

УТВЕРЖДАЮ
 Декан факультета

_____ Матвеев П.В.

« ____ » _____ 20__

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ

Направление/специальность подготовки	45.03.03 Фундаментальная и прикладная лингвистика
Специализация/профиль/программа подготовки	Теоретическая и прикладная лингвистика
Уровень высшего образования	Бакалавриат
Форма обучения	Очная
Факультет	Р Международного промышленного менеджмента и коммуникации
Выпускающая кафедра	Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА
Кафедра-разработчик рабочей программы	О7 Информационные системы и программная инженерия

КУРС	СЕМЕСТР	ОБЩАЯ ТРУДОЁМКОСТЬ (ЗАЧЕТНЫХ ЕДИНИЦ)	ЧАСЫ (по наличию видов занятий)									ВИД ПРОМЕЖУТОЧНОГО КОНТРОЛЯ
			ОБЩАЯ ТРУДОЁМКОСТЬ	АУДИТОРНЫЕ ЗАНЯТИЯ				САМОСТОЯТЕЛЬНАЯ РАБОТА				
				ВСЕГО	ЛЕКЦИИ	ЛАБОРАТОРНЫЙ ПРАКТИКУМ	ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	ВСЕГО	КУРСОВОЙ ПРОЕКТ	КУРСОВАЯ РАБОТА	ДРУГИЕ ВИДЫ САМОСТ. РАБОТЫ	
4	7	3	108	34	17	0	17	74	0	18	56	диф. зач.

ЛИСТ СОГЛАСОВАНИЯ

РАБОЧАЯ ПРОГРАММА СОСТАВЛЕНА В СООТВЕТСТВИИ С ТРЕБОВАНИЯМИ ФЕДЕРАЛЬНОГО
ГОСУДАРСТВЕННОГО ОБРАЗОВАТЕЛЬНОГО СТАНДАРТА ВЫСШЕГО ОБРАЗОВАНИЯ (ФГОС ВО)

45.03.03 Фундаментальная и прикладная лингвистика

год набора группы: 2025

Программу составил:

Кафедра О7 Информационные системы и программная инженерия
Удовиченко Андрей Сергеевич, к.ф.-м.н., доцент

Программа рассмотрена
на заседании кафедры-разработчика
рабочей программы **О7 Информационные системы и программная инженерия**

Заведующий кафедрой Семенова Е.Г., д.т.н., проф.

Программа рассмотрена
на заседании выпускающей кафедры

Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА

Заведующий кафедрой Невзорова Г.Д., к.ф.н., доц.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ

Разделы рабочей программы

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО
3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ
4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Приложения к рабочей программе дисциплины

- Приложение 1. Аннотация рабочей программы
- Приложение 2. Технологии и формы обучения
- Приложение 3. Фонды оценочных средств

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-7 — Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

Формированию компетенций служит достижение следующих результатов образования:

ОПК-7

знания:

на уровне представлений:

- основные понятия технологии обработки текста и звучащей речи;
- формы представления информации;
- основные задачи обработки информации, представленной с помощью естественных языков;
- основные виды программных средств обработки текста и звучащей речи;

на уровне воспроизведения:

- способы представления в цифровых электронно-вычислительных машинах текстов на естественных и искусственных языках;

- классификация формальных языков по уровням и типам формальных грамматик;
- способы представления в цифровых электронно-вычислительных машинах звучащей речи;
- общая постановка задачи поиска, индексированный и неиндексированный, четкий и нечеткий

поиск;

на уровне понимания:

- принципы обработки текстов на основе метрик близости;
- принципы функционирования глобальных информационно-поисковых систем;
- сходство и различие задач классификации, кластеризации и распознавания образов;

умения:

теоретические:

- определять необходимые средства для реализации процесса обработки информации, представленной с помощью естественных языков;

практические:

- использовать программные средства обработки звучащей речи в соответствии с их документацией;

навыки:

использовать языки расширенных запросов глобальных информационно-поисковых систем.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина **ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению 45.03.03 *Фундаментальная и прикладная лингвистика*.

