МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Рабочая программа дисциплины (модуля)

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

Рабочая программа дисциплины (модуля) «Компьютерная лингвистика» составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 09.04.03 Прикладная информатика (утв. приказом Минобрнауки России от 19.09.2017г. №916) и Порядком организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам бакалавриата, программам специалитета, программам магистратуры (утв. приказом Минобрнауки России от 06.04.2021 г. N245).

Составитель(и):

Клышинский Э.С.

Утверждена на заседании научно-образовательный центр "искусственный интеллект" от 27.05.2026 , протокол № 5

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)

Кригер А.Б.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1582918206
Номер транзакции	0000000000F71B54
Владелец	Кригер А.Б.

1 Цель, планируемые результаты обучения по дисциплине (модулю)

Целью освоения дисциплины «Компьютерная лингвистика» является формирование у студентов магистратуры представления о методах и средствах построения лингвистических процессов для различных прикладных задач по автоматической обработке текстов на естественном языке. Компьютерная лингвистика – это междисциплинарная область, которая возникла на стыке лингвистики, математики, информатики и искусственного интеллекта.

Задачи освоения дисциплины состоят формировании профессиональных компетенций, соответствующих виду профессиональной деятельности, на который ориентирована программа магистратуры.

Планируемыми результатами обучения по дисциплине (модулю), являются знания, умения, навыки. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы, представлен в таблице 1.

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины (модуля)

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине		
			Код результата	Формулировка результата	
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-2 : Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.1к : Решает профессиональные задачи, используя современные интеллектуальные технологии	РД1	Знание	этапов и существующих методик выполнения аналитических работ
			РД2	Умение	реализовывать методы обработки и анализа текстовой информации
			РД5	Навык	анализа текстовой информации
		ОПК-2.2к : Обосновывает выбор современных интеллектуальных технологий и программной среды при разработке оригинальных программных средств для решения профессиональных задач	РД3	Знание	различных методов обработки и анализа информации
			РД4	Умение	принимать решения по выбору используемых инструментов для обработки и анализа информации
			РД5	Умение	использовать возможности систем искусственного интеллекта при обработке и анализе текстовой информации
			РД6	Умение	использовать и модифицировать методы, модели, алгоритмы, технологии и инструментальные средства обработки и анализа текстовой информации при работе с большими данными
			РД6	Навык	компьютерной обработки данных

В процессе освоения дисциплины решаются задачи воспитания гармонично развитой, патриотичной и социально ответственной личности на основе традиционных российских духовно-нравственных и культурно-исторических ценностей, представленные в таблице 1.2.

Таблица 1.2 – Целевые ориентиры воспитания

Воспитательные задачи	**Формирование ценностей**	**Целевые ориентиры**
Формирование научного мировоззрения и культуры мышления		
Формирование осознания ценности научного мировоззрения и критического мышления	Гуманизм	Системное мышление
Формирование коммуникативных навыков и культуры общения		
Формирование навыков публичного выступления и презентации своих идей	Взаимопомощь и взаимоуважение	Умение работать в команде и взаимопомощь

2 Место дисциплины (модуля) в структуре ОПОП

Дисциплина относится к обязательной части блока 1 Дисциплины (модули) учебного плана.

3. Объем дисциплины (модуля)

Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу с обучающимися (по видам учебных занятий) и на самостоятельную работу, приведен в таблице 2.

Таблица 2 – Общая трудоемкость дисциплины

Название ОПОП ВО	Форма обуче-ния	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудо-емкость	Объем контактной работы (час)						СРС	Форма аттес-тации
				(З.Е.)	Всего	Аудиторная			Внеауди-торная			
						лек.	прак.	лаб.	ПА	КСР		
09.04.03 Прикладная информатика	ОФО	М01.Б	2	4	33	8	24	0	1	0	111	Э

4 Структура и содержание дисциплины (модуля)

4.1 Структура дисциплины (модуля) для ОФО

Тематический план, отражающий содержание дисциплины (перечень разделов и тем), структурированное по видам учебных занятий с указанием их объемов в соответствии с учебным планом, приведен в таблице 3.1

