

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

Кафедра № 63

УТВЕРЖДАЮ

Ответственный за образовательную
программу

доц., к.ф.н., доц.

(должность, уч. степень, звание)

М.А. Чиханова

(инициалы, фамилия)

(подпись)

«20» февраля 2025 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«Технологии обработки текста и звучащей речи»
(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	2025

Лист согласования рабочей программы дисциплины

Программу составил (а)

доц., к.э.н.

(должность, уч. степень, звание)



14.02.2025

(подпись, дата)

О.В. Злобина

(инициалы, фамилия)

Программа одобрена на заседании кафедры № 63

«14» февраля 2025 г, протокол № 7

Заведующий кафедрой № 63

доц., к.ф.н., доц.

(должность, уч. степень, звание)



14.02.2025

(подпись, дата)

М.А. Чиханова

(инициалы, фамилия)

Заместитель декана факультета №6 по методической работе

проф., д.и.н., доц.

(должность, уч. степень, звание)



14.02.2025

(подпись, дата)

Л.Ю. Гусман

(инициалы, фамилия)

Аннотация

Дисциплина «Технологии обработки текста и звучащей речи» входит в образовательную программу высшего образования – программу бакалавриата по направлению подготовки/ специальности 45.03.02 «Лингвистика» направленности «Теоретическая и прикладная лингвистика». Дисциплина реализуется кафедрой «№63».

Дисциплина нацелена на формирование у выпускника следующих компетенций:

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-4 «Способность использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, лингводидактики для решения профессиональных задач»

ПК-6 «Готовность овладеть методами формального и когнитивного моделирования естественного языка и методами создания метаязыков»

Содержание дисциплины охватывает круг вопросов, связанных с современными технологиями обработки естественного языка, рассмотрением особенностей систем русского и английского языков с точки зрения автоматической обработки текста, использования математических моделей, анализа и оценки возможностей существующих систем обработки звучащей речи – голосовых помощников, принципов их разработки, сферы применения и ограничений, современные подходы к верификации пользователя при помощи систем обработки звучащей речи, проблемы надежности, многоступенчатой верификации, компромиссов между степенью надежности и скоростью идентификации и т.д.

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.

Общая трудоемкость освоения дисциплины составляет 2 зачетных единицы, 72 часа.

Язык обучения по дисциплине «русский».

1. Перечень планируемых результатов обучения по дисциплине

1.1. Цели преподавания дисциплины состоит в получении знаний и формировании умений, связанных с системами обработки текста и речи; основными статистическими и математическими моделями, используемыми при анализе и синтезе речи; современными системами обработки и порождения текста и речи и принципами их разработки, а также видами языковых данных, использующихся при разработке таких систем.

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
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.В.3 владеть навыками использования цифровых средств для решения поставленной задачи
Профессиональные компетенции	ПК-4 Способность использовать понятийный аппарат теоретической и прикладной лингвистики, переводоведения, лингводидактики для решения профессиональных задач	ПК-4.3.1 знать современный понятийный аппарат теоретической и прикладной лингвистики, переводоведения, лингводидактики
Профессиональные компетенции	ПК-6 Готовность овладеть методами формального и когнитивного моделирования естественного языка и методами создания метаязыков	ПК-6.3.1 знать принципы построения больших языковых моделей, принципы функционирования нейросетей и систем машинного перевода ПК-6.У.1 уметь анализировать вербальные и невербальные компоненты речевой деятельности, различать основные типы формальных моделей описания, формальных грамматик, использовать методы

		когнитивного и формального моделирования естественного языка; разрабатывать автоматизированные средства анализа текста
--	--	--

2. Место дисциплины в структуре ОП

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

- Компьютерные программы в лингвистических исследованиях
- Математика. Теория вероятностей и математическая статистика
- Информационные технологии в лингвистике
- Теоретическая фонетика

Знания, полученные при изучении материала данной дисциплины, имеют как самостоятельное значение, так и используются при изучении других дисциплин:

- Когнитивные технологии в лингвистике,
- Обработка текстовых массивов,
- Корпусная лингвистика;
- Производственная практика

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
		№6
1	2	3
Общая трудоемкость дисциплины, 3Э/ (час)	2/ 72	2/ 72
Из них часов практической подготовки	6	6
Аудиторные занятия, всего час.	12	12
в том числе:		
лекции (Л), (час)	4	4
практические/семинарские занятия (ПЗ), (час)		
лабораторные работы (ЛР), (час)	8	8
курсовой проект (работа) (КП, КР), (час)		
экзамен, (час)		
Самостоятельная работа, всего (час)	60	60
Вид промежуточной аттестации: зачет, дифф. зачет, экзамен (Зачет, Дифф. зач, Экз.**)	Дифф. Зач.	Дифф. Зач.