Содержание дисциплины является логическим продолжением дисциплин: **ЛЕКСИКОЛОГИЯ, ФОНЕТИКА, БАЗЫ ДАННЫХ, ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ПРОГРАММИРОВАНИЕ, ОБЩАЯ МОРФОЛОГИЯ, ВВЕДЕНИЕ В ТЕОРИЮ ЯЗЫКА, СИНТАКСИС**.

Содержание дисциплины является основой для освоения дисциплин: **ТЕХНОЛОГИИ КОРПУСНОЙ ЛИНГВИСТИКИ, ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ**.

Предварительные компетенции, сформированные у обучающегося до начала изучения дисциплины:

- ОПК-1 — Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики
- ОПК-6 — Способен свободно говорить и понимать речь на первом изучаемом иностранном языке в его литературной форме, включая профессиональное письменное и устное общение; владеть вторым иностранным языком в объеме, достаточном для профессионального общения и чтения научной литературы
- ОПК-7 — Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
- ПК-1.5 — Способен использовать различные цифровые средства, позволяющие во взаимодействии с другими людьми достигать поставленных целей
- ПК-93 — Способен генерировать новые идеи для решения задач цифровой экономики, абстрагироваться от стандартных моделей, перестраивать сложившиеся способы решения задач, выдвигать альтернативные варианты действий с целью выработки новых оптимальных алгоритмов

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч.

3.1. Содержание (дидактика) дисциплины

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %
				ВСЕГО	Лекции	Практические занятия		ОПК-7
4	7	Раздел 1. Раздел 1. Введение в технологию обработки текста и звучащей речи. 1.1. Основные понятия технологии обработки текста и звучащей речи. 1.2. Формы представления информации. 1.3. Основные задачи обработки информации, представленной с помощью естественных языков.	7	2	2	0	5	5
4	7	Раздел 2. Раздел 2. Принципы автоматизации обработки текстов. 2.1. Представление в цифровых электронно-вычислительных машинах текстов на естественных и искусственных языках. 2.2. Формальные языки. 2.3. Формальные грамматики.	21	7	2	5	14	20
4	7	Раздел 3. Раздел 3. Поиск в текстах на естественных языках. Обработка текстов на основе метрик близости. Информационно-поисковые системы. 3.1. Общая формулировка задачи поиска. Поиск единственного образца. 3.2. Поиск множества образцов. Регулярные выражения. 3.3. Неиндексированный и индексированный поиск. 3.4. Автоматизация проверки орфографии и грамматики текстов на естественных языках. Нечеткий поиск. Метрики и расстояния близости слов. 3.5. Информационно-поисковые системы. Оценка близости текстовых документов.	41	13	7	6	28	30
4	7	Раздел 4. Раздел 4. Принципы обработки звучащей речи. Классификация, кластеризация и распознавание образов. 4.1. Представление в цифровых электронно-вычислительных машинах звучащей речи. 4.2. Классификация и распознавание образов. 4.3. Кластеризация. 4.4. Методы распознавания текста. 4.5. Основные методы распознавания символов. 4.6. Синтез и распознавание речи.	25	8	4	4	17	35
4	7	Раздел 5. Раздел 5. Специализированные задачи обработки текста и звучащей речи. 5.1. Анализ текста без априорной информации. Статистические методы. 5.2. Анализ текста с привлечением априорной информации. Содержательные методы. 5.3. Системы машинного перевода. 5.4. Фильтрация незапрашиваемой информации. 5.5. Интерфейсы информационных систем на естественном языке. Речевые интерфейсы.	14	4	2	2	10	10
Всего за 7 семестр			108	34	17	17	74	100
Всего по дисциплине			108	34	17	17	74	100

