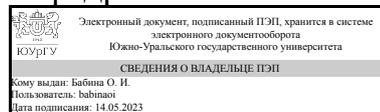


# ЮЖНО-УРАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

УТВЕРЖДАЮ:  
Заведующий выпускающей  
кафедрой



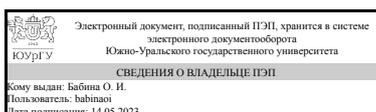
О. И. Бабина

## РАБОЧАЯ ПРОГРАММА

**дисциплины 1.Ф.М0.04.02 Извлечение информации из текста  
для направления 45.04.03 Фундаментальная и прикладная лингвистика  
уровень Магистратура  
магистерская программа Искусственный интеллект в обработке естественного  
языка  
форма обучения очная  
кафедра-разработчик Лингвистика и перевод**

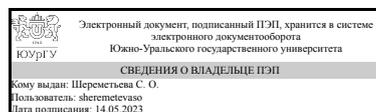
Рабочая программа составлена в соответствии с ФГОС ВО по направлению подготовки 45.04.03 Фундаментальная и прикладная лингвистика, утверждённым приказом Минобрнауки от 12.08.2020 № 993

Зав.кафедрой разработчика,  
к.филол.н., доц.



О. И. Бабина

Разработчик программы,  
д.филол.н., доц., профессор



С. О. Шереметьева

## 1. Цели и задачи дисциплины

Целью данного курса является ознакомление студентов с современными техническими средствами и информационными технологиями, служащими для задач автоматического извлечения информации из текста. Задачами курса являются приобретение студентами навыков работы с основными методами и основными существующими техническими средствами автоматического извлечения информации из текстов, а также умение изучать и осваивать новые технические средства и информационные технологии этого вида научно-технической деятельности.

### Краткое содержание дисциплины

1. Типы задач, решаемых автоматическим извлечением информации из текстов. Корпус текстов как объект извлечения информации. Типы и способы разметки корпуса текстов. Существующие ресурсы. 2. Современные подходы к извлечению лингвистической и фактографической информации из текста. Подход, основанный на словарях и правилах: этапы работы, проблемы. Подход на основе машинного обучения: этапы работы, проблемы. 3. Методы разрешения лексической неоднозначности. Метод PageRank в задаче разрешения лексической многозначности. Методы машинного обучения в задаче автоматического разрешения лексической неоднозначности. Метод Decision List. 4. Задачи извлечения именованных сущностей и извлечения отношений: подходы к подготовке обучающей коллекции. Bootstrapping. Distant supervision 5. Автоматический информационный поиск. Задача извлечения ключевых слов, многокомпонентной лексики из текстов. 6. Контент-анализ в компьютерной лингвистике. Задача анализа тональностей. Задача анализа эмоций. Задача извлечения и фактографической информации. Программа Word2vec и ее применения. 7. Задачи и методы автоматического сравнения текстов. Задача агрегации новостей. Задача проверки текстов на плагиат.

## 2. Компетенции обучающегося, формируемые в результате освоения дисциплины

Планируемые результаты освоения ОП ВО (компетенции)	Планируемые результаты обучения по дисциплине
ПК-2 Способен управлять получением, хранением и обработкой языковых данных для решения профессиональных задач	Знает: существующие электронные языковые ресурсы (корпусы текстов, словари, тезаурусы, онтологии, лексические, грамматические и иные базы данных и знаний) для решения задач извлечения информации из текста; современные методы создания и использования электронных языковых ресурсов для создания систем извлечения информации из текста Умеет: осуществлять самостоятельный поиск, изучение и критический анализ существующих ресурсов для разработки систем извлечения информации из текста; использовать современные технические средства для получения, хранения и обработки данных для задач извлечения информации из текста;

	<p>самостоятельно создавать и пополнять электронные ресурсы для извлечения информации из текста</p> <p>Имеет практический опыт: разработки электронных языковых ресурсов (корпусов аннотированных текстов, словарей, тезаурусов, онтологии, лексические, грамматические и иные базы данных и знаний) для извлечения информации из текста</p>
<p>ПК-4 Способен осуществлять организацию разработки лингвистического программного обеспечения с применением методов обработки письменного текста и звучащей речи на естественном языке, в том числе методов машинного обучения, искусственных нейронных сетей</p>	<p>Знает: типы информации, передаваемой текстом на естественном языке, и их релевантность для решения различных информационных задач; существующие лингвистические технологии, применяемые в электронных системах извлечения информации из текста, ориентированные как на представление знаний о функционировании языка на различных уровнях, так и на создание систем контент-анализа, принятия решений и т.п.</p> <p>Умеет: осуществлять самостоятельный поиск, изучение и критический анализ существующих исследований и методик создания приложений извлечения разного типа информации из текста; осваивать и применять современные технические средства и информационные технологии, служащие для автоматизации извлечения информации из текста; организовывать и проводить исследования в области извлечения разного типа информации из текста</p> <p>Имеет практический опыт: организации разработки дизайна лингвистического программного обеспечения извлечения разного типа информации из текста; создания баз данных и знаний для прототипов систем извлечения разного типа информации из текстов определенных предметных областей; создания прототипов систем извлечения разного типа информации из текстов определенных предметных областей</p>

### 3. Место дисциплины в структуре ОП ВО

Перечень предшествующих дисциплин, видов работ учебного плана	Перечень последующих дисциплин, видов работ
Искусственные нейронные сети, Искусственный интеллект и машинное обучение, Лингвистические базы знаний и языковые ресурсы	Производственная практика (преддипломная) (4 семестр)