Примечание: ** кандидатский экзамен

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 6					

Раздел 1. Анализ звучащей речи Тема 1.1. Артикуляторная и акустическая классификация фонем Тема 1.2. Формантная структура звуков речи, спектральный анализ звучащей речи Тема 1.3. Индивидуальные произносительные характеристики Тема 1.4. Супraseгментные характеристики и возможность их использования для идентификации.	1		2		10
Раздел 2 Системы голосовой идентификации пользователя Тема 2.1. Этапы процедуры идентификации Тема 2.2 Особенности открытых и закрытых систем идентификации Тема 2.3. Вариативность реализации как фактор идентификации Тема 2.4. Методы статистического анализа, используемые в системах голосовой идентификации Тема 2.5. Проблема защиты персональных данных	1		2		15
Раздел 3. Текстовые и речевые интерфейсы Тема 1.1. Классификация интерфейсов Тема 1.1. Возможности интерфейсов Тема 1.1. Обработка естественного языка как основа интерфейса Тема 1.1. Особенности существующих систем речевых интерфейсов	1		2		10
Раздел 4. Обработка естественного языка Тема 1.1. Виды систем обработки естественного языка Тема 1.1. Этапы обработки Тема 1.1. Различия языков с точки зрения процедуры обработки Тема 1.1. Решение проблемы вариативности речевых средств			2		15
Раздел 5. Перспективы и направления развития технологий обработки текста и звучащей речи	1				10
Итого в семестре:	4		8		60
Итого	4	0	8	0	60

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	Артикуляторная и акустическая классификация фонем, формантная структура звуков речи как основа машинного анализа и распознавания, спектральный анализ звучащей речи, проблема индивидуальной вариативности формантной структуры

	Системы голосовой идентификации и верификации, различие в подходах к решению задач в этих системах, проблема многофакторности и способы ее решения, критерии оценки эффективности работы системы
2	Интерфейсы, виды интерфейсов, их возможности и ограничения, сферы применения. Существующие на сегодняшний день интерфейсы и их возможности. Проблема вариативности речевых средств и ее влияние на эффективность работы интерфейса. Порождающий ИИ и перспективы его развития

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 6				
1	Формантная структура звуков речи; спектральный анализ звучащей речи Индивидуальные произносительные характеристики	2	1	1
2	Вариативность реализации как фактор идентификации Методы статистического анализа, используемые в системах голосовой идентификации	2	1	2
3	Особенности существующих систем интерфейсов	2	2	3
4	Решение проблемы вариативности речевых средств	2	2	4
Всего		8	6	

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 6, час
1	2	3
Изучение теоретического материала дисциплины (ТО)	20	20
Курсовое проектирование (КП, КР)		
Расчетно-графические задания (РГЗ)		
Выполнение реферата (Р)		
Подготовка к текущему контролю успеваемости (ТКУ)	15	15
Домашнее задание (ДЗ)		
Контрольные работы заочников (КРЗ)	20	20
Подготовка к промежуточной аттестации (ПА)	5	5
Всего:	60	60

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)
[004.8 В29 004]	Венцов А.В., Касевич В.Б. Проблемы восприятия речи. - 2- е изд. - М.: УРСС, 2003. - 237 с.	ФО(2), ГС(10).
[81(075) К28 8]	Касевич В.Б. Введение в языкознание: учебник для СПО. М.: Академия; СПб.: СПбГУ, 2011. –229 с.	ФО(4), СО(70)
https://www.pdfdrive.com/learning-python-5th-edition-e39205692.html	Learning Python, Fifth Edition by Mark Lutz, 2013.	Электронная версия

