



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение
высшего образования

«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ФГБОУ ВО «ИГУ»

Факультет бизнес-коммуникаций и информатики

Кафедра естественнонаучных дисциплин

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

по дисциплине Б1.В.09 Нейронные сети и обработка текста

направление подготовки 09.03.03 Прикладная информатика

направленность (профиль) Прикладная информатика (разработка программного обеспечения)

Одобрено
УМК факультета бизнес-коммуникаций
и информатики

Разработан в соответствии с ФГОС ВО

с учетом требований проф. стандарта

Председатель УМК

В.К. Карнаухова

ФИО, должность, ученая степень, звание

подпись, печать

Разработчики:

(подпись)

доцент

(занимаемая должность)

О.И. Бернгардт

(инициалы, фамилия)

Цель фонда оценочных средств. Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Б1.В.09 Нейронные сети и обработка текста». Перечень видов оценочных средств соответствует рабочей программе дисциплины.

Фонд оценочных средств включает контрольные материалы для проведения текущего контроля (в следующих формах: тест, устный опрос, практическое задание) и промежуточной аттестации в форме вопросов и заданий к зачету с оценкой.

Структура и содержание заданий – задания разработаны в соответствии с рабочей программой дисциплины «Б1.В.09 Нейронные сети и обработка текста».

1. Паспорт фонда оценочных средств

Компетенция	Индикаторы компетенций	Результаты обучения
ПК-1 Способность воспринимать математические, естественнонаучные, социально-экономические, инженерные знания, самостоятельно приобретать, развивать и применять их для формализации и решения задач разработки и модификации программного обеспечения	ПК-1.1	Знать специальные разделы математики, естественнонаучных и социально-экономических дисциплин необходимые для логического осмысления и обработки информации в профессиональной деятельности
	ПК-1.2	Уметь применять системный подход и математические методы в формализации решения прикладных задач разработки и модификации программных приложений
	ПК-1.3	Владеть навыками использования математических, естественнонаучных, социально-экономических, инженерных знаний в разработке компьютерных моделей и прототипов программного обеспечения для решения проектных и научно-исследовательских задач

Компетенция	Индикаторы компетенций	Результаты обучения
ПК-2 Способность разрабатывать, внедрять и адаптировать прикладное программное обеспечение	ПК-2.1	Знать технологии разработки программного обеспечения: методы, средства, процедуры и инструменты
	ПК-2.2	Уметь внедрять и адаптировать прикладное программное обеспечение
	ПК-2.3	Владеть навыками решения задач реализации и модификации ПО: планирования и оценки проекта по разработке ПО; анализа системных и программных требований; проектирования алгоритмов, структур данных и программных структур; кодирования с использованием различных языков программирования и разметки; рефакторинга ПО; тестирования и отладки программного кода; сопровождения

2. Показатели и критерии оценивания компетенций, шкалы оценивания

2.1. Показатели и критерии оценивания компетенций

№ п/п	Раздел, тема	Код индикатора компетенции	Наименование ОС	
			ТК	ПА
1	Введение	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО	Тест
2	Решающие деревья	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
3	Метод главных компонент	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
4	Элементарный нейрон	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
5	Основные понятия нейронных сетей	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
6	Метод градиентного спуска и его модификации	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
7	Токенизация текстов. Регулярные выражения	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
8	Векторные представления и метод мешка слов	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест

№ п/п	Раздел, тема	Код индикатора компетенции	Наименование ОС	
			ТК	ПА
9	Оптимальные векторные представления	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
10	Тематическое моделирование	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
11	Матричный и авторегрессионный подходы к морфологическому анализу текстов	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
12	Синтаксический анализ текстов и распознавание именованных сущностей	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
13	Семантический анализ текстов, обобщенные модели и трансфер знаний	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
14	Понимание и аннотирование текстов	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
15	Диалоговые системы и чат-боты	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
16	Вопросно-ответные системы	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
17	Сбор фактов	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест
18	Распознавание речи	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3	Тест, УО, Пз	Тест

2.2. Критерии оценивания результатов обучения для текущего контроля успеваемости и промежуточной аттестации

Оценочное средство	Критерии оценивания	Шкала оценивания
Тест	Студентом даны правильные ответы на 91-100% заданий	Отлично
	Студентом даны правильные ответы на 81-90% заданий	Хорошо
	Студентом даны правильные ответы на 71-80% заданий	Удовлетворительно
	Студентом даны правильные ответы менее чем на 70% заданий	Неудовлетворительно