Требования к «входным» знаниям, умениям, навыкам студента, необходимым при освоении данной дисциплины и приобретенным в результате освоения предшествующих дисциплин:

Дисциплина	Требования
------------	------------

Лингвистические базы знаний и языковые ресурсы	<p>Знает: виды лингво-информационных ресурсов и их характеристики; принципы, методы и средства проектирования и построения архитектуры лингво-информационных ресурсов; типовые решения, библиотеки программных модулей, шаблоны, классы объектов, используемые при разработке лингво-информационных ресурсов, основные методы, способы и средства управления получением, хранением и обработкой языковых данных; источники языковых данных; методы и средства извлечения и преобразования языковых данных; правовые основы сбора и использования языковых данных для решения профессиональных задач</p> <p>Умеет: использовать существующие типовые решения и шаблоны лингво-информационных ресурсов; применять принципы, методы и средства проектирования лингво-информационных ресурсов, взаимодействовать с пользователями информации и поставщиками языковых данных; пользоваться методами и средствами поиска, извлечения, хранения, передачи, преобразования и обработки языковых данных для решения профессиональных задач</p> <p>Имеет практический опыт: разработки, изменения архитектуры лингво-информационных ресурсов; проектирования структур данных, баз данных и баз знаний, поиска, извлечения, хранения, передачи, преобразования и обработки языковых данных для решения профессиональных задач</p>
Искусственные нейронные сети	<p>Знает: модели искусственных нейронных сетей, основные подходы решения профессиональных задач на базе искусственных нейронных сетей</p> <p>Умеет: осуществлять настройку параметров и обучение искусственных нейронных сетей для решения задач в профессиональной области</p> <p>Имеет практический опыт: решения профессиональных задач с использованием искусственных нейронных сетей</p>
Искусственный интеллект и машинное обучение	<p>Знает: основные понятия искусственного интеллекта, алгоритмы машинного обучения</p> <p>Умеет: выбирать комплексы методов и инструментальных средств искусственного интеллекта для решения профессиональных задач</p> <p>Имеет практический опыт: формирования обучающих наборов данных в области решения профессиональных задач для систем искусственного интеллекта</p>

#### 4. Объём и виды учебной работы

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч., 56,5 ч. контактной работы

Вид учебной работы	Всего часов	Распределение по семестрам в часах	
		Номер семестра	
		3	
Общая трудоёмкость дисциплины	108	108	
<i>Аудиторные занятия:</i>	48	48	
Лекции (Л)	16	16	
Практические занятия, семинары и (или) другие виды аудиторных занятий (ПЗ)	32	32	
Лабораторные работы (ЛР)	0	0	
<i>Самостоятельная работа (СРС)</i>	51,5	51,5	
Изучение литературы по темам курса	24,5	24,5	
Подготовка к экзамену	27	27	
Консультации и промежуточная аттестация	8,5	8,5	
Вид контроля (зачет, диф.зачет, экзамен)	-	экзамен	

## 5. Содержание дисциплины

№ раздела	Наименование разделов дисциплины	Объем аудиторных занятий по видам в часах			
		Всего	Л	ПЗ	ЛР
1	Типы задач, решаемых автоматическим извлечением информации из текстов	2	2	0	0
2	Корпус текстов как объект извлечения информации	6	2	4	0
3	Современные подходы к извлечению лингвистической и фактографической информации из текста	8	2	6	0
4	Задачи извлечения именованных сущностей и извлечения отношений	8	2	6	0
5	Автоматический информационный поиск	4	2	2	0
6	Извлечение основного содержания текста (аннотирование и реферирование)	4	2	2	0
7	Контент-анализ	10	2	8	0
8	Задачи и методы автоматического сравнения текстов	6	2	4	0

### 5.1. Лекции

№ лекции	№ раздела	Наименование или краткое содержание лекционного занятия	Кол-во часов
1	1	Типы задач, решаемых автоматическим извлечением информации из текстов	2
2	2	Корпус текстов как объект извлечения информации. Типы и способы разметки корпуса текстов. Существующие ресурсы. Методы разрешения лексической неоднозначности.	2
3	3	Современные подходы к извлечению лингвистической и фактографической информации из текста. Подход, основанный на словарях и правилах: этапы работы, проблемы. Подход на основе машинного обучения: этапы работы, проблемы	2
4	4	Задачи извлечения именованных сущностей и извлечения отношений:	2
5	5	Автоматический информационный поиск. Задача извлечения ключевых слов, многокомпонентной лексики из текстов	2

6	6	Извлечение основного содержания текста (аннотирование и реферирование)	2
7	7	Контент-анализ в компьютерной лингвистике. Задача анализа тональности. Задача анализа эмоций. Задача извлечения и фактографической информации	2
8	8	Задачи и методы автоматического сравнения текстов. Задача агрегации новостей. Задача проверки текстов на плагиат.	2

