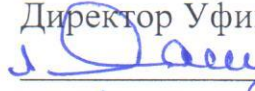


Федеральное государственное образовательное бюджетное
учреждение высшего образования
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**
(Финансовый университет)

Уфимский филиал Финуниверситета
Кафедра «Математика и информатика»

УТВЕРЖДАЮ

Директор Уфимского филиала

 Р.М. Сафуанов

« 1 » сентября 2021 г.

ОБРАБОТКА ТЕКСТОВ НА ЕСТЕСТВЕННЫХ ЯЗЫКАХ

Рабочая программа дисциплины

для студентов, обучающихся по направлению подготовки
09.03.03 Прикладная информатика,
образовательная программа «Прикладная информатика»,
(ИТ-сервисы и технологии обработки данных в экономике и финансах)

Рекомендовано Ученым советом филиала
(протокол № 39 от « 31 » августа 2021г.)

Одобрено кафедрой «Математика и информатика»
(протокол № 16 от « 30 » июня 2021г.)

Уфа 2021

Составитель **рабочей** программы дисциплины «Обработка текстов на естественных языках» для студентов, обучающихся по направлению подготовки 09.03.03 «Прикладная информатика», образовательная программа: «Прикладная информатика» (ИТ-сервисы и технологии обработки данных в экономике и финансах), доцент кафедры «Математика и информатика» Уфимского филиала Финуниверситета, канд. техн. наук, доцент Фархиева С.А.

ОГЛАВЛЕНИЕ

	Стр.
1. Наименование дисциплины	4
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	4
3. Место дисциплины в структуре образовательной программы	5
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся	5
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий	6
6. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины	9
7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины	10
8. Методические указания для обучающихся по освоению дисциплины	10
9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем	11
10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине	11

1. Наименование дисциплины

Обработка текстов на естественных языках

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ОПК- 7	Способен разрабатывать алгоритмы и программы, пригодные для практического применения	1. Разрабатывает алгоритмы решения прикладных задач с использованием математических и современных аналитических методов.	Знать: методы разработки алгоритмов решения прикладных задач с использованием математических и современных аналитических методов обработки текстов на естественных языках. Уметь: разрабатывать алгоритмы решения прикладных задач с использованием математических и современных аналитических методов обработки текстов на естественных языках.
		2. Реализует алгоритмы с использованием современных средств разработки прикладного программного обеспечения.	Знать: алгоритмы с использованием современных средств разработки прикладного программного обеспечения для обработки текстов на естественных языках. Уметь: реализовывать алгоритмы с использованием современных средств разработки прикладного программного обеспечения для обработки текстов на естественных языках.
		3. Владеет навыками тестирования программного обеспечения, необходимыми для создания программных продуктов промышленного качества.	Знать: методы тестирования программного обеспечения, необходимыми для создания программных продуктов обработки текстов на естественных языках промышленного качества. Уметь: тестировать программное обеспечение, необходимое для создания программных продуктов обработки текстов на естественных языках промышленного качества.

3. Место дисциплины в структуре образовательных программ

Дисциплина «Обработка текстов на естественных языках» относится к общепрофессиональному циклу обязательной части образовательной программы «Прикладная информатика» (ИТ-сервисы и технологии обработки данных в экономике и финансах) по направлению подготовки 09.03.03 Прикладная информатика.

4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 5 (в часах)
Общая трудоемкость дисциплины	4/144	144
Контактная работа - Аудиторные занятия	68	68
<i>Лекции</i>	16	16
<i>Семинары, практические занятия</i>	52	52
Самостоятельная работа	76	76
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

Заочная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6 (в часах)
Общая трудоемкость дисциплины	4/144	144
Контактная работа - Аудиторные занятия	14	14
<i>Лекции</i>	2	2
<i>Семинары, практические занятия</i>	12	12
Самостоятельная работа	130	130
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Экзамен	Экзамен

5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

5.1. Содержание дисциплины

Тема 1. Введение в автоматическую обработку естественного языка

Структурные особенности текстов на естественном языке. Неоднозначность на всех уровнях языка. Основные задачи автоматического анализа текстов. Основные подходы к решению задач: правила, написанные вручную и машинное обучение. Показатели качества: точность, полнота, F-мера.