Таблица 3.1 – Разделы дисциплины (модуля), виды учебной деятельности и формы текущего контроля для ОФО

№	Название темы		Кол-во часов, отведенное на	Форма

		Код результата обучения	Лек	Практ	Лаб	СРС	текущего контроля
1	Язык программирования Python и среда разработки Jupyter Notebook		0	0	0	10	практическое задание
2	Общие этапы и модули обработки текстов	РД1, РД2, РД3, РД4, РД5, РД6	2	2	0	10	отчет по практической работе
3	Извлечение имен собственных, фактов	РД1, РД2, РД3, РД4, РД4, РД5, РД6, РД6, РД7	2	4	0	10	отчет по практической работе
4	Выделение коллокаций	РД1, РД2, РД4, РД4, РД5, РД6, РД6, РД7	2	4	0	10	отчет по практической работе
5	Использование модели Word2Vec при обработке текстов	РД2, РД3, РД4, РД5, РД6, РД6, РД7	0	2	0	20	отчет по практической работе
6	Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	РД2, РД3, РД4, РД5, РД6, РД6, РД7	0	4	0	20	отчет по практической работе
7	Классификаторы на основе деревьев принятия решений	РД2, РД3, РД4, РД5, РД6, РД6, РД7	2	4	0	20	отчет по практической работе
8	Анализ тональности текстов. Тематическое моделирование	РД2, РД4, РД5, РД6, РД6, РД7	0	4	0	11	отчет по практической работе
Итого по таблице			**8**	**24**	**0**	**111**	

4.2 Содержание разделов и тем дисциплины (модуля) для ОФО

Тема 1 Язык программирования Python и среда разработки Jupyter Notebook.

Содержание темы: Основные операторы языка Python. Арифметические операции. Логические операции. Списки, кортежи, множества. Индексирование списков. Словари. Условный оператор. Цикл while. Цикл for. Итераторы и генераторы. Функции. Импорт библиотек. Работа с файлами. Установка библиотек. Понятие регулярного выражения. Конструкции регулярных выражений. Библиотека Re.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практическое занятие.

Виды самостоятельной подготовки студентов по теме: изучение теоретических вопросов, компьютерное моделирование.

Тема 2 Общие этапы и модули обработки текстов.

Содержание темы: Проведение информационного поиска методом requests.get(url) с использованием библиотек requests, BeautyfulSoup и html5lib, а также с использованием регулярных выражений. Графематический и морфологический анализ. Морфопроцессы: стемминг и поиск по словарю. Система pymorphy2 и библиотека nltk. Мера расстояния между объектами: Евклидово расстояние, Манхэттенское расстояние, расстояние Жаккарда, корреляция, дивергенция Кулльбака-Лейблера, косинусная мера сходства.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе, домашняя работа.

Тема 3 Извлечение имен собственных, фактов.

Содержание темы: Задача извлечения именованных сущностей. Библиотека Natasha класс NamesExtractor. Библиотека Stanford NER. Библиотека networkx и построение графа связей между участниками событий. Мера кластерности. Синтаксический анализ (парсинг): дерево зависимостей или дерево составляющих. Библиотека UDPipe (токенизатор, морфология и система снятия омонимии (тэггинг)). Отображение дерева зависимостей при помощи библиотеки NetworkX.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе.

Тема 4 Выделение коллокаций.

Содержание темы: Методы коллокаций: анализ частоты сочетаний. Распределение Ципфа. Формула странности. Работа с графикой Matplotlib. Библиотека ipywidgets. Взаимодействие функции с элементом управления методом interact и без использования interact.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе.

Тема 5 Использование модели Word2Vec при обработке текстов.

Содержание темы: Векторное представление слов. Уменьшение размерности пространства. Преобразованием точек старого пространства в новое. Векторные операции - сложение и вычитание.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе.

Тема 6 Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм.

