

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение высшего
образования «Вятский государственный университет»
(ВятГУ)
г. Киров

Утверждаю
Директор/Декан Бушмелева Н. А.



Номер регистрации
РПД_3-09.04.03.03_2020_113701
Актуализировано: 27.02.2021

Рабочая программа дисциплины
Автоматическая обработка текстов

	наименование дисциплины
Квалификация выпускника	Магистр
Направление подготовки	09.04.03 шифр
	Прикладная информатика наименование
Направленность (профиль)	3-09.04.03.03 шифр
	Машинное обучение и анализ данных наименование
Формы обучения	Очная наименование
Кафедра-разработчик	Кафедра прикладной математики и информатики (ОРУ) наименование
Выпускающая кафедра	Кафедра прикладной математики и информатики (ОРУ) наименование

Сведения о разработчиках рабочей программы дисциплины

Котельников Евгений Вячеславович

ФИО

Цели и задачи дисциплины

Цель дисциплины	<p>формирование знаний и представлений о принципах и методах, используемых при автоматической обработке текстов на естественном языке;</p> <p>подготовка к их грамотному и эффективному применению в профессиональной деятельности.</p>
Задачи дисциплины	<ul style="list-style-type: none"> • изучение базовых понятий автоматической обработки текстов; • формирование представления об основных направлениях компьютерной лингвистики; • изучение основных методов и технологий автоматической обработки текстов на естественном языке; • формирование умений и навыков по эффективному применению инструментов автоматической обработки текстов; • развитие всех видов мышления в процессе творческого исследования методов и технологий автоматической обработки текста; • обучение самостоятельному поиску и использованию нормативно-технической и справочной литературы и электронных источников информации; • воспитание творческого подхода к решению проблем, возникающих в процессе профессиональной деятельности.

Перечень планируемых результатов обучения по дисциплине, соотнесенных с планируемыми результатами освоения образовательной программы

Компетенция ПК-7

Способен осуществлять оценку качества разработанных процедур отладки программного кода, процедур сбора диагностических данных, измерения требуемых характеристик программного обеспечения, тестовых наборов данных в соответствии с выбранной методикой, оценку результатов проверки работоспособности программного обеспечения, рефакторинг и оптимизацию кода

Знает	Умеет	Владеет
<p>основные проблемы компьютерной обработки текстов и современные подходы к их решению;</p> <p>фундаментальные понятия и идеи, современные исследования в данной области нормативные документы, определяющие требования к проверке работоспособности программного кода;</p> <p>основные принципы отладки программного кода; основные виды диагностических данных и</p>	<p>использовать основные, наиболее употребительные в лингвистике методы математической обработки результатов исследований;</p> <p>производить подготовку тестовых наборов данных и проверку работоспособности программного обеспечения на их основе; применять методы и средства проверки работоспособности программного обеспечения; анализировать значения полученных характеристик</p>	<p>навыками использования современных средств обработки лингвистических данных; навыки использования современных средств обработки лингвистических данных; способностью осуществлять оценку качества разработанных процедур отладки программного кода;</p> <p>процедурами сбора диагностических данных, измерения требуемых характеристик</p>

<p>способы их представления; основные методы измерения и оценки характеристик программного обеспечения; методы подготовки тестовых наборов данных; методы и средства проверки работоспособности программного обеспечения; методы и средства рефакторинга и оптимизации программного кода; компоненты программно-технических архитектур, существующие приложения и интерфейсы взаимодействия с ними</p>	<p>программного обеспечения</p>	<p>программного обеспечения, тестовых наборов данных в соответствии с выбранной методикой; навыком оценки результатов проверки работоспособности программного обеспечения, рефакторинг и оптимизацию кода</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Компетенция ПК-8

<p>Способен осуществлять исследование и изучение мировых практик выполнения аналитических работ, выявление проблем и сложностей в существующих практиках, описание методик проведения аналитических работ, их апробацию на выбранных проектах и доработку</p>		
Знает	Умеет	Владеет
<p>существующие мировые практики автоматической обработки текстов их достоинства и недостатки</p>	<p>выявлять проблемы и сложности в существующих практиках автоматической обработки текстов; планировать аналитические работы, планировать ресурсы, управлять проектами</p>	<p>навыком исследования и изучения мировых практик выполнения аналитических работ; навыком выявления проблем и сложностей в существующих практиках; навыком описания методик проведения аналитических работ, их апробации на выбранных проектах и доработки</p>

