

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ЛУГАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ ВЛАДИМИРА ДАЛЯ»**

**Факультет компьютерных систем и информационных технологий
Кафедра информатики и программной инженерии**

УТВЕРЖДАЮ

**Декан факультета компьютерных
систем и информационных технологий**

Кочевский А. А.

« 19 » 04 2023 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

«Обработка естественного языка»

по направлению подготовки 09.03.03 Прикладная информатика
профиль подготовки «Прикладная информатика в экономике»

Луганск – 2023

Лист согласования РПУД

Рабочая программа учебной дисциплины «Обработка естественного языка» по направлению подготовки 09.03.03 Программная инженерия. – 9 с.

Рабочая программа учебной дисциплины «Обработка естественного языка» составлена на основе федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.03 Прикладная информатика, утвержденного приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 года № 922 (с изменениями и дополнениями), зарегистрированным в Министерстве юстиции Российской Федерации от 12 октября 2017 года № 48531, учебного плана по направлению подготовки 09.03.03 Прикладная информатика, (профиль «Прикладная информатика в экономике») и Положения о рабочей программе учебной дисциплины в ФГБОУ ВО «ЛГУ им. В. Даля».

СОСТАВИТЕЛЬ

канд. техн. наук, доцент, доцент кафедры информатики и программной инженерии
Петрущенко Т. В.

Рабочая программа дисциплины утверждена на заседании кафедры информатики и программной инженерии

18 апреля 2023 г., протокол № 17

Заведующий кафедрой информатики и программной инженерии
_____ Кочевский А. А.

Переутверждена: «__» _____ 20__ г., протокол № _____

Согласована:

Декан факультета компьютерных систем и информационных технологий
_____ Кочевский А. А.

Рекомендована на заседании учебно-методической комиссии факультета компьютерных систем и информационных технологий

19 апреля 2023 г., протокол № 8

Председатель учебно-методической комиссии факультета компьютерных систем и информационных технологий
_____ Ветрова Н. Н.

© Петрущенко Т. В., 2023 год

© ФГБОУ ВО «ЛГУ им. В. Даля», 2023 год

Структура и содержание дисциплины

1. Цели и задачи дисциплины, ее место в учебном процессе

Цель изучения дисциплины – ознакомление с методами обработки текста на естественном языке, а также методами обработки слабоструктурированных данных и извлечения информации; знакомство с методами извлечения отношений, анализа тональности, аннотирования и кластеризации текстов, а также с существующими программными реализациями этих методов.

Задачи: получение теоретических знаний и практических навыков обработки естественно-языковых текстов; получение представления о сложностях, связанных с применением существующих методов обработки естественно-языковых текстов; развитие умений использовать полученные знания по разработке, адаптации и использованию новейших средств информатики для обработки текстов на естественных языках.

2. Место дисциплины в структуре ОПОП ВО

Дисциплина входит в часть, формируемой участниками образовательных отношений учебного плана по направлению подготовки 09.03.03 Прикладная информатика

Дисциплина реализуется кафедрой информатики и программной инженерии.

Основывается на базе дисциплин: «Введение в прикладную информатику».

Является основой для изучения следующих дисциплин: «Интеллектуальные информационные системы», выполнение квалификационной работы.

3. Требования к результатам освоения содержания дисциплины

Студенты, завершившие изучение дисциплины «Разработка программно-информационных систем», должны

знать: как использовать современные языки и системы программирования, технологии проектирования программного обеспечения.

уметь: сформулировать требования к разрабатываемому программному обеспечению, выполнить его реализацию и оформить техническую документацию на его компоненты.

владеть: навыками осуществления проектирования программного обеспечения конкретной ИС и разработку технической документации на ее компоненты.

