

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

Кафедра № 43

УТВЕРЖДАЮ

Ответственный за образовательную  
программу

к.ф.н., доц.

(должность, уч. степень, звание)

М.А. Чиханова

(инициалы, фамилия)



(подпись)

«27» июня 2024 г

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Обработка текстовых массивов»  
(Наименование дисциплины)

Код направления подготовки/ специальности	45.03.02
Наименование направления подготовки/ специальности	Лингвистика
Наименование направленности	Теоретическая и прикладная лингвистика
Форма обучения	заочная
Год приема	2024

Лист согласования рабочей программы дисциплины

Программу составил (а)

старший преподаватель  
(должность, уч. степень, звание)

14.06.24  
(подпись, дата)

Н.А. Соловьева  
(инициалы, фамилия)

Программа одобрена на заседании кафедры № 43

«17» июня 2024 г, протокол № 05/2024

Заведующий кафедрой № 43

д.т.н., проф.  
(уч. степень, звание)



17.06.2024  
(подпись, дата)

М.Ю. Охтилев  
(инициалы, фамилия)

Заместитель декана факультета №6 по методической работе

проф., д.и.н., доц.  
(должность, уч. степень, звание)



27.06.24  
(подпись, дата)

Л.Ю. Гусман  
(инициалы, фамилия)

## Аннотация

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

Дисциплина нацелена на формирование у выпускника следующих компетенций:

УК-1 «Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач»

УК-2 «Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений»

ПК-1 «Способность работать с основными информационно-поисковыми и экспертными системами, системами представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации и автоматизированного перевода, автоматизированными системами идентификации и верификации личности»

ПК-2 «Владение системой лингвистических знаний, включающей в себя знание основных фонетических, лексических, грамматических, словообразовательных явлений и закономерностей функционирования изучаемого иностранного языка и русского языка, их функциональных разновидностей»

ПК-3 «Владение методами проведения лингвистических экспертиз»

ПК-5 «Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности»

ПК-7 «Владение основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов»

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, формированием компетенций в области проектирования и использования программных средств обработки 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 экзамена.

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы, 144 часа.

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

1.2. 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
Универсальные компетенции	УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	УК-1.В.2 владеть навыками системного подхода для решения поставленных задач
Универсальные компетенции	УК-2 Способен определять круг задач в рамках поставленной цели и выбирать оптимальные способы их решения, исходя из действующих правовых норм, имеющихся ресурсов и ограничений	УК-2.У.1 уметь проводить анализ поставленной цели и формулировать задачи, которые необходимо решить для ее достижения
Профессиональные компетенции	ПК-1 Способность работать с основными информационно-поисковыми и экспертными системами, системами представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической	ПК-1.3.1 знать основные информационно-поисковые и экспертные системы

	информации и автоматизированного перевода, автоматизированными системами идентификации и верификации личности	
Профессиональные компетенции	ПК-2 Владение системой лингвистических знаний, включающей в себя знание основных фонетических, лексических, грамматических, словообразовательных явлений и закономерностей функционирования изучаемого иностранного языка и русского языка, их функциональных разновидностей	ПК-2.У.1 уметь пользоваться методами формального и когнитивного моделирования естественного языка и методами создания метаязыков
Профессиональные компетенции	ПК-3 Владение методами проведения лингвистических экспертиз	ПК-3.У.1 уметь применять на практике различные методы лингвистических экспертиз, оценивая соответствие лингвистического объекта кодифицированным нормам современного русского языка
Профессиональные компетенции	ПК-5 Способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-лингвистических технологий и с учетом основных требований информационной безопасности	ПК-5.В.1 владеть технологиями построения тезаурусов и онтологий
Профессиональные компетенции	ПК-7 Владение основными математико-статистическими	ПК-7.У.1 уметь с учетом элементов программирования и автоматической обработки лингвистических корпусов обрабатывать лингвистическую

	методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов	информацию ПК-7.В.1 владеть навыками программирования и навыками автоматической обработки корпусов
--	---	--

## 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 как самостоятельное значение, так и могут использоваться при изучении других дисциплин:

- Информационные языки

## 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	
		№7	№8
1	2	3	4
<b>Общая трудоемкость дисциплины, ЗЕ/ (час)</b>	4/ 144	2/ 72	2/ 72
<b>Из них часов практической подготовки</b>	18	10	8
<b>Аудиторные занятия, всего час.</b>	52	28	24
в том числе:			
лекции (Л), (час)	26	14	12
практические/семинарские занятия (ПЗ), (час)			
лабораторные работы (ЛР), (час)	26	14	12
курсовой проект (работа) (КП, КР), (час)			
экзамен, (час)	9	9	
<b>Самостоятельная работа, всего (час)</b>	83	35	48
<b>Вид промежуточной аттестации:</b> зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.)	Экз., Зачет	Экз.	Зачет

## 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) (час)	ЛР (час)	КП (час)	СРС (час)
Семестр 7					
Раздел 1. Системный подход и задачи компьютерной лингвистики	4				5

Раздел 2. Структуры данных языка питон и их применение для хранения текстовой информации	4		11		10
Раздел 3. Тезаурусы и онтологии	2				10
Раздел 4. Формальные грамматики и метаязык	1				5
Раздел 5. Работа с текстовыми файлами на языке python	3		3		5
Итого в семестре:	14		14		35
Семестр 8					
Раздел 6. Работа с табличными файлами на языке python	2		5		8
Раздел 7. Автоматический анализ текста с помощью библиотеки NLTK языка python	3		7		10
Раздел 8. Искусственный интеллект	2				10
Раздел 9. Информационно-поисковые системы	3				10
Раздел 10. Лингвистический корпус	2				10
Итого в семестре:	12		12		48
Итого	26	0	26	0	83

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. Системный подход и задачи компьютерной лингвистики Темы: Система, классификация систем, свойства системы, системный подход, системный анализ, язык как система, цели и задачи, декомпозиция, средства анализа систем: DPD и UML, предметная область, задачи компьютерной лингвистики
2	Раздел 2. Структуры данных языка питон и их применение для хранения текстовой информации Темы: список, словарь, кортеж, строка, составные структуры данных
3	Раздел 3. Тезаурусы и онтологии Темы: онтология, принципы классификации онтологий, языки описания онтологий, тезаурус
4	Раздел 4. Формальные грамматики и метаязык Темы: состав формальной грамматики, виды формальных грамматик, форма Бэкуса-Наура, контекстно-свободная грамматика, метаязык
5	Раздел 5. Работа с текстовыми файлами на языке python Темы: способы чтения и записи текстовых файлов
6	Раздел 6. Работа с табличными файлами на языке python Темы: библиотеки для чтения/записи файлов .exl
7	Раздел 7. 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
8	Раздел 8. Искусственный интеллект Темы: приемы разработки программ с искусственным интеллектом, база знаний,

	экспертные системы, нейронные сети, голосовой помощник
9	Раздел 9. Информационно-поисковые системы Темы: информационно-поисковые системы, поисковые системы, индексация, пузырь фильтров
10	Раздел 10. Лингвистический корпус Темы: корпусная лингвистика, создание корпуса, известные корпуса, разметка корпуса

#### 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 7				
1	Использование в программе структуры данных список из списков	3	2	2
2	Использование в программе структуры данных словарь из словарей	3	2	2
3	Сохранение в файл структур данных разного типа	2	2	2
4	Преобразование структур данных из одного типа в другой	3	2	2
5	Работа с текстовыми файлами	3	2	5
Семестр 8				
1	Использование библиотеки NLTK	2	2	7
2	Экспертиза авторства текста	3	1	7
3	Отбор ключевых слов на основе частотного словаря	2	1	7
4	Работа с файлами табличного редактора Excel с помощью языка питон	2	2	6
5	Сохранение структур данных разного типа в таблице редактора Excel	3	2	6
Всего		26		



4.5. Курсовое проектирование/ выполнение 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 7, час	Семестр 8, час
1	2	3	4
Изучение теоретического материала дисциплины (ТО)		5	10
Подготовка к текущему контролю (ТК)		5	5
Подготовка отчетов по лабораторным работам (ЛР)		5	5
Контрольные работы заочников (КРЗ)		10	10
Подготовка к промежуточной аттестации (ПА)		10	18
Всего:	83	35	48

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)  
Учебно-методические материалы для самостоятельной работы обучающихся указаны в п.п. 7-11.

