

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ им. В.Г. ШУХОВА»**
(БГТУ им. В.Г. Шухова)

СОГЛАСОВАНО
Директор института
магистратуры

« 17 » мая 2021 г.



УТВЕРЖДАЮ
Директор института

« 17 » мая 2021 г.



РАБОЧАЯ ПРОГРАММА
дисциплины

Основы семантического анализа

Направление подготовки

09.04.01 Информатика и вычислительная техника

Направленность программы

Интеллектуальный анализ данных и процессов

Квалификация

Магистр

Форма обучения

очная

Институт энергетики, информационных технологий и управляющих систем

Кафедра программного обеспечения вычислительной техники и
автоматизированных систем

Белгород – 2021

Рабочая программа составлена на основании требований:

- Федерального государственного образовательного стандарта высшего образования по направлению 09.04.01 «Информатика и вычислительная техника», утвержденного приказом Министерства образования и науки Российской Федерации от 19 сентября 2017 г. N 918
- учебного плана, утвержденного ученым советом БГТУ им. В.Г. Шухова в 20__ году.

Составители: к.т.н. ЖК (П.С. Кабелянц)
(ученая степень и звание, подпись) (инициалы, фамилия)

Рабочая программа обсуждена на заседании кафедры
Программного обеспечения вычислительной техники и автоматизированных систем

« 14 » мая 2021 г., протокол № 8

Заведующий кафедрой: к.т.н, доцент (В.М.Поляков)
(ученая степень и звание, подпись) (инициалы, фамилия)

Рабочая программа согласована на заседании кафедры
Программного обеспечения вычислительной техники и автоматизированных систем

« 14 » мая 2021 г., протокол № 8

Заведующий кафедрой: к.т.н, доцент (В.М.Поляков)

Рабочая программа одобрена методической комиссией института

« 17 » мая 2021 г., протокол № 9

Председатель к.т.н., доцент (А.Н. Семернин)
(ученая степень и звание, подпись) (инициалы, фамилия)

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Категория (группа) компетенций	Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Наименование показателя оценивания
	ПК-2 Способен к разработке и сопровождению программно-информационных систем интеллектуального анализа данных и процессов	ПК-2.1. Знать принципы и методики создания интеллектуальных систем	Знания
		ПК-2.2. Уметь осуществлять организационное и технологическое обеспечение разработки программно-информационных систем интеллектуального анализа данных и процессов	Умения
		ПК-2.3. Владеть навыками разработки инструментов и методов анализа программно-информационных систем интеллектуального анализа данных и процессов	Навыки

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

1. Компетенция ПК-2

Данная компетенция формируется следующими дисциплинами.

Стадия	Наименования дисциплины
1.	Технологии искусственного интеллекта
2.	Машинное обучение
3.	Цифровая обработка сигналов
4.	Глубокое обучение
5.	Нейронные сети и системы
6.	Нечеткое моделирование
7.	Системы поддержки принятия решений
8.	Основы семантического анализа
9.	Проектное обучение
10.	Учебная технологическая (проектно-технологическая) практика
11.	Производственная преддипломная практика
12.	Государственная итоговая аттестация

3. ОБЪЕМ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 4 зач. единиц, 144 часа.

Форма промежуточной аттестации: зачет.

Вид учебной работы	Всего часов	Семестр № 2
Общая трудоемкость дисциплины, час	144	144
Контактная работа (аудиторные занятия), в т.ч.:	55	55
лекции	17	17
лабораторные	34	34

групповые консультации в период теоретического обучения и промежуточной аттестации	4	4
Самостоятельная работа студентов, включая индивидуальные и групповые консультации, в том числе:	89	89
Курсовой проект	-	-
Курсовая работа	-	-
Расчетно-графическое задание	-	-
Индивидуальное домашнее задание	-	-
Самостоятельная работа на подготовку к аудиторным занятиям (лекции, практические занятия, лабораторные занятия)	89	89
Экзамен	-	-

4. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

4.1 Наименование тем, их содержание и объем Курс 1 Семестр 2

№ п/п	Наименование раздела (краткое содержание)	Объем на тематический раздел по видам учебной нагрузки, час			
		Лекции	Практические занятия	Лабораторные занятия	Самостоятельная работа на подготовку к аудиторным занятиям
1	Введение в курс. Основные задачи обработки текстов. Основные этапы построения систем автоматической обработки текстовой информации.	2		4	6
2	Методы задания синтаксической структуры предложений. Деревья составляющих. Деревья зависимостей.	2		4	12
3	Порождающие грамматики. Распознающие грамматики. Регулярные выражения и конечные автоматы.	2		4	14
4	Предварительная обработка коллекций текстов: выделение токенов, нормализация и лемматизация, разметка частей речи.	2		4	10
5	Поверхностный (shallow) и полный (deep) синтаксический анализ, чанкинг и парсинг.	1		2	8
6	Задача сегментации текста. Использование решающих деревьев и генетических алгоритмов.	1		2	14
7	Задача тематического моделирования и классификация текстов.	1		2	7
8	Рекуррентные нейронные сети. Применение к	2		4	6