3.2. Аудиторный практикум

№ п/п	Номер и наименование раздела дисциплины	Тема практического занятия	Объем, ауд. часов
1	Раздел 2. Раздел 2. Принципы автоматизации обработки текстов.	Способы кодирования текстовой информации	2
2		Выполнение индивидуальной практической работы 1 (ИПР-1): Представление многоязычного текста с помощью различных способов кодирования.	1
3		Способы представления метаинформации в текстовых файлах. Текстовые редакторы и текстовые процессоры.	1
4		Выполнение индивидуальной практической работы 2 (ИПР-2): Представление текста, метаинформации и информации о форматировании с помощью текстовых редакторов и текстовых процессоров.	1
5	Раздел 3. Раздел 3. Поиск в текстах на естественных языках. Обработка текстов на основе метрик близости. Информационно-поисковые системы.	Неиндексированный поиск. Способы представления поисковых образцов.	1
6		Выполнение индивидуальной практической работы 3 (ИПР-3): Поиск в текстовых файлах средствами текстового процессора с использованием регулярных выражений.	1
7		Нечеткий поиск. Метрики и расстояния близости слов.	1
8		Выполнение индивидуальной практической работы 4 (ИПР-4): Сравнение средств проверки орфографии и грамматики в текстовых редакторах и текстовых процессорах.	1

9		Принципы построения информационно-поисковых систем по массивам связанных документов.	1
10		Выполнение индивидуальной практической работы 5 (ИПР-5): Сравнение возможностей языков формирования поисковых запросов информационно-поисковых систем Яндекс и Google.	1
11	Раздел 4. Раздел 4. Принципы обработки звучащей речи.	Принципы построения программных средств классификации, кластеризации и распознавания образов.	2
12	Классификация, кластеризация и распознавание образов.	Программные средства распознавания текста.	1
13		Программные средства синтеза и распознавания речи.	1
14	Раздел 5. Раздел 5. Специализированные задачи обработки текста и звучащей речи.	Программные средства машинного перевода. Сервисы машинного перевода в сети "Интернет".	2
Всего за 7 семестр			17

3.3. Самостоятельная работа студента (СРС)

№ п/п	Номер и наименование раздела дисциплины	Содержание учебного задания	Объем, часов
1	Раздел 1. Раздел 1. Введение в технологию обработки текста и звучащей речи.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	3
2		Выполнение этапа 1 курсовой работы	2
3	Раздел 2. Раздел 2. Принципы автоматизации обработки текстов.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	4
4		Подготовка к практическим занятиям	4
5		Оформление отчета по ИПР-1	2
6		Оформление отчета по ИПР-2	2
7		Выполнение этапа 2 курсовой работы	2
8	Раздел 3. Раздел 3. Поиск в текстах на естественных языках. Обработка текстов на основе метрик близости. Информационно-поисковые системы.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	8
9		Подготовка к практическим занятиям	8
10		Оформление отчета по ИПР-3	4
11		Оформление отчета по ИПР-4	4
12		Оформление отчета по ИПР-5	2
13		Выполнение этапа 2 курсовой работы	2
14	Раздел 4. Раздел 4. Принципы обработки звучащей речи. Классификация, кластеризация и распознавание образов.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	9
15		Выполнение этапа 3 курсовой работы	8
16	Раздел 5. Раздел 5. Специализированные задачи обработки текста и звучащей речи.	Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	4

17	Оформление и подготовка к защите курсовой работы	6
Всего за 7 семестр		74

3.4. Курсовая работа

СОДЕРЖАНИЕ ЭТАПА	ПЕРИОД ИСПОЛНЕНИЯ (недели семестра)	ПЛАНИРУЕМОЕ ВРЕМЯ (час)
Этап 1. Этап 1. Анализ задачи, формулирование ограничений для конкретного варианта выполнения курсовой работы. Выявление пользователей, описание их целей применительно к рассматриваемым процессам обработки информации.	1 - 2	2
Этап 2. Этап 2. Выбор способов представления информации.	3 - 4	2
Этап 3. Этап 3. Анализ существующих программных средств решения аналогичных задач. Выбор способа построения технологической цепочки обработки информации. Обоснование выбора средств реализации рассматриваемого процесса обработки информации.	5 - 9	5
Этап 4. Этап 4. Построение технологической цепочки рассматриваемого процесса обработки информации. Разработка алгоритма использования выбранных средств реализации процесса.	10 - 15	6
Этап 5. Этап 5. Оформление пояснительной записки, подготовка к защите и защита курсовой работы.	16 - 17	3
Всего за 7 семестр		18

