расставлены знаки препинания, среди них могут быть только запятая, точка, вопросительный и восклицательный знаки. Предложения разделены переносом строки.

На выходе: для из входного предложения на выходе получаем последовательность вида: "токен1 {лемма1=тег1} токен2 {лемма2=тег2} ... токенN {леммаN=тегN}". Полученная последовательность не содержит исходных знаков препинания.

При выполнении лемматизации буквы е и ё считать равноправными, а также не учитывать регистр. Частеречные теги должны быть приведены к следующему инвентарю:

существительные (S),

прилагательные (A),

глаголы, в том числе причастия и деепричастия (V),

предлоги (PR),

союзы (CONJ),

сборная категория (ADV), включающая наречия, вводные слова, частицы, междометия.

Любым образом могут быть местоимения (включая наречные и предикативные), числительные, составные предлоги и союзы.

Требования к отчету

Структура отчета должна соответствовать приведенным выше этапам:

- Титульный лист.

- Алгоритм решения задачи.
- Листинг программы.
- Контрольный тест.
- Выводы.

Задание 3.

Автоматически построить рефераты текстовых документов.

Указания к выполнению работы.

На вход подается массив текстов в формате JSON. Примеры текстов представлены в файле дополнительные материалы.json.

На выходе: массив рефератов в формате JSON, порядок которых рефератов соответствует порядку текстов во входных данных.

Максимальный размер каждого из рефератов составляет 300 вместе с пробелами.

Пример входа:

["Первый текст...", "Второй текст..."]

Пример выхода:

["Реферат первого текста...", "Реферат второго текста..."]

Требования к отчету

Структура отчета должна соответствовать приведённым выше этапам:

- Титульный лист.
- Алгоритм решения задачи.
- Листинг программы.
- Выводы.

Задание 4.

Определить тональность текста по шкале от 1 (негативный) до 10 (позитивный).

Указания к выполнению работы.

На вводе: тексты, разделенные переводом строки (\n).

На выводе: тональная оценка от 1 до 10 для каждого текста, разделенные переводом строки (\n).

Для обучения можно сети можно воспользоваться коллекцией текстов из файла texts_train4.txt, а соответствующие им оценки находятся в файле scores_train.txt. В качестве оценки используется евклидово (декартово) расстояние (RMSE).

Пример ввода:

Очень хорошо прописаны характеры. Занусси словно под микроскопом разглядывает человека, препарирует ситуацию, разбирает случай на детали, устраивает качели игр разума. Жанр не мой, но досмотрела — захватывает.

на мой взгляд самая неудачная книга у этого автора...

Пример выхода:

8

5

Требования к отчету

Структура отчета должна соответствовать приведённым выше этапам:

- Титульный лист.
- Алгоритм решения задачи.
- Листинг программы.
- Выводы.

Задание 5.

Выделить в тексте именованные сущности двух типов – организация и персона.

Указания к выполнению работы.

На вход передаются предложения, разделенные переносом строки.

На выходе для каждого входного предложения имеем последовательность вида "индекс1 длина1 тег1 индекс2 длина2 тег2 ... индексi длинаi теги ... индексN длинаN тегN EOL", где индексi – индекс начала токена, длинаi – длина токена, теги -- тег токена. В качестве тегов могут выступать PERSON (персона) и ORG (организация). Если именованная сущность состоит из нескольких токенов, тег приписывается к каждому токenu. Во входных предложениях могут отсутствовать именованные сущности этих типов. Разделитель в выводе в рамках одного предложения – пробельный символ, между выводами для различных предложений – перенос строки.

Для обучения сети воспользуйтесь коллекцией предложений из файла train_sentences.txt, соответствующие им теги можно найти в файле train_nes.txt. Также для вас доступны токены (файл train_tokens.txt) и предложения со вставленными тегами (файл train_sentences_enhanced.txt). Форматы предложенных файлов с обучающей коллекцией предложений, тегами, токенами соответствуют форматам ввода, вывода и формату файла с тегами с тем исключением, что перечисляются все токены, и вместо тегов для токенов указана используемая в предложении словоформа. В качестве метрики качества используется F1 по тегам обоих типов.

Требования к отчету

Структура отчета должна соответствовать приведённым выше этапам:

- Титульный лист.
- Алгоритм решения задачи.
- Листинг программы.
- Выводы.

Задание 6.

Реализуйте генератор юмора по мотивам работы Alessandro Valitutti et al. “Let Everything Turn Well in Your Wife”: Generation of Adult Humor Using Lexical Constraints, 2013.

Указания к выполнению работы.

На входе генератора – новостной заголовок, в котором надо заменить одно слово.

Для выполнения данного задания предлагаем использовать алгоритм:

1. Проведите разбор предложения с помощью библиотеки Stanza. На основе результатов разбора выберете слово-кандидат на замену.
2. Найдите антоним для слова в WordNet (используйте интерфейс библиотеки NLTK).
3. Если антоним не нашелся, то найдите несколько слов, близких по звучанию или рифму с помощью datamuse api.
4. Получите вектора fasttext для начального слова и вариантов замены. Среди этих слов найдите самое далекое по косинусному расстоянию.