Перечисленные результаты образования являются основой для формирования следующих компетенций (в соответствии с ФГОС ВО и требованиями к результатам освоения основной профессиональной образовательной программы (ОПОП ВО):

профессиональных:

ПК-2 Способен осуществлять проектирование программного обеспечения ИС и разрабатывать техническую документацию на его компоненты

4. Структура и содержание дисциплины

4.1. Объем учебной дисциплины и виды учебной работы

Вид учебной работы	Объем часов (з.е.)		
	Очная форма	Очно-заочная форма	Заочная форма
Объем учебной дисциплины (всего)	72 (2 з.е.)	-	108 (3 з.е.)
Обязательная аудиторная учебная нагрузка дисциплины (всего) в том числе:	56	-	12
Лекции	28	-	6
Семинарские занятия	-	-	-
Практические занятия	-	-	-
Лабораторные работы	28	-	6
Курсовая работа (курсовой проект)	-	-	-
Индивидуальное задание	-	-	-
Самостоятельная работа студента (всего)	16	-	56
Форма аттестации	зачет	-	зачет

4.2. Содержание разделов дисциплины

Семестр 8

- Тема 1. Введение в обработку естественно-языковых текстов
Лингвистика как наука о языке. Представление об уровнях представления языка - фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика. Лингвистическое моделирование. Действующие модели языка. Теория «Смысл - Текст» как фундамент для построения систем автоматической обработки текста.
- Тема 2. Методы обработки естественных языков
Анализ и синтез текста. Морфологический и синтаксический анализ. Парсинг. Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх». Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Интерактивное разрешение лексической и синтаксической неоднозначности. Правильные и статистические подходы к автоматической обработке текста. Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы. Синтаксическая структура предложения.
- Тема 3. Вопросно-ответные системы
Вопросно-ответные системы: основы вопросно-ответной системы,

архитектуры вопросно-ответной системы, установление смысла вопроса и порождение ответов. Распознавание имён людей, географических названий и других сущностей, различные подходы к распознаванию именованных сущностей

Тема 4. Программирование и проектирование систем обработки естественных языков

Задачи морфологического анализа, морфологический разбор, стемминг, лемматизация. Понятия лексемы, словоформы, леммы, морфемы, псевдо-основы и псевдо-окончания. Грамматические категории. Словоизменительная парадигма. Морфотактика. Структура данных морфологического словаря, лексикона. Грамматические модели русского языка в контексте автоматической обработки. Минимальное расстояние редактирования. Алгоритм подсчета расстояния Левенштейна. Практика по подсчету минимального расстояния Левенштейна. Понятие статистической языковой модели. Области применения. N-граммы.

4.3. Лекции

№ п/п	Название темы	Объем часов		
		Очная форма	Очно- заочная форма	Заочная форма
1	Введение в обработку естественно- языковых текстов	7	-	1
2	Методы обработки естественных языков	7	-	1
3	Вопросно-ответные системы	7	-	2
4	Введение в обработку естественно- языковых текстов	7	-	2
Итого:		28	-	6

4.4. Практические (семинарские) занятия

Не предусмотрены.

4.5. Лабораторные работы

№ п/п	Название темы	Объем часов		
		Очная форма	Очно- заочная форма	Заочная форма
1	Построение транскриптора слов русского языка	7	-	2
2	Обработка естественно-языковых текстов на морфологическом уровне.	7	-	1
3	Морфологический анализ слов русского языка.	7	-	1
4	Элементы синтаксического анализа.	7	-	2
Итого:		28	-	6

4.6. Самостоятельная работа студентов

№ п/п	Название темы	Вид СРС	Объем часов		
			Очная форма	Очно- заочная форма	Заочная форма
1	Введение в обработку естественно- языковых текстов	подготовка к лабораторным работам и оформление отчетов	4	-	14
2	Методы обработки естественных языков	подготовка к лабораторным работам и оформление отчетов	4	-	14
3	Вопросно-ответные системы	подготовка к лабораторным работам и оформление отчетов	4	-	14
4	Введение в обработку естественно- языковых текстов	подготовка к лабораторным работам и оформление отчетов	4		14
Итого:			16	-	56

4.7. Курсовые работы/проекты

Не предусмотрены.