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)
Электронный ресурс	Сузи, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. — 2-е изд. — Москва : ИНТУИТ, 2016. — 350 с. — ISBN 5-9556-0058-2. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/100546">https://e.lanbook.com/book/100546</a>	
Электронный ресурс	Ингерсолл, Г. С. Обработка неструктурированных текстов. Поиск, организация и манипулирование / Г. С. Ингерсолл, Т. С. Мортон, Э. Л. Фэррис. — Москва : ДМК Пресс, 2015. — 414 с. — ISBN 978-5-97060-144-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/73069">https://e.lanbook.com/book/73069</a>	
Электронный ресурс	Щипицина, Л.Ю. Информационные технологии в лингвистике. [Электронный ресурс] : учеб.	

	пособие — Электрон. дан. — М. : ФЛИНТА, 2013. — 128 с. — Режим доступа: <a href="http://e.lanbook.com/book/44291">http://e.lanbook.com/book/44291</a>	
Электронный ресурс	Боярский, К.К. Введение в компьютерную лингвистику. [Электронный ресурс] : учеб. пособие — Электрон. дан. — СПб. : НИУ ИТМО, 2013. — 72 с. — Режим доступа: <a href="http://e.lanbook.com/book/70822">http://e.lanbook.com/book/70822</a>	
Электронный ресурс	Гребенщикова, А.В. Основы количественной лингвистики и новых информационных технологий: учеб. Пособие. [Электронный ресурс] : учеб. пособие — Электрон. дан. — М. : ФЛИНТА, 2015. — 152 с. — Режим доступа: <a href="http://e.lanbook.com/book/70334">http://e.lanbook.com/book/70334</a>	
Электронный ресурс	Короткова, М.А. Задачник по курсу "Математическая лингвистика и теория автоматов": учебное пособие для вузов. [Электронный ресурс] : учеб. пособие / М.А. Короткова, Е.Е. Трифонова. — Электрон. дан. — М. : НИЯУ МИФИ, 2012. — 92 с. — Режим доступа: <a href="http://e.lanbook.com/book/75843">http://e.lanbook.com/book/75843</a>	
Электронный ресурс	Кайсарова, Д.В. Математическая лингвистика. Практикум. [Электронный ресурс] : учеб. пособие / Д.В. Кайсарова, И.Ю. Коцюба. — Электрон. дан. — СПб. : НИУ ИТМО, 2016. — 67 с. — Режим доступа: <a href="http://e.lanbook.com/book/91462">http://e.lanbook.com/book/91462</a>	

#### 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
<a href="http://aot.ru">http://aot.ru</a>	Сайт «Автоматическая обработка текста»
<a href="http://python.org">python.org</a>	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	Оболочка Python IDLE, библиотеки xlrd и xlwt
2	Табличный редактор MS Excel
3	Spider - оболочка для языка python

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
Экзамен	Список вопросов к экзамену; Тесты.
Зачет	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 5-балльная шкала	Характеристика сформированных компетенций
«отлично» «зачтено»	– обучающийся глубоко и всесторонне усвоил программный материал; – уверенно, логично, последовательно и грамотно его излагает; – опираясь на знания основной и дополнительной литературы, тесно привязывает усвоенные научные положения с практической деятельностью направления; – умело обосновывает и аргументирует выдвигаемые им идеи; – делает выводы и обобщения; – свободно владеет системой специализированных понятий.
«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;

Оценка компетенции	Характеристика сформированных компетенций
5-балльная шкала	
	<ul style="list-style-type: none"> <li>– аргументирует научные положения;</li> <li>– делает выводы и обобщения;</li> <li>– владеет системой специализированных понятий.</li> </ul>
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> <li>– обучающийся усвоил только основной программный материал, по существу излагает его, опираясь на знания только основной литературы;</li> <li>– допускает несущественные ошибки и неточности;</li> <li>– испытывает затруднения в практическом применении знаний направления;</li> <li>– слабо аргументирует научные положения;</li> <li>– затрудняется в формулировании выводов и обобщений;</li> <li>– частично владеет системой специализированных понятий.</li> </ul>
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> <li>– обучающийся не усвоил значительной части программного материала;</li> <li>– допускает существенные ошибки и неточности при рассмотрении проблем в конкретном направлении;</li> <li>– испытывает трудности в практическом применении знаний;</li> <li>– не может аргументировать научные положения;</li> <li>– не формулирует выводов и обобщений.</li> </ul>