Содержание темы: Метод k-средних. Метод спектральной кластеризации. Метод DBSCAN. Иерархическая (аггломеративная) кластеризация. Дендрограмма. Кофенетическое расстояние между объектами. Кофенетическая корреляция. Методы нечеткой кластеризации: c-средних, FLAME. Многопоточность и создание ботов для Слака. Прогресс в циклах (tqdm). Библиотека multiprocessing. Очередь, события и семафоры.

Токенизатор и лемматизатор. Библиотека для создания бота для Телеграмм telebot. Библиотеки для многопоточного выполнения программы eventlet и multiprocessing.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе.

Тема 7 Классификаторы на основе деревьев принятия решений.

Содержание темы: Пример дерева принятия решения. Определение лучших разбиений. Меры неопределенности. Алгоритмы построения деревьев. Преимущества и недостатки построения деревьев.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе.

Тема 8 Анализ тональности текстов. Тематическое моделирование.

Содержание темы: Подходы к оценке тональности текстов: словари оценочной лексики и методы машинного обучения. Понятие тематического моделирования. Представление матрицы в виде произведения трех матриц: матрицы слово на тематику, матрицы тематик и матрицы тематика на документ.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическое занятие.

Виды самостоятельной подготовки студентов по теме: подготовка к практической работе.

5 Методические указания для обучающихся по изучению и реализации дисциплины (модуля)

5.1 Методические рекомендации обучающимся по изучению дисциплины и по обеспечению самостоятельной работы

Дисциплина ориентирована на изучение различных методов компьютерного анализа текста. На занятиях предлагаются задачи и спектр возможных решений. Это касается как методов и подходов, так и инструментария изучаемой среды. В конце каждой темы проводится обобщение материала и анализ полученных результатов.

Основные формы учебной работы: лекции, практические занятия и самостоятельная работа студентов. Студент должен посещать занятия, слушать преподавателя, осмысляя и конспектируя теоретическую часть занятия, и выполнять предложенные практические работы. Анализируя пройденный материал, понимать возможность использования полученных знаний, умений и навыков.

5.2 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

При необходимости обучающимся из числа лиц с ограниченными возможностями здоровья и инвалидов (по заявлению обучающегося) предоставляется учебная информация в доступных формах с учетом их индивидуальных психофизических особенностей:

- для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания, консультации и др.

- для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; индивидуальные задания, консультации и др.

6 Фонд оценочных средств для проведения текущего контроля и промежуточной аттестации обучающихся по дисциплине (модулю)

В соответствии с требованиями ФГОС ВО для аттестации обучающихся на соответствие их персональных достижений планируемым результатам обучения по дисциплине (модулю) созданы фонды оценочных средств. Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие этапы формирования компетенций в процессе освоения образовательной программы, представлены в Приложении 1.

7 Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература

1. Гуриков, С. Р. Основы алгоритмизации и программирования на Python : учебное пособие / С.Р. Гуриков. — Москва : ИНФРА-М, 2025. — 343 с. — (Высшее образование). - ISBN 978-5-16-020255-6. - Текст : электронный. - URL: https://znanium.ru/catalog/product/2166199 (дата обращения: 31.05.2026)

2. Загорулько, Ю. А. Искусственный интеллект. Инженерия знаний : учебное пособие для вузов / Ю. А. Загорулько, Г. Б. Загорулько. — Москва : Издательство Юрайт, 2024. — 93 с. — (Высшее образование). — ISBN 978-5-534-07198-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/540987 (дата обращения: 19.05.2026).

3. Максименко, О. И. Формализованная и цифровая лингвистика : монография / О. И. Максименко. — Москва : ГУП, 2024. — 148 с. — ISBN 978-5-7017-3499-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/512917 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.2 Дополнительная литература

1. Буторина, Е. П. Прикладная лингвистика : учебник для вузов / Е. П. Буторина. — Москва : Издательство Юрайт, 2026. — 148 с. — (Высшее образование). — ISBN 978-5-534-16994-2. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/600290 (дата обращения: 19.05.2026).