Тема 2. Базовая обработка текста и дистанция редактирования

Предобработка текста: токенизация и сегментация. Нормализация слов: стеммеры, лемматизаторы, морфологические анализаторы. Регулярные выражения, дистанция редактирования.

Тема 3. Языковые модели

N-граммы. Перплексия. Методы сглаживания. Линейная интерполяция. Применение языковых моделей: предсказание ввода, исправление ошибок правописания, распознавание речи, порождение текста.

Тема 4. Классификация текстов и анализ тональности

Задачи классификации, наивный байесовский классификатор. Проблемы классификации текстов. Анализ тональности. Извлечение аспектов.

Тема 5. Информационный поиск

Векторные модели текстов Матричное представление. Обратный индекс;

Фразовые запросы. Ранжированный информационный поиск. Методы оценки поисковых машин.

Тема 6. Машинный перевод

Классические подходы: пословный, трансферный, интерлингвальный. Статистический машинный перевод. Модели IBM. Выравнивание текстов. Оценка параметров в моделях IBM. Фразовые модели. Извлечение фразовых лексиконов, алгоритм декодирования.

Тема 7. Компьютерная семантика

Значение и смысл. WordNet и аналогичные лексические базы данных. Измерение семантической близости. Тезаурусные методы. Дистрибуционные (корпусные) методы.

Тема 8. Автоматическое реферирование

Экстрактивное и абстрактное реферирование. Реферирование нескольких документов. Реферирование, основанное на запросе. Обучение с учителем и без учителя в контексте автоматического реферирования. Оценка

систем реферирования, ROUGE.

5.2. Учебно-тематический план

Очная/Заочная форма обучения

№ п/п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах					Формы теку- щего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Само- стоя- тельная работа	
			Об- щая	Лек- ции	Семина- ры, практи- ческие занятия		
1	Введение в авто- матическую обра- ботку естествен- ного языка	14	6/1	2/1	4/0	8/13	Самостоятель- ные работы. Участие в реше- нии задач на практических занятиях. Собеседования по домашним заданиям.
2	Базовая обработка текста и дистанция редактирования	16	8/3	2/1	6/2	8/13	
3	Языковые модели	18	8/2	2/0	6/2	10/16	
4	Классификация текстов и анализ тональности	18	8/3	2/0	6/3	10/15	
5	Информационный поиск	20	10/3	2/0	8/3	10/17	
6	Машинный пере- вод	20	10/2	2/0	8/2	10/18	
7	Компьютерная семантика	20	10/0	2/0	8/0	10/20	
8	Автоматическое реферирование	18	8/0	2/0	6/0	10/18	
	В целом по дис- циплине	144	68/14	16/2	52/12	76/130	Согласно учеб- ному плану: контрольная ра- бота

5.3. Содержание семинаров, практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 6,7 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Введение в автоматическую обработку естественного языка	Структурные особенности текстов на естественном языке. Неоднозначность на всех уровнях языка. Основные задачи автоматического анализа текстов. Основные подходы к решению задач: правила, написанные вручную и машинное обучение. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Базовая обработка текста и дистанция редактирования	Предобработка текста: токенизация и сегментация. Регулярные выражения; дистанция редактирования. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Языковые модели	N-граммы. Перплексия. Методы сглаживания. Линейная интерполяция. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Классификация текстов и анализ тональности	Задачи классификации, наивный байесовский классификатор. Проблемы классификации текстов. Анализ тональности. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