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
http://icame.uib.no/corpora.html	ICAME Corpus Collection
http://narusco.ru/	Корпус русского литературного языка
http://ruscorpora.ru/	Национальный корпус русского языка
http://www.aot.ru/onlinedemo.html	Программное обеспечение в области автоматической обработки текста
http://starling.rinet.ru/	Сайт «Вавилонская башня»
http://www.slav.helsinki.fi/hanco	Хельсинский аннотированный корпус
http://corpus.byu.edu/	Mark Davies Collection (six online corpora).
http://lands.let.kun.nl/cgn/	The Spoken Dutch Corpus.
www.natcorp.ox.ac.uk/archive/vault/tgcw30.pdf	CDIF (Corpus Document Interchange Format)
http://www.cs.vassar.edu/CES/CES1.html#Contents	CES (Corpus Encoding Standard)
http://www.ninjal.ac.jp/english/products/csj/	Corpus of Spontaneous Japanese
http://mirjamernestus.ruhosting.nl/Ernestus/NCCFr/index.html	Nijmegen Corpus of Casual French

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
	Не предусмотрено

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	Лаборатория формальной лингвистики	

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 5-балльная шкала	Характеристика сформированных компетенций
«отлично» «зачтено»	<ul style="list-style-type: none"> – 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»	<ul style="list-style-type: none"> – 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.
«удовлетворительно» «зачтено»	<ul style="list-style-type: none"> – 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.
«неудовлетворительно» «не зачтено»	<ul style="list-style-type: none"> – 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.

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	1) Какие методы статистического анализа используются в идентификации по голосу? 2) Как используется вариативность реализации фонем (аллофоны и фонны) в распознавании речи? 3) На каких этапах применяется NLP в речевых интерфейсах? 4) Какие виды чат-ботов существуют? На чем основано их различие? 5) Как определяется оценочность текста при обработке текста? Возможна ли разработка универсального алгоритма определения оценки? 6) Как должна учитываться речевая вариативность при создании интерфейсов? 7) Возможно ли соединение систем распознавания речи (речевого интерфейса, обеспечивающего открытое взаимодействие) и верификации пользователя (с неограниченным числом пользователей)? Почему? 8) Существуют ли в России, США, странах Евросоюза законодательные ограничения на сбор голосовых данных?	УК-2.В.3 ПК-4.3.1 ПК-6.3.1 ПК-6.У.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 идентификатора
1	Выберите правильный ответ При выборе цифровых средств необходимо учитывать 1) эффективность использования цифровых средств; 2) общий уровень развития технологий; 3) уровень цифровой грамотности населения.	УК-2.В.3
2	Выберите правильный ответ Преимущество системного подхода состоит в том, что он 1) позволяет увидеть всю картину целиком;	УК-2.В.3

	2) позволяет избежать ошибок при решении задач; 3) позволяет быстро найти решение задачи.	
3	Выберите правильный ответ При использовании цифровых средств могут возникнуть риски 1) утечки конфиденциальной информации; 2) нарушения авторских прав; 3) все вышеперечисленное.	УК-2.В.3
4	Выберите правильный ответ Ограничения при использовании цифровых средств связаны с 1) недостатком данных; 2) ограниченностью когнитивных способностей персонала; 3) все вышеперечисленные.	УК-2.В.3
5	Выберите верное утверждение 1) Любая цифровая технология в области обработки, распознавания или синтеза речи/текста представляет собой компромисс между стоимостью и качеством. 2) Цифровые технологии в области обработки, распознавания или синтеза речи/текста позволяют обеспечить высокое качество при низкой стоимости.	УК-2.В.3
6	Выберите верное утверждение 1) Законодательство в области использования биометрических данных практически не отличается в разных странах. 2) Законодательство в области использования биометрических данных значительно отличается в разных странах	УК-2.В.3
7	Выберите правильный ответ Основными факторами, определяющими эффективность систем автоматической обработки, распознавания или синтеза речи/текста, являются 1) человеческие ресурсы; 2) финансовые ресурсы; 3) объем и характеристика базы обучающих данных	УК-2.В.3
8	Выберите правильный ответ Согласно законодательству РФ биометрические данные граждан хранятся 1) в базе данных компании, которая их собирает; 2) в базе данных государственных услуг; 3) в базе данных Министерства внутренних дел; 4) в базе данных Центрального банка РФ.	УК-2.В.3
9	Приведите 2-3 примера биометрических данных	УК-2.В.3
10	Перечислите формы, в которых языковая информация сохраняется в цифровой форме	УК-2.В.3
11	Перечислите формы языковых данных, обработка которых может выполняться с использованием цифровых технологий	УК-2.В.3

