

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ
федеральное государственное автономное образовательное учреждение высшего
образования
"САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
АЭРОКОСМИЧЕСКОГО ПРИБОРОСТРОЕНИЯ"

Кафедра № 43

УТВЕРЖДАЮ

Руководитель образовательной программы

доц.,к.ф.н.

(должность, уч. степень, звание)

Е.А. Рудая

(инициалы, фамилия)



(подпись)

«16» февраля 2026 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Обработка текстовых массивов»

(Наименование дисциплины)

Код направления подготовки/ специальности	45.03.02
Наименование направления подготовки/ специальности	Лингвистика
Наименование направленности/ специализации	Теоретическая и прикладная лингвистика
Форма обучения	заочная
Год приема	2026

Лист согласования рабочей программы дисциплины

Программу составил (а)

старший преподаватель
(должность, уч. степень, звание)

16.02.26
(подпись, дата)

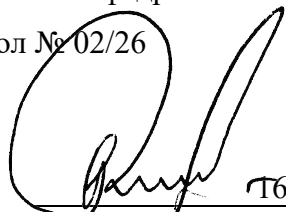
Н.А. Соловьева
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 43

«16» февраля 2026 г, протокол № 02/26

Заведующий кафедрой № 43

д.т.н., проф.
(уч. степень, звание)

16.02.26
(подпись, дата)

М.Ю. Охтилев
(инициалы, фамилия)

Заместитель директора института №8 по методической работе

доц., к.э.н., доц.
(должность, уч. степень, звание)


(подпись, дата)

Л.В. Рудакова
(инициалы, фамилия)

Аннотация

Дисциплина «Обработка текстовых массивов» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/специальности 45.03.02 «Лингвистика» направленности/специализации «Теоретическая и прикладная лингвистика». Дисциплина реализуется кафедрой «№43».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

ПК-9 «Способность работать с основными информационно-поисковыми и экспертными системами, системами представления знаний, системами автоматизированного перевода»

Содержание дисциплины охватывает круг вопросов, связанных с представлением текстовых данных и методами их автоматизированной обработки, развитием умений и навыков в применении программного обеспечения для редакции, коррекции и анализа текстов.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, лабораторные работы, самостоятельная работа обучающегося.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме дифференцированного зачета (9 семестр).

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц, 180 часов.

Язык обучения по дисциплине «русский»

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины

Дисциплина «Обработка текстовых массивов» ставит своей целью усвоение обучающимися знаний о форматах представления текстовых данных и методах их автоматизированной обработки, развитие умений и навыков в применении программного обеспечения для редакции, коррекции и анализа текстов, формирование компетенций в области проектирования и использования программных средств обработки текстов.

1.2. Дисциплина входит в состав части, формируемой участниками образовательных отношений, образовательной программы высшего образования (далее – ОП ВО).

1.3. Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения ОП ВО.

В результате изучения дисциплины обучающийся должен обладать следующими компетенциями или их частями. Компетенции и индикаторы их достижения приведены в таблице 1.

Таблица 1 – Перечень компетенций и индикаторов их достижения

Категория (группа) компетенции	Код и наименование компетенции	Код и наименование индикатора достижения компетенции
Профессиональные компетенции	ПК-9 Способность работать с основными информационно-поисковыми и экспертными системами, системами представления знаний, системами автоматизированного перевода	ПК-9.3.1 знать принципы, положенные в основу работы существующих программных продуктов, знать системы обработки естественного языка и машинного перевода ПК-9.У.1 уметь использовать результаты поиска для научно-исследовательской работы и в профессиональной деятельности ПК-9.В.1 владеть навыками поиска, анализа, систематизации и обобщения полученной научной информации

2. Место дисциплины в структуре ОП

Дисциплина может базироваться на знаниях, ранее приобретенных обучающимися при изучении следующих дисциплин:

– «Информатика»

Знания, полученные при изучении материала данной дисциплины, имеют самостоятельное значение.

3. Объем и трудоемкость дисциплины

Данные об общем объеме дисциплины, трудоемкости отдельных видов учебной работы по дисциплине (и распределение этой трудоемкости по семестрам) представлены в таблице 2.