	обработке текста.				
9	Алгоритмы обучения с подкреплением и обработка текстов.	2		4	6
10	Задача автоматического аннотирования текстов.	2		4	6
	ВСЕГО	17		34	89

4.2. Содержание практических (семинарских) занятий

Учебным планом не предусмотрены

4.3. Содержание лабораторных занятий

№ п/п	Наименование раздела дисциплины	Тема лабораторного занятия	К-во часов	Самостоятельная работа на подготовку к аудиторным занятиям
семестр № 3				
1	Введение в курс. Основные задачи обработки текстов. Основные этапы построения систем автоматической обработки текстовой информации.	Основы работы с библиотекой nltk языка Python.	4	5
2	Методы задания синтаксической структуры предложений. Деревья составляющих. Деревья зависимостей.	Задача построения синтаксических деревьев заданного текста.	4	11
3	Порождающие грамматики. Распознающие грамматики. Регулярные выражения и конечные автоматы.	Задача построения регулярных выражений для заданного текста.	4	12
4	Предварительная обработка коллекций текстов: выделение токенов, нормализация и лемматизация, разметка частей речи.	Токенизация и стемминг заданного текста.	4	8
5	Поверхностный (shallow) и полный (deep) синтаксический анализ, чанкинг и парсинг.	Задача парсинга заданного текста.	2	12
6	Задача сегментации текста. Использование решающих деревьев и генетических алгоритмов.	Задача разбиения текста на сегменты.	2	10
7	Задача тематического моделирования и классификация текстов.	Задача определения явных и скрытых тем в заданном тексте.	2	6

8	Рекуррентные нейронные сети. Применение к обработке текста.	Особенности работы с искусственной нейронной сетью LSTM.	4	5
9	Алгоритмы обучения с подкреплением и обработка текстов.	Использование модели агент-среда для обучения рекуррентных нейронных сетей.	4	5
10	Задача автоматического аннотирования текстов.	Задача выделения ключевых слов в тексте и его пересказа.	4	5
ИТОГО:			34	79

4.4. Содержание курсового проекта/работы

Учебным планом не предусмотрены.

4.5. Содержание расчетно-графического задания, индивидуальных домашних заданий

Учебным планом не предусмотрены.

5. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕГО КОНТРОЛЯ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

5.1. Реализация компетенций

1. Компетенция ПК-2 Способен к разработке и сопровождению программно-информационных систем интеллектуального анализа данных и процессов

Наименование индикатора (показателя оценивания)	Используемые средства оценивания
ПК-2.1. Знать принципы и методики создания интеллектуальных систем	Защита лабораторной работы, устный опрос, экзамен
ПК-2.2. Уметь осуществлять организационное и технологическое обеспечение разработки программно-информационных систем интеллектуального анализа данных и процессов	Защита лабораторной работы
ПК-2.3. Владеть навыками разработки инструментов и методов анализа программно-информационных систем интеллектуального анализа данных и процессов	Защита лабораторной работы, устный опрос, экзамен

5.2. Типовые контрольные задания для промежуточной аттестации

5.2.1. Перечень контрольных вопросов (типовых заданий) для зачета

№ п/п	Наименование раздела дисциплины	Содержание вопросов (типовых заданий)
1	Введение в курс. Основные задачи обработки текстов.	Перечислить направления компьютерной лингвистики. Перечислить основные этапы построения автоматизированных систем обработки текстов.