12	Охарактеризуйте направления использования цифровых технологий в отношении текста и звучащей речи	УК-2.В.3
13	Охарактеризуйте основные виды ресурсных ограничений при разработке систем автоматического распознавания текста/речи	УК-2.В.3
14	Выберите правильный ответ Лингвистическая экспертиза звучащей речи основана на анализе 1) речевого контекста; 2) синтаксических явлений в тексте; 3) фонетических явлений в речи.	ПК-4.3.1
15	Выберите правильный ответ К видам лингвистической экспертизы относятся 1) почерковедческая экспертиза, 2) технико-криминалистическая экспертиза; 3) автороведческая экспертиза.	ПК-4.3.1
16	Выберите правильный ответ Автороведческая экспертиза направлена на 1) установление подлинности документа; 2) установление автора текста; 3) установление времени написания документа.	ПК-4.3.1
17	Выберите правильный ответ Фоноскопическая экспертиза направлена на 1) изучение звуковой среды; 2) это изучение звуковых сигналов; 3) это изучение звучащей речи с целью идентификации человека по голосу и речи.	ПК-4.3.1
18	Выберите правильный ответ В фоносемантическом анализе производится 1) анализ грамматической формы слова 2) анализ восприятия звуковой формы слова; 3) анализ значения слова.	ПК-4.3.1
19	Выберите правильный ответ В качестве объекта фоносемантического анализа выступают, как правило, 1) научные работы; 2) договоры; 3) реклама.	ПК-4.3.1
20	Выберите определение Под кодифицированными нормами русского языка понимаются 1) правила, зафиксированные в словарях, грамматиках и т.п.; 2) правила написания текста.	ПК-4.3.1
21	Выберите правильный ответ Кодифицированные нормы языка используются в 1) системах обработки текста.; 2) системах порождения текста; 3) системах автоматической проверки.	ПК-4.3.1

22	<p>Выберите правильный ответ</p> <p>Морфологические нормы языка определяют</p> <p>1) образование формы слова;</p> <p>2) произношение;</p> <p>3) построение предложений;</p> <p>4) употребление слов и выражений.</p>	ПК-4.3.1
23	<p>Выберите правильный ответ</p> <p>Орфоэпические нормы языка определяют</p> <p>1) образование формы слова;</p> <p>2) произношение;</p> <p>3) построение предложений;</p> <p>4) употребление слов и выражений.</p>	ПК-4.3.1
24	<p>Выберите и запишите единицы лексического уровня, позволяющие отнести приведенный фрагмент к просторечию</p> <p>«Ты чё, совсем с катушек слетел?!»</p>	ПК-4.3.1
25	<p>Выберите и запишите единицы лексического уровня, позволяющие отнести приведенный фрагмент к научному стилю</p> <p>«В этих методах первоначально по зашумленному речевому сигналу оценивается передаточная характеристика “оптимального” фильтра.»</p>	ПК-4.3.1
26	<p>Определите, для какого стиля наиболее характерно наличие сложной многоуровневой синтаксической структуры с длинными цепочками однородных членов, отсутствие качественных прилагательных и наличие отглагольных существительных</p>	ПК-4.3.1
27	<p>На основе анализа синтаксической структуры определите стилистическую принадлежность следующего фрагмента</p> <p>«Основная энергия речевого сигнала сосредоточена на определенных частотах, называемых формантами, а энергия шума, вообще говоря, может быть распределена по всему диапазону звуковых частот. Таким образом, используя соответствующий фильтр, реализованный во временной или спектральной области и точно настроенный на формантные частоты, можно уменьшать шум, сохраняя при этом полезный сигнал.»</p>	ПК-4.3.1
28	<p>На основе анализа синтаксической структуры определите стилистическую принадлежность следующего фрагмента</p> <p>«Распознавание и обработка человеческих языков давно внедрена: именно поэтому обычный IVR с жестко заданными опциями ответов постепенно уходит в прошлое, чатботы начинают все адекватнее общаться без участия живого оператора, фильтры в почте работают на ура и т.д. Как же происходит распознавание записанной речи, то есть текста?»</p>	ПК-4.3.1
29	<p>Выберите правильный ответ</p>	ПК-6.3.1

