

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Российский государственный гуманитарный университет»  
(ФГБОУ ВО «РГГУ»)

Институт лингвистики  
Кафедра теоретической и прикладной лингвистики

**Инструменты искусственного интеллекта для анализа языка и текста**  
Рабочая программа дисциплины

Направление подготовки:  
**45.03.02 «ЛИНГВИСТИКА»**

**Профиль: Лингвистика**

---

Уровень высшего образования:  
**БАКАЛАВРИАТ**

**Форма обучения: очная**

---

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

**Москва 2022**

Инструменты искусственного интеллекта для анализа языка и текста  
Рабочая программа дисциплины

Составители:

Составители: д.ф.н., доцент, заведующий кафедрой теоретической и прикладной лингвистики Воронежского государственного университета Шилихина Ксения Михайловна, к.ф.н., доцент кафедры теоретической и прикладной лингвистики Воронежского государственного университета Донина Ольга Валерьевна

УТВЕРЖДЕНО

Протокол заседания кафедры ТиПЛ

№ 3 от 18.03.2022

## ОГЛАВЛЕНИЕ

1. Пояснительная записка .....	4
1.1. Цель и задачи дисциплины .....	4
1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций .....	4
1.3. Место дисциплины в структуре образовательной программы	<b>Error! Bookmark not defined.</b>
2. Структура дисциплины .....	5
3. Содержание дисциплины .....	<b>Error! Bookmark not defined.</b>
4. Образовательные технологии .....	<b>Error! Bookmark not defined.</b>
5. Оценка планируемых результатов обучения .....	7
5.1 Система оценивания .....	<b>Error! Bookmark not defined.</b>
5.2 Критерии выставления оценки по дисциплине .....	<b>Error! Bookmark not defined.</b>
5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине .....	<b>Error! Bookmark not defined.</b>
6. Учебно-методическое и информационное обеспечение дисциплины .....	10
6.1 Список источников и литературы .....	<b>Error! Bookmark not defined.</b>
6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».	<b>Error! Bookmark not defined.</b>
6.3 Профессиональные базы данных и информационно-справочные системы .....	12
7. Материально-техническое обеспечение дисциплины .....	<b>Error! Bookmark not defined.</b>
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов .....	<b>Error! Bookmark not defined.</b>
9. Методические материалы .....	<b>Error! Bookmark not defined.</b>
9.1 Планы семинарских/ практических/ лабораторных занятий	<b>Error! Bookmark not defined.</b>
9.2 Методические рекомендации по подготовке письменных работ	<b>Error! Bookmark not defined.</b>
9.3 Иные материалы .....	<b>Error! Bookmark not defined.</b>
Приложение 1. Аннотация рабочей программы дисциплины .....	22

## 1. Пояснительная записка

### 1.1. Цель и задачи дисциплины

**Целью** данного курса является знакомство с технологиями автоматической обработки естественного языка для обучения применению таких технологий в различных сферах филологической и лингвистической деятельности. В рамках курса предусматривается ознакомление учащихся с основными областями и задачами применения технологий искусственного интеллекта, а также с современными программными средствами для решения базовых филологических и лингвистических прикладных задач.

**Задачами** курса является приобретение студентами навыков работы с приложениями для обработки естественного языка, системами автоматического перевода, анализа тональности, а также получение представлений о возможном применении данных средств и ресурсов при проведении исследований.

### 1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ОПК-6 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности	ОПК-6.1 Умеет отбирать наиболее подходящие для решения задачи информационные технологии и продукты	<b>знать</b> существующие информационные технологии и продукты, <b>уметь</b> выбрать наиболее адекватные текущей задаче
	ОПК-6.2 Умеет применять информационные технологии и продукты в зависимости от текущей задачи	<b>уметь</b> применять оптимальные для текущей рабочей задачи информационные технологии и продукты, <b>владеть</b> их пользовательским интерфейсом
ПК-3 – Способен к научно-исследовательской деятельности	<i>ПК-3.1</i> – Владеет основами методов научного исследования, информационной и библиографической культурой  <i>ПК-3.2</i> – Владеет стандартными методиками поиска, анализа и обработки материала исследования	По окончании курса студент должен <b>знать</b> различные типы электронных лингвистических ресурсов, <b>уметь</b> работать с технологиями искусственного интеллекта в области обработки естественного языка и грамотно применять их в соответствии с типом задачи, <b>владеть</b> навыками работы с электронными словарями, лингвистическими базами данных, приложениями для автоматического анализа морфологии и синтаксиса, системами автоматизированного перевода.

### 1.3 Место дисциплины (модуля) в структуре образовательной программы, связь с другими дисциплинами (модулями) программы

Курс относится к части дисциплин, формируемой участниками образовательного процесса ОПОП ВО бакалавриата «Лингвистика» по направлению подготовки 45.03.02 – Лингвистика.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения курсов «Прикладная лингвистика», «Технологии искусственного интеллекта в гуманитарных сферах», учебной и производственной практик.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для подготовки к преддипломной практике, к государственной итоговой аттестации.

