МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ЭКОНОМИКИ И СЕРВИСА

КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

Рабочая программа дисциплины (модуля)

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Искусственный интеллект и машинное обучение в управлении и принятии решений

Год набора на ОПОП
2020

Форма обучения
очная

Владивосток 2021

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 05.04.2017 г. N301).

Составитель(и):

Клышинский Э.С., кандидат технических наук, доцент, Кафедра информационных технологий и систем

Утверждена на заседании кафедры информационных технологий и систем от 31.05.2021 , протокол № 9

СОГЛАСОВАНО:

Заведующий кафедрой (разработчика)
Кийкова Е.В.

ДОКУМЕНТ ПОДПИСАН ЭЛЕКТРОННОЙ ПОДПИСЬЮ	
Сертификат	1575633692
Номер транзакции	000000000071794D
Владелец	Кийкова Е.В.

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 профессиональной компетенции, соответствующей 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-1 : Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.2к : Решает нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных социально-экономических и профессиональных знаний	РД1	Знание	современных методов и инструментальных средств компьютернй обработки текстов
	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	РД3	Навыки	информационного поиска и компьютерного анализа текста

	ОПК-3 : Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем	РД2	Умение	анализировать текст с использованием современных информационных технологий и систем

2 Место дисциплины (модуля) в структуре ОПОП

Освоение дисциплины формирует у обучающихся компетенцию, необходимую для подготовки магистранта в соответствии с требованиями ФГОС ВО.

Входными требованиями к изучению дисциплины «Компьютерная лингвистика» является наличие у студентов компетенций, сформированных на предыдущем уровне образования.

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 обучения	Часть УП	Семестр (ОФО) или курс (ЗФО, ОЗФО)	Трудоемкость	Объем контактной работы (час)						СРС	Форма аттестации
				(З.Е.)	Всего	Аудиторная			Внеаудиторная			
						лек.	прак.	лаб.	ПА	КСР		
09.04.03 Прикладная информатика	ОФО	М01.Б	2	4	17	4	12	0	1	0	127	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	Код результата обучения	Кол-во часов, отведенное на				Форма текущего контроля
			Лек	Практ	Лаб	СРС	

1	Язык программирования Python и среда разработки Jupyter Notebook.	РД1	0	0	0	15	проверка индивидуального домашнего задания
2	Общие этапы и модули обработки текстов	РД1	1	1	0	12	выполнение практического задания
3	Извлечение имен собственных, фактов	РД2	0	1	0	12	выполнение практического задания
4	Выделение коллокаций	РД2	0	2	0	12	выполнение практического задания
5	Методы классификации	РД2	1	1	0	12	выполнение практического задания
6	Снижение размерности пространства	РД3	0	2	0	12	выполнение практического задания
7	Использование модели Word2Vec при обработке текстов	РД3	0	1	0	12	выполнение практического задания
8	Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	РД3	0	2	0	12	выполнение практического задания
9	Классификаторы на основе деревьев принятия решений	РД3	1	1	0	13	выполнение практического задания
10	Анализ тональности текстов. Тематическое моделирование	РД3	1	1	0	15	выполнение практического задания
Итого по таблице			**4**	**12**	**0**	**127**	

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: самостоятельная работа.

Виды самостоятельной подготовки студентов по теме: Повторение основ работы с языком Python. Выполнение индивидуальног домашнего задания.

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, практическая работа.

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

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: практическая работа.

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: практическая работа.

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

Тема 5 Методы классификации.

Содержание темы: Метод k ближайших соседей (k-NN). Метод fit. Метод predict. Линейная регрессия. Логистическая регрессия. Библиотека для обработки и анализа данных Pandas. Чтение данных из различных источников. Загрузка и запись данных. Настройка читаемых данных. Описание данных. Выборка данных. Индексирование по позиции. Визуализация данных с использованием библиотеки Seaborn .

