

Документ подписан проставив электронную подпись
Информация о владельце:
ФИО: Шамрай-Курбатова Лидия Викторовна
Должность: Ректор
Дата подписания: 09.06.2026 10:08:49
Уникальный программный ключ:
b1e4399771b07e18f31755456972d73b2ccfc531

Автономная некоммерческая организация высшего образования
«Волгоградский институт бизнеса»

Рабочая программа учебной дисциплины

Компьютерная лингвистика

(Наименование дисциплины)

09.03.03 Прикладная информатика, направленность (профиль) «Прикладной искусственный интеллект»

(Направление подготовки / Профиль)

Бакалавр

(Квалификация)

Кафедра разработчик

Экономики и управления

Год набора

2026

Вид учебной деятельности	Трудоемкость (объем) дисциплины	
	Очная форма	Очно-заочная форма
	д	в
Зачетные единицы	3	3
Общее количество часов	108	108
Аудиторные часы контактной работы обучающегося с преподавателями:	32	16
– Лекционные (Л)	16	8
– Практические (ПЗ)	16	8
– Лабораторные (ЛЗ)		
– Семинарские (СЗ)		
Самостоятельная работа обучающихся (СРО)	76	92
К (Р-Г) Р (П) (+;-)		
Тестирование (+;-)		
ДКР (+;-)		
Зачет (+;-)	+	+
Зачет с оценкой (+;- (Кол-во часов))		
Экзамен (+;- (Кол-во часов))		

Волгоград 2026

Содержание

Раздел 1. Организационно-методический раздел	3
Раздел 2. Тематический план.....	5
Раздел 3. Содержание дисциплины.....	6
Раздел 4. Организация самостоятельной работы обучающихся	12
Раздел 5. Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся	14
Раздел 6. Оценочные средства промежуточной аттестации (с ключами)	19
Раздел 7. Перечень учебной литературы, необходимой для освоения дисциплины	16
Раздел 8. Материально-техническая база и информационные технологии.....	19
Раздел 9. Методические указания для обучающихся по освоению дисциплины	21

Раздел 1. Организационно-методический раздел

1.1. Цели освоения дисциплины

Дисциплина «Компьютерная лингвистика» входит в Элективные дисциплины Б1.В.ДЭ.01.01 подготовки обучающихся по направлению Прикладная информатика, направленность (профиль) «Прикладной искусственный интеллект».

Целью дисциплины является формирование **компетенций** (в соответствии с ФГОС ВО и требованиями к результатам освоения основной профессиональной образовательной программы высшего образования (ОПОП ВО)):

ПК-1. Способен анализировать возможности реализации требований к компьютерному программному обеспечению.

Дескрипторы профессиональных компетенций:

ПК-1.1. Способен анализировать и обосновывать технические решения при разработке компьютерного программного обеспечения, включая программные системы с элементами искусственного интеллекта

ПК-1.2. Способен анализировать требования к данным и проектным ограничениям при разработке программного обеспечения, включая системы анализа данных и искусственного интеллекта.

Перечисленные компетенции формируются в процессе достижения **индикаторов компетенций**:

Обобщенная трудовая функция/ трудовая функция	Код и наименование дескриптора компетенций	Код и наименование индикатора достижения компетенций (из ПС)
ПС 06.001 Программист Д Разработка требований и проектирование программного обеспечения D/01.6. Анализ возможностей реализации требований к компьютерному программному обеспечению	ПК-1.1. Способен анализировать и обосновывать технические решения при разработке компьютерного программного обеспечения, включая программные системы с элементами искусственного интеллекта ПК-1.2. Способен анализировать требования к данным и проектным ограничениям при разработке программного обеспечения, включая системы анализа данных и искусственного интеллекта	Знает: ИД-1 ПК-1.1. Методологии разработки компьютерного программного обеспечения и технологии программирования D/01.6 ИД-2 ПК-1.2. Методологии и технологии проектирования и использования баз данных D/01.6 Умеет: ИД-3 ПК-1.1. Проводить оценку и обоснование рекомендуемых решений D/01.6 ИД-4 ПК-1.2. Осуществлять коммуникации с заинтересованными сторонами D/01.6 Имеет навыки и (или) опыт: ИД-5 ПК-1.1. Согласования требований к компьютерному программному обеспечению с заинтересованными сторонами D/01.6 ИД-6 ПК-1.2. Оценки и согласования сроков выполнения поставленных задач D/01.6

**1.2. Место дисциплины в структуре ОПОП ВО
направления подготовки «09.03.03 Прикладная информатика», направленность (профиль) «Прикладной искусственный интеллект»**

№	Предшествующие дисциплины (дисциплины, изучаемые параллельно)	Последующие дисциплины
<i>1</i>	<i>2</i>	<i>3</i>
1	Компьютерная лингвистика	Основы систем искусственного интеллекта
2	Информатика и программирование	Модели искусственного интеллекта
3	Базы данных	Выполнение и защита выпускной квалификационной работы
4	Линейная алгебра	Прикладной искусственный интеллект в БАС
5	Теория вероятностей и математическая статистика	Учебная практика (Научно-исследовательская работа (получение первичных навыков научно-исследовательской работы))
6	Математический анализ и моделирование	Производственная практика (Технологическая (проектно-технологическая) практика)
7		Производственная практика (Эксплуатационная практика)
8		Производственная практика (Преддипломная практика)

Последовательность формирования компетенций в указанных дисциплинах может быть изменена в зависимости от формы и срока обучения, а также преподавания с использованием дистанционных технологий обучения.

1.3. Нормативная документация

Рабочая программа учебной дисциплины составлена на основе:

- Федерального государственного образовательного стандарта высшего образования по направлению подготовки **09.03.03 Прикладная информатика**;
- Учебного плана направления подготовки **09.03.03 Прикладная информатика, направленность (профиль) «Прикладной искусственный интеллект»** 2026 года набора;
- Образца рабочей программы учебной дисциплины (приказ № 113-О от 01.09.2021 г.).