	<p>Когнитивное моделирование естественного языка используется для</p> <ol style="list-style-type: none"> 1) моделирования естественного языка с учетом когнитивных способностей человека; 2) моделирования естественного языка с использованием когнитивных методов; 3) моделирования естественного языка на основе когнитивной психологии. 	
30	<p>Выберите правильный ответ</p> <p>Для когнитивного моделирования естественного языка могут использоваться методы</p> <ol style="list-style-type: none"> 1) психолингвистики; 2) когнитивной лингвистики; 3) все вышеперечисленные. 	ПК-6.3.1
31	<p>Выберите правильный ответ</p> <p>Когнитивное моделирование естественного языка позволяет решать задачи</p> <ol style="list-style-type: none"> 1) восприятия языка; 2) изучения механизмов порождения и производства языка 	ПК-6.3.1
32	<p>Выберите правильный ответ</p> <p>Преимущества когнитивных методов моделирования языка заключается в</p> <ol style="list-style-type: none"> 1) возможности моделирования процессов, связанных с языком; 2) возможности анализа сложных синтаксических конструкций. 	ПК-6.3.1
33	<p>Выберите правильный ответ</p> <p>Недостатки когнитивных методов моделирования языка связаны с</p> <ol style="list-style-type: none"> 1) сложностью применения к большим объемам данных; 2) ограниченностью применения к конкретным языкам. 	ПК-6.3.1
34	<p>Выберите правильный ответ</p> <p>Формальное моделирование естественного языка представляет собой</p> <ol style="list-style-type: none"> 1) моделирование естественного языка с использованием формальных методов; 2) моделирование естественного языка на основе математических формул; 3) моделирование естественного языка с учетом формальных правил. 	ПК-6.3.1
35	<p>Выберите правильный ответ</p> <p>Формальное моделирование языка используется в системах</p> <ol style="list-style-type: none"> 1) машинного перевода; 2) генерация текста; 3) все вышеперечисленное. 	ПК-6.3.1
36	<p>Выберите правильный ответ</p> <p>Преимущества формальных методов моделирования связаны с</p> <ol style="list-style-type: none"> 1) отсутствием ошибок; 2) возможностью автоматизации речевых процессов. 	ПК-6.3.1
37	<p>Выберите правильный ответ</p>	ПК-6.3.1

	Недостатки формальных методов моделирования связаны с 1) необходимостью создания сложной архитектуры; 2) ограниченной выразительностью порождаемых текстов.	
38	Выберите правильный ответ Проблемы, возникающие при использовании формальных методов моделирования естественного языка для генерации текста, могут быть вызваны 1) отсутствием достаточного объема обучающих данных; 2) отсутствием достаточного числа разработчиков.	ПК-6.3.1
39	Выберите 2 правильных ответа Системы обработки естественного языка (NLP) включают в себя подсистемы 1) морфологического анализа, 2) синтаксического анализа, 3) семантического анализа, 4) прагматического анализа.	ПК-6.3.1
40	Выберите правильный ответ Машинный перевод является одной из областей 1) нейролингвистики; 2) математической лингвистики; 3) обработки естественного языка.	ПК-6.3.1
41	Выберите правильный ответ В системах машинного перевода используются 1) статистические методы, 2) методы, основанные на правилах; 3) гибридные методы; 4) все вышеперечисленное.	ПК-6.У.1
42	Выберите правильный ответ В системах машинного перевода снижение качества перевода вызвано проблемами 1) понимания контекста, 2) понимания идиоматических выражений; 3) понимания культурных особенностей; 4) все вышеперечисленное	ПК-6.У.1
43	Выберите правильный ответ Принципы построения больших языковых моделей основаны 1) на использовании глубокого обучения и нейронных сетей; 2) на использовании правил из учебников и грамматик.	ПК-6.У.1
44	Определите основные факторы, влияющие на качество результатов больших языковых моделей	ПК-6.У.1
45	Охарактеризуйте этап валидации большой языковой модели	ПК-6.У.1
46	Охарактеризуйте, к какому виду компонентов речи (вербальные или невербальные) относятся интонация и мимика	ПК-6.У.1
47	Охарактеризуйте и обоснуйте применимость машинного перевода к языковой паре «язык банту – вепсский язык»	ПК-6.У.1
48	Охарактеризуйте тип проблемы, лежащей в основе неадекватного машинного перевода ‘Professor Barrow befriended young Newton.’ – «Профессор Тележка опекал молодого Ньютона».	ПК-6.У.1