4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

СЕМЕСТР	НЕДЕЛИ СЕМЕСТРА																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
7	КР			ИПЗ, КР		ДР			ИПЗ, КР	ДР	ИПЗ			ИПЗ	КР	ДР	КР, Тест, диф. зач.

Условные обозначения:

- ДР – диагностическая работа;
- КР – курсовая работа;
- ИПЗ – индивидуальное практическое задание;
- Тест – тест;
- диф. зач. – дифференцированный зачет.

Текущий контроль успеваемости студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- курсовая работа;
- индивидуальное практическое задание;
- тест.

Промежуточная аттестация проводится в формах:

- дифференцированный зачет.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1. Основная литература по дисциплине:

1. А. Н. Гуцин. . Технология обработки текста и звучащей речи. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018, 66 экз.
2. В. Н. Каминский. . Информатика. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2016, 90 экз.
3. С. Г. Толмачёв. . Алгоритмы поиска в системах искусственного интеллекта. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2012, эл. рес.

5.2. Дополнительная литература по дисциплине:

не требуется.

5.3. Периодические издания:

не требуются.

5.4. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины, электронные библиотечные системы:

1. <http://www.library.voenmeh.ru> — Фундаментальная библиотека БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова;
2. <https://urait.ru/> — Главная – Образовательная платформа Юрайт. Для вузов и ссузов..

Современные профессиональные базы данных:

1. <https://rusneb.ru> – Национальная электронная библиотека (НЭБ);
2. <https://cyberleninka.ru/> - Научная электронная библиотека «Киберленинка»;
<http://www.rfbr.ru/rffi/ru/library> - Полнотекстовая электронная библиотека Российского фонда фундаментальных исследований.

Информационные справочные системы:

1. Техэксперт – Информационный портал технического регулирования: Нормы, правила, стандарты РФ;
2. http://library.voenmeh.ru/jirbis2/index.php?option=com_irbis&view=irbis&Itemid=457 - БД ГОСТов собственной генерации БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова;
3. <http://www.consultant.ru/>- КонсультантПлюс- информационный портал правовой информации.

5.5. Программное обеспечение:

1. Far 2.0;
2. LibreOffice;
3. Mozilla Firefox;
4. Opera;
5. Open Office;
6. Файловый менеджер Midnight Commander.

5.6. Информационные технологии:

взаимодействие с обучающимися посредством ЭИОС Moodle БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова.

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Лекционные занятия:

специализированные требования по оборудованию отсутствуют; аудитория с посадочными местами по количеству студентов; доска.

6.2. Практические занятия:

1. Far 2.0;
2. LibreOffice;
3. Mozilla Firefox;
4. Opera;
5. Open Office;
6. Файловый менеджер Midnight Commander.

6.3. Прочее:

1. рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;
2. рабочие места студентов, оснащенные компьютерами с доступом в Интернет, предназначенные для работы в электронной образовательной среде.

Аннотация рабочей программы

Дисциплина **ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению 45.03.03 *Фундаментальная и прикладная лингвистика*. Дисциплина реализуется на факультете О Естественнотехнический БГТУ "ВОЕНМЕХ" им. Д.Ф. Устинова кафедрой О7 Информационные системы и программная инженерия.

Дисциплина нацелена на формирование *компетенций*:

ОПК-7 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности.