Раздел 2. Тематический план

Очная форма обучения (полный срок)

№	Тема дисциплины	Трудоемкость				Код индикатора и дескриптора достижения компетенций
		Всего	Аудиторные занятия		СРО	
			Л	ПЗ (ЛЗ, СЗ)		
1	2	3	4	5	6	7
1	Введение в компьютерную лингвистику: задачи, методы, области применения, связь с искусственным интеллектом	27	4	4	19	ИД-1 ПК- 1.1 ИД-2 ПК- 1.2
2	Обработка естественного языка (NLP): токенизация, лемматизация, морфологический и синтаксический анализ	27	4	4	19	ИД-2 ПК- 1.2 ИД-3 ПК- 1.1
3	Модели представления текста: n-граммы, векторные представления слов (Word2Vec, TF-IDF), эмбединги	27	4	4	19	ИД-4 ПК- 1.2 ИД-5 ПК- 1.1
4	Применение компьютерной лингвистики: машинный перевод, анализ тональности текста, чат-боты и диалоговые системы	27	4	4	19	ИД-5 ПК- 1.1 ИД-6 ПК- 1.2
Вид промежуточной аттестации (Зачет)		+				
Итого		108	16	16	76	

Очно-заочная форма обучения (полный срок)

№	Тема дисциплины	Трудоемкость				Код индикатора и дескриптора достижения компетенций
		Всего	Аудиторные занятия		СРО	
			Л	ПЗ (ЛЗ, СЗ)		
1	2	3	4	5	6	7
1	Введение в компьютерную лингвистику: задачи, методы, области применения, связь с искусственным интеллектом	28	2	2	24	ИД-1 ПК- 1.1 ИД-2 ПК- 1.2
2	Обработка естественного языка (NLP): токенизация, лемматизация, морфологический и синтаксический анализ	28	2	2	24	ИД-2 ПК- 1.2 ИД-3 ПК- 1.1
3	Модели представления текста: n-граммы, векторные представления слов (Word2Vec, TF-IDF), эмбединги	26	2	2	22	ИД-4 ПК- 1.2 ИД-5 ПК- 1.1
4	Применение компьютерной лингвистики: машинный перевод, анализ тональности текста, чат-боты и диалоговые системы	26	2	2	22	ИД-5 ПК- 1.1 ИД-6 ПК- 1.2
Вид промежуточной аттестации (Зачет)		+				
Итого		108	8	8	92	

Раздел 3. Содержание дисциплины

3.1. Содержание дисциплины

Тема 1. Введение в компьютерную лингвистику

Определение компьютерной лингвистики как междисциплинарной области на стыке лингвистики, информатики и искусственного интеллекта. Основные задачи: автоматический анализ и синтез текста, извлечение информации, машинный перевод, речевые технологии. Связь с обработкой естественного языка (NLP), отличие от вычислительной лингвистики. История развития: от работ Ноама Хомского (формальные грамматики) до современных нейросетевых подходов. Ключевые этапы: правила-ориентированный подход (1950-1980), статистический подход (1990-2010), нейросетевой и трансформерный подход (2010-настоящее время). Области применения: поисковые системы (ранжирование, релевантность), голосовые ассистенты (распознавание и синтез речи), антиспам-фильтры, автоматическое реферирование, системы поддержки принятия решений. Примеры: Google Translate, Яндекс.Алиса, Siri, ChatGPT. Этика компьютерной лингвистики: предвзятость моделей, приватность данных, проблема «галлюцинаций» языковых моделей.

Тема 2. Обработка естественного языка (NLP)

Предобработка текста: очистка данных (удаление шумов, нормализация регистра, исправление опечаток). Токенизация: разделение текста на токены (слова, знаки препинания, числа). Проблемы токенизации (аббревиатуры, дефисы, языки без пробелов). Лемматизация и стемминг: приведение слов к нормальной форме, отличие лемматизации (с учётом контекста) от стемминга (отсечение аффиксов). Морфологический анализ: определение части речи, рода, числа, падежа, времени. Использование морфологических словарей и правил. Синтаксический анализ: построение деревьев зависимостей (dependency parsing) и составляющих (constituency parsing). Роль синтаксиса в понимании смысла. Работа с корпусами текстов: типы корпусов (сбалансированные, аннотированные, параллельные), разметка корпуса (POS-теги, синтаксические структуры). Известные корпуса: Brown Corpus, Treebank, Russian National Corpus. Особенности обработки русского языка: богатая морфология, свободный порядок слов, флективность. Проблемы омонимии, многозначности и эллипсиса.

Тема 3. Модели представления текста

Необходимость численного представления текста для машинного обучения. Дискретные методы: n-граммы (последовательности символов или слов) для учёта контекста. Мешок слов (BoW): игнорирование порядка слов, проблема разреженности, нормализация длин документов. TF-IDF (Term Frequency-Inverse Document Frequency): мера важности слова в документе относительно коллекции, компоненты (term frequency, document frequency). Векторные представления слов: распределённая семантика (гипотеза распределения Харриса). Word2Vec (CBOW, Skip-gram), GloVe, FastText. Понятие эмбедингов: отображение слов в непрерывное векторное пространство, сохранение семантических и синтаксических отношений (аналогии: король — мужчина + женщина \approx королева). Контекстуальные эмбединги: ELMo, BERT (би-направленные контекстуальные представления). Сравнение методов: разреженность (BoW, TF-IDF) vs плотность (Word2Vec), статичность (Word2Vec) vs контекстуальность (BERT). Применение в задачах классификации текстов, кластеризации, поиска семантических близостей. Оценка качества векторных представлений: аналогии, задачи замены слов.