2. Машинное обучение с использованием Python : учебно-методическое пособие / составители А. В. Осин,К. А. Хализев. — Москва :МТУСИ, 2025. — 20 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/501209 (дата обращения: 25.05.2026). — Режим доступа: для авториз. пользователей.

7.3 Ресурсы информационно-телекоммуникационной сети "Интернет", включая профессиональные базы данных и информационно-справочные системы (при необходимости):

1. Образовательная платформа "ЮРАЙТ"
2. Электронно-библиотечная система "ZNANIUM.COM"
3. Электронно-библиотечная система "ЛАНЬ"
4. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/
5. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/
6. Информационно-справочная система "Консультант Плюс" - Режим доступа: http://www.consultant.ru/

8 Материально-техническое обеспечение дисциплины (модуля) и перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

Основное оборудование:

- Ист.бесп.эл.питания Smart-UPS 3000VA
- Мульт проектор №1 Panasonic PT-LX26HE
- Облачный монитор 23" LG CAV42K
- Облачный монитор LG Electronics черный +клавиатура+мышь
- Сетевой монитор:Нулевой клиент Samsung SyncMaster NC240
- Усилитель-распределитель VGA/XGA Kramer VP-200

Программное обеспечение:

- Firefox
- Microsoft Office 2010 Standart
- Microsoft Windows Profesional 8 Russian

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

НАУЧНО-ОБРАЗОВАТЕЛЬНЫЙ ЦЕНТР "ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ"

Фонд оценочных средств
для проведения текущего контроля
и промежуточной аттестации по дисциплине (модулю)

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Интеллектуальный анализ данных

Год набора на ОПОП
2025

Форма обучения
очная

Владивосток 2026

1 Перечень формируемых компетенций

Название ОПОП ВО, сокращенное	Код и формулировка компетенции	Код и формулировка индикатора достижения компетенции
09.04.03 «Прикладная информатика» (М-ПИ)	ОПК-2 : Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач	ОПК-2.1к : Решает профессиональные задачи, используя современные интеллектуальные технологии
		ОПК-2.2к : Обосновывает выбор современных интеллектуальных технологий и программной среды при разработке оригинальных программных средств для решения профессиональных задач

Компетенция считается сформированной на данном этапе в случае, если полученные результаты обучения по дисциплине оценены положительно (диапазон критериев оценивания результатов обучения «зачтено», «удовлетворительно», «хорошо», «отлично»). В случае отсутствия положительной оценки компетенция на данном этапе считается несформированной.

2 Показатели оценивания планируемых результатов обучения

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

Таблица 2.1 – Критерии оценки индикаторов достижения компетенции

Код и формулировка индикатора достижения компетенции	Результаты обучения по дисциплине			Критерии оценивания результатов обучения
	Код рез-та	Тип рез-та	Результат	
ОПК-2.1к : Решает профессиональные задачи, используя современные интеллектуальные технологии	РД 1	Знание	этапов и существующих методик выполнения аналитических работ	сформировавшееся знание этапов и существующих методик выполнения аналитических работ
	РД 2	Умение	реализовывать методы обработки и анализа текстовой информации	Сформировавшееся умение реализовывать методы обработки и анализа текстовой информации
	РД 5	Навык	анализа текстовой информации	Сформировавшееся владение навыками анализа текстовой информации
ОПК-2.2к : Обосновывает выбор современных интеллектуальных технологий и программной среды при разработке оригинальных программных средств для решения профессиональных задач	РД 3	Знание	различных методов обработки и анализа информации	сформировавшееся знание методов для обработки и анализа информации
	РД 4	Умение	принимать решения по выбору используемых инструментов для обработки и анализа информации	сформировавшееся умение принимать решения по выбору используемых инструментов для обработки и анализа информации
	РД 5	Умение	использовать возможности систем искусственного интеллекта при обработке и анализе текстовой информации	Сформировавшееся умение использовать возможности систем искусственного интеллек

				та при обработке и анализе текстовой информации
	РД 6	Навык	компьютерной обработки данных	Сформировавшееся владение навыками компьютерной обработки данных
	РД 6	Умение	использовать и модифицировать методы, модели, алгоритмы, технологии и инструментальные средства обработки и анализа текстовой информации и при работе с большими данными	Сформировавшееся умение и спользовать и модифицировать методы, модели, алгоритмы, технологии и инструментальные средства обработки и анализа текстовой информации и при работе с большими данными

Таблица заполняется в соответствии с разделом 1 Рабочей программы дисциплины (модуля).