## 5.2. Практические занятия, семинары

№ занятия	№ раздела	Наименование или краткое содержание практического занятия, семинара	Кол-во часов
1-2	2	Изучение существующих корпусных ресурсов. Освоение онлайн приложений для разметки корпуса текстов. Существующие ресурсы. Изучение онлайн инструментария и Метода PageRank разрешения лексической неоднозначности. Знакомство с методами машинного обучения в задаче автоматического разрешения лексической неоднозначности. Метод Decision List.	4
3-5	3	Изучение онлайн статей по извлечению информации из текстов в рамках различных подходов. Извлечение требуемой информации по теме собственной НИР в рамках подхода, основанных на словарях и правилах	6
6	4	Оценить работу систем извлечения именованных сущностей: <a href="https://natasha.github.io/">https://natasha.github.io/</a> (основанная на правилах) <a href="http://demo.ipavlov.ai/#ru">http://demo.ipavlov.ai/#ru</a> (на машинном обучении) Указать, в чем проявляется различие между подходами, привести примеры текстов, на которых это видно.	2
7	4	Проанализировать работу системы Text Runner ( <a href="http://openie.allenai.org/">http://openie.allenai.org/</a> ): задать несколько сущностей в разных областях и посмотреть, какие выводятся отношения. Представить результат работы в виде короткого отчета с примерами.	2
8	4	Подготовка обучающего корпуса текстов предметной области собственной НИР и извлечение именованных сущностей и отношений с помощью онлайн инструментария или собственных алгоритмов	2
9	5	Освоение методов и онлайн инструментария извлечения однокомпонентных и многокомпонентных ключевых слов из документов и эмпирическая оценка из эффективности при поиске информации	2
10	6	Извлечение основного содержания текста (аннотирование и реферирование) с помощью онлайн инструментария и с помощью собственных формальных критериев. Оценка эффективности использованных инструментов и формализмов.	2
11	7	Ознакомление с существующими онлайн ресурсами ресурсами и с их использованием построение базы знаний и алгоритма извлечения тональности из оганиченного карпуса блогов по оценке работы магазинов/отелей/компаний и т.д.	2
12	7	Ознакомление с существующими онлайн ресурсами ресурсами и с их использованием построение базы знаний и алгоритма извлечения эмоций из оганиченного корпуса блогов по оценке работы магазинов/отелей/компаний и т.д.	2
13-14	7	Ознакомление с имеющимися на кафедре инструментарием и ресурсами для контент-анализа предметной области "Терроризм" и с их использованием извлечение контента, соответствующего различным запросам из корпуса текстов о террористической деятельности по существующей модели.	4
15	8	Решить задачу агрегации последних новостей по определенной тематике на основе определения близости соответствующих новостных сообщений	2
16	8	Решение задачи проверки текстов на плагиат с использованием имеющегося	2

	онлайн инструментария. Определение принципов работы использованных инструментов и их эффективности.	
--	---	--

### 5.3. Лабораторные работы

Не предусмотрены

### 5.4. Самостоятельная работа студента

Выполнение СРС			
Подвид СРС	Список литературы (с указанием разделов, глав, страниц) / ссылка на ресурс	Семестр	Кол-во часов
Изучение литературы по темам курса	<p>Шипицына, Л.Ю. Информационные технологии в лингвистике: учебное пособие. - М.: Флинта : Наука, 2013.</p> <p>Ермакова, Л.М. Методы извлечения информации из текста / Л.М. Ермакова // Вестник Пермского университета. Серия: Математика. Механика. Информатика. — 2012. — № 1. — С. 77-84. — ISSN 1993-0550. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/294160">https://e.lanbook.com/journal/issue/294160</a></p> <p>Сидорова, Е.А. ПОДХОД К МОДЕЛИРОВАНИЮ ПРОЦЕССА ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ТЕКСТА НА ОСНОВЕ ОНТОЛОГИИ / Е.А. Сидорова // Онтология проектирования. — 2018. — № 1. — С. 134-151. — ISSN 2223-9537. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/307292">https://e.lanbook.com/journal/issue/307292</a></p> <p>Шереметьева, С.О. ИНТЕРАКТИВНОЕ РЕФЕРИРОВАНИЕ, ОРИЕНТИРОВАННОЕ НА МАШИННЫЙ ПЕРЕВОД / С.О. Шереметьева // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2013. — № 1. — С. 89-92. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/296760">https://e.lanbook.com/journal/issue/296760</a></p> <p>Шереметьева, С.О. МЕТОДЫ И МОДЕЛИ АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ КЛЮЧЕВЫХ СЛОВ / С.О. Шереметьева, П.Г. Осминин // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2015. — № 1. — С. 76-81. — ISSN 1991-9751. — Текст : электронный // Лань :</p>	3	24,5

электронно-библиотечная система. — URL:  
<https://e.lanbook.com/journal/issue/295408>  
Грудева, Е. В. Корпусная лингвистика : учебное пособие / Е. В. Грудева. — 3-е изд. — Москва : ФЛИНТА, 2017. — 165 с. — ISBN 978-5-9765-1497-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL:  
<https://e.lanbook.com/book/106859>  
Самигулин, Т.Р. Анализ тональности текста методами машинного обучения / Т. Р. Самигулин, А. Э. Джурабаев // Научный результат. Информационные технологии. — 2021. — № 1. — С. 55-63. — ISSN 2518-1092. — Текст : электронный // Лань : электронно-библиотечная система. — URL:  
<https://e.lanbook.com/journal/issue/314475>  
Ермаков, С.А. Методы оценки эмоциональной окраски текста / С.А. Ермаков, Л.М. Ермакова // Вестник Пермского университета. Серия: Математика. Механика. Информатика. — 2012. — № 1. — С. 85-90. — ISSN 1993-0550. — Текст : электронный // Лань : электронно-библиотечная система. — URL:  
<https://e.lanbook.com/journal/issue/294160>  
Воронина, И.Е. АНАЛИЗ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ (НА ПРИМЕРЕ СЕТИ «ВКОНТАКТЕ») / И.Е. Воронина, В.А. Гончаров // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. — 2015. — № 4. — С. 151-158. — ISSN 1995-5499. — Текст : электронный // Лань : электронно-библиотечная система. — URL:  
<https://e.lanbook.com/journal/issue/297982>  
Кузнецов, Л.А. ТЕХНОЛОГИЯ АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ СОДЕРЖАТЕЛЬНОЙ БЛИЗОСТИ ТЕКСТОВ / Л.А. Кузнецов, В.Ф. Кузнецова // Программные продукты и системы. — 2013. — № 1. — С. 34-42. — ISSN 2311-2735. — Текст : электронный // Лань : электронно-библиотечная система. — URL:  
<https://e.lanbook.com/journal/issue/290183>  
Звонникова, Е.Д. Сравнение информационных систем автоматической проверки оригинальности текстов / Е.Д. Звонникова // Концепт. — 2015. — № 8.