Тема 4. Применение компьютерной лингвистики

Машинный перевод: статистический подход (фразовый перевод, выравнивание), нейросетевой подход (encoder-decoder, механизм внимания). Современные архитектуры (Transformer, GPT) для перевода. Оценка качества перевода: автоматические метрики BLEU, TER, chrF. Анализ тональности (sentiment analysis): классификация эмоциональной окраски (положительная, отрицательная, нейтральная). Уровни: на уровне документа, предложения, аспекта. Методы: лексические подходы (словари эмоций), машинное обучение (наивный Байес, SVM), глубокое обучение (LSTM, BERT). Информационный поиск (IR): задачи ранжирования документов по релевантности

запросу. Модели поиска (boolean, векторная, вероятностная). Индексация и оценка качества поиска (precision, recall, MAP, nDCG). Чат-боты и диалоговые системы: классификация (ретриверные, генеративные, гибридные). Управление диалогом (DST), отслеживание состояния диалога. Интеграция с базами знаний и API. Реальные кейсы: автоматическая модерация контента, извлечение сущностей (NER) из медицинских текстов, суммаризация новостей, системы ответов на вопросы (QA). Проблемы и ограничения: нехватка размеченных данных, мультиязычность, обработка редких явлений и иронии.

3.2. Содержание практического блока дисциплины

Очная форма обучения (полный срок)

№	Тема практического (семинарского, лабораторного) занятия
1	2
ПЗ 1	<p>Анализ текстовых данных: сбор корпуса текстов, очистка и предварительная обработка данных. Выполнение токенизации и нормализации текста. Подготовка данных для дальнейшего анализа.</p> <p>Сбор корпуса текстов из открытых источников с использованием веб-скрапинга (библиотеки BeautifulSoup, Scrapy) или загрузка готовых датасетов (например, новостные ленты, отзывы, литературные произведения). Очистка данных: удаление HTML-тегов, специальных символов, цифр (если не значимы), лишних пробелов, приведение текста к нижнему регистру. Удаление стоп-слов с использованием стандартных списков (nltk, spacy) или создание собственного списка. Реализация токенизации на уровне слов и предложений с помощью библиотек NLTK, spaCy, или регекспов. Сравнение результатов токенизации разных инструментов. Выполнение стемминга (PorterStemmer, SnowballStemmer) и лемматизации (WordNetLemmatizer, rymorphy2 для русского языка). Анализ качества нормализации: выявление ошибок (например, чрезмерное усечение при стемминге). Создание pipeline предобработки для автоматической очистки и нормализации текстов. Сохранение обработанных данных в структурированном формате (CSV, JSON) для использования в следующих работах. Визуализация результатов предобработки: сравнение исходных и обработанных текстов, облако слов до и после удаления стоп-слов.</p>
ПЗ 2	<p>Морфологический и синтаксический анализ текста: определение частей речи, построение синтаксических деревьев. Использование библиотек для обработки естественного языка. Анализ структуры предложений.</p> <p>Применение библиотек spaCy, NLTK, Stanza или rymorphy2 для морфологического анализа. Определение частей речи (POS-тегирование) с использованием предобученных моделей для русского и английского языков. Сравнение точности тегирования разными библиотеками. Разбор форматов POS-тегов (Universal Dependencies, Penn Treebank). Анализ морфологических характеристик: род, число, падеж, время, лицо.</p> <p>Визуализация распределения частей речи в корпусе текстов. Построение синтаксических деревьев зависимостей (dependency parsing) с помощью spaCy. Изучение типов синтаксических связей: подлежащее, сказуемое, дополнение, определение, обстоятельство. Визуализация синтаксических деревьев с использованием displacy. Извлечение типовых синтаксических конструкций (например, «субъект — действие — объект»). Анализ структуры сложных предложений: выделение главной и придаточных частей. Применение синтаксического анализа для извлечения ключевых фраз и фактов из текста. Сравнение результатов синтаксического анализа для текстов разных стилей (научный, публицистический, разговорный). Оценка качества морфологического и синтаксического анализа на размеченных тестовых наборах.</p>
ПЗ 3	<p>Построение моделей представления текста: реализация методов Bag of Words и TF-</p>

	<p>IDF. Создание векторных представлений текстов. Сравнение различных подходов к представлению данных.</p> <p>Реализация модели мешка слов (Bag of Words) с использованием CountVectorizer из scikit-learn. Создание словаря (vocabulary) по обучающему корпусу. Построение матрицы термин-документ. Анализ разреженности матрицы и способов её уменьшения.</p> <p>Реализация модели TF-IDF с использованием TfidfVectorizer: вычисление term frequency и inverse document frequency. Сравнение BoW и TF-IDF на примере поиска ключевых слов в документах. Построение векторных представлений слов с помощью предобученных моделей Word2Vec (Gensim) и FastText. Загрузка готовых эмбеддингов для русского языка (например, RusVectōrēs, Navec). Изучение семантических свойств: поиск аналогий (король — мужчина + женщина), вычисление семантической близости между словами. Визуализация векторных представлений в 2D с использованием t-SNE или PCA. Реализация усреднения векторов слов для получения векторных представлений документов. Сравнение различных подходов к векторизации текста на задаче классификации или кластеризации. Оценка качества: внутренние метрики (косинусная близость для похожих/разных документов), внешние метрики (ассигасу на размеченном датасете). Анализ вычислительной сложности и требований к памяти для разных методов.</p>
ПЗ 4	<p>Применение методов компьютерной лингвистики: реализация анализа тональности текста. Разработка простого чат-бота или системы классификации текстов. Оценка качества моделей и интерпретация результатов.</p> <p>Реализация анализа тональности (сентимент-анализа) на размеченном датасете отзывов. Сравнение нескольких подходов: лексический подход (словари эмоций, например, SentiWordNet или Linis Crowd), классические ML-алгоритмы (логистическая регрессия, наивный Байес) на признаках BoW/TF-IDF, нейросетевой подход (fine-tuning BERT или использование ruBERT). Разделение данных на обучающую, валидационную и тестовую выборки. Вычисление метрик качества: accuracy, precision, recall, F1-score, построение матрицы ошибок (confusion matrix). Анализ ошибок модели: какие типы текстов классифицируются неверно (ирония, сложные конструкции, сленг).</p> <p>Разработка системы классификации текстов по жанрам или темам с использованием методов из предыдущих занятий. Построение пайплайна: предобработка → векторизация → классификация. Подбор гиперпараметров с помощью GridSearchCV. Разработка простого чат-бота на основе ретриверной архитектуры: создание базы вопросов и ответов, векторизация вопросов, поиск ближайшего вопроса по косинусной близости (Faiss или Scikit-learn). Добавление эвристик для обработки неизвестных запросов. Реализация чат-бота с использованием библиотеки Rasa (базовые шаги) или через простой веб-интерфейс на Streamlit. Оценка качества чат-бота: метрики точности ответов, процент неудачных ответов. Интерпретация результатов: визуализация важности признаков для модели классификации, анализ наиболее информативных слов для каждого класса. Формулирование выводов о применимости различных методов для конкретной задачи.</p>