Таблица 2 – Объем и трудоемкость дисциплины

Вид учебной работы	Всего	Трудоемкость по семестрам
		№9
1	2	3
Общая трудоемкость дисциплины, ЗЕ/ (час)	5/ 180	5/ 180
Из них часов практической подготовки	12	12

Аудиторные занятия , всего час.	24	24
в том числе:		
лекции (Л), (час)	12	12
практические/семинарские занятия (ПЗ), (час)		
лабораторные работы (ЛР), (час)	12	12
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)		
Самостоятельная работа , всего (час)	156	156
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.)	Дифф. зач.,	Дифф. зач.,

4. Содержание дисциплины

4.1. Распределение трудоемкости дисциплины по разделам и видам занятий.

Разделы, темы дисциплины и их трудоемкость приведены в таблице 3.

Таблица 3 – Разделы, темы дисциплины, их трудоемкость

Разделы, темы дисциплины	Лекции (час)	ПЗ (СЗ)	ЛР (час)	КП/КР (час)	СР (час)
Семестр 9					
Раздел 1. Формальные грамматики и метаязык	2				16
Раздел 2. Структуры данных языка python и их применение для хранения текстовой информации	4		4		50
Раздел 3. Работа с текстовыми и табличными файлами на языке python	2		6		50
Раздел 4. Автоматический анализ текста с помощью библиотеки NLTK языка python	2		2		20
Раздел 5. Тезаурусы и онтологии	2				20
Итого в семестре:	12		12		156
Итого	12	0	12	0	156

Практическая подготовка заключается в непосредственном выполнении обучающимися определенных трудовых функций, связанных с будущей профессиональной деятельностью.

4.2. Содержание разделов и тем лекционных занятий.

Содержание разделов и тем лекционных занятий приведено в таблице 4.

Таблица 4 – Содержание разделов и тем лекционного цикла

Номер раздела	Название и содержание разделов и тем лекционных занятий
1	Раздел 1. Формальные грамматики и метаязык Темы: состав формальной грамматики, виды формальных грамматик, форма Бэкуса-Наура, контекстно-свободная грамматика, метаязык
2	Раздел 2. Структуры данных языка python и их применение для хранения текстовой информации Темы: список, словарь, кортеж, строка, составные структуры данных
3	Раздел 3. Работа с текстовыми и табличными файлами на языке python

	Темы: способы чтения и записи текстовых файлов (формат .txt), файлов формата json и табличных файлов .exl
4	Раздел 4. Автоматический анализ текста с помощью библиотеки NLTK языка python Темы: обзор библиотек для анализа текстов на естественном языке, применение библиотеки NLTK, стоп-слова, токенизация, стемминг, лемматизация, векторизация, мешок слов, морфологический анализ
5	Раздел 5. Тезаурусы и онтологии Темы: онтология, принципы классификации онтологий, языки описания онтологий, тезаурус

4.3. Практические (семинарские) занятия

Темы практических занятий и их трудоемкость приведены в таблице 5.

Таблица 5 – Практические занятия и их трудоемкость

№ п/п	Темы практических занятий	Формы практических занятий	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Учебным планом не предусмотрено					
Всего					

4.4. Лабораторные занятия

Темы лабораторных занятий и их трудоемкость приведены в таблице 6.

Таблица 6 – Лабораторные занятия и их трудоемкость

№ п/п	Наименование лабораторных работ	Трудоемкость, (час)	Из них практической подготовки, (час)	№ раздела дисциплины
Семестр 9				
1	Использование в программе на языке python составных структуры данных	4	4	2
2	Сохранение/чтение в файл формата .txt структур данных разного типа	2	2	3
3	Сохранение/чтение в файл формата .json структур данных разного типа	2	2	3
4	Работа с файлами табличного редактора Excel с помощью языка python	2	2	3
5	Использование библиотеки NLTK	2	2	4
Всего		12		

4.5. Выполнение курсового проекта/ курсовой работы

Учебным планом не предусмотрено

4.6. Самостоятельная работа обучающихся

Виды самостоятельной работы и ее трудоемкость приведены в таблице 7.