Формы и методы проведения занятий по теме, применяемые образовательные технологии: лекция, практическая работа.

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

Тема 6 Снижение размерности пространства.

Содержание темы: Метод главных компонент PCA (Principle Component Analysis). Метод MDS (Multidimensional Scaling). Метод t-SNE (t-distributed stochastic neighbor embedding). Метод UMAP (Uniform Manifold Approximation and Projection). Графическая интерпретация результатов работы методов снижения размерности пространства.

Формы и методы проведения занятий по теме, применяемые образовательные технологии: практическая работа.

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

Тема 7 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: практическая работа.

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

Тема 8 Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм.

Содержание темы: Метод k-средних. Метод спектральной кластеризации. Метод DBSCAN. Иерархическая (агломеративная) кластеризация. Дендрограмма. Кофенетическое расстояние между объектами. Кофенетическая корреляция. Методы нечеткой кластеризации:

с-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: практическая работа.

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

Тема 9 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, практическая работа.

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

Тема 10 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: лекция, практическая работа.

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 среды R и 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, лабораторные работы и самостоятельная работа студентов. Студент должен посещать занятия, слушать преподавателя, осмысляя и конспектируя теоретическую часть занятия, и выполнять предложенные лабораторные 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. Буйначев С. К., Боклаг Н. Ю. Основы программирования на языке Python : Учебники и учебные пособия для ВУЗов [Электронный ресурс] - Екатеринбург : Издательство Уральского университета , 2014 - 92 - Режим доступа: http://biblioclub.ru/index.php?page=book_red&id=275962

7.2 Дополнительная литература

1. Гуриков Сергей Ростиславович. Основы алгоритмизации и программирования на Python : Учебное пособие [Электронный ресурс] , 2018 - 343 - Режим доступа: http://znanium.com/go.php?id=924699

2. Сараев П. В. Методы машинного обучения [Электронный ресурс] , 2017 - 48 - Режим доступа: https://lib.rucont.ru/efd/670997

3. Федоров Д. Ю. ПРОГРАММИРОВАНИЕ НА ЯЗЫКЕ ВЫСОКОГО УРОВНЯ PYTHON. Учебное пособие для прикладного бакалавриата [Электронный ресурс] : М.:Издательство Юрайт , 2019 - 126 - Режим доступа: https://biblio-online.ru/book/programmirovanie-na-yazyke-vysokogo-urovnya-python-444065

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

1. Электронная библиотечная система «РУКОНТ» - Режим доступа: http://biblioclub.ru/
2. Электронная библиотечная система «РУКОНТ» - Режим доступа: https://lib.rucont.ru/
3. Электронная библиотечная система ZNANIUM.COM - Режим доступа: http://znanium.com/
4. Электронно-библиотечная система издательства "Юрайт" - Режим доступа: https://biblio-online.ru/
5. Open Academic Journals Index (OAJI). Профессиональная база данных - Режим доступа: http://oaji.net/

6. Президентская библиотека им. Б.Н.Ельцина (база данных различных профессиональных областей) - Режим доступа: https://www.prlib.ru/

7. 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:

- Microsoft Office Professional Plus 2013 Russian
- Microsoft Windows Profesional 8 Russian

МИНОБРНАУКИ РОССИИ

ВЛАДИВОСТОКСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ЭКОНОМИКИ И СЕРВИСА

КАФЕДРА ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ И СИСТЕМ

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

КОМПЬЮТЕРНАЯ ЛИНГВИСТИКА

Направление и направленность (профиль)
09.04.03 Прикладная информатика. Искусственный интеллект и машинное обучение в управлении и принятии решений

Год набора на ОПОП
2020

Форма обучения
очная

Владивосток 2021

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-1 : Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте	ОПК-1.2к : Решает нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных социально-экономических и профессиональных знаний
	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-3 : Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями	ОПК-3.1к : Анализирует требования и создает сценарии использования технических и программных систем