Очно-заочная форма обучения (полный срок)

№	Тема практического (семинарского, лабораторного) занятия
1	2
ПЗ 1	<p>Анализ текстовых данных: сбор корпуса текстов, очистка и предварительная обработка данных. Выполнение токенизации и нормализации текста. Подготовка данных для дальнейшего анализа.</p> <p>Сбор корпуса текстов из открытых источников с использованием веб-скрапинга (библиотеки BeautifulSoup, Scrapy) или загрузка готовых датасетов (например, новостные ленты, отзывы, литературные произведения). Очистка данных: удаление HTML-тегов, специальных символов, цифр (если не значимы), лишних пробелов, приведение текста к нижнему регистру. Удаление стоп-слов с использованием стан-</p>

	<p>дартных списков (nltk, spacy) или создание собственного списка. Реализация токенизации на уровне слов и предложений с помощью библиотек NLTK, spaCy, или регекспов. Сравнение результатов токенизации разных инструментов. Выполнение стемминга (PorterStemmer, SnowballStemmer) и лемматизации (WordNetLemmatizer, rymorphy2 для русского языка). Анализ качества нормализации: выявление ошибок (например, чрезмерное усечение при стемминге). Создание pipeline предобработки для автоматической очистки и нормализации текстов. Сохранение обработанных данных в структурированном формате (CSV, JSON) для использования в следующих работах. Визуализация результатов предобработки: сравнение исходных и обработанных текстов, облако слов до и после удаления стоп-слов.</p>
ПЗ 2	<p>Морфологический и синтаксический анализ текста: определение частей речи, построение синтаксических деревьев. Использование библиотек для обработки естественного языка. Анализ структуры предложений.</p> <p>Применение библиотек spaCy, NLTK, Stanza или rymorphy2 для морфологического анализа. Определение частей речи (POS-тегирование) с использованием предобученных моделей для русского и английского языков. Сравнение точности тегирования разными библиотеками. Разбор форматов POS-тегов (Universal Dependencies, Penn Treebank). Анализ морфологических характеристик: род, число, падеж, время, лицо. Визуализация распределения частей речи в корпусе текстов. Построение синтаксических деревьев зависимостей (dependency parsing) с помощью spaCy. Изучение типов синтаксических связей: подлежащее, сказуемое, дополнение, определение, обстоятельство. Визуализация синтаксических деревьев с использованием displacy. Извлечение типовых синтаксических конструкций (например, «субъект — действие — объект»). Анализ структуры сложных предложений: выделение главной и придаточных частей. Применение синтаксического анализа для извлечения ключевых фраз и фактов из текста. Сравнение результатов синтаксического анализа для текстов разных стилей (научный, публицистический, разговорный). Оценка качества морфологического и синтаксического анализа на размеченных тестовых наборах.</p>
ПЗ 3	<p>Построение моделей представления текста: реализация методов Bag of Words и TF-IDF. Создание векторных представлений текстов. Сравнение различных подходов к представлению данных.</p> <p>Реализация модели мешка слов (Bag of Words) с использованием CountVectorizer из scikit-learn. Создание словаря (vocabulary) по обучающему корпусу. Построение матрицы термин-документ. Анализ разреженности матрицы и способов её уменьшения. Реализация модели TF-IDF с использованием TfidfVectorizer: вычисление term frequency и inverse document frequency. Сравнение BoW и TF-IDF на примере поиска ключевых слов в документах. Построение векторных представлений слов с помощью предобученных моделей Word2Vec (Gensim) и FastText. Загрузка готовых эмбеддингов для русского языка (например, RusVectōrēs, Navec). Изучение семантических свойств: поиск аналогий (король — мужчина + женщина), вычисление семантической близости между словами. Визуализация векторных представлений в 2D с использованием t-SNE или PCA. Реализация усреднения векторов слов для получения векторных представлений документов. Сравнение различных подходов к векторизации текста на задаче классификации или кластеризации. Оценка качества: внутренние метрики (косинусная близость для похожих/разных документов), внешние метрики (ассигасу на размеченном датасете). Анализ вычислительной сложности и требований к памяти для разных методов.</p>
ПЗ 4	<p>Применение методов компьютерной лингвистики: реализация анализа тональности текста. Разработка простого чат-бота или системы классификации текстов. Оценка качества моделей и интерпретация результатов.</p> <p>Реализация анализа тональности (сентимент-анализа) на размеченном датасете отзывов. Сравнение нескольких подходов: лексический подход (словари эмоций, например, SentiWordNet или Linis Crowd), классические ML-алгоритмы (логистическая ре-</p>