### 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
1	Структура данных «Список из списков»	ПК-7.В.1 ПК-7.У.1
2	Структура данных «Список из словарей»	ПК-7.В.1 ПК-7.У.1
3	Структура данных «Словарь из списков»	ПК-7.В.1 ПК-7.У.1
4	Структура данных «Словарь из словарей»	ПК-7.В.1 ПК-7.У.1
5	Операторы языка питон общего назначения	ПК-7.В.1
6	Генераторы списков	ПК-7.В.1
7	Написание функций на языке питон: структура, входные и выходные данные, вызов	ПК-7.В.1
8	Стандартные операторы, функции и методы языка питон для работы со строками	ПК-7.В.1 ПК-7.У.1
9	Задачи компьютерной лингвистики	ПК-3.У.1
10	Работа с сайтом aot.ru	ПК-7.У.1 ПК-3.У.1
11	Формальные грамматики: классификация Хомского, форма Бэкуса-Наура, символы метаязыка, рекурсивное правило	ПК-2.У.1
12	Пример формальной грамматики	ПК-2.У.1
13	Тезаурус	ПК-5.В.1
14	Онтология	ПК-5.В.1
15	Морфологический анализатор	ПК-2.У.1

		ПК-3.У.1
16	Теория систем: определение системы, классификация систем, свойства системы	УК-1.В.2 УК-2.У.1
17	Системный подход, системный анализ, язык как система, цели и задачи, декомпозиция, предметная область	УК-1.В.2 УК-2.У.1
18	Средства анализа систем: DPD и UML	УК-1.В.2 УК-2.У.1
19	Методы файловых объектов языка питон	ПК-7.В.1

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	Доступ к файлам пакета excel из языка питон: библиотеки, чтение файла	ПК-7.В.1
2	Доступ к файлам пакета excel из языка питон: применение стилия, запись файла	ПК-7.В.1
3	Искусственный интеллект: сфера задач, проблемы, способы решения	УК-1.В.2
4	Экспертные системы	ПК-1.3.1
5	Нейронные сети	УК-1.В.2
6	Экспертиза авторства текста	ПК-3.У.1
7	Решение проблемы определения ключевых слов текста	ПК-3.У.1
8	Обзор библиотек для анализа текстов на естественном языке	ПК-3.У.1
9	Применение библиотеки NLTK	ПК-3.У.1
10	Векторизация	ПК-3.У.1
11	Токенизация, стемминг, лемматизация	ПК-3.У.1
12	Морфологический анализ	ПК-3.У.1
13	Мешок слов	ПК-3.У.1
14	Голосовой помощник	ПК-1.3.1
15	Информационно-поисковые системы	ПК-1.3.1
16	Поисковые системы	ПК-1.3.1
17	Индексация в поисковых системах	ПК-1.3.1
18	Корпусная лингвистика	ПК-7.У.1 ПК-7.В.1
19	Создание корпуса	ПК-7.У.1 ПК-7.В.1
20	Разметка корпуса	ПК-7.У.1 ПК-7.В.1
21	Обзор лингвистических корпусов	ПК-7.У.1 ПК-7.В.1