Оценочное средство	Критерии оценивания	Шкала оценивания
Устный опрос	<p>Ответ соответствует поставленной теме и содержит ответы на поставленные задачи, имеет четкую структуру, логически сопоставляемую с поставленными вопросами. Ответ демонстрирует способность анализировать и обобщать информацию, опираясь на знания, полученные в ходе изучения темы, а также демонстрировать самостоятельность автора в решении поставленных задач. Ответ содержит качественную речь и аргументацию, которая убедительно подтверждает выводы и ответы на поставленные вопросы</p>	Отлично
	<p>Ответ должен быть направлен на ответ на поставленные вопросы и соответствовать поставленной теме, иметь логическую цепочку рассуждений и четко демонстрировать связь между поставленными вопросами. Ответ выдержан в четкой форме, быть грамотно и без ошибок озвучен, выделены ключевые термины. Ответ должен демонстрировать способность анализировать и критически оценивать информацию, выбирая ключевые аспекты и выделяя главные выводы</p>	Хорошо
	<p>Ответ должен соответствовать поставленной теме и содержать ответы на поставленные вопросы, должен содержать существенную информацию, ясно передавать ответы и идеи. Ответ должен содержать достаточное количество аргументов и примеров, связанных с темой работы и позволяющих изложить свою точку зрения. Ответ должен быть грамотно сформулирован</p>	Удовлетворительно
	<p>Ответ не соответствует поставленной теме или не содержит ответов на поставленные задачи, содержит недостаточно аргументации и примеров, которые подтверждают высказанные в ответе идеи и выводы. Ответ не соответствует логической цепочке рассуждений и не выполняет требования логической последовательности высказывания, затрудняющей понимание ответа. Ответ содержит грубые ошибки, что затрудняет понимание высказывания</p>	Неудовлетворительно

Оценочное средство	Критерии оценивания	Шкала оценивания
Практическое задание	Задание выполнено верно. Выбран оптимальный путь решения. Присутствует развернутое описание алгоритма решения	Отлично
	Задание выполнено верно. Допущены негрубые логические ошибки при описании алгоритма решения. Отсутствуют пояснения к решению задания	Хорошо
	Ход решения задания верный, но допущены ошибки приведшие к неправильному ответу	Удовлетворительно
	В работе получен неверный ответ, связанный с грубыми ошибками допущенными в ходе решения, либо решение отсутствует полностью	Неудовлетворительно

2.3. Оценочные средства для текущего контроля (примеры)

2.3.1. Материалы для компьютерного тестирования обучающихся

Общие критерии оценивания

Процент правильных ответов	Оценка
91% – 100%	5 (отлично)
81% – 90%	4 (хорошо)
71% – 80%	3 (удовлетворительно)
Менее 70%	2 (неудовлетворительно)

Соответствие вопросов теста индикаторам формируемых и оцениваемых компетенций

№ вопроса в тесте	Код индикатора компетенции
1	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
2	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
3	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
4	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
5	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
6	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
7	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
8	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
9	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
10	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
11	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
12	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
13	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
14	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
15	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3

№ вопроса в тесте	Код индикатора компетенции
16	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
17	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
18	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3

Ключ ответов

№ вопроса в тесте	Номер ответа (или ответ, или соответствие)
1	c
2	b
3	c
4	a, b
5	d
6	b
7	a
8	b
9	a
10	b
11	a
12	b
13	d
14	c
15	b
16	b
17	a, b, c, d, e, f
18	a

Перечень тестовых вопросов

№ 1. Задание с единичным выбором. Выберите один правильный ответ.

Что обычно важнее всего для работы FAQ чат-бота?

- a. Большой объем обучающих данных для обучения
- b. Языковая модель для понимания вопроса
- c. База вопрос-ответ
- d. Языковая модель для генерации ответа

№ 2. Задание с единичным выбором. Выберите один правильный ответ.

Какую из этих задач можно успешно решать с помощью решающих деревьев?

- a. задачи кластеризации
- b. задачи классификации

№ 3. Задание с единичным выбором. Выберите один правильный ответ.

Для чего нужно преобразовывать токен (или слово) в его векторное представление?

- a. Чтобы уменьшить его размер
- b. Чтобы увеличить его размер

с. Чтобы работать с числами, а не с символами

№ 4. Задание с множественным выбором. Выберите 2 правильных ответа.

Я знаю такие обобщенные языковые модели

- a. BERT
- b. GPT3
- c. VGG19

№ 5. Задание с единичным выбором. Выберите один правильный ответ.

На вход нейрона подается два числа, сколько чисел будет на выходе нейрона

- a. 4
- b. 3
- c. 2
- d. 1

№ 6. Задание с единичным выбором. Выберите один правильный ответ.

На какие возможные токены можно разбить русскоязычный текст?

- a. Слова
- b. Все из перечисленных
- c. N-граммы

№ 7. Задание с единичным выбором. Выберите один правильный ответ.

Можно-ли использовать регулярные выражения для сбора фактов в тексте?

- a. Да, для извлечения данных по маске, например телефонов
- b. Нет, факты можно извлекать только с помощью нейронных сетей, потому-что это очень сложно

№ 8. Задание с единичным выбором. Выберите один правильный ответ.

Для чего используется метод градиентного спуска в машинном обучении?

- a. Для поиска оптимального размера обучающей выборки
- b. Для поиска минимума функции потерь нейронной сети

№ 9. Задание с единичным выбором. Выберите один правильный ответ.

Какую задачу решают, когда обучают нейронную сеть?

- a. минимизацию функции потерь
- b. максимизацию функции активации

№ 10. Задание с единичным выбором. Выберите один правильный ответ.