	<p>грессия, наивный Байес) на признаках BoW/TF-IDF, нейросетевой подход (fine-tuning BERT или использование ruBERT). Разделение данных на обучающую, валидационную и тестовую выборки. Вычисление метрик качества: accuracy, precision, recall, F1-score, построение матрицы ошибок (confusion matrix). Анализ ошибок модели: какие типы текстов классифицируются неверно (ирония, сложные конструкции, сленг). Разработка системы классификации текстов по жанрам или темам с использованием методов из предыдущих занятий. Построение пайплайна: предобработка → векторизация → классификация. Подбор гиперпараметров с помощью GridSearchCV. Разработка простого чат-бота на основе ретриверной архитектуры: создание базы вопросов и ответов, векторизация вопросов, поиск ближайшего вопроса по косинусной близости (Faiss или Scikit-learn). Добавление эвристик для обработки неизвестных запросов. Реализация чат-бота с использованием библиотеки Rasa (базовые шаги) или через простой веб-интерфейс на Streamlit. Оценка качества чат-бота: метрики точности ответов, процент неудачных ответов. Интерпретация результатов: визуализация важности признаков для модели классификации, анализ наиболее информативных слов для каждого класса. Формулирование выводов о применимости различных методов для конкретной задачи.</p>
--	--

3.3. Образовательные технологии Очная форма обучения (полный срок)

№	Тема занятия	Вид учебного занятия	Форма / Методы интерактивного обучения	% учебного времени
1	2	3	4	5
1	Введение в компьютерную лингвистику: задачи, методы, области применения, связь с искусственным интеллектом	ПЗ	Дискуссионные технологии, Работа в малых группах, Взаимопроверка, Мозговой штурм, Интерактивные тренажёры	25
2	Обработка естественного языка (NLP): токенизация, лемматизация, морфологический и синтаксический анализ	ПЗ	Дискуссионные технологии, Групповое решение проблемных задач, Кейс-стади, Взаимообучение, Проектно-ориентированное обучение	25
3	Модели представления текста: n-граммы, векторные представления слов (Word2Vec, TF-IDF), эмбединги	ПЗ	Работа в парах, Деловая игра, Дискуссионные технологии, Мозговой штурм, Интерактивная визуализация с коллективным обсуждением	25
4	Применение компьютерной лингвистики: машинный перевод, анализ тональности текста, чат-боты и диалоговые системы	ПЗ	Семинар-дискуссия, Работа в группах с презентацией, Ролевая игра, Интерактивная доска, Кейс-стади	25
Итого				25%

Очно-заочная форма обучения (полный срок)

№	Тема занятия	Вид учебного занятия	Форма / Методы интерактивного обучения	% учебного времени
---	--------------	----------------------	--	--------------------

		тия		
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>
1	Введение в компьютерную лингвистику: задачи, методы, области применения, связь с искусственным интеллектом	ПЗ	Дискуссионные технологии, Работа в малых группах, Взаимопроверка, Мозговой штурм, Интерактивные тренажёры	25
2	Обработка естественного языка (NLP): токенизация, лемматизация, морфологический и синтаксический анализ	ПЗ	Дискуссионные технологии, Групповое решение проблемных задач, Кейс-стади, Взаимообучение, Проектно-ориентированное обучение	25
3	Модели представления текста: n-граммы, векторные представления слов (Word2Vec, TF-IDF), эмбединги	ПЗ	Работа в парах, Деловая игра, Дискуссионные технологии, Мозговой штурм, Интерактивная визуализация с коллективным обсуждением	25
4	Применение компьютерной лингвистики: машинный перевод, анализ тональности текста, чат-боты и диалоговые системы	ПЗ	Семинар-дискуссия, Работа в группах с презентацией, Ролевая игра, Интерактивная доска, Кейс-стади	25
Итого				25%

Раздел 4. Организация самостоятельной работы обучающихся

4.1. Организация самостоятельной работы обучающихся

№	Тема дисциплины	№ вопро-сов	№ рекоменду-емой литерату-ры
1	2	3	4
1	Введение в компьютерную лингвистику: задачи, методы, области применения, связь с искусственным интеллектом	1-4	1, 2, 3, 5, 11, 12
2	Обработка естественного языка (NLP): токенизация, лемматизация, морфологический и синтаксический анализ	5-12	1, 2, 4, 5, 6, 13
3	Модели представления текста: n-граммы, векторные представления слов (Word2Vec, TF-IDF), эмбединги	13-20	2, 3, 4, 7, 8, 13
4	Применение компьютерной лингвистики: машинный перевод, анализ тональности текста, чат-боты и диалоговые системы	20-30	1, 3, 4, 9, 10, 14, 15

Перечень вопросов, выносимых на самостоятельную работу обучающихся

1. Что такое компьютерная лингвистика и какие задачи она решает?
2. Какова роль компьютерной лингвистики в современных информационных системах?
3. Какие основные направления включает компьютерная лингвистика?
4. Как компьютерная лингвистика связана с искусственным интеллектом?
5. Что такое обработка естественного языка (NLP)?
6. Какие этапы включает обработка текстовых данных?
7. Что такое токенизация и какие существуют методы её реализации?
8. Что такое лемматизация и чем она отличается от стемминга?
9. Что включает морфологический анализ текста?
10. Что такое синтаксический анализ и как он выполняется?
11. Какие существуют подходы к построению синтаксических деревьев?
12. Что такое корпус текстов и как он используется?
13. Какие методы представления текста существуют?
14. Что такое модель «мешок слов» (Bag of Words)?
15. Что такое n-граммы и где они применяются?
16. Что такое TF-IDF и как он рассчитывается?
17. Что такое векторные представления слов?
18. Что такое эмбединги и в чём их преимущества?
19. Какие существуют модели Word Embeddings?
20. Как выполняется сравнение текстов с помощью векторных моделей?
21. Что такое машинный перевод и какие подходы используются?
22. Что такое анализ тональности текста?
23. Какие методы используются для классификации текстов?
24. Что такое информационный поиск?
25. Как работают поисковые системы с текстами?
26. Что такое чат-бот и как он реализуется?
27. Какие методы используются в диалоговых системах?
28. Как оценивается качество моделей обработки текста?
29. Какие проблемы возникают при обработке естественного языка?
30. Какие перспективы развития компьютерной лингвистики существуют?