3 Перечень оценочных средств

Таблица 3 – Перечень оценочных средств по дисциплине (модулю)

Контролируемые планируемые результаты обучения		Контролируемые темы дисциплины	Наименование оценочного средства и представление его в ФОС	
			Текущий контроль	Промежуточная аттестация
Очная форма обучения				
РД1	Знание : этапов и существующих методик выполнения аналитических работ	1.2. Общие этапы и модули обработки текстов	Практическая работа	Тест
		1.3. Извлечение имен собственных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
РД2	Умение : реализовывать методы обработки и анализа текстовой информации	1.2. Общие этапы и модули обработки текстов	Практическая работа	Тест
		1.3. Извлечение имен собственных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
		Практическая работа	Тест	
		1.5. Снижение размерности пространства	Практическая работа	Тест
		Практическая работа	Тест	
		1.6. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Тест

		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
		1.8. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Тест
РД3	Умение : систематизировать и обобщать информацию	1.2. Общие этапы и модули обработки текстов		
		1.3. Извлечение именованных, фактов		
РД3	Знание : различных методов обработки и анализа информации	1.5. Использование модели Word2Vec при обработке текстов	Практическая работа	Тест
		1.6. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Тест
		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
РД4	Умение : принимать решения по выбору используемых инструментов для обработки и анализа информации	1.2. Общие этапы и модули обработки текстов	Практическая работа	Тест
		1.3. Извлечение именованных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
		Практическая работа	Тест	
		1.5. Снижение размерности пространства	Практическая работа	Тест
		Практическая работа	Тест	
		1.6. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Тест
		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
		1.8. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Тест
РД4	Умение : учитывать особенности обработки текстовой информации и возможностей доступа к ней	1.3. Извлечение именованных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
		Практическая работа	Тест	
РД5	Навык : анализа текстовой информации	1.2. Общие этапы и модули обработки текстов	Практическая работа	Тест
РД5	Умение : использовать возможности систем иску	1.3. Извлечение именованных, фактов	Практическая работа	Тест

	сственного интеллекта при обработке и анализе текстовой информации	1.4. Выделение коллокаций	Практическая работа	Тест
		1.5. Снижение размерности пространства	Практическая работа	Тест
		Практическая работа	Тест	
		1.6. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Тест
		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
		1.8. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Тест
РД6	Навык : компьютерной обработки данных	1.2. Общие этапы и модули обработки текстов	Практическая работа	Тест
		1.3. Извлечение имен собственных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
		Практическая работа	Тест	
		1.5. Снижение размерности пространства	Практическая работа	Тест
		Практическая работа	Тест	
		1.6. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Тест
		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
		1.8. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Тест
РД6	Умение : использовать и модифицировать методы, модели, алгоритмы, технологии и инструментальные средства обработки и анализа текстовой информации при работе с большими данными	1.3. Извлечение имен собственных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
		Практическая работа	Тест	
		1.5. Снижение размерности пространства	Практическая работа	Тест
		Практическая работа	Тест	
		1.6. Кластеризация. Параллельная обработка данных. Создание бота дл	Практическая работа	Тест

		я Slack. Разработка бота для Телеграмм		
		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
		1.8. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Тест
РД7	Умение : разрабатывать методы обработки и анализа текстовой информации при работе с большими данными	1.3. Извлечение имен собственных, фактов	Практическая работа	Тест
		1.4. Выделение коллокаций	Практическая работа	Тест
		Практическая работа	Тест	
		1.5. Снижение размерности пространства	Практическая работа	Тест
		Практическая работа	Тест	
		1.6. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Тест
		1.7. Классификаторы на основе деревьев принятия решений	Практическая работа	Тест
		1.8. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Тест

4 Описание процедуры оценивания

Качество сформированности компетенций на данном этапе оценивается по результатам текущих и промежуточных аттестаций при помощи количественной оценки, выраженной в баллах. Максимальная сумма баллов по дисциплине (модулю) равна 100 баллам.