В каком смысле оптимальные векторные представления слов оптимальны?

- a. Они имеют минимальный размер
- b. Они позволяют наиболее точно решить нужную нам задачу

№ 11. Задание с единичным выбором. Выберите один правильный ответ.

Для решения какой задачи лучше применять метод главных компонент?

- a. Уменьшения размерности данных
- b. Увеличения размерности данных

№ 12. Задание с единичным выбором. Выберите один правильный ответ.

Для решения какой задачи можно использовать тематическое моделирование?

- a. Для автоматической генерации текста на заданную тематику

б. Для определения того, на какие группы(темы) можно разделить множество имеющихся у нас текстов

№ 13. Задание с единственным выбором. Выберите один правильный ответ.

В каких задачах часто используется синтаксический анализ русскоязычного текста?

- а. Все перечисленное
- б. Векторизация токенов
- с. Токенизация текста
- д. Распознавание именованных сущностей

№ 14. Задание с единственным выбором. Выберите один правильный ответ.

Для чего обычно используют преобразования Фурье и спектрограммы при распознавании речи?

- а. Для преобразования звука в число и обработки случайным решающим лесом
- б. Для преобразования звука в вектор и его обработки линейной регрессией
- с. Для преобразования звука в матрицу и ее обработки сверточными сетями

№ 15. Задание с единственным выбором. Выберите один правильный ответ.

Отметьте верное о генеративном и экстрактивном аннотировании текстов.

- а. При экстрактивном извлекаются части из старого текста, и новых слов не используется.
- б. Верно все
- с. При генеративном создается новый текст на основе понимания старого и возможно использование новых слов

№ 16. Задание с единственным выбором. Выберите один правильный ответ.

Что относится к морфологическому анализу текста?

- а. Генерация краткого содержания текста
- б. Определение части речи неизвестного слова
- с. Определение эмоционального окраса (негативный, позитивный, нейтральный) текста

№ 17. Задание с множественным выбором. Выберите 6 правильных ответов.

Для решения каких задач сегодня используют нейронные сети?

- а. создание изображений по текстовому описанию
- б. автоматический перевод
- с. определение эмоционального окраса текста (позитивный, негативный, нейтральный)
- д. автоматическое управление автомобилем
- е. распознавание изображений
- ф. классификация изображений

№ 18. Задание с единственным выбором. Выберите один правильный ответ.

Какими могут быть вопросно-ответные системы?

- а. Все перечисленное
- б. Общие - для ответов на любые вопросы
- с. Узкоспециализированные - для ответов на вопросы в конкретной области

2.3.2. Практические задания для оценки компетенции «ПК-1.1»

№ 1. Решающие деревья.

На основе заданного размеченного датасета построить и обучить решающее дерево (python, библиотека sklearn, DecisionTreeClassifier)

№ 2. Метод главных компонент.

Найти оптимальное количество главных компонент для заданного многомерного датасета

№ 3. Элементарный нейрон.

На основе заданного размеченного датасета обучить нейронную сеть из одного нейрона

(python, библиотека sklearn, класс MLPClassifier)

№ 4. Основные понятия нейронных сетей.

На основе заданного датасета обучить полносвязную классифицирующую нейронную сеть с одним скрытым слоем. (библиотека sklearn, класс MLPClassifier)

№ 5. Метод градиентного спуска и его модификации.

На основе заданного размеченного датасета обучить нейронную сеть в течение 10 эпох. Выбрать модификацию метода градиентного спуска (оптимизаторы ADAM, RMSProp, SGD), обучающий нейронную сеть с наилучшей точностью за 10 эпох.

№ 6. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений разбить заданный текст на слова.

№ 7. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений выбрать из текста пары 'наречие+глагол'

№ 8. Векторные представления и метод мешка слов.

Создать векторное представление для слов в заданном тексте. В качестве векторного представления использовать количество употреблений этого слова в тексте.

№ 9. Оптимальные векторные представления.

При помощи заданного текста, библиотеки gensim и класса word2vec обучить и найти оптимальные векторные представления для слов текста.

№ 10. Тематическое моделирование.

Построить матрицу числа появлений различных слов в различных текстах.

№ 11. Матричный и авторегрессионный подходы к морфологическому анализу текстов.

На основе размеченного набора слов и частей их речи и библиотеки Tensorflow:

разделить слова на буквы, создать датасет (последовательность букв -> часть речи)

Создать нейронную сеть, которая каждую букву закодирует оптимальным эмбедингом размера 8, и на основе сверточной сети обучится классифицировать тексты (пример разбирается в классе)

№ 12. Синтаксический анализ текстов и распознавание именованных сущностей.

С помощью библиотеки spacy выделить в тексте именованные сущности в заданном тексте (пример разбирается в классе)

№ 13. Семантический анализ текстов, обобщенные модели и трансфер знаний.

Получить векторное представление слова после прохождения модели BERT (пример разобрать в классе)

№ 14. Понимание и аннотирование текстов.