Таблица 7 – Виды самостоятельной работы и ее трудоемкость

Вид самостоятельной работы	Всего, час	Семестр 9, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	56	56
Подготовка к текущему контролю успеваемости (ТКУ)	50	50
Подготовка к промежуточной аттестации (ПА)	50	50
Всего:	156	156

5. Перечень учебно-методического обеспечения

для самостоятельной работы обучающихся по дисциплине (модулю)

Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. разделов 6-11.

6. Перечень печатных и электронных учебных изданий

Перечень печатных и электронных учебных изданий приведен в таблице 8.

Таблица 8– Перечень печатных и электронных учебных изданий

Шифр/ URL адрес	Библиографическая ссылка	Количество экземпляров в библиотеке (кроме электронных экземпляров)
URL: https://e.lanbook.com/book/454175 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Всеволодова, А. В. Компьютерная обработка лингвистических данных : учебное пособие / А. В. Всеволодова. — Ярославль : МУБиНТ, 2024. — 102 с. — ISBN 978-5-93002-406-7. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/195144 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Лутошкина, Н. В. Модели знаний и онтологии : учебное пособие / Н. В. Лутошкина. — Красноярск : СибГУ им. академика М. Ф. Решетнёва, 2021. — 80 с. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/246464 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Абдуллаева, З. М. Лингвистическое обеспечение информационных систем : учебное пособие / З. М. Абдуллаева, Ю. И. Родионова, С. В. Удахина. — Санкт-Петербург : ИЭО СПбУТУиЭ, 2021. — 163 с. — ISBN 978-5-94047- 835-5. — Текст : электронный // Лань	

	: электронно-библиотечная система. —	
URL: https://e.lanbook.com/book/462479 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Андропова, О. Ю. Искусственный интеллект и язык программирования Python : учебное пособие / О. Ю. Андропова, И. И. Васильева, Н. А. Гнездилова. — Елец : ЕГУ им. И.А. Бунина, 2024. — 106 с. — ISBN 978-5-00151-413-8. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/437747 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Аргунеев, Э. П. Основы прикладных лингвистических информационных технологий : учебное пособие / Э. П. Аргунеев. — Оренбург : ОГУ, 2024. — 109 с. — ISBN 978-5-7410-3255-8. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/512917 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	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/454175 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Всеволодова, А. В. Компьютерная обработка лингвистических данных : учебное пособие / А. В. Всеволодова. — Ярославль : МУБиНТ, 2024. — 102 с. — ISBN 978-5-93002-406-7. — Текст : электронный // Лань : электронно-библиотечная система.	
— URL: https://e.lanbook.com/book/106069 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Гуслякова, А. В. Информационные технологии и лингвистика XXI века : учебное пособие / А. В. Гуслякова. — Москва : МПГУ, 2016. — 96 с. — ISBN 978-5-4263-0398-0. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/497525 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Чернышев, С. А. Python: продвинутый уровень : учебное пособие / С. А. Чернышев. — Санкт-Петербург : ГУАП, 2024. — 244 с. — ISBN 978-5-8088-1969-6. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/504279 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Канева, О. Н. Введение в программирование на языке Python : учебное пособие / О. Н. Канева, Т. Ю. Финк. — Омск : ОмГТУ, 2024. — 149 с. — ISBN 978 5 8149 3864 0. —	

	Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/518818 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей	Алюнина, Ю. М. Цифровые технологии в переводе : учебное пособие для вузов / Ю. М. Алюнина. — 2-е изд., стер. — Санкт-Петербург : Лань, 2026. — 144 с. — ISBN 978-5- 507-56717-1. — Текст : электронный // Лань : электронно-библиотечная система..	
URL: https://e.lanbook.com/book/75843 (дата обращения: 11.02.2026) Режим доступа: для авториз. пользователей.	Короткова, М. А. Задачник по курсу "Математическая лингвистика и теория автоматов : учебное пособие / М. А. Короткова, Е. Е. Трифонова. — Москва : НИЯУ МИФИ, 2012. — 92 с. — ISBN 978-5-7262-1702-4. — Текст : электронный // Лань : электронно- библиотечная система.	
URL: https://e.lanbook.com/book/452363 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Двоенко, С. Д. Введение в методы трансляции : учебное пособие / С. Д. Двоенко. — Тула : ТулГУ, 2024. — 150 с. — ISBN 978-5-7679-5521-3. — Текст : электронный // Лань : электронно-библиотечная система.	
URL: https://e.lanbook.com/book/169596 (дата обращения: 11.02.2026). — Режим доступа: для авториз. пользователей.	Каледин, В. О. Концепции языков программирования : учебное пособие / В. О. Каледин. — Новокузнецк : КГПИ КемГУ, 2012. — 141 с. — ISBN 978-5-8353-0788-3. — Текст : электронный // Лань : электронно- библиотечная система.	

7. Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины приведен в таблице 9.

Таблица 9 – Перечень электронных образовательных ресурсов информационно-телекоммуникационной сети «Интернет»

URL адрес	Наименование
aot.ru	Сайт «Автоматическая обработка текста»
python.org	Сайт, посвященный языку python

8. Перечень информационных технологий

8.1. Перечень программного обеспечения, используемого при осуществлении образовательного процесса по дисциплине.

Перечень используемого программного обеспечения представлен в таблице 10.

Таблица 10– Перечень программного обеспечения

№ п/п	Наименование
1	Spider - оболочка для языка python. Распространяется свободно

8.2. Перечень информационно-справочных систем, используемых при осуществлении образовательного процесса по дисциплине

Перечень используемых информационно-справочных систем представлен в таблице 11.

Таблица 11– Перечень информационно-справочных систем

№ п/п	Наименование
	Не предусмотрено

9. Материально-техническая база

Состав материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине, представлен в таблице 12.

Таблица 12 – Состав материально-технической базы

№ п/п	Наименование составной части материально-технической базы	Номер аудитории (при необходимости)
1	Лекционная аудитория	
2	Компьютерный класс	

10. Оценочные средства для проведения промежуточной аттестации

10.1. Состав оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине приведен в таблице 13.

Таблица 13 – Состав оценочных средств для проведения промежуточной аттестации

Вид промежуточной аттестации	Перечень оценочных средств
Дифференцированный зачёт	Список вопросов; Тесты.

10.2. В качестве критериев оценки уровня сформированности (освоения) компетенций обучающимися применяется 5-балльная шкала оценки сформированности компетенций, которая приведена в таблице 14. В течение семестра может использоваться 100-балльная шкала модульно-рейтинговой системы Университета, правила использования которой, установлены соответствующим локальным нормативным актом ГУАП.

Таблица 14 –Критерии оценки уровня сформированности компетенций

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«отлично» «зачтено»	Обучающийся: – глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно связывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий. – правильно выполнил от 90% до 100% тестовых заданий.

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
«хорошо» «зачтено»	Обучающийся: – твердо усвоил программный материал, грамотно и по существу излагает его, опираясь на знания основной литературы; – не допускает существенных неточностей; – увязывает усвоенные знания с практической деятельностью направления; – аргументирует научные положения; – делает выводы и обобщения; – владеет системой специализированных понятий. – правильно выполнил от 70% до 89% тестовых заданий.
«удовлетворительно» «зачтено»	– обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы; – допускает несущественные ошибки и неточности; – испытывает затруднения в практическом применении знаний направления; – слабо аргументирует научные положения; – затрудняется в формулировании выводов и обобщений; – частично владеет системой специализированных понятий. – правильно выполнил от 51% до 69% тестовых заданий.
«неудовлетворительно» «не зачтено»	– обучающийся не усвоил значительной части программного материала; – допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении; – испытывает трудности в практическом применении знаний; – не может аргументировать научные положения; – не формулирует выводов и обобщений. – правильно выполнил менее 51% тестовых заданий**.

10.3. Типовые контрольные задания или иные материалы.

Вопросы (задачи) для экзамена представлены в таблице 15.

