



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Волгоградский государственный технический университет»

Факультет электроники и вычислительной техники

УТВЕРЖДЕНО

Факультет электроники и вычислительной
техники

Декан Авдеюк О.А.
г.

Компьютерная лингвистика и анализ текста

рабочая программа дисциплины (модуля, практики)

Закреплена за кафедрой	Системы автоматизированного проектирования и поискового конструирования
Учебный план	Направление 09.04.01 Информатика и вычислительная техника Программа "Инженерия производственных программно-информационных систем"
Профиль	
Квалификация	Магистр
Срок обучения	2 года

Форма обучения	очная	Общая трудоемкость	4 ЗЕТ
Виды контроля в семестрах:	экзамены 2		

Семестр(Курс.Номер семестра на курсе)	2(1.2)		Итого	
	УП	ПП	УП	ПП
Лекции	16	16	16	16
Лабораторные	16	16	16	16
Итого ауд.	32	32	32	32
Контактная работа	32.35	32.35	32.35	32.35
Сам. работа	67	87.65	67	87.65
Часы на контроль	44.65	24	44.65	24
Практическая подготовка	0	0	0	0
Итого трудоемкость в часах	144	144	0	0

ЛИСТ ОДОБРЕНИЯ, СОГЛАСОВАНИЯ И АКТУАЛИЗАЦИИ РАБОЧЕЙ ПРОГРАММЫ

Разработчик(и) программы:

Доцент Коробкин Д.М. к.т.н.

Рецензент(ы):

(при наличии)

главный программист департамента, Натров В.В.

Рабочая программа дисциплины (модуля, практики)

Компьютерная лингвистика и анализ текста

разработана в соответствии с ФГОС ВО:

Федеральный государственный образовательный стандарт высшего образования - магистратура по направлению подготовки 09.04.01 Информатика и вычислительная техника (приказ Минобрнауки России от 19.09.2017 г. № 918)

составлена на основании учебного плана:

Направление 09.04.01 Информатика и вычислительная техника

Программа "Инженерия производственных программно-информационных систем"

Профиль:

утвержденного учёным советом вуза от 05.06.2019 протокол № 12.

Рабочая программа одобрена на заседании кафедры

Системы автоматизированного проектирования и поискового конструирования

номер протокола 2019 г.

Зав. кафедрой Щербаков Максим Владимирович

СОГЛАСОВАНО:

Факультет электроники и вычислительной техники

Председатель НМС факультета: Авдеюк О.А.

Протокол заседания НМС от

г. №

Рабочая программа дисциплины (модуля, практики) актуализирована 31.08.2023

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ). ВИД, ТИП ПРАКТИКИ, СПОСОБ И ФОРМА (ФОРМЫ) ЕЕ ПРОВЕДЕНИЯ.	
Цель: формирование у студентов теоретических знаний, практических умений и навыков по применению современных технологий анализа текстовых данных	
Задачи дисциплины:	
Анализ основных задач обработки естественно-языковых текстов (Natural Language Processing), лингвистических основ автоматической обработки текста.	
Изучение статистического анализа текстовых данных.	
Формирование у студентов знаний о корпусной лингвистике и корпусной разметке	
Изучение аспектов машинного обучения синтаксических и семантических анализаторов текста	
Изучение принципов формирования компьютерных моделей естественного языка, грамматик связей, деревьев зависимостей.	

2. МЕСТО ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ) В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ				
Цикл (раздел) ОП:		Б1.В		
2.1	Требования к предварительной подготовке обучающегося:			
2.1.1	Мобильные и сетевые технологии			
2.1.2	Системы обработки больших данных			
2.1.3	Технологии программирования			
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:			
2.2.1	Междисциплинарный курсовой проект			
2.2.2	Производственная практика: Научно-исследовательская работа			
2.2.3	Производственная практика: Педагогическая практика			
2.2.4	Профессиональная иноязычная коммуникация			
2.2.5	Учебная практика: Технологическая (проектно-технологическая) практика			
2.2.6	Выполнение и защита выпускной квалификационной работы			
3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)				
УК-4: Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия				
УК-4.1: Знать: правила и закономерности личной и деловой устной и письменной коммуникации; современные коммуникативные технологии на русском и иностранном языках; существующие профессиональные сообщества для профессионального взаимодействия				
Результаты обучения:				
УК-4.2: Уметь: применять на практике коммуникативные технологии, методы и способы делового общения для академического и профессионального взаимодействия.				
Результаты обучения:				
УК-4.3: Владеть: методикой межличностного делового общения на русском и иностранном языках, с применением профессиональных языковых форм, средств и современных коммуникативных технологий.				
Результаты обучения:				
ПК-2: Технологическая поддержка подготовки технических публикаций				
ПК-2.1: Знает: основы подготовки технических публикаций.				
Результаты обучения:				
ПК-2.2: Умеет: создавать технические публикации.				
Результаты обучения:				
ПК-2.3: Владеет навыками: применения средств технологической поддержки подготовки технических публикаций				
Результаты обучения:				
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)				
Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Форма контроля
1	Раздел 1. Компьютерная морфология и синтаксис			
1.1	Компьютерная морфология /Тема/	2	0	
1.1.1	Компьютерная морфология /Лек/	2	2	Эк
1.2	Компьютерный синтаксис /Тема/	2	0	
1.2.1	Компьютерная морфология /Лек/	2	2	Эк