Извлечь из текста наиболее важные предложения методом экстрактивного аннотирования текста с использованием TextRank (пример разбирается в классе)

№ 15. Диалоговые системы и чат-боты.

На основе фреймворка DeepPavlov создать чат-бота с 10 вариантами ответов (пример разбирается в классе)

№ 16. Вопросно-ответные системы.

Реализовать простейшую вопросно-ответную систему

№ 17. Сбор фактов.

Написать программу для извлечения сотовых телефонных номеров из заданного текста.

№ 18. Распознавание речи.

Определить, какая звуковая команда распознается нейронной сетью наименее точно в алгоритме распознавания команд, разбираемом в классе.

2.3.3. Практические задания для оценки компетенции «ПК-1.2»

№ 19. Решающие деревья.

На основе заданного размеченного датасета построить и обучить решающее дерево (python, библиотека sklearn, DecisionTreeClassifier)

№ 20. Метод главных компонент.

Найти оптимальное количество главных компонент для заданного многомерного датасета

№ 21. Элементарный нейрон.

На основе заданного размеченного датасета обучить нейронную сеть из одного нейрона

(python, библиотека sklearn, класс MLPClassifier)

№ 22. Основные понятия нейронных сетей.

На основе заданного датасета обучить полносвязную классифицирующую нейронную сеть с одним скрытым слоем. (библиотека sklearn, класс MLPClassifier)

№ 23. Метод градиентного спуска и его модификации.

На основе заданного размеченного датасета обучить нейронную сеть в течение 10 эпох. Выбрать модификацию метода градиентного спуска (оптимизаторы ADAM, RMSProp, SGD), обучающий нейронную сеть с наилучшей точностью за 10 эпох.

№ 24. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений разбить заданный текст на слова.

№ 25. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений выбрать из текста пары 'наречие+глагол'

№ 26. Векторные представления и метод мешка слов.

Создать векторное представление для слов в заданном тексте. В качестве векторного представления использовать количество употреблений этого слова в тексте.

№ 27. Оптимальные векторные представления.

При помощи заданного текста, библиотеки gensim и класса word2vec обучить и найти оптимальные векторные представления для слов текста.

№ 28. Тематическое моделирование.

Построить матрицу числа появлений различных слов в различных текстах.

№ 29. Матричный и авторегрессионный подходы к морфологическому анализу текстов.

На основе размеченного набора слов и частей их речи и библиотеки Tensorflow:
разделить слова на буквы, создать датасет (последовательность букв -> часть речи)

Создать нейронную сеть, которая каждую букву закодирует оптимальным эмбедингом размера 8, и на основе сверточной сети обучится классифицировать тексты (пример разбирается в классе)

№ 30. Синтаксический анализ текстов и распознавание именованных сущностей.

С помощью библиотеки spacy выделить в тексте именованные сущности в заданном тексте (пример разбирается в классе)

№ 31. Семантический анализ текстов, обобщенные модели и трансфер знаний.

Получить векторное представление слова после прохождения модели BERT (пример разобрать в классе)

№ 32. Понимание и аннотирование текстов.

Извлечь из текста наиболее важные предложения методом экстрактивного аннотирования текста с использованием TextRank (пример разбирается в классе)

№ 33. Диалоговые системы и чат-боты.

На основе фреймворка DeepPavlov создать чат-бота с 10 вариантами ответов (пример разбирается в классе)

№ 34. Вопросно-ответные системы.

Реализовать простейшую вопросно-ответную систему

№ 35. Сбор фактов.

Написать программу для извлечения сотовых телефонных номеров из заданного текста.

№ 36. Распознавание речи.

Определить, какая звуковая команда распознается нейронной сетью наименее точно в алгоритме распознавания команд, разбираемом в классе.

2.3.4. Практические задания для оценки компетенции «ПК-1.3»

№ 37. Решающие деревья.

На основе заданного размеченного датасета построить и обучить решающее дерево (python, библиотека sklearn, DecisionTreeClassifier)

№ 38. Метод главных компонент.

Найти оптимальное количество главных компонент для заданного многомерного датасета

№ 39. Элементарный нейрон.

На основе заданного размеченного датасета обучить нейронную сеть из одного нейрона

(python, библиотека sklearn, класс MLPClassifier)

№ 40. Основные понятия нейронных сетей.

На основе заданного датасета обучить полносвязную классифицирующую нейронную сеть с одним скрытым слоем. (библиотека sklearn, класс MLPClassifier)

№ 41. Метод градиентного спуска и его модификации.

На основе заданного размеченного датасета обучить нейронную сеть в течение 10 эпох. Выбрать модификацию метода градиентного спуска (оптимизаторы ADAM, RMSProp, SGD), обучающий нейронную сеть с наилучшей точностью за 10 эпох.

№ 42. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений разбить заданный текст на слова.

№ 43. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений выбрать из текста пары 'наречие+глагол'

№ 44. Векторные представления и метод мешка слов.

Создать векторное представление для слов в заданном тексте. В качестве векторного представления использовать количество употреблений этого слова в тексте.

№ 45. Оптимальные векторные представления.