Таблица 15 – Вопросы (задачи) для экзамена

№ п/п	Перечень вопросов (задач) для экзамена	Код индикатора
	Учебным планом не предусмотрено	

Вопросы (задачи) для зачета / дифф. зачета представлены в таблице 16.

Таблица 16 – Вопросы (задачи) для зачета / дифф. зачета

№ п/п	Перечень вопросов (задач) для зачета / дифф. зачета	Код индикатора
1	Язык python. Структура данных «Список из списков»	ПК-9.3.1
2	Язык python. Структура данных «Список из словарей»	ПК-9.3.1
3	Язык python. Структура данных «Словарь из списков»	ПК-9.3.1
4	Язык python. Структура данных «Словарь из словарей»	ПК-9.3.1
5	Методы файловых объектов языка python	ПК-9.3.1
6	Доступ к файлам табличного редактора из языка питон: библиотеки, чтение файла	ПК-9.3.1 ПК-9.У.1 ПК-9.В.1
7	Файлы формата json и их запись/чтение с помощью языка python	ПК-9.3.1 ПК-9.У.1 ПК-9.В.1
8	Решение проблемы определения ключевых слов текста	ПК-9.3.1

9	Обзор библиотек для анализа текстов на естественном языке	ПК-9.3.1 ПК-9.У.1 ПК-9.В.1
10	Применение библиотеки NLTK	ПК-9.3.1 ПК-9.У.1
11	Векторизация	ПК-9.3.1
12	Токенизация, стемминг, лемматизация	ПК-9.3.1
13	Морфологический анализ	ПК-9.3.1
14	Мешок слов	ПК-9.3.1
15	Формальные грамматики: классификация Хомского, символы метаязыка, рекурсивное правило	ПК-9.3.1
16	Пример формальной грамматики	ПК-9.3.1
17	форма Бэкуса-Наура	ПК-9.3.1
18	Работа с сайтом aot.ru	ПК-9.3.1 ПК-9.У.1 ПК-9.В.1
19	Тезаурус	ПК-9.3.1 ПК-9.В.1
20	Онтология	ПК-9.3.1 ПК-9.В.1

Перечень тем для выполнения курсового проекта/ курсовой работы представлены в таблице 17.

Таблица 17 – Перечень тем для выполнения курсового проекта / курсовой работы

№ п/п	Примерный перечень тем для выполнения курсового проекта/ курсовой работы
	Учебным планом не предусмотрено

Вопросы для проведения промежуточной аттестации в виде тестирования представлены в таблице 18.

Таблица 18 – Примерный перечень вопросов для тестов

№ п/п	Примерный перечень вопросов для тестов	Код индикатора
1	Инструкция: выберите несколько ответов Какие существуют режимы открытия файлов? <ol style="list-style-type: none"> 1. открыть файл для добавления записи в конец файла 2. открыть файл для добавления записи в середину файла 3. открыть файл для записи 4. открыть файл для чтения 5. открыть файл для дублирования 	ПК-9.3.1 ПК-9.У.1 ПК-9.В.1
2	Инструкция: Прочитайте текст и запишите развернутый обоснованный ответ. Из каких элементов состоит текстовый файл?	ПК-9.3.1 ПК-9.У.1
3	Инструкция: выберите один ответ Какое понятие задает следующее определение: «система, состоящая из набора понятий и набора утверждений об этих понятиях»?	ПК-9.3.1 ПК-9.В.1