Оцените 20 модификаций по шкале от 0 (совсем не смешно) до 3 (очень смешно), приведите среднюю оценку.

Требования к отчету

Структура отчета должна соответствовать приведённым выше этапам:

- Титульный лист.
- Алгоритм решения задачи.
- Листинг программы.
- Выводы.

Критерии оценки:

4 балла – задание выполнено полностью;

в логических рассуждениях и обосновании решения нет пробелов и ошибок;

в решении нет математических ошибок (возможна одна неточность, описка, не являющаяся следствием незнания или непонимания учебного материала).

2-3 балл – задание выполнено полностью, но обоснования шагов решения недостаточны (если умение обосновывать рассуждения не являлось специальным объектом проверки);

допущена одна ошибка или два-три недочета в выкладках, рисунках, чертежах или графиках (если эти виды работы не являлись специальным объектом проверки);

допущены более одной ошибки или более двух-трех недочетов в выкладках, чертежах или графиках, но учащийся владеет обязательными умениями по проверяемой теме.

0-1 баллов - допущены существенные ошибки, показавшие, что студент не владеет обязательными умениями по данной теме в полной мере.

7.2.3. Типовые задания для итогового теста

(наименование оценочного средства)

Задание №1		
Что такое обработка естественного языка?		
Выберите один из 4 вариантов ответа:		
1)	+	Это технология машинного обучения, которая дает компьютерам возможность интерпретировать, манипулировать и понимать человеческий язык
2)	-	Это разработка программ на естественном языке
3)	-	Это технология машинного обучения, которая дает компьютерам возможность понимать естественный язык
4)	-	Это технология машинного обучения, которая дает компьютерам возможность выполнять всё на естественном языке

Задание №2		
На какие представления о языке опирается НЛП?		
Выберите несколько из 4 вариантов ответа:		
1)	+	Формальные модели, или представления
2)	+	Знание языка на всех уровнях
3)	-	Правила правописания
4)	-	Правила орфографии

Задание №3		
Выберите задачи НЛП.		
Выберите несколько из 4 вариантов ответа:		
1)	+	Диалоговые агенты / диалоговые системы
2)	+	Машинный перевод
3)	-	Дедукция и мышление
4)	-	Представление знаний и планирование

Задание №4

Назовите две проблемы НЛП, которые «в основном решены».

Выберите несколько из 4 вариантов ответа:

1)	+	Обнаружение спама
2)	+	POS-маркировка
3)	-	Языковое моделирование
4)	-	Преобразование текста в речь

Задание №5

Что вообще делает НЛП сложным?

Выберите один из 4 вариантов ответа:

1)	+	Неоднозначность человеческого языка на многих уровнях
2)	-	Разнообразие языков
3)	-	Существование различных языковых правил
4)	-	Множество различных правил составления предложения для разных языков

Задание №6

Ключевое слово реализация?

Выберите обозначение ключевого слова «реализация».

Выберите один из 5 вариантов ответа:

1)	-	domains
2)	+	implement
3)	-	constant
4)	-	clauses
5)	-	goal

Задание №7

Ключевое слово цель?

Выберите один из 5 вариантов ответа:

1)	-	domains
2)	-	implement
3)	-	constant
4)	-	clauses
5)	+	goal

Задание №8

Правила вывода

Выберите один из 4 вариантов ответа:

1)	-	не позволяют логически выводить одну информацию из другой
----	---	---

2)	+	позволяют логически выводить одну информацию из другой
3)	-	означают способности восприятия
4)	-	означают механизмы ввода

Задание №9

Текстовый редактор является одним из примеров программных средств, применяемых

Выберите один из 4 вариантов ответа:

1)	-	для осуществления различных регистрирующих функций
2)	+	для приобретения знаний
3)	-	для создания интерфейса
4)	-	для выполнения хода решения задачи

Задание №10

Понятие предметной области обладает следующими свойствами:

Выберите один из 4 вариантов ответа:

1)	-	уникальность, полнота, достоверность, противоречивость
2)	+	уникальность, полнота, достоверность, непротиворечивость
3)	-	уникальность, полнота, недостоверность, непротиворечивость
4)	-	уникальность, полнота, достоверность, наличие омонимии

7.3. Оценочные средства для промежуточной аттестации по итогам освоения дисциплины

7.3.1. Вопросы к промежуточной аттестации

Семестр 5

№ п/п	Вопросы
1	Какие основные инструменты и приложения используются при обработке естественного языка?
2	История развития методов обработки естественного языка?
3	Что такое конечный автомат? Как он используется в обработке естественного языка?
4	Что такое регулярное выражение?
5	Каковы основные подходы к морфологическому анализу?
6	Какова основная методика отбора данных для морфологического анализа?
7	Каковы основные инструменты морфологического анализа?
8	Структурные особенности текстов на естественном языке?
9	Базовая обработка текста и предварительная обработка расстояния редактирования?
10	Языковые модели N-грамм; недоумение; методы сглаживания?
11	Использование языковых моделей?
12	Что такое TF-IDF?
13	Какие показатели можно использовать для анализа документов?
14	В суть латентно-семантического анализа?
15	Каковы методы латентно-семантического анализа?