	— С. 1-6. — ISSN 2304-120X. — Текст : электронный // Лань : электронно- библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/297264">https://e.lanbook.com/journal/issue/297264</a>		
Подготовка к экзамену	<p>Шипицына, Л.Ю. Информационные технологии в лингвистике: учебное пособие. - М.: Флинта : Наука, 2013.</p> <p>Ермакова, Л.М. Методы извлечения информации из текста / Л.М. Ермакова // Вестник Пермского университета. Серия: Математика. Механика. Информатика. — 2012. — № 1. — С. 77-84. — ISSN 1993-0550. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/294160">https://e.lanbook.com/journal/issue/294160</a></p> <p>Сидорова, Е.А. ПОДХОД К МОДЕЛИРОВАНИЮ ПРОЦЕССА ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ТЕКСТА НА ОСНОВЕ ОНТОЛОГИИ / Е.А. Сидорова // Онтология проектирования. — 2018. — № 1. — С. 134-151. — ISSN 2223-9537. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/307292">https://e.lanbook.com/journal/issue/307292</a></p> <p>Шереметьева, С.О. ИНТЕРАКТИВНОЕ РЕФЕРИРОВАНИЕ, ОРИЕНТИРОВАННОЕ НА МАШИННЫЙ ПЕРЕВОД / С.О. Шереметьева // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2013. — № 1. — С. 89-92. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/296760">https://e.lanbook.com/journal/issue/296760</a></p> <p>Шереметьева, С.О. МЕТОДЫ И МОДЕЛИ АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ КЛЮЧЕВЫХ СЛОВ / С.О. Шереметьева, П.Г. Осминин // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2015. — № 1. — С. 76-81. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/295408">https://e.lanbook.com/journal/issue/295408</a></p> <p>Грудева, Е. В. Корпусная лингвистика : учебное пособие / Е. В. Грудева. — 3-е изд. — Москва : ФЛИНТА, 2017. — 165 с. — ISBN 978-5-9765-1497-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/106859">https://e.lanbook.com/book/106859</a></p>	3	27

	<p>Самигулин, Т.Р. Анализ тональности текста методами машинного обучения / Т. Р. Самигулин, А. Э. Джурабаев // Научный результат. Информационные технологии. — 2021. — № 1. — С. 55-63. — ISSN 2518-1092. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/314475">https://e.lanbook.com/journal/issue/314475</a></p> <p>Ермаков, С.А. Методы оценки эмоциональной окраски текста / С.А. Ермаков, Л.М. Ермакова // Вестник Пермского университета. Серия: Математика. Механика. Информатика. — 2012. — № 1. — С. 85-90. — ISSN 1993-0550. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/294160">https://e.lanbook.com/journal/issue/294160</a></p> <p>Воронина, И.Е. АНАЛИЗ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ (НА ПРИМЕРЕ СЕТИ «ВКОНТАКТЕ») / И.Е. Воронина, В.А. Гончаров // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. — 2015. — № 4. — С. 151-158. — ISSN 1995-5499. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/297982">https://e.lanbook.com/journal/issue/297982</a></p> <p>Кузнецов, Л.А. ТЕХНОЛОГИЯ АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ СОДЕРЖАТЕЛЬНОЙ БЛИЗОСТИ ТЕКСТОВ / Л.А. Кузнецов, В.Ф. Кузнецова // Программные продукты и системы. — 2013. — № 1. — С. 34-42. — ISSN 2311-2735. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/290183">https://e.lanbook.com/journal/issue/290183</a></p> <p>Звонникова, Е.Д. Сравнение информационных систем автоматической проверки оригинальности текстов / Е.Д. Звонникова // Концепт. — 2015. — № 8. — С. 1-6. — ISSN 2304-120X. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/297264">https://e.lanbook.com/journal/issue/297264</a></p>		
--	--	--	--

**6. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации**

Контроль качества освоения образовательной программы осуществляется в соответствии с Положением о балльно-рейтинговой системе оценивания результатов учебной деятельности обучающихся.