	1. Онтология 2. Аксиома 3. Иерархия 4. Когнитивность																	
4	Инструкция: выберите несколько ответов Какие существуют принципы классификаций онтологий? 1. По степени формальности 2. По содержанию 3. По объему 4. По цели создания 5. По языку	ПК-9.3.1 ПК-9.B.1																
5	Инструкция: Каждому термину из левого столбца подберите определение из правого столбца. <table><tr><td>A</td><td>RDF</td><td>1</td><td>Расширяемый язык разметки</td></tr><tr><td>B</td><td>OWL</td><td>2</td><td>Язык описания ресурсов, основанный на триплетях «субъект-предикат-объект»</td></tr><tr><td>C</td><td>XML</td><td>3</td><td>Электронный словарь-тезаурус и набор семантических сетей для английского языка.</td></tr><tr><td>D</td><td>WordNet</td><td>4</td><td>Язык описания онтологий в терминах классов и свойств</td></tr></table>	A	RDF	1	Расширяемый язык разметки	B	OWL	2	Язык описания ресурсов, основанный на триплетях «субъект-предикат-объект»	C	XML	3	Электронный словарь-тезаурус и набор семантических сетей для английского языка.	D	WordNet	4	Язык описания онтологий в терминах классов и свойств	ПК-9.3.1 ПК-9.Y.1 ПК-9.B.1
A	RDF	1	Расширяемый язык разметки															
B	OWL	2	Язык описания ресурсов, основанный на триплетях «субъект-предикат-объект»															
C	XML	3	Электронный словарь-тезаурус и набор семантических сетей для английского языка.															
D	WordNet	4	Язык описания онтологий в терминах классов и свойств															
6	Инструкция: Запишите в правильном порядке иерархию онтологий с точки зрения цели создания: 1. онтология верхнего уровня 2. онтология представления 3. прикладная онтология 4. онтология предметной области	ПК-9.3.1 ПК-9.B.1																
7	Инструкция: Прочитайте текст и запишите развернутый обоснованный ответ. Что такое тезаурус?	ПК-9.3.1 ПК-9.B.1																
8	Инструкция: выберите один ответ Как называется большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач? 1. База данных 2. Лингвистический корпус 3. База знаний	ПК-9.3.1 ПК-9.B.1																

	4. Тезаурус																	
9	<p>Инструкция: выберите несколько ответов</p> <p>Как называются библиотеки для обработки естественного языка для python?</p> <p>1. pandas 2. SciPy 3. NLTK 4. spaCy 5. Gensim</p>	ПК-9.3.1 ПК-9.У.1 ПК-9.В.1																
10	<p>Инструкция: для каждого термина из левого столбца подберите его правильное описание из правого столбца.</p> <table><tr><td>A</td><td>Токенизация</td><td>1</td><td>Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.</td></tr><tr><td>B</td><td>Стемминг</td><td>2</td><td>Процесс приведения словоформы к её нормальной (словарной) форме.</td></tr><tr><td>C</td><td>Лемматизация</td><td>3</td><td>Преобразование текста в числовой формат.</td></tr><tr><td>D</td><td>Векторизация</td><td>4</td><td>Процесс разбиения текста на более мелкие части (слова или предложения)</td></tr></table>	A	Токенизация	1	Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.	B	Стемминг	2	Процесс приведения словоформы к её нормальной (словарной) форме.	C	Лемматизация	3	Преобразование текста в числовой формат.	D	Векторизация	4	Процесс разбиения текста на более мелкие части (слова или предложения)	ПК-9.3.1
A	Токенизация	1	Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.															
B	Стемминг	2	Процесс приведения словоформы к её нормальной (словарной) форме.															
C	Лемматизация	3	Преобразование текста в числовой формат.															
D	Векторизация	4	Процесс разбиения текста на более мелкие части (слова или предложения)															

Перечень тем контрольных работ по дисциплине обучающихся заочной формы обучения, представлены в таблице 19.

Таблица 19 – Перечень контрольных работ

№ п/п	Перечень контрольных работ
1	Структуры данных языка python – список, словарь, кортеж, строка, множество

10.4. Методические материалы, определяющие процедуры оценивания индикаторов, характеризующих этапы формирования компетенций, содержатся в локальных нормативных актах ГУАП, регламентирующих порядок и процедуру проведения текущего контроля успеваемости и промежуточной аттестации обучающихся ГУАП.

11. Методические указания для обучающихся по освоению дисциплины

11.1. Методические указания для обучающихся по освоению лекционного материала.

Основное назначение лекционного материала – логически стройное, системное, глубокое и ясное изложение учебного материала. Назначение современной лекции в рамках дисциплины не в том, чтобы получить всю информацию по теме, а в освоении фундаментальных проблем дисциплины, методов научного познания, новейших достижений научной мысли. В учебном процессе лекция выполняет методологическую, организационную и информационную функции. Лекция раскрывает понятийный аппарат

конкретной области знания, её проблемы, дает цельное представление о дисциплине, показывает взаимосвязь с другими дисциплинами.