1.2.2	СРС Раздел 1 контрольной работы /Ср/	2	16	К
1.2.3	Морфологический и синтаксический анализ /Лаб/	2	4	Эк
2	Раздел 2. Машинное обучение в лингвистике			
2.1	Машинное обучение в лингвистике /Тема/	2	0	
2.1.1	Машинное обучение в лингвистике /Лек/	2	2	Эк
2.2	Корпусная лингвистика /Тема/	2	0	
2.2.1	Корпусная лингвистика /Лек/	2	2	Эк
2.2.2	СРС Раздел 2 контрольной работы /Ср/	2	16	К
2.2.3	Корпусная лингвистика и машинное обучени /Лаб/	2	4	Эк
3	Раздел 3. Информационный поиск и извлечение информации			
3.1	Информационный поиск /Тема/	2	0	
3.1.1	Информационный поиск /Лек/	2	2	Эк
3.2	Извлечение информации из текста /Тема/	2	0	
3.2.1	Извлечение информации из текста. /Лек/	2	2	Эк
3.2.2	СРС Раздел 3 контрольной работы /Ср/	2	16	К
3.2.3	Статистический анализ текстовых данных /Лаб/	2	4	Эк
4	Раздел 4. Применение компьютерной лингвистики			
4.1	Диалоги и чат-боты /Тема/	2	0	
4.1.1	Диалоги и чат-боты /Лек/	2	2	Эк
4.2	Анализ тональности текста /Тема/	2	0	
4.2.1	Анализ тональности текста /Лек/	2	2	Эк
4.2.2	СРС Раздел 4 контрольной работы «Анализ тональности текста» /Ср/	2	19	К
4.2.3	Лаб. работа по разделу 4 «Анализ тональности текста» /Лаб/	2	4	Эк
4.2.4	Опрос /КоРа/	2	0.35	Ко
4.2.5	/Контр.раб./	2	20.65	к
4.2.6	/Экзамен/	2	24	Эк

Примечание. Формы контроля: Эк – экзамен, К- контрольная работа, Ко- контрольный опрос, Сз- семестровое задание, 3-зачет, ОП- отчет по практике.

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Оценочные средства планируемых результатов обучения представлены в виде фондов оценочных средств (ФОС), разработанных в соответствии с локальным нормативным актом университета. ФОС может быть представлен в Приложении к рабочей программе.

Виды анализа слов
Анализ морфологии на основе правил
Статистические методы анализа слов
Морфологический анализатор Rymorphy2
Анализатор Mystem
Понятие парсинга
Грамматики зависимостей
Статистический парсинг
Лингвистический процессор ЭТАП
Парсер Stanford NLP
Парсер Link Grammar Parser
Парсер NLTK
Методы машинного обучения
Регрессионные методы
Искусственные нейронные сети
Кластеризация
Основные свойства корпуса
Разметка корпусов.
Поисковый индекс
Релевантность, полнота, точность
Фильтрация и ранжирование
Какую информацию извлекаем?
Распознавание сущностей
Разрешение анафоры и кореферентности
Установление референта. Автоматические подходы.
Извлечение отношений. Обучение моделей на размеченных текстах.

Извлечение событий
Особенности диалога на естественном языке
Архитектура диалоговых систем
Принципы работы чат-ботов
Обучение диалоговых систем на реальных диалогах.
Как измерить тональность текста
Оценка качества работы алгоритмов

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)

6.1. Рекомендуемая литература

6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Онлайн-курс Компьютерная лингвистика и анализ текста в СДО «Moodle»
----	---

6.3 Перечень программного обеспечения

6.3.1.1	Adobe Acrobat Reader DC — бесплатное решение для просмотра файлов PDF
---------	---

6.3.1.2	LibreOffice — офисный пакет
---------	-----------------------------

6.4 Перечень информационных справочных систем и электронных библиотечных систем (ЭБС)

6.3.2.1	Библиотека (НТБ), http://library.vstu.ru/sci-nci
---------	---

6.3.2.2	Электронная информационно-образовательная среда университета, http://eos.vstu.ru
---------	---