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

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

№ п/п	Примерный перечень тем для курсового проектирования/выполнения 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																
<b>Компетенция УК-1</b>																		
1	<p><b>Инструкция: выберите один ответ</b></p> <p>Кто является автором фразы: «...Язык есть система, все элементы которой образуют целое, а значимость одного элемента проистекает только от одновременного наличия прочих...»</p> <ol style="list-style-type: none"> <li>1. Л. Блумфилд</li> <li>2. Н.С. Трубецкой</li> <li>3. Ф. де Соссюр</li> <li>4. Л. Т. Ельмслев</li> </ol>	УК-1.В.2																
2	<p><b>Инструкция: выберите несколько ответов</b></p> <p>Какими признаками обладают сложные системы?</p> <ol style="list-style-type: none"> <li>1. Система как целое обладает свойствами, которых нет у ее частей (эмерджентность)</li> <li>2. Систему можно разделить на части и изучать их отдельно</li> <li>3. Система выбирает свое поведение целенаправленно</li> <li>4. Систему нельзя поделить на части</li> <li>5. Части системы не связаны между собой</li> </ol>	УК-1.В.2																
3	<p><b>Инструкция:</b></p> <p>Для каждого свойства системы из левого столбца подберите определение из правого столбца</p> <table border="1" data-bbox="347 1267 1107 2074"> <tbody> <tr> <td data-bbox="347 1267 427 1464">А</td> <td data-bbox="427 1267 727 1464">Иерархичность</td> <td data-bbox="727 1267 807 1464">1</td> <td data-bbox="807 1267 1107 1464">Упорядоченность частей системы и наличие связей между ними</td> </tr> <tr> <td data-bbox="347 1464 427 1706">В</td> <td data-bbox="427 1464 727 1706">Неоднозначность</td> <td data-bbox="727 1464 807 1706">2</td> <td data-bbox="807 1464 1107 1706">Построение частей чего-либо в порядке от высшего к низшему</td> </tr> <tr> <td data-bbox="347 1706 427 1948">С</td> <td data-bbox="427 1706 727 1948">Структурность</td> <td data-bbox="727 1706 807 1948">3</td> <td data-bbox="807 1706 1107 1948">После принятия решения возможны различные варианты поведения системы</td> </tr> <tr> <td data-bbox="347 1948 427 2074">D</td> <td data-bbox="427 1948 727 2074">Самоорганизация</td> <td data-bbox="727 1948 807 2074">4</td> <td data-bbox="807 1948 1107 2074">Образование пространственной, временной,</td> </tr> </tbody> </table>	А	Иерархичность	1	Упорядоченность частей системы и наличие связей между ними	В	Неоднозначность	2	Построение частей чего-либо в порядке от высшего к низшему	С	Структурность	3	После принятия решения возможны различные варианты поведения системы	D	Самоорганизация	4	Образование пространственной, временной,	УК-1.В.2
А	Иерархичность	1	Упорядоченность частей системы и наличие связей между ними															
В	Неоднозначность	2	Построение частей чего-либо в порядке от высшего к низшему															
С	Структурность	3	После принятия решения возможны различные варианты поведения системы															
D	Самоорганизация	4	Образование пространственной, временной,															

				информационной или функциональной организации, структуры за счет внутренних ресурсов системы в результате целеполагающих взаимодействий с окружением	
4	<b>Инструкция:</b> Расставьте этапы системного анализа в правильном порядке <ol style="list-style-type: none"> <li>1. Разработка вариантов решения</li> <li>2. Оценка найденных вариантов и способов реализации решения</li> <li>3. Выявление проблем и постановка целей</li> </ol>				УК-1.В.2 УК-2.У.1
5	<b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.  Что такое эмерджентность?				УК-1.В.2
<b>Компетенция УК-2</b>					
1	<b>Инструкция: выберите один ответ</b>  Как называется прием, при котором большая задача разбивается на более мелкие? <ol style="list-style-type: none"> <li>1. децимация</li> <li>2. девальвация</li> <li>3. декомпозиция</li> <li>4. демпинг</li> </ol>				УК-2.У.1
2	<b>Инструкция: выберите несколько ответов</b>  Как называются средства для анализа систем? <ol style="list-style-type: none"> <li>1. URL</li> <li>2. DFD</li> <li>3. NLTK</li> <li>4. UML</li> <li>5. DNS</li> </ol>				УК-2.У.1
3	<b>Инструкция:</b> для каждого термина из левого столбца подберите его правильное описание из правого столбца.				УК-2.У.1
	А	Цель	1	Отвечает на вопрос «Какими	

			действиями этого можно достигнуть желаемого?»		
	В	Задача	2	Часть реального мира, рассматриваемая в рамках проведения некоторого исследования, выполнения проекта	
	С	Система	3	Представление о будущем или желаемом результате, который человек или группа людей представляют себе, планируют и обязуются достичь.	
	Д	Предметная область	4	Множество элементов, находящихся в отношениях и связях друг с другом, которое образует определённую целостность, единство.	
4	<b>Инструкция:</b> Расставьте этапы системного анализа в правильном порядке <ol style="list-style-type: none"> <li>1. Разработка вариантов решения</li> <li>2. Оценка найденных вариантов и способов реализации решения</li> <li>3. Выявление проблем и постановка целей</li> </ol>				УК-2.У.1 УК-1.В.2
5	<b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.  Чем цель отличается от задачи?				УК-2.У.1
<b>Компетенция ПК-1</b>					
1	<b>Инструкция: выберите один ответ</b>  Как называется феномен поисковых систем, когда пользователь видит только те результаты, которые согласуются с его предпочтениями? <ol style="list-style-type: none"> <li>1. пузырь фильтров</li> <li>2. мешок слов</li> <li>3. релевантность</li> <li>4. информационная сфера</li> </ol>				ПК-1.3.1