При помощи заданного текста, библиотеки gensim и класса word2vec обучить и найти оптимальные векторные представления для слов текста.

№ 46. Тематическое моделирование.

Построить матрицу числа появлений различных слов в различных текстах.

№ 47. Матричный и авторегрессионный подходы к морфологическому анализу текстов.

На основе размеченного набора слов и частей их речи и библиотеки Tensorflow:

разделить слова на буквы, создать датасет (последовательность букв -> часть речи)

Создать нейронную сеть, которая каждую букву закодирует оптимальным эмбедингом размера 8, и на основе сверточной сети обучится классифицировать тексты (пример разбирается в классе)

№ 48. Синтаксический анализ текстов и распознавание именованных сущностей.

С помощью библиотеки spacy выделить в тексте именованные сущности в заданном тексте (пример разбирается в классе)

№ 49. Семантический анализ текстов, обобщенные модели и трансфер знаний.

Получить векторное представление слова после прохождения модели BERT (пример разобрать в классе)

№ 50. Понимание и аннотирование текстов.

Извлечь из текста наиболее важные предложения методом экстрактивного аннотирования текста с использованием TextRank (пример разбирается в классе)

№ 51. Диалоговые системы и чат-боты.

На основе фреймворка DeepPavlov создать чат-бота с 10 вариантами ответов (пример разбирается в классе)

№ 52. Вопросно-ответные системы.

Реализовать простейшую вопросно-ответную систему

№ 53. Сбор фактов.

Написать программу для извлечения сотовых телефонных номеров из заданного текста.

№ 54. Распознавание речи.

Определить, какая звуковая команда распознается нейронной сетью наименее точно в алгоритме распознавания команд, разбираемом в классе.

2.3.5. Практические задания для оценки компетенции «ПК-2.1»

№ 55. Решающие деревья.

На основе заданного размеченного датасета построить и обучить решающее дерево (python, библиотека sklearn, DecisionTreeClassifier)

№ 56. Метод главных компонент.

Найти оптимальное количество главных компонент для заданного многомерного датасета

№ 57. Элементарный нейрон.

На основе заданного размеченного датасета обучить нейронную сеть из одного нейрона

(python, библиотека sklearn, класс MLPClassifier)

№ 58. Основные понятия нейронных сетей.

На основе заданного датасета обучить полносвязную классифицирующую нейронную сеть с одним скрытым слоем. (библиотека sklearn, класс MLPClassifier)

№ 59. Метод градиентного спуска и его модификации.

На основе заданного размеченного датасета обучить нейронную сеть в течение 10 эпох. Выбрать модификацию метода градиентного спуска (оптимизаторы ADAM, RMSProp, SGD), обучающий нейронную сеть с наилучшей точностью за 10 эпох.

№ 60. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений разбить заданный текст на слова.

№ 61. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений выбрать из текста пары 'наречие+глагол'

№ 62. Векторные представления и метод мешка слов.

Создать векторное представление для слов в заданном тексте. В качестве векторного представления использовать количество употреблений этого слова в тексте.

№ 63. Оптимальные векторные представления.

При помощи заданного текста, библиотеки gensim и класса word2vec обучить и найти оптимальные векторные представления для слов текста.

№ 64. Тематическое моделирование.

Построить матрицу числа появлений различных слов в различных текстах.

№ 65. Матричный и авторегрессионный подходы к морфологическому анализу текстов.

На основе размеченного набора слов и частей их речи и библиотеки Tensorflow:

разделить слова на буквы, создать датасет (последовательность букв -> часть речи)

Создать нейронную сеть, которая каждую букву закодирует оптимальным эмбедингом размера 8, и на основе сверточной сети обучится классифицировать тексты (пример разбирается в классе)

№ 66. Синтаксический анализ текстов и распознавание именованных сущностей.

С помощью библиотеки spacy выделить в тексте именованные сущности в заданном тексте (пример разбирается в классе)

№ 67. Семантический анализ текстов, обобщенные модели и трансфер знаний.

Получить векторное представление слова после прохождения модели BERT (пример

разобрать в классе)

№ 68. Понимание и аннотирование текстов.

Извлечь из текста наиболее важные предложения методом экстрактивного аннотирования текста с использованием TextRank (пример разбирается в классе)

№ 69. Диалоговые системы и чат-боты.

На основе фреймворка DeepPavlov создать чат-бота с 10 вариантами ответов (пример разбирается в классе)

№ 70. Вопросно-ответные системы.

Реализовать простейшую вопросно-ответную систему

№ 71. Сбор фактов.

Написать программу для извлечения сотовых телефонных номеров из заданного текста.

№ 72. Распознавание речи.

Определить, какая звуковая команда распознается нейронной сетью наименее точно в алгоритме распознавания команд, разбираемом в классе.

2.3.6. Практические задания для оценки компетенции «ПК-2.2»

№ 73. Решающие деревья.

На основе заданного размеченного датасета построить и обучить решающее дерево (python, библиотека sklearn, DecisionTreeClassifier)

№ 74. Метод главных компонент.