5. Образовательные технологии

Лекционные занятия проводятся с использованием педагогической технологии продукционного обучения.

Проведение лекционного занятия - демонстрация слайдов лекционного материала с подробным объяснением излагаемого учебного материала.

С целью формирования и развития профессиональных навыков обучающихся применяются интерактивные формы обучения на аудиторных занятиях. Учебный процесс, опирающийся на использование интерактивных методов обучения, организуется с учетом включенности в процесс познания всех студентов группы без исключения. Совместная деятельность означает, что каждый вносит свой особый индивидуальный вклад, в ходе работы идет обмен знаниями, идеями, способами деятельности. Создается среда образовательного общения, которая характеризуется открытостью, взаимодействием участников, равенством их аргументов, накоплением совместного знания, возможностью взаимной оценки и контроля.

6. Формы контроля освоения дисциплины

Текущая аттестация студентов производится в дискретные временные интервалы лектором или преподавателем, ведущим лабораторные занятия по дисциплине, в следующих формах:

- фронтальные и индивидуальные опросы;
- контрольные работы;

Фонды оценочных средств, включающие типовые контрольные работы, позволяющие оценить результаты текущей и промежуточной аттестации

обучающихся по данной дисциплине, помещаются в приложении к рабочей программе в соответствии с «Положением о фонде оценочных средств».

Промежуточная аттестация по результатам освоения дисциплины проходит в форме зачета (предполагает выполнение всех видов учебной работы, предусмотренных рабочей программой учебной дисциплины).

В экзаменационную ведомость и зачетную книжку выставляются оценки по шкале, приведенной в таблице.

Зачеты	Характеристика знания предмета и ответов
зачтено	Студент глубоко и в полном объеме владеет программным материалом. Грамотно, исчерпывающе и логично его излагает в устной или письменной форме. При этом знает рекомендованную литературу, проявляет творческий подход в ответах на вопросы и правильно обосновывает принятые решения, хорошо владеет умениями и навыками при выполнении практических задач.
	Студент знает программный материал, грамотно и по сути излагает его в устной или письменной форме, допуская незначительные неточности в утверждениях, трактовках, определениях и категориях или незначительное количество ошибок. При этом владеет необходимыми умениями и навыками при выполнении практических задач.
	Студент знает только основной программный материал, допускает неточности, недостаточно четкие формулировки, непоследовательность в ответах, излагаемых в устной или письменной форме. При этом недостаточно владеет умениями и навыками при выполнении практических задач. Допускает до 30% ошибок в излагаемых ответах.
не зачтено	Студент не знает значительной части программного материала. При этом допускает принципиальные ошибки в доказательствах, в трактовке понятий и категорий, проявляет низкую культуру знаний, не владеет основными умениями и навыками при выполнении практических задач. Студент отказывается от ответов на дополнительные вопросы.

7. Учебно-методическое и программно-информационное обеспечение дисциплины:

а) основная литература:

1. Винокурский Д. Л. Инструментальные средства информационных систем: курс лекций: учебное пособие / Д.Л. Винокурский, Б.В. Крахоткина - Ставрополь: СКФУ, 2018. - 165 с. <http://biblioclub.ru/index.php?page=book&id=562702>

2. Программные системы статистического анализа: обнаружение закономерностей в данных с использованием системы R и языка Python: учебное пособие / В.М. Волкова, М.А. Семенова, Е.С. Четвертакова, С.С. Вожов; Новосибирский государственный технический университет - Новосибирск: Новосибирский государственный технический университет, 2017. - 74 с. <http://biblioclub.ru/index.php?page=book&id=576496>

3. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч.2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же