Вид учебной деятельности	Практическая работа	Тест	Итого
Лекции		10	10
Практические занятия		40	40
Самостоятельная работа	30		30
Промежуточная аттестация		20	20
Итого	30	70	100

Сумма баллов, набранных студентом по всем видам учебной деятельности в рамках дисциплины, переводится в оценку в соответствии с таблицей.

Сумма баллов по дисциплине	Оценка по промежуточной аттестации	Характеристика качества сформированности компетенции
от 91 до 100	«зачтено» / «отлично»	Студент демонстрирует сформированность дисциплинарных компетенций, обнаруживает всестороннее, систематическое и глубокое знание учебного материала, усвоил основную литературу и знаком с дополнительной литературой, рекомендованной программой, умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными знаниями, умениями, применяет их в ситуациях повышенной сложности.

от 76 до 90	«зачтено» / «хорошо»	Студент демонстрирует сформированность дисциплинарных компетенций: основные знания, умения освоены, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе знаний и умений на новые, нестандартные ситуации.
от 61 до 75	«зачтено» / «удовлетворительно»	Студент демонстрирует сформированность дисциплинарных компетенций: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных знаний, умений, навыков по некоторым дисциплинарным компетенциям, студент испытывает значительные затруднения при оперировании знаниями и умениями при их переносе на новые ситуации.
от 41 до 60	«не зачтено» / «неудовлетворительно»	У студента не сформированы дисциплинарные компетенции, проявляется недостаточность знаний, умений, навыков.
от 0 до 40	«не зачтено» / «неудовлетворительно»	Дисциплинарные компетенции не сформированы. Проявляется полное или практически полное отсутствие знаний, умений, навыков.

5 Примерные оценочные средства

5.1 Примеры заданий для выполнения практических работ

Тема 1. «*Язык программирования Python и среда разработки Jupyter Notebook*»

Тема вынесена полностью на самостоятельное изучение, точнее повторение, т.к. в учебном плане предусмотрена дисциплина «*Методы статистического анализа и прогнозирования на языке R*», которая и формирует у студента компетенции по базовым знаниям, умениям и навыкам работы с языком R: *Основные операторы языка Python. Арифметические операции. Логические операции. Списки, кортежи, множества. Индексирование списков. Словари. Условный оператор. Цикл while. Цикл for. Итераторы и генераторы. Функции. Импорт библиотек. Работа с файлами. Установка библиотек. Регулярные выражения.*

Проверка знаний по Теме 1 осуществляется по результату выполненного индивидуального ***домашнего задания:***

1. Написать программу, которая открывает текстовый файл и считывает его построчно.
2. В каждой строке необходимо найти один из фрагментов…..
3. Найденные фрагменты необходимо сохранить в другой текстовый файл, один фрагмент - одна строка.
4. Основной алгоритм должен быть оформлен как функция.

Тема 2. «Общие этапы и модули обработки текстов». Демонстрация понимания принципа обработки текстов. Д***омашнее задание:***

1. Выгрузить текстовые страницы с какого-нибудь сайта (не используемого на занятии), разметить их выбранной морфологией.
2. Написать функцию для построения векторов частот слов.
3. Посчитать меру сходства между несколькими страницами.

Краткие методические указания

Самостоятельная работа студента заключается в освоении теоретического и практического материала по использованию среды в своей профессиональной деятельности.