### 6.1. Контрольные мероприятия (КМ)

№ КМ	Се-местр	Вид контроля	Название контрольного мероприятия	Вес	Макс. балл	Порядок начисления баллов	Учитывается в ПА
1	3	Текущий контроль	Контрольная работа 1	1	18	Контрольная работа содержит шесть заданий. Каждое задание оценивается максимум в три балла. Три балла выставляется, если задание выполнено в полном объеме и корректно. Два балла выставляется, если задание выполнено в полном объеме и с незначительными ошибками. 1 балл выставляется, если задание выполнено с многочисленными ошибками. 0 баллов выставляется, если задание не выполнено .	экзамен
2	3	Текущий контроль	Контрольная работа 2	1	18	Контрольная работа содержит шесть заданий. Каждое задание оценивается максимум в три балла. Три балла выставляется, если задание выполнено в полном объеме и корректно. Два балла выставляется, если задание выполнено в полном объеме и с незначительными ошибками. 1 балл выставляется, если задание выполнено с многочисленными ошибками. 0 баллов выставляется, если задание не выполнено .	экзамен
3	3	Текущий контроль	Контрольная работа 3	1	18	Контрольная работа содержит шесть заданий. Каждое задание оценивается максимум в три балла. Три балла выставляется, если задание выполнено в полном объеме и корректно. Два балла выставляется, если задание выполнено в полном объеме и с незначительными ошибками. 1 балл выставляется, если задание выполнено с многочисленными ошибками. 0 баллов выставляется, если задание не выполнено .	экзамен
4	3	Промежуточная аттестация	экзамен	-	20	Экзаменационная работа для каждого студента содержит четыре вопроса из общего списка. Ответ на каждый вопрос оценивается максимум в пять баллов. Пять баллов выставляется, если ответ дан в полном объеме и корректно. Четыре балла выставляется, если ответ дан в полном объеме и с незначительными ошибками. Три балла выставляется, если ответ дан с	экзамен

						немногочисленными, но серьезными ошибками. Два балла выставляется, если ответ дан неполностью с многочисленными ошибками. 0 баллов выставляется, если ответ отсутствует .	
--	--	--	--	--	--	---	--

## 6.2. Процедура проведения, критерии оценивания

Вид промежуточной аттестации	Процедура проведения	Критерии оценивания
экзамен	Экзамен проводится в письменной форме. Студенты получают 4 вопроса из общего списка вопросов по всему курсу и отвечают на их в аудитории. По выполнении сдают преподавателю, который проверяет выполнение в присутствии студента и выставляет оценку.	В соответствии с пп. 2.5, 2.6 Положения

## 6.3. Паспорт фонда оценочных средств

Компетенции	Результаты обучения	№ КМ			
		1	2	3	4
ПК-2	Знает: существующие электронные языковые ресурсы (корпусы текстов, словари, тезаурусы, онтологии, лексические, грамматические и иные базы данных и знаний) для решения задач извлечения информации из текста; современные методы создания и использования электронных языковых ресурсов для создания систем извлечения информации из текста	+	+		+
ПК-2	Умеет: осуществлять самостоятельный поиск, изучение и критический анализ существующих ресурсов для разработки систем извлечения информации из текста; использовать современные технические средства для получения, хранения и обработки данных для задач извлечения информации из текста; самостоятельно создавать и пополнять электронные ресурсы для извлечения информации из текста	+	+		+
ПК-2	Имеет практический опыт: разработки электронных языковых ресурсов (корпусов аннотированных текстов, словарей, тезаурусов, онтологии, лексические, грамматические и иные базы данных и знаний) для извлечения информации из текста	+	+		+
ПК-4	Знает: типы информации, передаваемой текстом на естественном языке, и их релевантность для решения различных информационных задач; существующие лингвистические технологии, применяемые в электронных системах извлечения информации из текста, ориентированные как на представление знаний о функционировании языка на различных уровнях, так и на создание систем контент-анализа, принятия решений и т.п.				+
ПК-4	Умеет: осуществлять самостоятельный поиск, изучение и критический анализ существующих исследований и методик создания приложений извлечения разного типа информации из текста; осваивать и применять современные технические средства и информационные технологии, служащие для автоматизации извлечения информации из текста; организовывать и проводить исследования в области извлечения разного типа информации из текста				+
ПК-4	Имеет практический опыт: организации разработки дизайна лингвистического программного обеспечения извлечения разного типа информации из текста; создания баз данных и знаний для прототипов систем извлечения разного типа информации из текстов определенных предметных областей; создания прототипов систем извлечения разного типа информации из текстов				+

Типовые контрольные задания по каждому мероприятию находятся в приложениях.

## 7. Учебно-методическое и информационное обеспечение дисциплины

### Печатная учебно-методическая документация

#### а) основная литература:

1. Леонтьева, Н. Н. Автоматическое понимание текстов : системы, модели, ресурсы [Текст] учеб. пособие для лингвист. фак. вузов Н. Н. Леонтьева. - М.: Академия, 2006. - 303 с.
2. Англо-русский синонимический словарь [Текст] Ю. Д. Апресян, В. В. Ботякова, Т. Э. Латышева и др.; под рук.: А. И. Розенмана, Ю. Д. Апресяна. - 5-е изд., стер. - М.: Русский язык, 2000. - 542, [1] с.

#### б) дополнительная литература:

1. Шаров, Ю. В. Введение в базы данных: Знакомство с компьютером. Обработка текстов. Электронные таблицы. Банки данных. - М.: АБФ, 1995. - 383 с.
2. Апресян, Ю. Д. Избранные труды [Текст] Т. 1 Лексическая семантика : Синонимические средства языка Ю. Д. Апресян. - 2-е изд., испр. и доп. - М.: Школа "Языки русской культуры" : Восточная литература, 1995. - 472 с. ил.