Содержание дисциплины охватывает круг вопросов, связанных с основными средствами и методами обработки информации, представленной преимущественно с помощью естественных языков в текстовой и речевой формах, в основном с использованием современных информационно-коммуникационных технологий, подразумевающих использование средств вычислительной техники.

Программой дисциплины предусмотрены следующие **виды контроля**:

Текущий контроль успеваемости студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- курсовая работа;
- индивидуальное практическое задание;
- тест.

Промежуточная аттестация проводится в формах:

- дифференцированный зачет.

Общая трудоемкость освоения дисциплины составляет 3 з.е., **108 ч**. Программой дисциплины предусмотрены лекционные занятия (**17 ч.**), практические занятия (**17 ч.**), самостоятельная работа студента (**74 ч**).

ТЕХНОЛОГИИ И ФОРМЫ ОБУЧЕНИЯ

Рекомендации по освоению дисциплины для студента

Трудоемкость освоения дисциплины составляет 108 ч., из них 34 ч. аудиторных занятий, и 74 ч., отведенных на самостоятельную работу студента.

Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины приведены в таблице.

Контроль освоения дисциплины производится в соответствии с Положением о текущем, рубежном контроле успеваемости и промежуточной аттестации обучающихся.

Формы контроля и критерии оценивания приведены в приложении 3 к Рабочей программе.

Наименование работы	Рекомендуемая литература	Трудоемкость, час.
Раздел 1. Раздел 1. Введение в технологию обработки текста и звучащей речи.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	В. Н. Каминский. . Информатика: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2016 (1) А. Н. Гуцин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (1,2)	3
Выполнение этапа 1 курсовой работы		2
Итого по разделу 1		5
Раздел 2. Раздел 2. Принципы автоматизации обработки текстов.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. Н. Гуцин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (3,4) В. Н. Каминский. . Информатика: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2016 (1)	4
Подготовка к практическим занятиям		4
Оформление отчета по ИПР-1		2
Оформление отчета по ИПР-2		2
Выполнение этапа 2 курсовой работы		2
Итого по разделу 2		14
Раздел 3. Раздел 3. Поиск в текстах на естественных языках. Обработка текстов на основе метрик близости. Информационно-поисковые системы.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. Н. Гуцин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (5,6,7,8,9) С. Г. Толмачёв. . Алгоритмы поиска в системах искусственного интеллекта: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2012 (2,3)	8
Подготовка к практическим занятиям		8
Оформление отчета по ИПР-3		4
Оформление отчета по ИПР-4		4
Оформление отчета по ИПР-5		2
Выполнение этапа 2 курсовой работы		2
Итого по разделу 3		28
Раздел 4. Раздел 4. Принципы обработки звучащей речи. Классификация, кластеризация и распознавание образов.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. Н. Гуцин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (10,11,12,13,14,15)	9
Выполнение этапа 3 курсовой работы		8
Итого по разделу 4		17

Раздел 5. Раздел 5. Специализированные задачи обработки текста и звучащей речи.		
Изучение предусмотренных программой дидактических единиц по рекомендуемой литературе	А. Н. Гуцин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (16,17)	4
Оформление и подготовка к защите курсовой работы		6
Итого по разделу 5		10

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств, позволяющие оценить результаты обучения по данной дисциплине, включают в себя:

- диагностическая работа
- индивидуальное практическое задание;
- тест;
- курсовая работа;
- дифференцированный зачет.

Критерии оценивания

Диагностическая работа

Диагностическая работа проводится в форме теста в ЭИОС Moodle:

- при правильном ответе менее чем на 60% вопросов - не аттестация;
- при правильном ответе на 60% вопросов и более - аттестация.

Индивидуальное практическое задание

Требования к выполнению ИПР:

По всем ИПР необходимо успешное выполнение требования общей и вариативной части задания, включая, если это предусмотрено заданием, предъявление процесса выполнения модельной практической задачи в диалоге с компьютером при использовании указанных в задании программных средств.