Шкала оценки

оценка	Баллы	Описание
5	23–30	Студент демонстрирует умения на итоговом уровне: умеет свободно выполнять практические задания, предусмотренные программой, свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.
4	15–22	Студент демонстрирует умения на среднем уровне: освоил основные умения, но допускаются незначительные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.
3	7–14	Студент демонстрирует умения и навыки на базовом уровне: в ходе контрольных мероприятий допускаются значительные ошибки, проявляется отсутствие отдельных умений, навыков по дисциплинарной компетенции, испытываются значительные затруднения при оперировании умениями и при их переносе на новые ситуации.
2	0–6	Студент демонстрирует умения и навыки на уровне ниже базового: проявляется недостаточность умений и навыков.

5.2 Примеры тестовых заданий

1. Объяснение фактов языка является задачей

а) компьютерной лингвистики

б) теоретической лингвистики

в) прикладной лингвистики

г) математической лингвистики

2. Большие банки хранения информации, которыми пользуются специалисты по гуманитарным наукам, являются достижением

а) компьютерной лингвистики

б) теоретической лингвистики

в) прикладной лингвистики

г) математической лингвистики

3. К новым методам анализа языка, которые использует прикладная лингвистика, относятся

а) методы классификации и систематизации языкового материала

б) метод моделирования

в) статистический метод

г) описательный метод

4. Компьютерная лингвистика как прикладная дисциплина выделяется, прежде всего

а) по новым методам анализа языка

б) по использованию компьютерных средств обработки языковых данных

в) по использованию математических методов исследования языковых данных

5. Машинным переводом занимается

а) математическая лингвистика

б) компьютерная лингвистика

в) структурная лингвистика

г) социолингвистика

6. Основным достоинством компьютерных словарей является

а) удобная компактная форма

б) возможность активного использования при изучении иностранного языка

в) скорость поиска информации

г) соответствуют текущей языковой и культурной ситуации

7. При изучении иностранных языков наиболее важным преимуществом компьютерного словаря является

а) возможность быстрого поиска слова

б) возможность практиковаться в произношении

в) возможность формировать «блокноты»

8. Найти слова, написания которых пользователю не известны, можно в компьютерном словаре

а) Partner

б) Collins

в) Longman

9. Показу контекстного значения слова в компьютерном словаре способствует

а) функция текстового редактора

б) функция «гипертекст»

в) функция озвучивания слова

г) функция графического изображения

10. Основной недостаток машинного перевода - это

а) дороговизна времени работы компьютера

б) невозможность получить полноценный перевод

в) невозможность оперативно его использовать

г) невозможность последующего редактирования текста

Краткие методические указания

Суть тестовых заданий заключается в проверке знаний по теоретическому и практическому материала по использованию среды в своей профессиональной деятельности.

Шкала оценки

№	Баллы	Описание
5	61–70	Процент правильных ответов от 95% до 100%
4	51–60	Процент правильных ответов от 80 до 94%
3	41–50	Процент правильных ответов от 65 до 79%
2	26–40	Процент правильных ответов от 45 до 64%
1	0–25	Процент правильных ответов менее 45%

ФОС и ключи для ФОС
по дисциплине «Компьютерная лингвистика»

5.1 Примеры заданий для выполнения практических работ

Тема 1. «Язык программирования Python и среда разработки Jupyter Notebook»
Тема вынесена полностью на самостоятельное изучение, точнее повторение, т.к. в учебном плане предусмотрена дисциплина «*Методы статистического анализа и прогнозирования на языке R*», которая и формирует у студента компетенции по базовым знаниям, умениям и навыкам работы с языком R: *Основные операторы языка Python. Арифметические операции. Логические операции. Списки, кортежи, множества. Индексирование списков. Словари. Условный оператор. Цикл while. Цикл for. Итераторы и генераторы. Функции. Импорт библиотек. Работа с файлами. Установка библиотек. Регулярные выражения.*
Проверка знаний по Теме 1 осуществляется по результату выполненного индивидуального **домашнего задания:**

1. Провести выгрузку текстов из новостного ресурса, указанного преподавателем.
2. Результат сохранить в файл с разделителями для заметок.