Вопросы на проверку остаточных знаний		
1	<p>Выберите правильный ответ</p> <p>Условия, которые необходимо соблюдать при решении задачи, относятся к</p> <ol style="list-style-type: none"> ограничениям в контексте решения задач понимаются; ресурсам в контексте решения задач понимаются; результатам, которые должны быть достигнуты при решении задачи. 	УК-2.В.3
2	<p>Выберите правильный ответ</p> <p>Основным критерием использования цифровых средств является</p> <ol style="list-style-type: none"> общий уровень развития технологий; уровень цифровой грамотности населения; эффективность использования цифровых средств для решения конкретной задачи. 	УК-2.В.3
3	<p>Выберите верное утверждение</p> <ol style="list-style-type: none"> Для разработки чат-бота необходимо привлечение специалистов. Чат-бот можно разработать самостоятельно, пользуясь специализированными платформами и системами. 	УК-2.В.3
4	<p>Выберите верное утверждение</p> <p>Улучшить качество работы систем обработки естественного языка можно</p> <ol style="list-style-type: none"> привлекая экспертов к оценке результатов в ходе обучения системы; привлекая ученых-лингвистов к разработке системы. 	ПК-4.3.1
5	<p>Выберите правильный ответ</p> <p>В системах идентификации по голосу используются</p> <ol style="list-style-type: none"> Гауссовы смеси, скрытые марковские модели; многослойные персептроны, машина опорных векторов; все вышеперечисленное. 	ПК-4.3.1
6	<p>Выберите правильный ответ</p> <p>Целью фоносемантического анализа является</p> <ol style="list-style-type: none"> оценка звукового состава слогана или названия бренда; оценка восприятия рекламного слогана или названия бренда аудиторией получателя 	ПК-4.3.1
7	<p>Выберите правильный ответ</p> <p>Фраза «Ты чё, совсем с катушек слетел?!» относится к</p> <ol style="list-style-type: none"> публицистическому стилю; разговорной речи; просторечию. 	ПК-4.3.1
8	<p>Выберите верное утверждение</p> <p>Лемматизация и стемминг</p> <ol style="list-style-type: none"> повышают эффективность обработки текстов, так как снижают количество уникальных токенов; снижают эффективность обработки текстов, так как повышают количество уникальных токенов. 	ПК-6.3.1
9	<p>Выберите верное утверждение</p>	ПК-6.3.1

	Наличие в тексте имен собственных, омонимов, многозначных слов, идиом 1) потенциально повышает качество машинного перевода 2) потенциально снижает качество машинного перевода.	
10	Выпишите леммы в следующем предложении 'Stephen Hawking is one of the best known scientists'	ПК-6.У.1
11	Выпишите стоп-слова в следующем предложении 'Stephen Hawking is one of the best known scientists'	ПК-6.У.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	Сравнение эффективности работы голосовых помощников (2 по выбору студента)
2	Оценка эффективности работы систем распознавания речи на видео/аудио

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 (*если предусмотрено учебным планом по данной дисциплине*).

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.

Планируемые результаты при освоении обучающимися лекционного материала:

- получение современных, целостных, взаимосвязанных знаний, уровень которых определяется целевой установкой к каждой конкретной теме;
- получение опыта творческой работы совместно с преподавателем;
- развитие профессионально-деловых качеств, любви к предмету и самостоятельного творческого мышления.
- появление необходимого интереса, необходимого для самостоятельной работы;
- получение знаний о современном уровне развития науки и техники и о прогнозе их развития на ближайшие годы;
- научиться методически обрабатывать материал (выделять главные мысли и положения, приходить к конкретным выводам, повторять их в различных формулировках);
- получение точного понимания всех необходимых терминов и понятий.