Информационный поиск	Векторные модели текстов. Матричное представление. Обратный индекс. Фразовые запросы. Ранжированный информационный поиск. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Машинный перевод	Классические подходы: пословный, трансферный, интерлингвальный. Статистический машинный перевод. Модели IBM. Выравнивание текстов. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Компьютерная семантика	Значение и смысл. WordNet и аналогичные лексические базы данных. Измерение семантической близости. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений
Автоматическое реферирование	Экстрактивное и абстрактное реферирование. Реферирование нескольких документов. Реферирование, основанное на запросе. Обучение с учителем и без учителя в контексте автоматического реферирования. Рекомендуемые источники: из раздела 6: 6.1,6.2, 6.3, 6.4. из раздела 7: 7.1- 7.10.	Интерактивная форма, Практикум по решению задач по тематике занятия в малых группах (2-4 студента) и коллективное обсуждение решений

6. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины

Основная литература:

1. Советов, Б. Я. Информационные технологии: теоретические основы : учебное пособие / Б. Я. Советов, В. В. Цехановский. — 2-е изд., стер. — Санкт-Петербург : Лань, 2021. — 444 с.— URL: <https://ezpro.fa.ru:3178/book/167404>
2. Станкевич, Л. А. Интеллектуальные системы и технологии : учебник и практикум для вузов / Л. А. Станкевич. — Москва : Издательство Юрайт, 2021. — 397 с. — URL: <https://ezpro.fa.ru:3217/bcode/469517>

Дополнительная литература:

3. Болотнова, Н. С. Методики смыслового и лингвопрагматического анализа медиатекста : учебное пособие : [16+] / Н. С. Болотнова. — 2-е изд., стер. —

Москва : ФЛИНТА, 2019. – 156 с. : – URL:
<https://biblioclub.ru/index.php?page=book&id=603174>

4. Воронина, В. В. Теория и практика машинного обучения : учебное пособие / В. В. Воронина. — Ульяновск : УлГТУ, 2017. — 290 с.— URL:
<https://ezpro.fa.ru:3178/book/165053>

7. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
3. Электронно-библиотечная система «Университетская библиотека ОН-ЛАЙН» <http://biblioclub.ru/>
4. Электронно-библиотечная система Znanium <http://www.znaniy.com>
5. Электронно-библиотечная система издательства «ЮРАЙТ»
<https://urait.ru/>
6. Электронно-библиотечная система издательства Проспект
<http://ebs.prospekt.org/books>
7. Электронно-библиотечная система издательства «Лань»
<https://e.lanbook.com/>
8. Электронная библиотека Издательского дома «Гребенников»
<https://grebennikon.ru/>
9. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>

8. Методические указания для обучающихся по освоению дисциплины

Наименование методических материалов для обучающихся	Год утверждения	Местонахождение материала (ссылка на ИОП, информационный стенд кафедры/филиала, др.)
Методические указания к лекциям	2021	http://www.fa.ru/fil/ufa/about/ums/Pages/info.aspx
Методические указания к практическим занятиям	2021	http://www.fa.ru/fil/ufa/about/ums/Pages/info.aspx
Методические указания самостоятельной работе	2021	http://www.fa.ru/fil/ufa/about/ums/Pages/info.aspx
Методические указания к контрольной работе	2021	http://www.fa.ru/fil/ufa/about/ums/Pages/info.aspx

9. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем (при необходимости)

9.1. Комплект лицензионного программного обеспечения

Продукты компании Microsoft, включая ОС Windows и Office.

9.2. Современные профессиональные базы данных и информационные справочные системы

Электронное периодическое издание Справочная Правовая Система Консультант Бюджетные организации: версия Проф.

9.3. Сертифицированные программные и аппаратные средства защиты информации

Сертифицированные программные и аппаратные средства защиты информации – не используются.

10. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Учебная аудитория для проведения всех видов занятий, предусмотренных программой бакалавриата, оснащенная оборудованием и техническими средствами обучения.