2	<p><b>Инструкция: выберите несколько ответов</b></p> <p>Какими средствами можно запретить индексацию веб-страницы?</p> <ol style="list-style-type: none"> <li>1. Файл index.txt</li> <li>2. Тег &lt;a&gt;</li> <li>3. Тег &lt;noindex&gt;</li> <li>4. Файл robots.txt</li> <li>5. DNS</li> </ol>	ПК-1.3.1																
3	<p><b>Инструкция:</b> для каждого термина из левого столбца подберите его правильное описание из правого столбца.</p> <table border="1"> <tr> <td>A</td> <td>wolframalpha.com</td> <td>1</td> <td>Крупнейшая в мире поисковая система, дополненная большим количеством служб</td> </tr> <tr> <td>B</td> <td>google.com</td> <td>2</td> <td>Система, которая имеет собственную базу знаний по точным наукам. Может решить уравнение и построить график.</td> </tr> <tr> <td>C</td> <td>yandex.ru</td> <td>3</td> <td>Веб-сервис, способный угадать задуманного персонажа</td> </tr> <tr> <td>D</td> <td>akinator.com</td> <td>4</td> <td>Российская интернет-компания, разработавшая голосовой помощник Алиса</td> </tr> </table>	A	wolframalpha.com	1	Крупнейшая в мире поисковая система, дополненная большим количеством служб	B	google.com	2	Система, которая имеет собственную базу знаний по точным наукам. Может решить уравнение и построить график.	C	yandex.ru	3	Веб-сервис, способный угадать задуманного персонажа	D	akinator.com	4	Российская интернет-компания, разработавшая голосовой помощник Алиса	ПК-1.3.1
A	wolframalpha.com	1	Крупнейшая в мире поисковая система, дополненная большим количеством служб															
B	google.com	2	Система, которая имеет собственную базу знаний по точным наукам. Может решить уравнение и построить график.															
C	yandex.ru	3	Веб-сервис, способный угадать задуманного персонажа															
D	akinator.com	4	Российская интернет-компания, разработавшая голосовой помощник Алиса															
4	<p><b>Инструкция:</b> Расставьте в правильном порядке этапы работы голосового помощника.</p> <ol style="list-style-type: none"> <li>1. Запрос к языковой модели</li> <li>2. Транскрибирование речи</li> <li>3. Синтез речи</li> <li>4. Обнаружение слова активации</li> </ol>	ПК-1.3.1																
5	<p><b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.</p> <p>Что такое экспертная система?</p>	ПК-1.3.1																
	<b>Компетенция ПК-2</b>																	
1	<b>Инструкция: выберите один ответ</b>	ПК-2.У.1																

	<p>Как называется формальная система описания синтаксиса, используемая в контекстно-свободных формальных грамматиках?</p> <ol style="list-style-type: none"> <li>1. Нормальная форма</li> <li>2. Контекстная диаграмма</li> <li>3. Форма Бэкуса-Наура</li> <li>4. Дерево вывода</li> </ol>																	
2	<p><b>Инструкция: выберите несколько ответов</b></p> <p>Какие бывают формальные грамматики?</p> <ol style="list-style-type: none"> <li>1. Дискретные</li> <li>2. Контекстно-зависимые</li> <li>3. Конечные</li> <li>4. Контекстно-свободные</li> <li>5. Статические</li> </ol>	ПК-2.У.1																
3	<p><b>Инструкция:</b> для каждого термина из левого столбца подберите его правильное описание из правого столбца.</p> <table border="1" data-bbox="347 864 1107 1335"> <tr> <td>A</td> <td>Терминал</td> <td>1</td> <td>Конечное непустое множество символов</td> </tr> <tr> <td>B</td> <td>Алфавит</td> <td>2</td> <td>Синтаксический элемент языка не имеющий конкретного символического значения</td> </tr> <tr> <td>C</td> <td>Нетерминал</td> <td>3</td> <td>Способ описания формального языка</td> </tr> <tr> <td>D</td> <td>Формальная грамматика</td> <td>4</td> <td>Элемент языка, присутствующий в каком-либо слове.</td> </tr> </table>	A	Терминал	1	Конечное непустое множество символов	B	Алфавит	2	Синтаксический элемент языка не имеющий конкретного символического значения	C	Нетерминал	3	Способ описания формального языка	D	Формальная грамматика	4	Элемент языка, присутствующий в каком-либо слове.	ПК-2.У.1
A	Терминал	1	Конечное непустое множество символов															
B	Алфавит	2	Синтаксический элемент языка не имеющий конкретного символического значения															
C	Нетерминал	3	Способ описания формального языка															
D	Формальная грамматика	4	Элемент языка, присутствующий в каком-либо слове.															
4	<p><b>Инструкция:</b> Расставьте части правила контекстно-свободной грамматики в правильном порядке</p> <ol style="list-style-type: none"> <li>1. Разделитель</li> <li>2. Нетерминал</li> <li>3. Терминал_нетерминал</li> </ol>	ПК-2.У.1																
5	<p><b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.</p> <p>Что такое метаязык?</p>	ПК-2.У.1																
<b>Компетенция ПК-3</b>																		
1	<p><b>Инструкция: выберите один ответ</b></p> <p>Что выведет на экран этот фрагмент кода на языке python?</p> <pre>from nltk.tokenize import word_tokenize text = "В саду созрели яблоки." word_tokens = word_tokenize(text)</pre>	ПК-3.У.1																