Лекционный материал может сопровождаться демонстрацией слайдов и использованием раздаточного материала при проведении коротких дискуссий об особенностях применения отдельных тематик по дисциплине.

Структура предоставления лекционного материала:

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

Начинать надо с изучения рекомендованной литературы. Необходимо помнить, что на лекции обычно рассматривается не весь материал, а только его часть. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. В процессе этой работы студент должен стремиться понять и запомнить основные положения рассматриваемого материала, примеры, поясняющие его, а также разобраться в иллюстративном материале.

Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам.

Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику и тем самым проникнуть в творческую лабораторию автора.

Ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд подсобных материалов для быстрого повторения прочитанного, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

11.2. Методические указания для обучающихся по участию в семинарах

Не предусмотрены

Основной целью для обучающегося является систематизация и обобщение знаний

11.3. Методические указания для обучающихся по прохождению практических занятий

Не предусмотрены

11.4. Методические указания для обучающихся по выполнению лабораторных работ

В ходе выполнения лабораторных работ обучающийся должен углубить и закрепить знания, практические навыки, овладеть современной методикой и техникой эксперимента в соответствии с квалификационной характеристикой обучающегося. Выполнение лабораторных работ состоит из экспериментально-практической, расчетно-аналитической частей и контрольных мероприятий.

Выполнение лабораторных работ обучающимся является неотъемлемой частью изучения дисциплины, определяемой учебным планом, и относится к средствам, обеспечивающим решение следующих основных задач обучающегося:

- приобретение навыков исследования процессов, явлений и объектов, изучаемых в рамках данной дисциплины;
- закрепление, развитие и детализация теоретических знаний, полученных на лекциях;

- получение новой информации по изучаемой дисциплине;
- приобретение навыков самостоятельной работы с лабораторным оборудованием и приборами.

Лабораторное занятие состоит из следующих элементов: вводная часть, основная и заключительная.

Вводная часть обеспечивает подготовку студентов к выполнению заданий работы. В ее состав входят:

- формулировка темы, цели и задач занятия, обоснование его значимости в профессиональной подготовке студентов;
- изложение теоретических основ работы; - характеристика состава и особенностей заданий работы и объяснение методов (способов, приемов) их выполнения;
- характеристика требований к результату работы;
- инструктаж по технике безопасности при эксплуатации технических средств;
- проверка готовности студентов выполнять задания работы;
- указания по самоконтролю результатов выполнения заданий студентами.

Основная часть включает процесс выполнения лабораторной работы, оформление отчета и его защиту. Она может сопровождаться дополнительными разъяснениями по ходу работы, устранением трудностей при ее выполнении, текущим контролем и оценкой результатов отдельных студентов, ответами на вопросы студентов. Возможно пробное выполнение задания(ий) под руководством преподавателя.

Заключительная часть содержит:

- сбор отчетов студентов для проверки, изложение сведений, касающихся подготовки к выполнению следующей работы.

Задание и требования к проведению лабораторных работ

В начале занятия студенты получают темы и материалы для выполнения лабораторной работы.

Структура и форма отчета о лабораторной работе

Теоретическая часть должна содержать минимум необходимых теоретических сведений о предметной области. Не следует копировать целиком или частично описание лабораторной работы или разделы учебника.

В разделе «Практическая часть» необходимо описать, с помощью каких инструментальных средств и каким образом были получены результаты. Рисунки, блок-схемы, описание модели и её особенностей, необходимость отладки – все это должно быть представлено в указанном разделе могут быть представлены рисунки, скриншоты и т.д.

Раздел «Результаты» включает в себя результаты, полученные при выполнении лабораторной работы. Рисунки, графики и таблицы (при наличии) нумеруются и подписываются заголовками.

Выводы не должны быть простым перечислением того, что сделано. Здесь важно отметить, какие новые знания о предмете исследования были получены при выполнении работы, к чему привело обсуждение результатов, насколько выполнена заявленная цель работы.