Тема 2. «Общие этапы и модули обработки текстов». Демонстрация понимания принципа обработки текстов. **Домашнее задание:**

1. Провести синтаксический анализ выгруженных текстов.
2. Выделить участников новостных сюжетов.
3. Рассчитать статистику встречаемости для участников новостных сюжетов, провести анализ статистики.

5.2 Примеры тестовых заданий

1. При помощи t-SNE и UMAP можно получить новое пространство, точки в котором ...

A. расположены точно так же, как и в исходном пространстве;

B. повернуты, отмасштабированы и спроецированы на новое пространство более низкой размерности;

C. расположены случайным образом;

D. по возможности остаются близкими, если были близки, и остаются далекими, если были далеко в исходном пространстве.

2. В результате снижения размерности пространство признаков качество классификации

A. увеличится

B. не изменится

C. уменьшится

D. изменится непредсказуемым образом

3. Нужна ли целевая переменная при кластеризации?

A. Конечно да

B. Конечно нет

C. Может пригодиться

4. Что позволяет оценить метод Силуэт? Речь идёт о кластеризации.

A. Количество кластеров, на которые делятся данные.

B. Границы кластеров.

C. Оба варианта.

5. Операция свёртки -- это

A. Преобразование данных в векторное пространство;

B. Операция снижения размерности векторного пространства;

C. Скалярное произведение двух векторов, один из кототрых -- вектор свёртки.

6. Полносвязанный слой нейронной сети используется для построения:

A. персептрона;

B. рекуррентной сети;

C. свёрточной сети;

D. трансформера;

E. всех перечисленных вариантов.

7. У вас имеется три непересекающихся облака точек в многомерном пространстве. Какой метод подойдёт для их классификации?

A. Любой линейный метод.

B. К-средних

C. Любой бустинг.

D. PCA

8. Модель трансформера использует:

A. механизм перекрестного внимания;

B. механизм самовнимания;

C. механизм усреднения векторов токенов;

D. несколько механизмов из вышеперечисленных.

9. Какой метод используется для визуализации многомерных данных, сохраняя локальные расстояния между точками?

A. PCA

B. LDA

C. t-SNE

D. SVM

10. Какой тип данных НЕ является стандартным в Python?

A. Список

B. Кортеж

C. Массив

D. Множество

11. Какой оператор в Python используется для проверки принадлежности элемента к множеству?

A. ==

B. in
C. contains
D. is

12. Что делает метод split() в Python?
A. Объединяет строки
B. Разделяет строку по указанному разделителю
C. Удаляет пробелы
D. Заменяет подстроку

13. Какой алгоритм кластеризации основан на идее "ближайших соседей"?
A. PCA
B. K-средних
C. SVM
D. Линейная регрессия

14. Какой инструмент используется для лемматизации текста в библиотеке NLTK?
A. nltk.tokenize
B. nltk.stem.WordNetLemmatizer
C. nltk.corpus
D. nltk.pos_tag

15. Какой метод снижения размерности сохраняет глобальную структуру данных?
A. PCA
B. t-SNE
C. UMAP
D. LDA

16. Какой тип нейронной сети наиболее подходит для обработки последовательностей?
A. Свёрточная сеть
B. Рекуррентная сеть
C. Перцептрон
D. Полносвязная сеть

17. Какой параметр в алгоритме K-средних определяет количество кластеров?
A. epsilon
B. n_clusters
C. max_iter
D. random_state

18. Какой метод используется для оценки качества кластеризации?
A. Точность (Accuracy)
B. Метод силуэта (Silhouette Score)
C. F1-мера
D. ROC-кривая

19. Какой из перечисленных методов НЕ является методом обучения с учителем?
A. Линейная регрессия
B. SVM
C. K-средних
D. Дерево решений

20. Какой механизм лежит в основе модели Transformer?
A. Свёртка
B. Самовнимание (Self-Attention)
C. Рекуррентные связи
D. Бустинг

5.3 Ключи к тесту

1. D
2. D
3. C
4. C
5. C
6. E
7. C
8. B
9. C
10. C
11. B
12. B
13. B
14. B
15. A
16. B
17. B
18. B
19. C
20. B