	Основные этапы построения систем автоматической обработки текстовой информации.	Разъяснить принципы работы графематического и морфологического анализаторов. Установка и подключение библиотеки nltk. Основные функциональные возможности библиотеки.
2	Методы задания синтаксической структуры предложений. Деревья составляющих. Деревья зависимостей.	Перечислить методы задания синтаксической структуры предложений. Перечислить проблемы автоматизации синтаксического анализа.
3	Порождающие грамматики. Распознающие грамматики. Регулярные выражения и конечные автоматы.	Объяснить принципы использования регулярных выражений в поисковых системах.
4	Предварительная обработка коллекций текстов: выделение токенов, нормализация и лемматизация, разметка частей речи.	Отличие стемминга от лемматизации. Что такое нормальная форма слова.
5	Поверхностный (shallow) и полный (deep) синтаксический анализ, чанкинг и парсинг.	Особенности синтаксического анализа, выделение пар подлежащие-сказуемое.
6	Задача сегментации текста. Использование решающих деревьев и генетических алгоритмов.	Основные алгоритмы фрагментации текста.
7	Задача тематического моделирования и классификация текстов.	В чем отличие между классификацией и кластеризацией текстов? Перечислить методы классификации и кластеризации текстовой информации. Сформулировать основные принципы.
8	Рекуррентные нейронные сети. Применение к обработке текста.	Основные виды рекуррентных нейронных сетей. Структура слоев LSTM сети.
9	Алгоритмы обучения с подкреплением и обработка текстов.	Основные алгоритмы обучения рекуррентных сетей. Особенности алгоритмов с подкреплением.
10	Задача автоматического аннотирования текстов.	Задача тематического моделирования коллекции текстовых документов. Модель агент-среда.

**5.2.2. Перечень контрольных материалов
для защиты курсового проекта/ курсовой работы**
Учебным планом не предусмотрены.

**5.3. Типовые контрольные задания (материалы)
для текущего контроля в семестре**

Текущий контроль осуществляется в течение семестра в форме защиты лабораторных работ.

Защита лабораторной работы проводится в форме устного опроса студента и направлена на проверку степени усвоения материала и понимания теоретических сведений, используемых в процессе выполнения работы; для защиты необходимо представить в печатной (рукописной) форме отчет по лабораторной работе, выполненный самостоятельно и в соответствии со всеми требованиями, приведёнными в методических указаниях к выполнению лабораторных работ. Примерные перечень контрольных вопросов для защиты лабораторных работ приведен в таблицах.

1. Компетенция ПК-2. Способен к разработке и сопровождению программно-информационных систем интеллектуального анализа данных и процессов.

Наименование индикатора (показателя оценивания)	Типовые контрольные задания для текущего контроля
Знать принципы и методики создания интеллектуальных систем	Вопросы при защите лабораторной работы: - методы построения синтаксических деревьев; - метрики (еклидова, редакционное, Махаланобиса, Хемминга и др.); расстояние между классами (внутриклассовое и междуклассовое расстояния); - архитектуры нейронных сетей; - методы сегментации текстов;
Уметь осуществлять организационное и технологическое обеспечение разработки программно-информационных систем интеллектуального анализа данных и процессов	Вопросы при защите лабораторной работы: - представление больших коллекций документов, разбиение коллекции данных на обучающую и тестовую выборку; - использование параллельных вычислений в алгоритмах обработки текстов; - оценивание качества разработки интеллектуальной системы.
Владеть навыками разработки инструментов и методов анализа программно-информационных систем интеллектуального анализа данных и процессов	Вопросы при защите лабораторной работы: - предварительная обработка текстовых данных, выделение токенов, нормализация данных – стемминг и лемматизация; - интерфейс основных библиотек работы с текстовыми данными, особенности их настройки; – особенности программирования систем интеллектуального анализа текстовых данных.

5.4. Описание критериев оценивания компетенций и шкалы оценивания

При промежуточной аттестации в форме зачета используется следующая шкала оценивания: зачтено, не зачтено.

Критериями оценивания достижений показателей являются:

Показатель оценивания	Критерий оценивания
Знания	Знание основных терминов, понятий, структур
	Объем усвоенного материала
	Полнота ответов на поставленные вопросы
	Четкость изложения ответов на вопросы

Умения	Умение полно ответить на поставленный вопрос по лабораторной работе
	Умение дополнить и изменить свою программу в соответствии с поставленной задачей
	Умение разработать алгоритм по поставленному заданию
Навыки	Навык самостоятельной разработки приложения
	Навык работы с известными библиотеками обработки текстовых данных
	Навык правильности выбора алгоритма обработки текстовых данных

Оценка преподавателем выставляется интегрально по всем показателям и критериям оценивания.

Оценка сформированности компетенций по показателю Знания.

Критерий	Уровень освоения и оценка	
	Не зачтено	Зачтено
Знание основных терминов, понятий, структур	Не знает терминов и определений	Знает термины и определения, может корректно сформулировать их самостоятельно
Объем усвоенного материала	Не знает значительной части материала дисциплины	Обладает твердым и полным знанием материала дисциплины, владеет дополнительными знаниями
Полнота ответов на поставленные вопросы	Не дает ответы на большинство вопросов	Дает полные, развернутые ответы на поставленные вопросы
Четкость изложения ответов на вопросы	Излагает знания без логической последовательности	Излагает знания в логической последовательности, самостоятельно их интерпретируя и анализируя

Оценка сформированности компетенций по показателю Умения.