В лекционной части курса дается объяснение изучаемого материала, рассматриваются примеры составления программ для решения типовых задач лингвистики. Для успешного освоения дисциплины обучающиеся должны перед каждой лекцией просматривать конспект и повторять материал, рассмотренный на предыдущих занятиях.

11.2. Методические указания для обучающихся по выполнению лабораторных работ

Задание и требования к проведению лабораторных работ

Задание заключается в написании программы на языке программирования питон. После проверки работоспособности программы следует оформить отчет и устно ответить преподавателю на вопросы по тексту отчета.

Темы лабораторных занятий и их трудоемкость приведены в таблице 5.

Структура и форма отчета о лабораторной работе

Отчет должен содержать следующие разделы: титульный лист, цель работы, постановка задачи, результаты выполнения заданий по лабораторной работе (структуры данных, текст программы, копии экранов с работающим приложением, реализованным в рамках задания).

Требования к оформлению отчета о лабораторной работе

Отчет о выполнении лабораторной работы должен быть представлен в электронном виде (файл). В отчет включается текст подготовленной программы с комментариями.

Методические указания по выполнению лабораторных работ приведены в электронных ресурсах кафедры 43 в локальной сети кафедры:

\\dcbm\Методическое обеспечение кафедры 43\Обработка текстовых массивов

11.3. Методические указания для обучающихся по прохождению самостоятельной работы

В ходе выполнения самостоятельной работы, обучающийся выполняет работу по заданию и при методическом руководстве преподавателя, но без его непосредственного участия.

Для обучающихся по заочной форме обучения выполнение контрольных работ является элементом текущего контроля успеваемости и самостоятельной работы.

В процессе выполнения самостоятельной работы у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет ему развивать умения и навыки в усвоении и систематизации приобретаемых знаний, обеспечивает высокий уровень успеваемости в период обучения, помогает получить навыки повышения профессионального уровня.

Методическими материалами, направляющими самостоятельную работу обучающихся являются:

- учебно-методический материал по дисциплине;
- методические указания по выполнению контрольных работ (для обучающихся по заочной форме обучения).

Методические указания по прохождению самостоятельной работы приведены в электронных ресурсах кафедры 43 в локальной сети кафедры:

\\dcbm\Методическое обеспечение кафедры 43\Обработка текстовых массивов

11.4. Методические указания для обучающихся по прохождению текущего контроля успеваемости.

Текущий контроль успеваемости предусматривает контроль качества знаний обучающихся, осуществляемого в течение семестра с целью оценивания хода освоения дисциплины.

Текущий контроль успеваемости заключается в устной защите контрольной и лабораторных работ. На защите обучающийся должен: прокомментировать все термины, встречающиеся в тексте отчета; объяснить алгоритм решения задачи; описать инструменты, с помощью которых был достигнут результат; по просьбе преподавателя внести изменения в текст программы, чтобы получить решение измененной задачи. Баллы, полученные за выполнение и защиту каждой работы, суммируются и учитываются при выставлении итоговой оценки.

11.5. Методические указания для обучающихся по прохождению промежуточной аттестации.

Промежуточная аттестация обучающихся предусматривает оценивание промежуточных и окончательных результатов обучения по дисциплине. Она включает в себя:

- дифференцированный зачет – это форма оценки знаний, полученных обучающимся при изучении дисциплины, при выполнении курсовых проектов, курсовых работ, научно-исследовательских работ и прохождении практик с аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

Оценка за дифференцированный зачет выставляется на основе суммы баллов, которые получены за контрольную работу, лабораторные работы и устный ответ на зачете на вопросы (см. Таблицу 16).

Лист внесения изменений в рабочую программу дисциплины

Дата внесения изменений и дополнений. Подпись внесшего изменения	Содержание изменений и дополнений	Дата и № протокола заседания кафедры	Подпись зав. кафедрой