4. Кокорина, И.В. Основы математической обработки информации в филологии: комбинаторика, теория вероятностей и математическая статистика : учебно-методическое пособие / И.В. Кокорина ; Министерство образования и науки Российской Федерации, Федеральное государственное автономное образовательное учреждение высшего профессионального образования Северный (Арктический) федеральный университет им. М.В. Ломоносова. - Архангельск : ИД САФУ, 2014. - 115 с. : ил. - Библиогр. в кн. - ISBN 978-5-261-00928-3 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=312317>

б) дополнительная литература:

5. Ромм Я. Е. Детерминированный информационный поиск на основе сортировки с распараллеливанием базовых операций / Я.Е. Ромм; С.С. Белоконова - М.: Издательство Научный мир, 2014. - 197 с. <http://biblioclub.ru/index.php?page=book&id=468725>

6. Миркин Б. Г. Введение в анализ данных. Учебник и практикум - М.: Издательство Юрайт, 2020. - 174 с. <https://urait.ru/book/vvedenie-v-analiz-dannyh-450262>

7. Костюкова, Н.И. Комбинаторные алгоритмы для программистов / Н.И. Костюкова. - 2-е изд./, исправ./ - Москва : Национальный Открытый Университет «ИНТУИТ», 2016. - 217 с. : ил. ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=429067>

8. Гуслякова, А.В. Информационные технологии и лингвистика XXI века : учебное пособие / А.В. Гуслякова ; Министерство образования и науки Российской Федерации. - Москва : МПГУ, 2016. - 96 с. : ил. - Библиогр. в кн. - ISBN 978-5-4263-0398-0 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=469675>

в) Internet-ресурсы:

Министерство образования и науки Российской Федерации – <http://минобрнауки.рф/>

Федеральная служба по надзору в сфере образования и науки – <http://obrnadzor.gov.ru/>

Министерство образования и науки Луганской Народной Республики – <https://minobr.su>

Народный совет Луганской Народной Республики – <https://nslnr.su>

Портал Федеральных государственных образовательных стандартов высшего образования – <http://fgosvo.ru>

Федеральный портал «Российское образование» – <http://www.edu.ru/>

Информационная система «Единое окно доступа к образовательным ресурсам» – <http://window.edu.ru/>

Федеральный центр информационно-образовательных ресурсов – <http://fcior.edu.ru/>

Электронные библиотечные системы и ресурсы

Электронно-библиотечная система «Консультант студента» –
<http://www.studentlibrary.ru/cgi-bin/mb4x>

Электронно-библиотечная система «StudMed.ru» –
<https://www.studmed.ru>

Информационный ресурс библиотеки образовательной организации
Научная библиотека имени А. Н. Коняева – <http://biblio.dahluniver.ru/>

8. Материально-техническое обеспечение дисциплины

Освоение дисциплины предполагает использование академических аудиторий, соответствующих действующим санитарным и противопожарным правилам и нормам.

Прочее: рабочее место преподавателя, оснащенное компьютером с доступом в Интернет.

Программное обеспечение:

Функциональное назначение	Бесплатное программное обеспечение	Ссылки
Офисный пакет	Libre Office 6.3.1	https://www.libreoffice.org/ https://ru.wikipedia.org/wiki/LibreOffice
Операционная система	UBUNTU 19.04	https://ubuntu.com/ https://ru.wikipedia.org/wiki/Ubuntu
Браузер	Firefox Mozilla	http://www.mozilla.org/ru/firefox/fx
Браузер	Opera	http://www.opera.com
Почтовый клиент	Mozilla Thunderbird	http://www.mozilla.org/ru/thunderbird
Файл-менеджер	Far Manager	http://www.farmanager.com/download.php
Архиватор	7Zip	http://www.7-zip.org/
Графический редактор	GIMP (GNU Image Manipulation Program)	http://www.gimp.org/ http://gimp.ru/viewpage.php?page_id=8 http://ru.wikipedia.org/wiki/GIMP
Редактор PDF	PDFCreator	http://www.pdfforge.org/pdfcreator
Аудиоплеер	VLC	http://www.videolan.org/vlc/