4.2. Перечень учебно-методического обеспечения самостоятельной работы обучающихся

Самостоятельная работа обучающихся обеспечивается следующими учебно-методическими материалами:

1. Указаниями в рабочей программе по дисциплине (п.4.1.)
2. Лекционные материалы в составе учебно-методического комплекса по дисциплине
3. Заданиями и методическими рекомендациями по организации самостоятельной работы обучающихся в составе учебно-методического комплекса по дисциплине.
4. Глоссарием по дисциплине в составе учебно-методического комплекса по дисциплине.

Раздел 5. Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся

Фонд оценочных средств по дисциплине представляет собой совокупность контролирующих материалов, предназначенных для измерения уровня достижения обучающимися установленных результатов образовательной программы. ФОС по дисциплине используется при проведении оперативного контроля и промежуточной аттестации обучающихся. Требования к структуре и содержанию ФОС дисциплины регламентируются Положением о фонде оценочных материалов по программам высшего образования – программам бакалавриата, магистратуры.

5.1. Паспорт фонда оценочных средств

Очная форма обучения (полный срок)

№	Контролируемые разделы (темы) дисциплины	Оценочные средства			
		Л	ПЗ (ЛЗ, СЗ)	СРО	Код индикатора и дескриптора достижения компетенций
1	2	3	4	5	6
1	Введение в гибридные системы поддержки принятия решений: понятие, структура, области применения, роль в современных информационных системах	УО	ЗЗ, МШ	ПРВ	ИД-1 ПК- 1.1 ИД-2 ПК- 1.2
2	Классические методы поддержки принятия решений: экспертные системы, базы знаний, логический вывод	УО	ЗЗ, МШ	ПРВ	ИД-2 ПК- 1.2 ИД-3 ПК- 1.1
3	Методы машинного обучения в СППР: классификация, регрессия, прогнозирование и их интеграция в системы принятия решений	УО	ЗЗ, Д, МШ	ПРВ	ИД-4 ПК- 1.2 ИД-5 ПК- 1.1
4	Нечёткая логика и нечёткие системы: представление неопределённости, функции принадлежности, правила вывода	УО	ЗЗ, Д, МП	ПРВ	ИД-5 ПК- 1.1 ИД-6 ПК- 1.2

Очно-заочная форма обучения (полный срок)

№	Контролируемые разделы (темы) дисциплины	Оценочные средства			
		Л	ПЗ (ЛЗ, СЗ)	СРО	Код индикатора и дескриптора достижения компетенций
1	2	3	4	5	6
1	Введение в гибридные системы поддержки принятия решений: понятие, структура, области применения, роль в современных информационных системах	УО	ЗЗ, МШ	ПРВ	ИД-1 ПК- 1.1 ИД-2 ПК- 1.2
2	Классические методы поддержки принятия решений: экспертные системы, базы знаний, логический вывод	УО	ЗЗ, МШ	ПРВ	ИД-2 ПК- 1.2 ИД-3 ПК- 1.1
3	Методы машинного обучения в СППР: классификация, регрессия, прогнозирование и их интеграция в системы принятия решений	УО	ЗЗ, Д, МШ	ПРВ	ИД-4 ПК- 1.2 ИД-5 ПК- 1.1

4	Нечёткая логика и нечёткие системы: представление неопределённости, функции принадлежности, правила вывода	УО	33, Д, МП	ПРВ	ИД-5 ПК- 1.1 ИД-6 ПК- 1.2
---	--	----	--------------	-----	------------------------------

Условные обозначения оценочных средств (Столбцы 3, 4, 5):

ЗЗ – защита выполненных заданий (творческих, расчетных и т.д.), представление презентаций;

ПРВ – проверка рефератов, отчетов, рецензий, аннотаций, конспектов, графического материала, эссе, переводов, решений заданий, выполненных заданий в электронном виде и т.д.;

МШ – Метод мозгового штурма;

Д – Дискуссия, полемика, диспут, дебаты;

МП – Метод проектов.

5.2. Тематика письменных работ обучающихся

1. Основные задачи и области применения компьютерной лингвистики
2. Связь компьютерной лингвистики с искусственным интеллектом
3. Методы обработки естественного языка (NLP)
4. Токенизация текста и её особенности
5. Лемматизация и стемминг: сравнение методов
6. Морфологический анализ текста
7. Синтаксический анализ и его применение
8. Корпусная лингвистика и её роль в NLP
9. Модель «мешок слов» (Bag of Words)
10. Использование n-грамм в обработке текста
11. Метод TF-IDF и его применение
12. Векторные представления слов и текстов
13. Эмбединги слов и их использование
14. Модели Word2Vec и их особенности
15. Машинный перевод: методы и проблемы
16. Анализ тональности текста
17. Классификация текстов с использованием машинного обучения
18. Информационный поиск и обработка текстов
19. Разработка чат-ботов и диалоговых систем
20. Проблемы и перспективы развития компьютерной лингвистики

5.3. Перечень вопросов промежуточной аттестации по дисциплине

Вопросы к экзамену:

1. Понятие компьютерной лингвистики и её задачи
2. Основные области применения компьютерной лингвистики
3. Связь компьютерной лингвистики с искусственным интеллектом
4. Понятие обработки естественного языка (NLP)
5. Этапы обработки текстовых данных
6. Токенизация: понятие и методы
7. Лемматизация и стемминг: различия
8. Морфологический анализ текста
9. Синтаксический анализ: основные подходы
10. Построение синтаксических деревьев
11. Корпус текстов: понятие и использование
12. Методы представления текстовой информации
13. Модель «мешок слов» (Bag of Words)
14. n-граммы и их применение
15. Метод TF-IDF
16. Векторные представления слов
17. Эмбединги слов и их особенности

18. Модели Word2Vec
19. Сравнение методов представления текста
20. Машинный перевод: принципы и подходы
21. Анализ тональности текста
22. Методы классификации текстов
23. Информационный поиск: основные принципы
24. Работа поисковых систем с текстами
25. Чат-боты и диалоговые системы
26. Методы обработки диалогов
27. Оценка качества моделей NLP
28. Проблемы обработки естественного языка
29. Многозначность и омонимия в языке
30. Перспективы развития компьютерной лингвистики

Раздел 6. Оценочные средства промежуточной аттестации (с ключами)

1. Что такое компьютерная лингвистика?