Найти оптимальное количество главных компонент для заданного многомерного датасета

№ 75. Элементарный нейрон.

На основе заданного размеченного датасета обучить нейронную сеть из одного нейрона

(python, библиотека sklearn, класс MLPClassifier)

№ 76. Основные понятия нейронных сетей.

На основе заданного датасета обучить полносвязную классифицирующую нейронную сеть с одним скрытым слоем. (библиотека sklearn, класс MLPClassifier)

№ 77. Метод градиентного спуска и его модификации.

На основе заданного размеченного датасета обучить нейронную сеть в течение 10 эпох. Выбрать модификацию метода градиентного спуска (оптимизаторы ADAM, RMSProp, SGD), обучающий нейронную сеть с наилучшей точностью за 10 эпох.

№ 78. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений разбить заданный текст на слова.

№ 79. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений выбрать из текста пары 'наречие+глагол'

№ 80. Векторные представления и метод мешка слов.

Создать векторное представление для слов в заданном тексте. В качестве векторного представления использовать количество употреблений этого слова в тексте.

№ 81. Оптимальные векторные представления.

При помощи заданного текста, библиотеки gensim и класса word2vec обучить и найти

оптимальные векторные представления для слов текста.

№ 82. Тематическое моделирование.

Построить матрицу числа появлений различных слов в различных текстах.

№ 83. Матричный и авторегрессионный подходы к морфологическому анализу текстов.

На основе размеченного набора слов и частей их речи и библиотеки Tensorflow:

разделить слова на буквы, создать датасет (последовательность букв -> часть речи)

Создать нейронную сеть, которая каждую букву закодирует оптимальным эмбедингом размера 8, и на основе сверточной сети обучится классифицировать тексты (пример разбирается в классе)

№ 84. Синтаксический анализ текстов и распознавание именованных сущностей.

С помощью библиотеки spacy выделить в тексте именованные сущности в заданном тексте (пример разбирается в классе)

№ 85. Семантический анализ текстов, обобщенные модели и трансфер знаний.

Получить векторное представление слова после прохождения модели BERT (пример разобрать в классе)

№ 86. Понимание и аннотирование текстов.

Извлечь из текста наиболее важные предложения методом экстрактивного аннотирования текста с использованием TextRank (пример разбирается в классе)

№ 87. Диалоговые системы и чат-боты.

На основе фреймворка DeepPavlov создать чат-бота с 10 вариантами ответов (пример разбирается в классе)

№ 88. Вопросно-ответные системы.

Реализовать простейшую вопросно-ответную систему

№ 89. Сбор фактов.

Написать программу для извлечения сотовых телефонных номеров из заданного текста.

№ 90. Распознавание речи.

Определить, какая звуковая команда распознается нейронной сетью наименее точно в алгоритме распознавания команд, разбираемом в классе.

2.3.7. Практические задания для оценки компетенции «ПК-2.3»

№ 91. Решающие деревья.

На основе заданного размеченного датасета построить и обучить решающее дерево (python, библиотека sklearn, DecisionTreeClassifier)

№ 92. Метод главных компонент.

Найти оптимальное количество главных компонент для заданного многомерного датасета

№ 93. Элементарный нейрон.

На основе заданного размеченного датасета обучить нейронную сеть из одного нейрона

(python, библиотека sklearn, класс MLPClassifier)

№ 94. Основные понятия нейронных сетей.

На основе заданного датасета обучить полносвязную классифицирующую нейронную сеть с одним скрытым слоем. (библиотека sklearn, класс MLPClassifier)

№ 95. Метод градиентного спуска и его модификации.

На основе заданного размеченного датасета обучить нейронную сеть в течение 10 эпох. Выбрать модификацию метода градиентного спуска (оптимизаторы ADAM, RMSProp, SGD), обучающий нейронную сеть с наилучшей точностью за 10 эпох.

№ 96. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений разбить заданный текст на слова.

№ 97. Токенизация текстов. Регулярные выражения.

С помощью регулярных выражений выбрать из текста пары 'наречие+глагол'

№ 98. Векторные представления и метод мешка слов.

Создать векторное представление для слов в заданном тексте. В качестве векторного представления использовать количество употреблений этого слова в тексте.

№ 99. Оптимальные векторные представления.

При помощи заданного текста, библиотеки gensim и класса word2vec обучить и найти оптимальные векторные представления для слов текста.

№ 100. Тематическое моделирование.

Построить матрицу числа появлений различных слов в различных текстах.

№ 101. Матричный и авторегрессионный подходы к морфологическому анализу текстов.

На основе размеченного набора слов и частей их речи и библиотеки Tensorflow:

разделить слова на буквы, создать датасет (последовательность букв -> часть речи)

Создать нейронную сеть, которая каждую букву закодирует оптимальным эмбеддингом размера 8, и на основе сверточной сети обучится классифицировать тексты (пример разбирается в классе)

№ 102. Синтаксический анализ текстов и распознавание именованных сущностей.

С помощью библиотеки spacy выделить в тексте именованные сущности в заданном тексте (пример разбирается в классе)

№ 103. Семантический анализ текстов, обобщенные модели и трансфер знаний.