16	В чем суть алгоритма LSA?
17	Каковы инструменты метода LSA?
18	В суть метода линейного дискриминантного анализа (LDA)?
19	В суть метода латентное размещение Дирихле (LDA)?
20	Виды расстояний редактирования?
21	Умозаключения на основе векторов слов (Word2vec)?
22	В чем суть алгоритма Skip-gram?
23	Концепция модели Skip-gram?
24	Реализация модели Skip-gram?
25	Методы аппроксимации функции потерь?
26	В чем суть алгоритма CBOW?
27	Использование сверточных нейронных сетей в NLP?
28	Операция растянутой субдискретизации в задаче классификации предложений?
29	Использование рекуррентных нейронных сетей в NLP?
30	Обратное распространение?
31	Использование сетей с долгой краткосрочной памятью в NLP?
32	Модели sequence-to-sequence и механизм внимания?
33	Как можно выделить информацию: выделение поименованных сущностей и формирование ответов на вопросы?
34	Каковы методы создания диалоговой системы?

7.3.2. Критерии и нормы оценки

Семестр	Форма проведения промежуточной аттестации	Критерии и нормы оценки	
		«не зачтено»	рейтинговый балл 55-100
10	Зачет	«не зачтено»	рейтинговый балл 0-54
		«не зачтено»	рейтинговый балл 0-54

8. Учебно-методическое и информационное обеспечение дисциплины

8.1. Обязательная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1.	Павлов С. Н.	Колмогорова, С. С. Основы искусственного интеллекта : учебное пособие для студентов / С. С. Колмогорова. — Санкт-Петербург : СПбГЛТУ, 2022. — 108 с. — ISBN 978-5-9239-1308-8. — Текст : электронный // Лань : электронно-библиотечная система. — Режим доступа: для авториз. пользователей.	учебное пособие	2022	ЭБС «Лань»
2.	Павлов С. Н.	Копырин, А. С. Программирование на Python : учебное пособие / А. С. Копырин, Т. Л. Салова. — Сочи : СГУ, 2018. — 48 с. — Текст : электронный // Лань : электронно-библиотечная система. — Режим доступа: для авториз. пользователей.	учебное пособие	2018	ЭБС «Лань»

8.2. Дополнительная литература

№ п/п	Авторы, составители	Заглавие (заголовок)	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Год издания	Количество в научной библиотеке / Наименование ЭБС
1.	Сысоев Д. В.	Введение в теорию искусственного интеллекта [Электронный ресурс] : учеб. пособие / Д. В. Сысоев, О. В. Курипта, Д. К. Проскурин. - Воронеж : ВГАСУ : ЭБС АСВ, 2014. - 171 с. - ISBN 978-5-89040-498-5..	Учебное пособие	2015	ЭБС «IPRbooks»

8.3. Перечень профессиональных баз данных и информационных справочных систем

№ пп	Наименование	Ссылка
1	Springer Nature (Полнотекстовая коллекция журналов)	https://www.springernature.com/gp/products
2	Springer eBooks (Полнотекстовая коллекция электронных книг издательства Springer Nature)	https://link.springer.com/
3	«Кодекс»	https://kodeks.ru/
4	Техэксперт	https://cntd.ru/

8.4. Перечень программного обеспечения

№ п/п	Наименование ПО	Реквизиты договора (дата, номер, срок действия)
1	WinPro 10 RUS Upgrd OLP NL Acdmc	Договор № 757 от 04.07.2018, срок действия - бессрочно; Контракт № 1653 от 14.12.2018, срок действия – бессрочно
2	Office Stdandard 2013 Russian OLP NL AcademicEdition	Контракт № 690 от 19.05.2015, срок действия - бессрочно
3	Digital 2.0.0	Лицензия: GNU General Public License (GPL) (свободное и бесплатное ПО)
4	GParted (GNOME Partition Editor)	Лицензия: GNU General Public License (GPL) (свободное и бесплатное ПО)
5	VirtualBox 7.1.6	Лицензия: PUEL (для личного и образовательного использования)
6	Ubuntu 24.04.1 LTS	Лицензия: GNU General Public License (GPL) (свободное и бесплатное ПО)

8.5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий, помещений для самостоятельной работы обучающихся (номер аудитории)	Перечень основного оборудования
1	Помещение для самостоятельной работы обучающихся (УЛК-406)	Столы компьютерные, стулья, микрокомпьютеры raspberry pi 32 bit