Структура дисциплины
Тематический план

№ п/п	Наименование разделов дисциплины	Шифр формируемых компетенций
1	Основные задачи и проблемы анализа естественно-языковых текстов	ПК-7, ПК-8
2	Прикладные задачи компьютерной лингвистики	ПК-7
3	Подготовка и прохождение промежуточной аттестации	ПК-7, ПК-8

Формы промежуточной аттестации

Зачет	Не предусмотрен (Очная форма обучения)
Экзамен	2 семестр (Очная форма обучения)
Курсовая работа	Не предусмотрена (Очная форма обучения)
Курсовой проект	Не предусмотрена (Очная форма обучения)

Трудоемкость дисциплины

Форма обучения	Курсы	Семестры	Общий объем (трудоемкость)		Контактная работа, час	в том числе аудиторная контактная работа обучающихся с преподавателем, час				Самостоятельная работа, час	Курсовая работа (проект), семестр	Зачет, семестр	Экзамен, семестр
			Часов	ЗЕТ		Всего	Лекции	Семинарские, практические занятия	Лабораторные занятия				
Очная форма обучения	1	2	216	6	113	54	18	0	36	103			2

Содержание дисциплины

Очная форма обучения

Код занятия	Наименование тем занятий	Трудоемкость, академических часов
Раздел 1 «Основные задачи и проблемы анализа естественно-языковых текстов»		84.50
Лекции		
Л1.1	Введение в автоматическую обработку текстов	2.00
Л1.2	Лингвистический анализ	4.00
Л1.3	Инструменты компьютерной лингвистики	4.00
Лабораторные занятия		
Р1.1	Морфологический анализатор	4.00
Р1.2	Программа бессловарного морфологического анализа	4.00
Р1.3	Программы синтаксического анализа	4.00
Самостоятельная работа		
С1.1	Подготовка к лекциям, разработка и оформление отчетов по лабораторным работам	36.00
Контактная внеаудиторная работа		
КВР1.1	Контактная внеаудиторная работа	26.50
Раздел 2 «Прикладные задачи компьютерной лингвистики»		104.50
Лекции		
Л2.1	Классификация и кластеризация текстов	2.00
Л2.2	Анализ тональности текстов	2.00
Л2.3	Автоматическое реферирование и аннотирование текстов	2.00
Л2.4	Машинный перевод	2.00
Лабораторные занятия		
Р2.1	Методы взвешивания терминов	6.00
Р2.2	Разработка программы классификации текстов	6.00
Р2.3	Системы анализа тональности текста	6.00
Р2.4	Системы автоматического реферирования текстов	6.00
Самостоятельная работа		
С2.1	Подготовка к лекциям, разработка и оформление отчетов по лабораторным работам	42.50
Контактная внеаудиторная работа		
КВР2.1	Контактная внеаудиторная работа	30.00
Раздел 3 «Подготовка и прохождение промежуточной аттестации»		27.00
ЭЗ.1	Подготовка к сдаче экзамена	24.50
КВР3.1	Консультация перед экзаменом	2.00
КВР3.2	Сдача экзамена	0.50
ИТОГО		216.00

Содержание дисциплины данной рабочей программы используется при обучении по индивидуальному учебному плану, при ускоренном обучении, при применении дистанционных образовательных технологий и электронном обучении (при наличии).

Методические указания для обучающихся по освоению дисциплины

Успешное освоение дисциплины предполагает активное, творческое участие обучающегося на всех этапах ее освоения путем планомерной, повседневной работы. Обучающийся обязан посещать лекции, семинарские, практические и лабораторные занятия (при их наличии), получать консультации преподавателя и выполнять самостоятельную работу.

Изучение дисциплины следует начинать с проработки настоящей рабочей программы, методических указаний и разработок, указанных в программе, особое внимание уделить целям, задачам, структуре и содержанию дисциплины.

Главной задачей каждой лекции является раскрытие сущности темы и анализ ее основных положений. Тематика лекций определяется настоящей рабочей программой дисциплины.

Лекции – это систематическое устное изложение учебного материала. На них обучающийся получает основной объем информации по каждой конкретной теме. Лекции обычно носят проблемный характер и нацелены на освещение наиболее трудных и дискуссионных вопросов.

Предполагается, что обучающиеся приходят на лекции, предварительно проработав соответствующий учебный материал по источникам, рекомендованным программой. Часто обучающимся трудно разобраться с дискуссионными вопросами, дать однозначный ответ. Преподаватель, сравнивая различные точки зрения, излагает свой взгляд и нацеливает их на дальнейшие исследования и поиск научных решений. После лекции желательно вечером перечитать и закрепить полученную информацию, тогда эффективность ее усвоения значительно возрастает. При работе с конспектом лекции необходимо отметить материал, который вызывает затруднения для понимания, попытаться найти ответы на затруднительные вопросы, используя предлагаемую литературу. Если самостоятельно не удалось разобраться в материале, сформулируйте вопросы и обратитесь за помощью к преподавателю.