Получить векторное представление слова после прохождения модели BERT (пример разобрать в классе)

№ 104. Понимание и аннотирование текстов.

Извлечь из текста наиболее важные предложения методом экстрактивного аннотирования текста с использованием TextRank (пример разбирается в классе)

№ 105. Диалоговые системы и чат-боты.

На основе фреймворка DeepPavlov создать чат-бота с 10 вариантами ответов (пример разбирается в классе)

№ 106. Вопросно-ответные системы.

Реализовать простейшую вопросно-ответную систему

№ 107. Сбор фактов.

Написать программу для извлечения сотовых телефонных номеров из заданного текста.

№ 108. Распознавание речи.

Определить, какая звуковая команда распознается нейронной сетью наименее точно в алгоритме распознавания команд, разбираемом в классе.

3. Промежуточная аттестация

3.1. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и опыта деятельности

Зачет является заключительным этапом процесса формирования компетенций обучающегося при изучении дисциплины и имеет целью проверку и оценку знаний обучающегося по теории, и применению полученных знаний, умений и навыков при решении практических задач.

Зачет проводится по расписанию, сформированному учебно-методическим управлением, в сроки, предусмотренные календарным учебным графиком. Зачет принимается преподавателем, ведущим лекционные занятия.

Зачет проводится только при предъявлении обучающимся зачетной книжки и при условии выполнения всех контрольных мероприятий, предусмотренных учебным планом и рабочей программой дисциплины. Обучающимся на зачете представляется право выбрать один из билетов. Время подготовки к ответу составляет 30 минут. По истечении установленного времени обучающийся должен ответить на вопросы билета. Результаты зачета оцениваются по четырехбалльной системе и заносятся в зачетно-экзаменационную ведомость и зачетную книжку. В зачетную книжку заносятся только положительные оценки. Подписанный преподавателем экземпляр ведомости сдается не позднее следующего дня в деканат.

В случае неявки обучающегося на зачет в зачетно-экзаменационную ведомость делается отметка «не явка». Обучающиеся, не прошедшие промежуточную аттестацию по дисциплине, должны ликвидировать академическую задолженность в установленном локальными нормативными актами порядке.

3.2. Вопросы к зачету с оценкой

№	Вопрос	Код компетенции
1.	Для решения каких задач сегодня используют нейронные сети?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
2.	Какие задачи можно решать с помощью решающих деревьев?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
3.	Для решения каких задач можно применять метод главных компонент?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
4.	Какую метаматическую операцию выполняет нейрон?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
5.	Если стоит задача классификации изображений на три класса (собака, кошка, мышка), сколько нейронов должно быть в выходном слое такой сети?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
6.	Для чего требуется использовать оптимизаторы (модификации метода градиентного спуска) при обучении?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
7.	Для чего используется метод градиентного спуска? Что такое скорость обучения?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
8.	Что такое токен? На какие токены можно разбить литературный текст?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3

№	Вопрос	Код компетенции
9.	Для чего нужно преобразовывать токены (слова, слоги) в векторные представления? Какие векторные представления вы знаете?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
10.	Для чего нужны векторные представления слов (токенов)? Какие варианты векторных представлений слов вы знаете?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
11.	Для решения каких задач на множестве текстов можно использовать тематическое моделирование?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
12.	Можно-ли создать нейронную сеть, которая по слову определяет часть речи, и как-бы вы решали такую задачу?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
13.	В каких задачах анализа текстов может использоваться распознавание именованных сущностей (NER - Named Entity Recognition)?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
14.	Какие обобщенные языковые модели вы знаете? Для чего и как их можно использовать?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
15.	Чем отличается генеративное от экстрактивного аннотирования текстов, какое проще реализовать?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
16.	Чем отличается FAQ-чатбот от генеративного чатбота. Какой проще реализовать?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
17.	Какими могут быть вопросно-ответные системы? Описать принципы построения вопросно-ответных систем.	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
18.	Почему обычно звуки переводят в спектрограммы в задачах распознавания звуковых команд и речи? Как идет дальнейшая обработка спектрограмм?	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3

3.3. Тематика курсовых работ

По данной дисциплине выполнение курсовых проектов (работ) не предусматривается.

3.4. Материалы для компьютерного тестирования обучающихся

Общие критерии оценивания

Процент правильных ответов	Оценка
91% – 100%	5 (отлично)
81% – 90%	4 (хорошо)
71% – 80%	3 (удовлетворительно)
Менее 70%	2 (неудовлетворительно)

Соответствие вопросов теста индикаторам формируемых и оцениваемых компетенций

№ вопроса в тесте	Код индикатора компетенции
1	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
2	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3

№ вопроса в тесте	Код индикатора компетенции
3	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
4	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
5	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
6	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
7	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
8	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
9	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
10	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
11	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
12	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
13	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
14	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
15	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
16	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
17	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3
18	ПК-1.1, ПК-1.2, ПК-1.3, ПК-2.1, ПК-2.2, ПК-2.3

Ключ ответов

№ вопроса в тесте	Номер ответа (или ответ, или соответствие)
1	c
2	b
3	c
4	a, b
5	d
6	b
7	a
8	b
9	a
10	b
11	a
12	b
13	d
14	c
15	b
16	b
17	a, b, c, d, e, f
18	a

Перечень тестовых вопросов

№ 1. Задание с единственным выбором. Выберите один правильный ответ.