Критерий	Уровень освоения и оценка	
	Не зачтено	Зачтено
Способность полно ответить на поставленный вопрос по лабораторной работе	Не способен ответить на поставленный вопрос по лабораторной работе	Способен полно ответить на поставленный вопрос по лабораторной работе
Умение дополнить и изменить свою программу в соответствии с поставленной задачей	Не умеет дополнить и изменить свою программу в соответствии с поставленной задачей	Умеет самостоятельно дополнить и изменить свою программу в соответствии с поставленной задачей
Умение разработать алгоритм по поставленному заданию	Не умеет разработать алгоритм по поставленному заданию	Умеет самостоятельно разработать алгоритм по поставленному заданию

Оценка сформированности компетенций по показателю Навыки.

Критерий	Уровень освоения и оценка	
	Не зачтено	Зачтено
Навык самостоятельной разработки приложения	Не способен самостоятельно разработать приложение	Способен самостоятельно разработать приложение
Навык работы с известными библиотеками обработки текстовых данных	Не имеет навыка работы с известными библиотеками обработки текстовых данных	Имеет твердые навыки работы с известными библиотеками обработки текстовых данных
Навык правильности выбора алгоритма обработки текстовых данных	Не способен правильно выбрать алгоритм обработки текстовых данных	Способен правильно выбрать алгоритм обработки текстовых данных

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ

6.1. Материально-техническое обеспечение

№	Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы
1.	Учебная аудитория для проведения лекционных занятий	Специализированная мебель. Мультимедийная установка, экран, доски
2.	Учебная аудитория для проведения лабораторных занятий	Специализированная мебель. Компьютеры на базе процессоров Intel или AMD.
3.	Читальный зал библиотеки для самостоятельной работы	Специализированная мебель. Компьютерная техника, подключенная к сети интернет и имеющая доступ в электронно-образовательную среду

6.2. Лицензионное и свободно распространяемое программное обеспечение

№	Перечень лицензионного программного обеспечения.	Реквизиты подтверждающего документа
1	Python 3.7 (CPython)	Свободно распространяемое ПО
2	Anaconda (дистрибутив Python)	Свободно распространяемое ПО
3	Microsoft Office Professional Plus 2016	Соглашение Microsoft Open Value Subscription V6328633 Соглашение действительно с 02.10.2017 по 31.10.2020
4	Google Chrome	Свободно распространяемое ПО
5	Mozilla Firefox	Свободно распространяемое ПО
6	Интегрированная среда разработки Microsoft Visual Studio	Подписка Microsoft Imagine Premium id: 6f22ecb4-6882-420b-a39b-afba0ace820c.

6.3. Перечень учебных изданий и учебно-методических материалов

1. Vapnik V. N. Statistical learning theory. — N.Y.: John Wiley & Sons, Inc., 1998.
2. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие / Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. — М.: МИЭМ, 2011. — 272 с.
3. Батура Т. В. Методы определения авторского стиля текстов и их программная реализация // Программные системы и вычислительные методы. М.: НБ-Медиа, 2014. № 2. С. 197–216. DOI: 10.7256/2305-6061.2014.2.11705.
4. Воронцов К. В. Лекции по алгоритмам кластеризации и многомерного шкалирования. [Электронный ресурс] — Электрон. текстовые данные. (доступ свободный: <http://www.ccas.ru/voron/download/Clustering.pdf>)
5. Дубров А. М., Мхитарян В. С., Трошин Л.И. Многомерные статистические методы. – М.: Финансы и статистика, 2003.
6. Кобзарева Т. Ю. В поисках синтаксической структуры: автоматический анализ русского предложения с опорой на сегментацию. – М.: РГГУ, 2015. – 371 с.

6.4. Перечень интернет ресурсов, профессиональных баз данных, информационно-справочных систем

1. Информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных, профессора МФТИ, д.ф.-м.н. Воронцова К.В. - <http://www.machinelearning.ru>
2. Блог профессора МГУ, д.ф.-м.н. Дьяконова А.Г. - <https://dyakonov.org>

7. УТВЕРЖДЕНИЕ РАБОЧЕЙ ПРОГРАММЫ

Рабочая программа утверждена на 20____ /20____ учебный год без изменений

Протокол № _____ заседания кафедры от «_»_____ 20____ г.

Заведующий кафедрой _____(Поляков В.М.)
подпись, ФИО

Директор института _____(Белоусов А.В.)
подпись, ФИО