Целью семинарских занятий является проверка уровня понимания обучающимися вопросов, рассмотренных на лекциях и в учебной литературе.

Целью практических и лабораторных занятий является формирование у обучающихся умений и навыков применения теоретических знаний в реальной практике решения задач; восполнение пробелов в пройденной теоретической части курса.

Семинарские, практические и лабораторные занятия в равной мере направлены на совершенствование индивидуальных навыков решения теоретических и прикладных задач, выработку навыков интеллектуальной работы, а также ведения дискуссий. Для успешного участия в семинарских, практических и лабораторных занятиях обучающемуся следует тщательно подготовиться.

Основной формой подготовки обучающихся к практическим (лабораторным) занятиям является самостоятельная работа с учебно-методическими материалами, научной литературой, статистическими данными и т.п.

Изучив конкретную тему, обучающийся может определить, насколько хорошо он в ней разобрался. Если какие-то моменты остались непонятными, целесообразно составить список вопросов и на занятии задать их преподавателю. Практические (лабораторные) занятия предоставляют обучающемуся возможность творчески раскрыться, проявить инициативу и развить навыки публичного ведения дискуссий и общения.

Самостоятельная работа обучающихся включает в себя выполнение различного рода заданий (изучение учебной и научной литературы, материалов лекций, систематизацию прочитанного материала, подготовку контрольной работы, решение

задач, подготовка докладов, написание рефератов, публикация тезисов, научных статей, подготовка и защита курсовой работы / проекта и другие), которые ориентированы на глубокое усвоение материала изучаемой дисциплины.

Обучающимся рекомендуется систематически отводить время для повторения пройденного материала, проверяя свои знания, умения и навыки.

Внутренняя система оценки качества освоения дисциплины включает входной контроль уровня подготовленности обучающихся, текущий контроль успеваемости, промежуточную аттестацию, направленную на оценивание промежуточных и окончательных результатов обучения по дисциплине (в том числе результатов курсового проектирования (выполнения курсовых работ) при наличии).

При проведении промежуточной аттестации обучающегося учитываются результаты текущего контроля, проводимого в течение освоения дисциплины.

Процедура оценивания результатов освоения дисциплины осуществляется на основе действующих локальных нормативных актов ФГБОУ ВО «Вятский государственный университет», с которыми обучающиеся знакомятся на официальном сайте университета www.vyatsu.ru.

Учебно-методическое обеспечение дисциплины, в том числе учебно-методическое обеспечение самостоятельной работы обучающегося по дисциплине

Учебная литература (основная)

1) Прикладная и компьютерная лингвистика : коллектив. монография / под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо. - 2-е изд. - Москва : URSS, 2017. - 315 с. : ил, портр. - Библиогр. в конце глав. - Тираж не указ. - ISBN 978-5-9710-4633-2 : 875.00 р. - Текст : непосредственный.

2) Кухаренко, Б. Г. Интеллектуальные системы и технологии : учебное пособие / Б.Г. Кухаренко. - Москва : Альтаир|МГАВТ, 2015. - 115 с. - Б. ц. - URL: <http://biblioclub.ru/index.php?page=book&id=429758/> (дата обращения: 24.03.2020). - Режим доступа: ЭБС Университетская библиотека ONLINE. - Текст : электронный.

Учебная литература (дополнительная)

1) Захаров, В. П. Корпусная лингвистика : учебник / В.П. Захаров, С.Ю. Богданова. - Иркутск : Иркутский государственный лингвистический университет, 2011. - 161 с. - ISBN 978-5-88267-316-0 : Б. ц. - URL: <http://biblioclub.ru/index.php?page=book&id=89753/> (дата обращения: 24.03.2020). - Режим доступа: ЭБС Университетская библиотека ONLINE. - Текст : электронный.

2) Северенс, Ч. Введение в программирование на Python / Ч. Северенс. - 2-е изд., испр. - Москва : Национальный Открытый Университет «ИНТУИТ», 2016. - 231 с. - Б. ц. - URL: <http://biblioclub.ru/index.php?page=book&id=429184/> (дата обращения: 24.03.2020). - Режим доступа: ЭБС Университетская библиотека ONLINE. - Текст : электронный.