#### в) отечественные и зарубежные журналы по дисциплине, имеющиеся в библиотеке:

1. Научно-техническая информация. Серия 1, Организация и методика информационной работы науч.-техн. сб. М-во науки и техн. политики Рос. Федерации, Рос. акад. наук, Всерос. ин-т науч. и техн. информ. (ВИНИТИ) сборник. - М.: ВИНТИ, 1961-
2. Научно-техническая информация. Серия 2, Информационные процессы и системы науч.-техн. сб. Рос. акад. наук, М-во науки и техн. политики РФ, Всерос. ин-т науч. и техн. информ. (ВИНИТИ) сборник. - М., 1961-
3. Южно-Уральский государственный университет (ЮУрГУ) Челябинск Вестник Южно-Уральского государственного университета Юж.-Урал. гос. ун-т; ЮУрГУ журнал. - Челябинск: Издательство ЮУрГУ, 2001-

#### г) методические указания для студентов по освоению дисциплины:

1. методические рекомендации для организации самостоятельной работы

#### из них: учебно-методическое обеспечение самостоятельной работы студента:

1. методические рекомендации для организации самостоятельной работы

### Электронная учебно-методическая документация

№	Вид литературы	Наименование ресурса в	Библиографическое описание
---	----------------	------------------------	----------------------------

		электронной форме	
1	Основная литература	Электронно-библиотечная система издательства Лань	Грудева, Е. В. Корпусная лингвистика : учебное пособие / Е. В. Грудева. — 3-е изд. — Москва : ФЛИНТА, 2017. — 165 с. — ISBN 978-5-9765-1497-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/106859">https://e.lanbook.com/book/106859</a> (дата обращения: 24.01.2022). — Режим доступа: для авториз. пользователей.
2	Основная литература	Электронно-библиотечная система издательства Лань	Шереметьева, С.О. МЕТОДЫ И МОДЕЛИ АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ КЛЮЧЕВЫХ СЛОВ / С.О. Шереметьева, П.Г. Осминин // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2015. — № 1. — С. 76-81. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/295408">https://e.lanbook.com/journal/issue/295408</a> (дата обращения: 26.01.2022). — Режим доступа: для авториз. пользователей.
3	Основная литература	Электронно-библиотечная система издательства Лань	Шереметьева, С.О. ИНТЕРАКТИВНОЕ РЕФЕРИРОВАНИЕ, ОРИЕНТИРОВАННОЕ НА МАШИННЫЙ ПЕРЕВОД / С.О. Шереметьева // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2013. — № 1. — С. 89-92. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/296760">https://e.lanbook.com/journal/issue/296760</a> (дата обращения: 26.01.2022). — Режим доступа: для авториз. пользователей.
4	Дополнительная литература	Электронно-библиотечная система издательства Лань	Шереметьева, С.О. К ВОПРОСУ О РАЗРАБОТКЕ ОНТОЛОГИЧЕСКИХ РЕСУРСОВ ПРЕДМЕТНОЙ ОБЛАСТИ «ТЕРРОРИЗМ» / С.О. Шереметьева, А.Ю. Зиновьева // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2017. — № 4. — С. 48-54. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/308128">https://e.lanbook.com/journal/issue/308128</a> (дата обращения: 26.01.2022). — Режим доступа: для авториз. пользователей.
5	Дополнительная литература	Электронно-библиотечная система издательства Лань	Шереметьева, С.О. К ВОПРОСУ ОБ ЭЛЕКТРОННЫХ РЕСУРСАХ ПРОФЕССИОНАЛЬНОЙ ЛЕКСИКИ / С.О. Шереметьева, П.Г. Осминин, Е.С. Щербаков // Вестник Южно-Уральского государственного университета. Серия: Лингвистика. — 2014. — № 1. — С. 57-63. — ISSN 1991-9751. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/296759">https://e.lanbook.com/journal/issue/296759</a> (дата обращения: 26.01.2022). — Режим доступа: для авториз. пользователей.
6	Основная литература	Электронно-библиотечная система издательства Лань	Шипицина, Л. Ю. Информационные технологии в лингвистике : учебное пособие / Л. Ю. Шипицина. — 3-е изд., стер. — Москва : ФЛИНТА, 2017. — 126 с. — ISBN 978-5-9765-1431-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/119463">https://e.lanbook.com/book/119463</a> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.
7	Основная литература	Электронно-библиотечная система издательства Лань	Кулешов, С.В. Ассоциативно-онтологический подход к обработке текстов на естественном языке / С.В. Кулешов, А.А. Зайцева, В.С. Марков // Интеллектуальные технологии на транспорте. — 2015. — № 4. — С. 40-45. — ISSN 2413-2527. — Текст : электронный // Лань : электронно-

			библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/302363">https://e.lanbook.com/journal/issue/302363</a> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.
8	Дополнительная литература	Электронно-библиотечная система издательства Лань	Ингерсолл, Г. С. Обработка неструктурированных текстов. Поиск, организация и манипулирование / Г. С. Ингерсолл, Т. С. Мортон, Э. Л. Фэррис. — Москва : ДМК Пресс, 2015. — 414 с. — ISBN 978-5-97060-144-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/book/73069">https://e.lanbook.com/book/73069</a> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.
9	Основная литература	Электронно-библиотечная система издательства Лань	Ермакова, Л.М. Методы извлечения информации из текста / Л.М. Ермакова // Вестник Пермского университета. Серия: Математика. Механика. Информатика. — 2012. — № 1. — С. 77-84. — ISSN 1993-0550. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/294160">https://e.lanbook.com/journal/issue/294160</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
10	Основная литература	Электронно-библиотечная система издательства Лань	Сидорова, Е.А. ПОДХОД К МОДЕЛИРОВАНИЮ ПРОЦЕССА ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ ИЗ ТЕКСТА НА ОСНОВЕ ОНТОЛОГИИ / Е.А. Сидорова // Онтология проектирования. — 2018. — № 1. — С. 134-151. — ISSN 2223-9537. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/307292">https://e.lanbook.com/journal/issue/307292</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
11	Дополнительная литература	Электронно-библиотечная система издательства Лань	Ботов, Д.С. Извлечение информации с использованием нейросетевых моделей языка на примере анализа вакансий в системах онлайн-рекрутмента / Д.С. Ботов, Ю.Д. Кленин, И.Е. Николаев // Вестник Югорского государственного университета. — 2018. — № 3. — С. 37-48. — ISSN 1816-9228. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/309649">https://e.lanbook.com/journal/issue/309649</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
12	Дополнительная литература	Электронно-библиотечная система издательства Лань	Осочкин, А.А. Метод частотно-морфологической классификации текстов / А.А. Осочкин, В.В. Фомин, А.В. Флегонтов // Программные продукты и системы. — 2017. — № 3. — С. 478-486. — ISSN 2311-2735. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/301897">https://e.lanbook.com/journal/issue/301897</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
13	Дополнительная литература	Электронно-библиотечная система издательства Лань	Ульянкин, Ф.В. ПОСТРОЕНИЕ ИНДЕКСА ПРЕДПРИНИМАТЕЛЬСКОЙ УВЕРЕННОСТИ В РОССИИ НА ОСНОВЕ АНАЛИЗА ТОНАЛЬНОСТИ НОВОСТНЫХ ТЕКСТОВ В ИНТЕРНЕТЕ / Ф. В. Ульянкин, А. В. Полбин // Экономическое развитие России. — 2020. — № 6. — С. 22-27. — ISSN 2306-5001. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/313305">https://e.lanbook.com/journal/issue/313305</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
14	Основная литература	Электронно-библиотечная система издательства	Самигулин, Т.Р. Анализ тональности текста методами машинного обучения / Т. Р. Самигулин, А. Э. Джурабаев // Научный результат. Информационные технологии. — 2021. — № 1. — С. 55-63. — ISSN 2518-1092. — Текст :

		Лань	электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/314475">https://e.lanbook.com/journal/issue/314475</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
15	Основная литература	Электронно-библиотечная система издательства Лань	Ермаков, С.А. Методы оценки эмоциональной окраски текста / С.А. Ермаков, Л.М. Ермакова // Вестник Пермского университета. Серия: Математика. Механика. Информатика. — 2012. — № 1. — С. 85-90. — ISSN 1993-0550. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/294160">https://e.lanbook.com/journal/issue/294160</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
16	Дополнительная литература	Электронно-библиотечная система издательства Лань	Семенкин, Е.С. СИСТЕМА АВТОМАТИЧЕСКОГО ИЗВЛЕЧЕНИЯ ИНФОРМАТИВНЫХ ПРИЗНАКОВ ДЛЯ РАСПОЗНАВАНИЯ ЭМОЦИЙ ЧЕЛОВЕКА В РЕЧЕВОЙ КОММУНИКАЦИИ / Е.С. Семенкин, К.Ю. Брестер, М.Ю. Сидоров // Программные продукты и системы. — 2014. — № 4. — С. 127-131. — ISSN 2311-2735. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/298069">https://e.lanbook.com/journal/issue/298069</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
17	Основная литература	Электронно-библиотечная система издательства Лань	Воронина, И.Е. АНАЛИЗ ЭМОЦИОНАЛЬНОЙ ОКРАСКИ СООБЩЕНИЙ В СОЦИАЛЬНЫХ СЕТЯХ (НА ПРИМЕРЕ СЕТИ «ВКОНТАКТЕ») / И.Е. Воронина, В.А. Гончаров // Вестник Воронежского государственного университета. Серия: Системный анализ и информационные технологии. — 2015. — № 4. — С. 151-158. — ISSN 1995-5499. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/297982">https://e.lanbook.com/journal/issue/297982</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
18	Основная литература	Электронно-библиотечная система издательства Лань	Кузнецов, Л.А. ТЕХНОЛОГИЯ АВТОМАТИЗИРОВАННОЙ ОЦЕНКИ СОДЕРЖАТЕЛЬНОЙ БЛИЗОСТИ ТЕКСТОВ / Л.А. Кузнецов, В.Ф. Кузнецова // Программные продукты и системы. — 2013. — № 1. — С. 34-42. — ISSN 2311-2735. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/290183">https://e.lanbook.com/journal/issue/290183</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
19	Основная литература	Электронно-библиотечная система издательства Лань	Гельман, В.Я. ПРОБЛЕМЫ ФОРМАЛЬНО-МЕХАНИСТИЧЕСКОГО ПОДХОДА К ВЫЯВЛЕНИЮ ПЛАГИАТА В НАУЧНЫХ РАБОТАХ / В. Я. Гельман // Экономика науки. — 2020. — № 3. — С. 180-185. — ISSN 2410-132X. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/313406">https://e.lanbook.com/journal/issue/313406</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.
20	Основная литература	Электронно-библиотечная система издательства Лань	Звонникова, Е.Д. Сравнение информационных систем автоматической проверки оригинальности текстов / Е.Д. Звонникова // Концепт. — 2015. — № 8. — С. 1-6. — ISSN 2304-120X. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <a href="https://e.lanbook.com/journal/issue/297264">https://e.lanbook.com/journal/issue/297264</a> (дата обращения: 30.01.2022). — Режим доступа: для авториз. пользователей.