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-1 «Способен самостоятельно приобретать, развивать и применять математические, естественнонаучные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте»

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-1.2к : Решает нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных социально-экономических и профессиональных знаний	РД1	Знание	современных методов и инструментальных средств компьютерной обработки текстов	сформировавшееся знание современных методов и инструментальных средств компьютерной обработки текстов

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

Таблица 2.2 – 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	РД3	Навыки	информационного поиска и компьютерного анализа текста	сформировавшиеся владение навыками информационного поиска и компьютерного анализа текста

Компетенция ОПК-3 «Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями»

Таблица 2.3 – 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-3.1к : Анализирует требования и создает сценарии использования технических и программных систем	РД2	Умение	анализировать текст с использованием современных информационных технологий и систем	сформировавшееся умение анализировать текст с использованием современных информационных технологий и систем

Таблица заполняется в соответствии с разделом 2 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 : современных методов и инструментальных средств компьютерной обработки текстов	1.1. Язык программирования Python и среда разработки Jupyter Notebook .	Практическая работа	Разноуровневые задачи и задания
		1.2. Общие этапы и модули обработки текстов	Практическая работа	Разноуровневые задачи и задания

РД2	Умение : анализировать текст с использованием современных информационных технологий и систем	1.3. Извлечение имен собственных, фактов	Практическая работа	Разноуровневые задачи и задания
		1.4. Выделение коллокаций	Практическая работа	Разноуровневые задачи и задания
		1.5. Методы классификации	Практическая работа	Разноуровневые задачи и задания
РД3	Навыки : информационного поиска и компьютерного анализа текста	1.6. Снижение размерности пространства	Практическая работа	Разноуровневые задачи и задания
		1.7. Использование модели Word2Vec при обработке текстов	Практическая работа	Разноуровневые задачи и задания
		1.8. Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм	Практическая работа	Разноуровневые задачи и задания
		1.9. Классификаторы на основе деревьев принятия решений	Практическая работа	Разноуровневые задачи и задания
		1.10. Анализ тональности текстов. Тематическое моделирование	Практическая работа	Разноуровневые задачи и задания

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	Индивидуальное домашнее задание	Итого
Лекции			
практические занятия	60		60
Самостоятельная работа		10	10
Промежуточная аттестация		20	20
Итого	70	30	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, нет очности, 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

Перечень тем и задач практической части курса

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

Самостоятельная работа студентов: повторение базовых принципов работы с языком программирования Python и выполнение домашней работы (на проверку остаточных знаний)

Тема 2. Общие этапы и модули обработки текстов.

Написание робота для скачивания новостей с сайта Лента.Ру и их фильтровки в зависимости от интересов пользователя. (Решение задачи разными способами). Проведение информационного поиска методом requests.get(url) с использованием библиотек requests, BeautyfulSoup и html5lib, а также с использованием регулярных выражений.

Мера расстояния между объектами. Расчет косинусной меры сходства для разных статей.

Тема 3 Извлечение имен собственных, фактов.

Задача извлечения именованных сущностей с помощью библиотеки Natasha. Оформление класса выгруженного с Лента.ру отдельным файлом. Построение графа связей между участниками событий. Расчет меры кластерности для всех вершин графа.

Определение связи между словами с помощью синтаксического анализа на примере загрузки ленты новостей. Построение дерева зависимостей и отображение его при помощи библиотеки NetworkX.

Тема 4 Выделение коллокаций.

Выделение коллокаций в новостях разными методами: анализ частоты сочетаний; распределение Ципфа; частота по контрастивному корпусу. Работа с графикой. Взаимодействие функции с элементами управления методом interact и без использования interact.

Тема 5 Методы классификации

Метод k ближайших соседей на примере классификации жертв и выживших на «Титанике». Линейная и логистическая регрессия для данных по «Титанику».