Требования к оформлению отчета о лабораторной работе

Объём отчёта должен быть оптимальным для понимания того, что и как сделал студент, выполняя работу. Обязательные требования к отчёту включают общую и специальную грамотность изложения, а также использование необходимых форм и аккуратность оформления. Отчет о лабораторной работе необходимо прикрепить в ЛК.

Форма отчета: <https://guap.ru/regdocs/docs/uch> (Для учебного процесса – Документация – Нормативная документация – ГУАП)

11.5. Методические указания для обучающихся по прохождению курсового проектирования/выполнения курсовой работы
Не предусмотрено

11.6. 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, самостоятельная работа может включать в себя контрольную работу.

В процессе выполнения самостоятельной работы, у обучающегося формируется целесообразное планирование рабочего времени, которое позволяет им 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).

Начинать надо с изучения рекомендованной литературы. Необходимо помнить, что на лекции обычно рассматривается не весь материал, а только его часть. Остальная его часть восполняется в процессе самостоятельной работы. В связи с этим работа с рекомендованной литературой обязательна. Особое внимание при этом необходимо обратить на содержание основных положений и выводов, объяснение явлений и фактов, уяснение практического приложения рассматриваемых теоретических вопросов. В процессе этой работы студент должен стремиться понять и запомнить основные положения рассматриваемого материала, примеры, поясняющие его, а также разобраться в иллюстративном материале.

Заканчивать подготовку следует составлением плана (конспекта) по изучаемому материалу (вопросу). Это позволяет составить концентрированное, сжатое представление по изучаемым вопросам.

Записи имеют первостепенное значение для самостоятельной работы студентов. Они помогают понять построение изучаемого материала, выделить основные положения, проследить их логику и тем самым проникнуть в творческую лабораторию автора.

Ведение записей способствует превращению чтения в активный процесс, мобилизует, наряду со зрительной, и моторную память. Следует помнить: у студента, систематически ведущего записи, создается свой индивидуальный фонд подсобных материалов для быстрого повторения прочитанного, для мобилизации накопленных знаний. Особенно важны и полезны записи тогда, когда в них находят отражение мысли, возникшие при самостоятельной работе.

11.7. 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.

При выполнении контрольной работы 1 необходимо выбрать два голосовых помощника (Алиса от Яндекс, Маруся от ВК, Елена от Сбербанка, Олег От Тинькофф-банка и т.д.). Оба помощника должны относиться к одной сфере, например, можно

сравнить два банковских голосовых помощника или два помощника, предназначены для поиска информации и выполнения общих функций. В работе необходимо произвести сравнение основных параметров, либо самостоятельно, либо пользуясь имеющими описаниями. В конце работы необходимо произвести сравнительную оценку помощников.

При выполнении контрольной работы 1 необходимо выбрать систему автоматического субтитрирования звучащей речи (например, автоматическое субтитрирование в видеохостинге Utube). Далее для анализа отбирается видео продолжительностью 5-7 минут, отобранное видео анализируется с точки зрения корректности передачи звучащей речи. Далее анализируются неточности и ошибки в передаче, делается вывод о частотных видах ошибок и неточностей, вероятных причинах их появления.

11.8. 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».

- 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».

Для допуска к промежуточной аттестации необходимо выполнение студентами контрольных и лабораторных работ, в ходе выполнения которых студенты демонстрируют приобретенные умения и навыки.

Промежуточная аттестация представляет собой письменный ответ на вопросы по материалу курса. Основой для выставления оценки за письменный ответ является демонстрация полученных в ходе курса знаний.

На промежуточную аттестацию по курсу выносятся вопросы по курсу. Использование каких-либо материалов при ответе на теоретический вопрос не допускается. Оценка ставится исходя из знаний, продемонстрированных студентом при ответе на предложенный вопрос:

«отлично» - студент свободно владеет материалом, умеет размышлять, способен ответить на все уточняющие вопросы;

«хорошо» - студент уверенно владеет материалом, умеет размышлять, способен ответить на уточняющие вопросы;

«удовлетворительно» - студент владеет основным материалом, способен ответить на 1-2 уточняющих вопроса;

«незачтено» - студент демонстрирует крайне ограниченное или фрагментарное владение материалом, неспособен ответить на уточняющие вопросы.

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