	<pre>print(word_tokens)</pre> <ol style="list-style-type: none"> <li>[1:'В', 2:'саду', 3:'созрели', 4:'яблоки', 5:'.']</li> <li>['В', 'саду', 'созрели', 'яблоки', '.']</li> <li>['В', 'саду', 'созрели', 'яблоки']</li> <li>['В', 'с', 'а', 'д', 'у', 'с', 'о', 'з', 'р', 'е', 'л', 'и', 'я', 'б', 'л', 'о', 'к', 'и', '.']</li> </ol>																	
2	<p><b>Инструкция: выберите несколько ответов</b></p> <p>Как называются библиотеки для обработки естественного языка для python?</p> <ol style="list-style-type: none"> <li>pandas</li> <li>SciPy</li> <li>NLTK</li> <li>spaCy</li> <li>Gensim</li> </ol>	<p>ПК-3.У.1</p> <p>ПК-7.У.1</p>																
3	<p><b>Инструкция:</b> для каждого термина из левого столбца подберите его правильное описание из правого столбца.</p> <table border="1"> <tr> <td>А</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>В</td> <td>Стемминг</td> <td>2</td> <td>Процесс приведения словоформы к её нормальной (словарной) форме.</td> </tr> <tr> <td>С</td> <td>Лемматизация</td> <td>3</td> <td>Преобразование текста в числовой формат.</td> </tr> <tr> <td>Д</td> <td>Векторизация</td> <td>4</td> <td>Процесс разбиения текста на более мелкие части (слова или предложения)</td> </tr> </table>	А	Токенизация	1	Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.	В	Стемминг	2	Процесс приведения словоформы к её нормальной (словарной) форме.	С	Лемматизация	3	Преобразование текста в числовой формат.	Д	Векторизация	4	Процесс разбиения текста на более мелкие части (слова или предложения)	<p>ПК-3.У.1</p> <p>ПК-7.У.1</p>
А	Токенизация	1	Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.															
В	Стемминг	2	Процесс приведения словоформы к её нормальной (словарной) форме.															
С	Лемматизация	3	Преобразование текста в числовой формат.															
Д	Векторизация	4	Процесс разбиения текста на более мелкие части (слова или предложения)															
4	<p><b>Инструкция:</b> Запишите операторы программы на языке python в правильном порядке:</p> <ol style="list-style-type: none"> <li>sentence_tokens = sent_tokenize(text)</li> <li>print(sentence_tokens)</li> <li>from nltk.tokenize import sent_tokenize text = "Раннее утро радовало светом."&gt;</li> </ol>	ПК-3.У.1																
5	<p><b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.</p> <p>Почему перед анализом из текстов часто удаляют стоп-слова?</p>	ПК-3.У.1																
	<b>Компетенция ПК-5</b>																	
1	<p><b>Инструкция: выберите один ответ</b></p> <p>Какое понятие задает следующее определение: «система, состоящая из набора понятий и набора утверждений об этих понятиях»?</p>	ПК-5.В.1																