Библиотека для обработки и анализа данных Pandas. Чтение данных из различных источников. Загрузка и запись данных. Настройка читаемых данных. Описание данных. Методы данного класса: pandas: head, tail; pandas: shape; pandas: columns, index; pandas: info; pandas: describe; pandas: Series; pandas: map. Выборка данных. Индексирование по позиции. Сортировка, группировка, агрегированность данных по функции.

Визуализация данных с использованием библиотеки Seaborn: распределение возраста пассажиров.

Тема 6 Снижение размерности пространства

Обработка данных (линия со случайными смещениями и плоская область, идущая вдоль той же линии; два непересекающихся кластера; данные о пассажирах «Титаника»; автоматическое определение языка статьи из потока научных статей на русском и украинском языках) разными методами: методом главных компонент PCA; методом MDS;

методом t-SNE; методом UMAP. Графическая интерпретация результатов работы методов снижения размерности пространства.

Тема 7 Использование модели Word2Vec при обработке текстов

Векторное представление слов в скачанных новостях с сайта Лента.Ру. Работа с моделью по уменьшению размерности пространства и преобразованию точек старого пространства в новое для выполнения векторных операций. Отбор новостей с помощью модели Word2Vec.

Тема 8 Кластеризация. Параллельная обработка данных. Создание бота для Slack. Разработка бота для Телеграмм.

Работа с разными методами кластеризации: метод k-средних; метод спектральной кластеризации; метод DBSCAN; иерархическая кластеризация; дендрограмма.

Получение новостей из нескольких источников. Синхронизация операций в потоках. Создание бота для Slack.

Работа с библиотекой telebot по созданию бота для Телеграмм для размещения новостных сообщений.

Тема 9 Классификаторы на основе деревьев принятия решений

Классификация научных статей разной тематики по разделам науки с использованием дерева принятия решения. Определение лучших разбиений. Меры неопределенности. Алгоритмы построения деревьев. Преимущества и недостатки построения деревьев

Тема 10 Анализ тональности текстов. Тематическое моделирование.

Подходы к оценке тональности текстов: словарный (словари оценочной лексики) и на основе методов машинного обучения (корпуса текстов, размеченные по тональности). Библиотека BeautifulSoup и BeautifulStoneSoup. Классы оценочной лексики: positive, negative, neutral и both.

Понятие тематического моделирования. Представление матрицы в виде произведения трех матриц: матрицы слово на тематику, матрицы тематик и матрицы тематика на документ. Библиотека BigARTM

Краткие методические указания

На выполнение практического задания отводится не более двух академических часов. После выполнения практической работы студент должен продемонстрировать преподавателю результаты выполнения работы.

Шкала оценки

оценка	Баллы	Описание
5	46–60	Студент демонстрирует умения на итоговом уровне: умеет свободно выполнять практические задания , предусмотренные программой, свободно оперирует приобретенными умениями, применяет их в ситуациях повышенной сложности.
4	31–45	Студент демонстрирует умения на среднем уровне: освоил основные умения, но допускаются незначи тельные ошибки, неточности, затруднения при аналитических операциях, переносе умений на новые, нестандартные ситуации.
3	16–30	Студент демонстрирует умения и навыки на базовом уровне: в ходе контрольных мероприятий допуск аются значительные ошибки, проявляется отсутствие отдельных умений, навыков по дисциплинарной компетенции, испытываются значительные затруднения при оперировании умениями и при их перено се на новые ситуации.
2	0–15	Студент демонстрирует умения и навыки на уровне ниже базового: проявляется недостаточность умен ий и навыков.

5.2 Варианты индивидуальных домашних заданий

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 незначи тельные 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 допуск аются значительные ошибки, проявляется отсутствие отдельных умений, навыков по дисциплинарной компетенции, испытываются значительные затруднения при оперировании умениями и при их перено се на новые ситуации.
2	0–6	Студент демонстрирует умения и навыки на уровне ниже базового: проявляется недостаточность умен ий и навыков.