Что обычно важнее всего для работы FAQ чат-бота?

- a. Большой объем обучающих данных для обучения
- b. Языковая модель для понимания вопроса
- c. База вопрос-ответ
- d. Языковая модель для генерации ответа

№ 2. Задание с единственным выбором. Выберите один правильный ответ.

Какую из этих задач можно успешно решать с помощью решающих деревьев?

- a. задачи кластеризации
- b. задачи классификации

№ 3. Задание с единственным выбором. Выберите один правильный ответ.

Для чего нужно преобразовывать токен (или слово) в его векторное представление?

- a. Чтобы уменьшить его размер
- b. Чтобы увеличить его размер
- c. Чтобы работать с числами, а не с символами

№ 4. Задание с множественным выбором. Выберите 2 правильных ответа.

Я знаю такие обобщенные языковые модели

- a. BERT
- b. GPT3
- c. VGG19

№ 5. Задание с единственным выбором. Выберите один правильный ответ.

На вход нейрона подается два числа, сколько чисел будет на выходе нейрона

- a. 4
- b. 3
- c. 2
- d. 1

№ 6. Задание с единственным выбором. Выберите один правильный ответ.

На какие возможные токены можно разбить русскоязычный текст?

- a. Слова
- b. Все из перечисленных
- c. N-граммы

№ 7. Задание с единственным выбором. Выберите один правильный ответ.

Можно-ли использовать регулярные выражения для сбора фактов в тексте?

- a. Да, для извлечения данных по маске, например телефонов
- b. Нет, факты можно извлекать только с помощью нейронных сетей, потому-что это очень сложно

№ 8. Задание с единственным выбором. Выберите один правильный ответ.

Для чего используется метод градиентного спуска в машинном обучении?

- a. Для поиска оптимального размера обучающей выборки
- b. Для поиска минимума функции потерь нейронной сети

№ 9. Задание с единственным выбором. Выберите один правильный ответ.

Какую задачу решают, когда обучают нейронную сеть?

- a. минимизацию функции потерь
- b. максимизацию функции активации

№ 10. Задание с единственным выбором. Выберите один правильный ответ.

В каком смысле оптимальные векторные представления слов оптимальны?

- a. Они имеют минимальный размер
- b. Они позволяют наиболее точно решить нужную нам задачу

№ 11. Задание с единственным выбором. Выберите один правильный ответ.

Для решения какой задачи лучше применять метод главных компонент?

- a. Уменьшение размерности данных
- b. Увеличения размерности данных

№ 12. Задание с единственным выбором. Выберите один правильный ответ.

Для решения какой задачи можно использовать тематическое моделирование?

- a. Для автоматической генерации текста на заданную тематику
- b. Для определения того, на какие группы(темы) можно разделить множество имеющихся у нас текстов

№ 13. Задание с единственным выбором. Выберите один правильный ответ.

В каких задачах часто используется синтаксический анализ русскоязычного текста?

- a. Все перечисленное
- b. Векторизация токенов
- c. Токенизация текста
- d. Распознавание именованных сущностей

№ 14. Задание с единственным выбором. Выберите один правильный ответ.

Для чего обычно используют преобразования Фурье и спектрограммы при распознавании речи?

- a. Для преобразования звука в число и обработки случайным решающим лесом
- b. Для преобразования звука в вектор и его обработки линейной регрессией
- c. Для преобразования звука в матрицу и ее обработки сверточными сетями

№ 15. Задание с единственным выбором. Выберите один правильный ответ.

Отметьте верное о генеративном и экстрактивном аннотировании текстов.

- a. При экстрактивном извлекаются части из старого текста, и новых слов не используется.
- b. Верно все
- c. При генеративном создается новый текст на основе понимания старого и возможно использование новых слов

№ 16. Задание с единственным выбором. Выберите один правильный ответ.

Что относится к морфологическому анализу текста?

- a. Генерация краткого содержания текста
- b. Определение части речи неизвестного слова
- c. Определение эмоционального окраса (негативный, позитивный, нейтральный) текста

№ 17. Задание с множественным выбором. Выберите 6 правильных ответов.

Для решения каких задач сегодня используют нейронные сети?

- a. создание изображений по текстовому описанию
- b. автоматический перевод
- c. определение эмоционального окраса текста (позитивный, негативный, нейтральный)
- d. автоматическое управление автомобилем
- e. распознавание изображений
- f. классификация изображений

№ 18. Задание с единственным выбором. Выберите один правильный ответ.

Какими могут быть вопросно-ответные системы?

- a. Все перечисленное
- b. Общие - для ответов на любые вопросы
- c. Узкоспециализированные - для ответов на вопросы в конкретной области