Отчет по ИПР:

Дополнительно к представлению всех результатов выполнения ИПР в электронной форме предусмотрено оформление печатных отчетов, состоящих из титульного листа, вариативной части задания и основных результатов выполнения данной работы.

Защита ИПР:

Защита ИПР предусматривает обсуждение порядка решения предусмотренных его тематикой задач, включая проверку усвоения студентом соответствующих сведений из теории и степени самостоятельности при выполнении работы.

Тест

Тест считается сданным, если результат теста более 60%.

Вопросы к тесту размещены в УМК дисциплины.

Курсовая работа

Курсовая работа допускается к защите при следующих условиях:

- все пункты работы выполнены в соответствии с индивидуальным заданием;
- электронная и печатная версии пояснительной записки соответствуют установленным требованиям;
- предлагаемые решения корректно обоснованы в тексте пояснительной записки.

Оценка защиты курсовой работы производится по четырехбалльной шкале («отлично», «хорошо», «удовлетворительно» и «неудовлетворительно») на основании обсуждения порядка решения предусмотренных ее тематикой задач, включая проверку усвоения студентом соответствующих сведений из теории и степени самостоятельности при обосновании предлагаемых решений.

Дифференцированный зачет

Итоговый контроль по дисциплине проходит в форме защиты курсовой работы и дифференцированного зачета, который оформляется на 17-й неделе семестра по результатам ИПР-1, ИПР-2, ИПР-3, ИПР-4, ИПР-5, КрР и итогового тестирования. Для получения оценки «зачтено-удовлетворительно» необходимо пройти итоговое тестирование с рейтингом не менее 60% и защитить курсовую работу на оценку «удовлетворительно», либо пройти итоговое тестирование с рейтингом 60% и выше, но менее 80%, и защитить курсовую работу на оценку «хорошо». Для получения оценки «зачтено-хорошо» необходимо пройти итоговое тестирование с рейтингом 60% и выше, но менее 80%, и защитить курсовую работу на оценку «отлично», либо пройти итоговое тестирование с рейтингом 80% и выше, но менее 90%, и защитить курсовую работу на оценку «хорошо». Для получения оценки «зачтено-отлично» необходимо пройти итоговое тестирование с рейтингом 80% и выше и защитить

курсовую работу на оценку «отлично», либо пройти итоговое тестирование с рейтингом 90% и выше и защитить курсовую работу на оценку «хорошо».

Паспорт фонда оценочных средств

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %	НАИМЕНОВАНИЕ ОЦЕНОЧНОГО СРЕДСТВА
				ВСЕГО	Лекции	Практические занятия		ОПК-7	
4	7	Раздел 1. Раздел 1. Введение в технологию обработки текста и звучащей речи.	7	2	2	0	5	5	Индивидуальное практическое задание
4	7	Раздел 2. Раздел 2. Принципы автоматизации обработки текстов.	21	7	2	5	14	20	Индивидуальное практическое задание
4	7	Раздел 3. Раздел 3. Поиск в текстах на естественных языках. Обработка текстов на основе метрик близости. Информационно-поисковые системы.	41	13	7	6	28	30	Индивидуальное практическое задание
4	7	Раздел 4. Раздел 4. Принципы обработки звучащей речи. Классификация, кластеризация и распознавание образов.	25	8	4	4	17	35	Индивидуальное практическое задание
4	7	Раздел 5. Раздел 5. Специализированные задачи обработки текста и звучащей речи.	14	4	2	2	10	10	Индивидуальное практическое задание, Курсовая работа, Тест
Всего за 7 семестр			108	34	17	17	74	100	
Всего по дисциплине			108	34	17	17	74	100	

Оценочные материалы по дисциплине ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ

ОПК-7 - Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности

№ 1 Прочитайте текст и запишите развернутый обоснованный ответ

Что такое формальная грамматика с точки зрения формальных языков?

№ 2 Прочитайте текст и запишите развернутый обоснованный ответ

Что такое терминал (терминальный символ) с точки зрения формальных языков?