Учебно-методические издания

1) Хахаев, И. А. Практикум по алгоритмизации и программированию на Python : курс / И.А. Хахаев. - 2-е изд., исправ. - Москва : Национальный Открытый Университет «ИНТУИТ», 2016. - 179 с. - Б. ц. - URL: <http://biblioclub.ru/index.php?page=book&id=429256/> (дата обращения: 24.03.2020). - Режим доступа: ЭБС Университетская библиотека ONLINE. - Текст : электронный.

2) Буйначев, С. К. Основы программирования на языке Python : учебное пособие / С.К. Буйначев. - Екатеринбург : Издательство Уральского университета, 2014. - 92 с. - ISBN 978-5-7996-1198-9 : Б. ц. - URL: <http://biblioclub.ru/index.php?page=book&id=275962/> (дата обращения: 24.03.2020). - Режим доступа: ЭБС Университетская библиотека ONLINE. - Текст : электронный.

Электронные образовательные ресурсы

- 1) Портал дистанционного обучения ВятГУ [электронный ресурс] / - Режим доступа: <http://mooc.do-kirov.ru/>
- 2) Раздел официального сайта ВятГУ, содержащий описание образовательной программы [электронный ресурс] / - Режим доступа: https://www.vyatsu.ru/php/programms/eduPrograms.php?Program_ID=3-09.04.03.03
- 3) Личный кабинет студента на официальном сайте ВятГУ [электронный ресурс] / - Режим доступа: <https://new.vyatsu.ru/account/>
- 4) Единое окно доступа к образовательным ресурсам <http://window.edu.ru/>

Электронные библиотечные системы (ЭБС)

- ЭБС «Научная электронная библиотека eLIBRARY» (<http://elibrary.ru/defaultx.asp>)
- ЭБС «Издательства Лань» (<http://e.lanbook.com/>)
- ЭБС «Университетская библиотека online» (www.biblioclub.ru)
- Внутренняя электронно-библиотечная система ВятГУ (<http://lib.vyatsu.ru/>)
- ЭБС «ЮРАЙТ» (<https://urait.ru>)

Современные профессиональные базы данных и информационные справочные системы

- ГАРАНТ
- КонсультантПлюс
- Техэксперт: Нормы, правила, стандарты
- Роспатент (<https://www1.fips.ru/elektronnye-servisy/informatsionno-poiskovaya-sistema>)
- Web of Science® (<http://webofscience.com>)

Материально-техническое обеспечение дисциплины

Демонстрационное оборудование

Перечень используемого оборудования
ПРОЕКТОР NEC V302H FULL 3D

Специализированное оборудование

Перечень используемого оборудования
ГРАФИЧЕСКАЯ РАБОЧАЯ СТАНЦИЯ DEPO Race G560S
ПЕРСОНАЛЬНЫЙ КОМПЬЮТЕР TP CORP OPTIMA
ПЕРСОНАЛЬНЫЙ КОМПЬЮТЕР В СБОРЕ

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, в том числе лицензионное и свободно распространяемое ПО (включая ПО отечественного производства)

№ п.п	Наименование ПО	Краткая характеристика назначения ПО
1	Программная система с модулями для обнаружения текстовых заимствований в учебных и научных работах «Антиплагиат.ВУЗ»	Программный комплекс для проверки текстов на предмет заимствования из Интернет-источников, в коллекции диссертация и авторефератов Российской государственной библиотеки (РГБ) и коллекции нормативно-правовой документации LEXPRO
2	Microsoft Office 365 ProPlusEdu ALNG SubsVL MVL AddOn toOPP	Набор веб-сервисов, предоставляющий доступ к различным программам и услугам на основе платформы Microsoft Office, электронной почте бизнес-класса, функционалу для общения и управления документами
3	Office Professional Plus 2016	Пакет приложений для работы с различными типами документов: текстами, электронными таблицами, базами данных, презентациями
4	Windows Professional	Операционная система
5	Kaspersky Endpoint Security для бизнеса	Антивирусное программное обеспечение
6	Справочная правовая система «Консультант Плюс»	Справочно-правовая система по законодательству Российской Федерации
7	Электронный периодический справочник ГАРАНТ Аналитик	Справочно-правовая система по законодательству Российской Федерации
8	Security Essentials (Защитник Windows)	Защита в режиме реального времени от шпионского программного обеспечения, вирусов.
9	МойОфис Стандартный	Набор приложений для работы с документами, почтой, календарями и контактами на компьютерах и веб браузерах

Обновленный список программного обеспечения данной рабочей программы находится по адресу:
https://www.vyatsu.ru/php/list_it/index.php?op_id=113701