	1. Онтология 2. Аксиома 3. Иерархия 4. Когнитивность																	
2	<b>Инструкция: выберите несколько ответов</b>  Какие существуют принципы классификаций онтологий?  1. По степени формальности 2. По содержанию 3. По объему 4. По цели создания 5. По языку	ПК-5.В.1																
3	<b>Инструкция:</b> Каждому термину из левого столбца подберите определение из правого столбца. <table border="1" data-bbox="347 862 1107 1413"> <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-5.В.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															
4	<b>Инструкция:</b> 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-5.В.1																
5	<b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.  Что такое тезаурус?	ПК-5.В.1																
	<b>Компетенция ПК-7</b>																	
1	<b>Инструкция: выберите один ответ</b>  Как называется большой, представленный в машиночитаемом	ПК-7.У.1																

	<p>формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач?</p> <ol style="list-style-type: none"> <li>1. База данных</li> <li>2. Лингвистический корпус</li> <li>3. База знаний</li> <li>4. Тезаурус</li> </ol>																	
2	<p><b>Инструкция:</b> выберите несколько ответов</p> <p>Как называются библиотеки для обработки естественного языка для python?</p> <ol style="list-style-type: none"> <li>1. pandas</li> <li>2. SciPy</li> <li>3. NLTK</li> <li>4. spaCy</li> <li>5. Gensim</li> </ol>	<p>ПК-3.У.1</p> <p>ПК-7.У.1</p>																
3	<p><b>Инструкция:</b> для каждого термина из левого столбца подберите его правильное описание из правого столбца.</p> <table border="1" data-bbox="347 936 1107 1391"> <tr> <td>А</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>В</td> <td>Стемминг</td> <td>2</td> <td>Процесс приведения словоформы к её нормальной (словарной) форме.</td> </tr> <tr> <td>С</td> <td>Лемматизация</td> <td>3</td> <td>Преобразование текста в числовой формат.</td> </tr> <tr> <td>Д</td> <td>Векторизация</td> <td>4</td> <td>Процесс разбиения текста на более мелкие части (слова или предложения)</td> </tr> </table>	А	Токенизация	1	Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.	В	Стемминг	2	Процесс приведения словоформы к её нормальной (словарной) форме.	С	Лемматизация	3	Преобразование текста в числовой формат.	Д	Векторизация	4	Процесс разбиения текста на более мелкие части (слова или предложения)	<p>ПК-3.У.1</p> <p>ПК-7.У.1</p>
А	Токенизация	1	Процесс сведения слов к их основной (корневой) форме путем удаления окончания и суффиксов.															
В	Стемминг	2	Процесс приведения словоформы к её нормальной (словарной) форме.															
С	Лемматизация	3	Преобразование текста в числовой формат.															
Д	Векторизация	4	Процесс разбиения текста на более мелкие части (слова или предложения)															
4	<p><b>Инструкция:</b> Запишите в правильном порядке действия, выполняемые при создании корпуса текстов:</p> <ol style="list-style-type: none"> <li>1. Сбор и оцифровка данных</li> <li>2. Разметка корпуса</li> <li>3. Планирование</li> <li>4. Приведение к единому формату</li> </ol>	ПК-7.В.1																
5	<p><b>Инструкция:</b> Прочитайте текст и запишите развернутый обоснованный ответ.</p> <p>Что такое морфологическая разметка корпуса текстов?</p>	ПК-7.В.1																

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	Анализ предметной области с применением диаграмм потоков данных или языка UML
2	Создание собственного лингвистического корпуса

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

В лекционной части курса дается объяснение изучаемого материала, рассматриваются примеры использования разнообразных алгоритмических конструкций. 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

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

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

### **Задание и требования к проведению лабораторных работ**

Цель выполнения лабораторных работ по дисциплине «Обработка текстовых массивов» – приобретение практических навыков работы с текстовыми данными на языке высокого уровня.

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.

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.

Методическими материалами, направляющими самостоятельную работу обучающихся является учебно-методический материал по дисциплине.

11.4. Методические указания для обучающихся по прохождению промежуточной аттестации.

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

– экзамен – форма оценки знаний, полученных обучающимся в процессе изучения всей дисциплины или ее части, навыков самостоятельной работы, способности применять их для решения практических задач. Экзамен, как правило, проводится в период экзаменационной сессии и завершается аттестационной оценкой «отлично», «хорошо», «удовлетворительно», «неудовлетворительно».

– зачет – это форма оценки знаний, полученных обучающимся в ходе изучения учебной дисциплины в целом или промежуточная (по окончании семестра) оценка знаний обучающимся по отдельным разделам дисциплины с аттестационной оценкой «зачтено» или «не зачтено».

На экзамене и на зачете студент отвечает на вопросы и выполняет практические задания.

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