- А) Раздел математики
- Б) Наука о компьютерной обработке языка
- В) Язык программирования
- Г) База данных

Правильный ответ: Б

2. Что такое обработка естественного языка (NLP)?

- А) Работа с изображениями
- Б) Анализ и обработка текстов и речи
- В) Программирование
- Г) Хранение данных

Правильный ответ: Б

3. Что такое токенизация?

- А) Удаление текста
- Б) Разделение текста на отдельные элементы
- В) Перевод текста
- Г) Сжатие данных

Правильный ответ: Б

4. Что такое лемматизация?

- А) Удаление слов
- Б) Приведение слова к начальной форме
- В) Перевод текста
- Г) Классификация текста

Правильный ответ: Б

5. Что такое стемминг?

- А) Определение смысла слова
- Б) Усечение слова до основы
- В) Перевод текста
- Г) Кодирование текста

Правильный ответ: Б

6. Что такое Bag of Words?

- А) Модель хранения файлов
- Б) Модель представления текста через частоты слов
- В) Алгоритм сортировки

Г) Тип базы данных

Правильный ответ: Б

7. Что такое TF-IDF?

А) Метод шифрования

Б) Метод оценки важности слов в тексте

В) Алгоритм кластеризации

Г) Тип нейронной сети

Правильный ответ: Б

8. Что такое эмбединги слов?

А) Хранение текста

Б) Векторное представление слов

В) Удаление слов

Г) Перевод текста

Правильный ответ: Б

9. Что такое анализ тональности текста?

А) Определение длины текста

Б) Определение эмоциональной окраски текста

В) Перевод текста

Г) Сжатие данных

Правильный ответ: Б

10. Что такое чат-бот?

А) Программа для хранения данных

Б) Программа для автоматического общения с пользователем

В) Операционная система

Г) Браузер

Правильный ответ: Б

Раздел 7. Перечень учебной литературы, необходимой для освоения дисциплины

7.1. Основная литература

1. Джурафски Д., Мартин Дж. Речевые и языковые технологии (Speech and Language Processing). 3-е изд. Москва : Вильямс, 2021. 1024 с.
2. Маннинг К., Шютце Х. Основы статистической обработки естественного языка. Москва : Вильямс, 2019. 704 с.
3. Голдберг Й. Нейронные сети и глубокое обучение для обработки естественного языка. Москва : ДМК Пресс, 2020. 512 с.
4. Бёрд С., Клейн Э., Лопер Э. Обработка естественного языка с Python. Москва : ДМК Пресс, 2018. 504 с.
5. Черняк В.Я. Введение в компьютерную лингвистику. Москва : Академия, 2020. 320 с.

7.2. Дополнительная литература

1. Гудфеллоу И., Бенджио Й., Курвилль А. Глубокое обучение. Москва : ДМК Пресс, 2018. 652 с.
2. Мёрфи К. Машинное обучение: вероятностный подход. Москва : Вильямс, 2020. 1100 с.
3. Джеймс Г., Уиттен Д., Хасты Т., Тибширани Р. Введение в статистическое обучение. Москва : ДМК Пресс, 2019. 456 с.
4. Мохри М., Ростамизаде А., Талвалькар А. Основы машинного обучения. Москва : ДМК Пресс, 2018. 432 с.
5. Юрьева Н.В. Компьютерная лингвистика и обработка текстов. Москва : Флинта, 2019. 288 с.

7.3. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Онлайн-курс «Natural Language Processing» (Coursera). Режим доступа: <https://www.coursera.org/specializations/natural-language-processing> (дата обращения: 10.03.2026).
2. Видеолекции по NLP (Stanford University). Режим доступа: <https://www.youtube.com/playlist?list=PLoROMvodv4rOf0G0y8Pp5kF1yRj0yHn2v> (дата обращения: 10.03.2026).
3. Документация библиотеки NLTK. Режим доступа: <https://www.nltk.org> (дата обращения: 10.03.2026).
4. Электронная библиотечная система «Лань». Режим доступа: <https://e.lanbook.com> (дата обращения: 10.03.2026).
5. Образовательная платформа «Юрайт». Режим доступа: <https://urait.ru> (дата обращения: 10.03.2026).

Раздел 8. Материально-техническая база и информационные технологии

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине:

Материально-техническое обеспечение дисциплины «Компьютерная лингвистика» включает в себя учебные аудитории для проведения занятий, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, самостоятельной работы обучающихся.

Учебные аудитории укомплектованы специализированной мебелью и техническими средствами обучения. Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети Интернет.

Дисциплина может реализовываться с применением дистанционных технологий обучения. Специфика реализации дисциплины с применением дистанционных технологий обучения устанавливается дополнением к рабочей программе. В части не противоречащей специфике, изложенной в дополнении к программе, применяется настоящая рабочая программа.

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине с применением дистанционных образовательных технологий включает в себя:

Компьютерная техника, расположенная в учебном корпусе Института (ул. Качинцев, 63, кабинет Центра дистанционного обучения):

1. Intel i 3 3.4Ghz\ОЗУ 4Gb\500GB\RadeonHD5450
2. Intel PENTIUM 2.9GHz\ОЗУ 4GB\500GB

3 личные электронные устройства (компьютеры, ноутбуки, планшеты и иное), а также средства связи преподавателей и студентов.