6.3.2.3	ЭБС "Лань", https://e.lanbook.com/
---------	---

6.3.2.4	ЭБС "Book.ru", https://www.book.ru/
---------	--

6.3.2.5	Электронная библиотека "Grebennikon", https://grebennikon.ru/
---------	---

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ) /ОБОРУДОВАНИЕ

7.1	Мультимедийная учебная аудитория для проведения занятий лекционного и семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации. /Учебная доска, учебная мебель, интерактивная трибуна, видеопроектор.
-----	---

7.2	Лаборатория информационных технологий. /Учебная мебель, компьютерная техника, оснащенная программным обеспечением, доступом в Интернет и в электронную информационно-образовательную среду университета.
-----	--

7.3	Аудитория для самостоятельной работы обучающихся./Учебная мебель, компьютерная техника с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета (читальный зал информационно-библиотечного центра)
-----	--

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ, ПРАКТИКИ)

Организация образовательного процесса по данной дисциплине регламентируется учебным планом и расписанием учебных занятий. При формировании своей индивидуальной образовательной траектории обучающийся имеет право на перезачет дисциплины (переаттестации ее части), если она была освоена в процессе предшествующего обучения. Перезачёт (переаттестации ее части)освобождает обучающегося от необходимости повторного освоения дисциплины (полностью или частично).

Учебный процесс при преподавании курса основывается на использовании традиционных, инновационных и информационных образовательных технологий. Традиционные образовательные технологии представлены лекциями и практическими занятиями. Инновационные образовательные технологии используются в виде широкого применения активных и интерактивных форм проведения занятий. Информационные образовательные технологии реализуются путем активизации самостоятельной работы студентов в электронной информационной образовательной среде.

Лекционный курс предполагает систематизированное изложение основных вопросов учебного плана. На первой лекции лектор информирует студентов о рекомендуемой литературе и электронных источниках информации по дисциплине, с указанием, какой учебник (учебное пособие) является базовым.

Практические занятия представляют собой детализацию лекционного теоретического материала, проводятся в целях закрепления курса и охватывают основные разделы дисциплины.

Основной формой проведения практических занятий является решение конкретных задач, аналогичных которым, будут выполнять студенты на лабораторных работах.

Лабораторные работы предполагают выполнение и отчет заданий по темам, рассмотренным на лекционных и закрепленных на практических занятиях. Каждому лабораторному занятию предшествует самостоятельная подготовка студента, включающая: ознакомление с содержанием лабораторной работы по методическим указаниям; проработку теоретической части по лекционному материалу и учебникам, рекомендованным в методических указаниях;

Самостоятельная работа студентов включает изучение законспектированного на лекционных занятиях материала, дополнение его с учетом рекомендованной по данной теме литературы, самостоятельную подготовку к лабораторным работам, самостоятельное выполнение и оформление заданий контрольной работы, аналогичных выполненным на занятиях.

Перечень методических указаний для освоения дисциплины представлен в таблице 6.1.3

В течении семестра для студентов проводятся групповые текущие консультации по учебной дисциплине, а также консультация перед экзаменом.

Методические рекомендации по обучению лиц с ограниченными возможностями здоровья и инвалидов

Профессорско-педагогический состав знакомится с психолого-физиологическими особенностями обучающихся инвалидов и лиц с ограниченными возможностями здоровья (ОВЗ), индивидуальными программами реабилитации инвалидов (при наличии). При необходимости осуществляется дополнительная поддержка преподавания тьюторами, психологами, социальными работниками, прошедшими подготовку ассистентами.

В соответствии с методическими рекомендациями Минобрнауки РФ (утв. 8 апреля 2014 г. N АК-44/05вн), в курсе предполагается использовать социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создании комфортного психологического климата в студенческой группе. Подбор и разработка учебных материалов производится с учетом предоставления материала в различных формах: аудиальной, визуальной, с использованием специальных технических средств и информационных систем.

Освоение дисциплины лицами с ОВЗ осуществляется с использованием средств обучения общего и специального назначения (персонального и коллективного использования). Материально-техническое обеспечение предусматривает приспособление аудиторий к нуждам лиц с ОВЗ (при необходимости).

Форма проведения аттестации для студентов-инвалидов устанавливается с учетом индивидуальных психофизических особенностей. Для студентов с ОВЗ предусматривается доступная форма предоставления заданий оценочных средств. Студентам с инвалидностью увеличивается время на подготовку ответов на контрольные вопросы. Для таких студентов предусматривается доступная форма предоставления ответов на задания.

При необходимости для обучающихся с инвалидностью процедура оценивания результатов обучения может проводиться в несколько этапов.