Перечень используемого программного обеспечения:

1. Microsoft-Windows(бессрочно)
2. Microsoft-Office(бессрочно)
3. -JUST AI Conversational Platform Ultimate (Developer)(бессрочно)

Перечень используемых профессиональных баз данных и информационных справочных систем:

Нет

## 8. Материально-техническое обеспечение дисциплины

Вид занятий	№ ауд.	Основное оборудование, стенды, макеты, компьютерная техника, предустановленное программное обеспечение, используемое для различных видов занятий
Лекции	478 (1)	Главный учебный корпус «Компьютерный класс», ауд. 478 Оборудование и технические средства обучения: 1. Персональный компьютер студента – 16 шт. 2. Персональный компьютер преподавателя – 1 шт. 3. Монитор – 17 шт. 4. Мышь – 17 шт. 5. Клавиатура – 17 шт. 6. Микротелефонная гарнитура – 17 шт. 7. Интерактивная панель планшет – 1 шт. 8. Мультимедиапроектор – 1 шт. 9. Настенно-потолочный экран с электроприводом – 1 шт. 10. Активная акустическая система (аудиоколонки) – 1 шт. 11. Коммутатор – 1 шт. 12. Сетевой фильтр – 1 шт. Имущество: 1. Стул «Стандарт» – 44 шт. 2. Шкаф открытый – 1 шт. 3. Антресоль с замком – 1 шт. 4. Аудиторная доска – 1 шт. 5. Жалюзи вертикальные блэкаут – 3 шт. 6. Кронштейн (крепление для аудиоколонок) – 2 шт. 7. Кронштейн (крепление для проектора) – 1 шт. 8. Рабочий стол студента – 16 шт. 9. Рабочий стол преподавателя – 1 шт. 10. Специализированный стол для практических занятий – 1 шт.
Практические занятия и семинары	478 (1)	Главный учебный корпус «Компьютерный класс», ауд. 478 Оборудование и технические средства обучения: 1. Персональный компьютер студента – 16 шт. 2. Персональный компьютер преподавателя – 1 шт. 3. Монитор – 17 шт. 4. Мышь – 17 шт. 5. Клавиатура – 17 шт. 6. Микротелефонная гарнитура – 17 шт. 7. Интерактивная панель планшет – 1 шт. 8. Мультимедиапроектор – 1 шт. 9. Настенно-потолочный экран с электроприводом – 1 шт. 10. Активная акустическая система (аудиоколонки) – 1 шт. 11. Коммутатор – 1 шт. 12. Сетевой фильтр – 1 шт. Имущество: 1. Стул «Стандарт» – 44 шт. 2. Шкаф открытый – 1 шт. 3. Антресоль с замком – 1 шт. 4. Аудиторная доска – 1 шт. 5. Жалюзи вертикальные блэкаут – 3 шт. 6. Кронштейн (крепление для аудиоколонок) – 2 шт. 7. Кронштейн (крепление для проектора) – 1 шт. 8. Рабочий стол студента – 16 шт. 9. Рабочий стол преподавателя – 1 шт. 10. Специализированный стол для практических занятий – 1 шт.
Самостоятельная работа студента	478 (1)	Главный учебный корпус «Компьютерный класс», ауд. 478 Оборудование и технические средства обучения: 1. Комплект компьютерного оборудования (монитор, системный блок, клавиатура, мышь, микротелефонная гарнитура) с возможностью подключения к сети «Интернет» и доступом к ЭИОС Университета - 17 шт. 2. Интерактивная панель планшет – 1 шт. 3. Мультимедиапроектор – 1 шт. 4. Настенно-потолочный экран с электроприводом – 1 шт. 5. Активная акустическая система (аудиоколонки) – 1 шт. 6. Коммутатор – 1 шт. 7. Сетевой фильтр – 1 шт. Имущество: 1. Стул «Стандарт» – 44 шт. 2. Шкаф открытый – 1 шт. 3. Антресоль с замком – 1 шт. 4. Аудиторная доска – 1 шт. 5. Жалюзи

		вертикальные блэкаут – 3 шт. 6. Кронштейн (крепление для аудиоколонок) – 2 шт. 7. Кронштейн (крепление для проектора) – 1 шт. 8. Рабочий стол студента – 16 шт. 9. Рабочий стол преподавателя – 1 шт. 10. Специализированный стол для практических занятий – 1 шт.
Экзамен	478 (1)	<p>Главный учебный корпус «Компьютерный класс», ауд. 478 Оборудование и технические средства обучения: 1. Персональный компьютер студента – 16 шт. 2. Персональный компьютер преподавателя – 1 шт. 3. Монитор – 17 шт. 4. Мышь – 17 шт. 5. Клавиатура – 17 шт. 6. Микротелефонная гарнитура – 17 шт. 7. Интерактивная панель планшет – 1 шт. 8. Мультимедиапроектор – 1 шт. 9. Настенно-потолочный экран с электроприводом – 1 шт. 10. Активная акустическая система (аудиоколонки) – 1 шт. 11. Коммутатор – 1 шт. 12. Сетевой фильтр – 1 шт.</p> <p>Имущество: 1. Стул «Стандарт» – 44 шт. 2. Шкаф открытый – 1 шт. 3. Антресоль с замком – 1 шт. 4. Аудиторная доска – 1 шт. 5. Жалюзи вертикальные блэкаут – 3 шт. 6. Кронштейн (крепление для аудиоколонок) – 2 шт. 7. Кронштейн (крепление для проектора) – 1 шт. 8. Рабочий стол студента – 16 шт. 9. Рабочий стол преподавателя – 1 шт. 10. Специализированный стол для практических занятий – 1 шт.</p>