Информационные технологии, необходимые для осуществления образовательного процесса по дисциплине с применением дистанционных образовательных технологий включают в себя:

- система дистанционного обучения (СДО) (Learning Management System) (LMS) Moodle (Modular Object-Oriented Dynamic Learning Environment);

- электронная почта;
- система компьютерного тестирования;
- Цифровой образовательный ресурс IPR SMART;
- система интернет-связи skype;
- телефонная связь;
- ПО для организации конференций.

Обучение обучающихся инвалидов и обучающихся с ограниченными возможностями здоровья осуществляется посредством применения специальных технических средств в зависимости от вида нозологии.

При проведении учебных занятий по дисциплине используются мультимедийные комплексы, электронные учебники и учебные пособия, адаптированные к ограничениям здоровья обучающихся.

Лекционные аудитории оборудованы мультимедийными кафедрами, подключенными к звуковым колонкам, позволяющими усилить звук для категории слабослышащих обучающихся, а также проекционными экранами, которые увеличивают изображение в несколько раз и позволяют воспринимать учебную информацию обучающимся с нарушениями зрения.

При обучении лиц с нарушениями слуха используется усилитель слуха для слабослышащих людей Cyber Ear модель НАР-40, помогающий обучаемым лучше воспринимать учебную информацию.

Обучающиеся с ограниченными возможностями здоровья, обеспечены печатными и электронными образовательными ресурсами (программы, учебники, учебные пособия, материалы для самостоятельной работы и т.д.) в формах, адаптированных к ограничениям их здоровья и восприятия информации:

для лиц с нарушениями зрения:

- в форме электронного документа;
- в форме аудиофайла;

для лиц с нарушениями слуха:

- в печатной форме;
- в форме электронного документа;

для лиц с нарушениями опорно-двигательного аппарата:

- в печатной форме;
- в форме электронного документа;
- в форме аудиофайла.

Раздел 9. Методические указания для обучающихся по освоению дисциплины

Дисциплина включает практические занятия, самостоятельную работу обучающегося.

В ходе изучения дисциплины «Компьютерная лингвистика» перед обучающимися стоит задача не только закрепить знания о сложных информационных явлениях, о чем свидетельствует содержание тематического плана, глубоко разобраться в объемном учебном материале, но и сформировать у себя на основе полученных компьютерных знаний соответствующие профессионально важные качества.

Практические занятия – один из самых эффективных видов учебных занятий, на которых обучающиеся учатся творчески работать с различной информацией, являются также действенной формой активизации самостоятельной работы обучающихся.

Целью практических занятий является закрепление полученных в ходе лекций, а также в ходе самостоятельной работы над учебной и специальной литературой, знаний, умений и навыков. На практических занятиях особо обращается внимание на умение обучающихся проявлять элементы творчества в процессе самостоятельной работы, применять полученные знания на практике.

Практические занятия занимают центральное место в учебном процессе, так как позволяют на завершающем этапе усвоения материала, после прослушанной лекции и самостоятельного поиска дополнительных сведений по рассматриваемой проблематике, окончательно уточнить, сформировать свои позиции в ходе работы в составе учебной группы.

Основное в подготовке и проведении практикума – это самостоятельная работа обучающегося над изучением темы лекционного материала. Практические занятия проводятся по специальным планам – заданиям, которые содержатся в материалах, подготовленных на кафедре. Обучающийся обязан точно знать план занятия либо конкретное задание к нему.

При подготовке к практическим занятиям следует чаще обращаться к справочной литературе, полнее использовать консультации (групповые и индивидуальные, устные и письменные) с преподавателями, которые читают лекции и проводят практикумы.

Таким образом, в процессе подготовке к практическому занятию рекомендуется:

- ознакомиться с вопросами плана;
- прочитать конспект лекции по изучаемой теме;
- прочитать соответствующие главы учебников, статьи;
- просмотреть перечень научных источников, предлагаемых в рабочей программе, выбрав несколько из них для углубленного изучения данной темы.

По каждому практическому заданию обучающиеся отчитываются преподавателю, оформляя письменный отчет, в котором сохраняют результаты своей работы в виде файлов. Результаты выполнения практических заданий оцениваются с учетом теоретических знаний по соответствующим вопросам дисциплины и уровнем владения практическими навыками при работе на компьютере.

Для углубленного изучения и освоения материала целесообразно выполнение практических работ, наряду с другими различными формами обучения обучающихся: тесты, задачи, упражнения, которые используются при проведении практических занятий, выполнении контрольных и аудиторных работ, а также при самостоятельном изучении данной дисциплины.

Одним из наиболее интенсивных способов изучения дисциплины является самостоятельное выполнение практических работ, на которых вырабатываются навыки по дисциплине «Компьютерная лингвистика».

СРО позволяет глубже освоить теоретические и практические вопросы, понять принципы дисциплины «Компьютерная лингвистика».

Основными задачами организации процесса самостоятельной работы по дисциплине являются:

- приобретение знаний по теоретическим основам дисциплины «Компьютерная лингвистика», являющихся дополнением к материалу лекционных аудиторных занятий;
- приобретение практических навыков по дисциплине «Компьютерная лингвистика».

Основные формы реализации СРО – изучение учебно-методической литературы по дисциплине «Компьютерная лингвистика». В качестве базовой литературы можно использовать учебники и учебные пособия, согласно приведенному списку в разделе 6 рабочей программы, а также

любые другие источники информации, такие как электронные учебники, обучающие и энциклопедические сайты, публикации журналов и конференций.

Обучающийся допускается к зачетному занятию по результатам успешного выполнения всех практических заданий и самостоятельной работы.

Учебно-методическое издание

Рабочая программа учебной дисциплины

Компьютерная лингвистика

(Наименование дисциплины в соответствии с учебным планом)

Скоробогатченко Дмитрий Анатольевич

(Фамилия, Имя, Отчество составителя)