№ 3 Прочитайте текст и установите соответствие

Сопоставьте звуковые форматы с их описаниями.

1.	MP3	Формат, который обеспечивает высокое качество звука А. без потерь и часто используется для хранения аудио в высоком качестве. Популярный формат сжатия, который уменьшает размер файлов с
2.	WAV	Б. минимальными потерями в качестве, часто используемый в потоковых сервисах
3.	FLAC	Формат, который поддерживает сжатие без В. потерь, идеально подходящий для аудиофилов и профессионалов
4.	AAC	Стандартный формат для хранения звука на компьютерах, Г. обеспечивает высокое качество, но имеет большой размер файла
5.	OGG	Д. Формат, который часто используется в играх и приложениях, поддерживает

многоканальный
звук и сжатие
Формат,
который
Е. обеспечивает
хорошее сжатие,
но может терять
качество

№ 4 Прочитайте текст и установите соответствие
Сопоставьте графические форматы с их описаниями.

1 JPEG

А Формат,
поддерживающий
анимацию и
прозрачность, часто
используется для
простых
изображений и
иконок.

2 PNG

Б Векторный формат,
идеально
подходящий для
логотипов и
графики, которая
требует
масштабируемости.

3 GIF

В Формат, который
обычно
используется для
фотографий,
обеспечивает
хорошее сжатие, но
может терять
качество

4 BMP

Г Формат, который
сохраняет
изображения без
потерь, подходит
для редактирования
и хранения
высококачественных
изображений.

Д Формат, который
сохраняет
изображения в
растровом виде, но
не поддерживает
сжатие, что
приводит к большим
размерам файлов

№ 5 Прочитайте текст и установите последовательность
Установите правильную последовательность этапов распознавания текста.

1. Распознавание символов
2. Предобработка изображения
3. Анализ и интерпретация текста

4. Сегментация
5. Постобработка
- № 6 Прочитайте текст и установите последовательность
Установите правильную последовательность этапов распознавания звучащей речи.
1. Предобработка сигнала
 2. Запись аудиосигнала
 3. Анализ акустических признаков
 4. Распознавание слов
 5. Постобработка и вывод текста
- № 7 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа
Какая из кодировок текстовой информации включает в себя наибольшее количество символов?
Запишите номер выбранного ответа без точки и обоснование выбора
1. ASCII
 2. KOI8
 3. Unicode
 4. CP866
- № 8 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа
Соответствие документа информационному запросу, определяемое неформальным путём –
1. Индекс цитирования
 2. Пертинентность
 3. Формальная релевантность
 4. Содержательная релевантность
- № 9 Прочитайте текст, выберите правильный ответ и запишите аргументы, обосновывающие выбор ответа
Соответствие, определяемое путём сравнения образа поискового запроса с поисковым образом документа по определённом алгоритму.
1. Индекс цитирования
 2. Пертинентность
 3. Формальная релевантность
 4. Содержательная релевантность
- № 10 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов
Какие этапы проходит аналоговый сигнал при преобразовании в цифровой при методе Импульсно-кодовой модуляции?
Запишите номера выбранных ответов без пробелов и точек и обоснование выбора
1. Дискретизация по времени
 2. Таблично-волновой синтез
 3. Кодирование

4. Квантование по уровню

№ 11 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Что характерно для алгоритмов сжатия данных без потерь?

Запишите номера выбранных ответов без пробелов и точек и обоснование выбора

1. Возможность полного восстановления исходных данных
2. Применяются для текста
3. Высокая степень сжатия за счет отбрасывания части информации
4. Уменьшают объем файла до 99%

№ 12 Прочитайте текст, выберите правильные ответы и запишите аргументы, обосновывающие выбор ответов

Как называется число позиций, в которых различны символы двух слов одинаковой длины?

1. Расстояние Хемминга
2. Расстояние Левенштейна
3. кодовое расстояние
4. Расстояние Хоара