## 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 2 з.е., 76\_ академических часов, в том числе контактная работа обучающихся с преподавателем 28 ч., самостоятельная работа обучающихся и промежуточная аттестация 48 ч.

Семестр	Тип учебных занятий	Количество часов
7	Лекции	10
7	Практические/лабораторные занятия	18
	Всего:	28

Объем дисциплины в форме самостоятельной работы обучающихся составляет 48 академических часов

## 3. Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
	Тема 1.	Искусственный интеллект в гуманитарной сфере
	Тема 2	Лингвистические и филологические ресурсы и программы. Big Data
	Тема 3	Компьютерная лексикография
	Тема 4	Корпусная лингвистика
	Тема 5	ИИ в переводческой деятельности. Автоматизированный перевод. Машинный перевод
	Тема 6	Визуализация текстовых данных. Стилеметрия
	Тема 7	Системы автоматической обработки письменного текста и звучащей речи
	Тема 8	Аспекты интеллектуального анализа текста. Text Mining

## 4. Образовательные технологии

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1.	Темы 1-2	Лекция 1. Практическое/	Лекция с использованием видеоматериалов Обсуждение, дискуссия

		<p>лабораторное занятие 1.</p> <p>Самостоятельная работа</p>	<p>Консультирование и проверка домашних заданий посредством электронной почты</p>
	Тема 3	<p>Лекция 2.</p> <p>Практическое/ лабораторное занятие 2.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов</p> <p>Практическое задание, обсуждение, дискуссия</p> <p>Консультирование и проверка домашних заданий посредством электронной почты</p>
	Тема 4	<p>Лекция 3.</p> <p>Практическое/ лабораторное занятие 3.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов</p> <p>Практическое задание, обсуждение, дискуссия</p> <p>Консультирование и проверка домашних заданий посредством электронной почты</p>
	Тема 5	<p>Лекция 4.</p> <p>Практическое/ лабораторное занятие 5.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов</p> <p>Практическое задание, обсуждение, дискуссия</p> <p>Консультирование и проверка домашних заданий посредством электронной почты</p>
	Тема 6	<p>Лекция 5.</p> <p>Практическое занятие 6.</p> <p>Самостоятельная работа</p>	<p>Лекция с использованием видеоматериалов</p> <p>Практическое задание, обсуждение, дискуссия</p> <p>Консультирование и проверка домашних заданий посредством электронной почты</p>
	Тема 7	<p>Практическое/ лабораторное занятие 7.</p> <p>Самостоятельная работа</p>	<p>Практические задания, обсуждение, дискуссия</p> <p>Консультирование и проверка домашних заданий посредством электронной почты</p>
	Тема 8	<p>Практическое/ лабораторное занятие 8.</p> <p>Самостоятельная работа</p>	<p>Практические задания, обсуждение, дискуссия</p> <p>Консультирование и проверка домашних заданий посредством электронной почты</p>
		<p>Практическое занятие 9</p>	<p>Отчетная презентация</p>

В преподавании дисциплины «Инструменты искусственного интеллекта для анализа языка и текста» используются разнообразные образовательные технологии как традиционного, так и инновационного характера, учитывающие смешанный, теоретико- и практикоориентированный характер дисциплины:

- лекции;
- практические занятия;
- дискуссии;
- выступления с докладами и сообщениями;
- аудиторные практические работы;
- внеаудиторные практические работы;
- презентация проекта;
- тестирование.

В период временного приостановления посещения обучающимися помещений и территории РГГУ для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

## 5. Оценка планируемых результатов обучения

### 5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- опрос	5 баллов	30 баллов
- участие в дискуссии на семинаре	10 баллов	30 баллов
Промежуточная аттестация (защита проекта)		40 баллов
<b>Итого за семестр зачёт</b>		<b>100 баллов</b>

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала	Шкала ECTS
95 – 100	отлично	A
83 – 94		B
68 – 82	хорошо	C
56 – 67		D
50 – 55		E
20 – 49	неудовлетворительно	не зачтено FX

## 5.2. Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ А,В	«отлично»/ «зачтено (отлично)»/«зачте но»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/«зачте но»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетворитель но»/ «зачтено (удовлетворитель но)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p>



Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

### 5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

#### Вопросы для подготовки к текущим аттестациям

1. История ИИ как научного направления
2. Основные разделы теории и приложений ИИ
3. Взаимодействие искусственного интеллекта с другими научными направлениями
4. Обработка речевого сигнала в системах распознавания речи.
5. Синтаксический анализ текста.
6. Системы машинного перевода: история разработок, нерешенные проблемы, перспективы.
7. Требования, предъявляемые к системам представления и обработки знаний
8. Семантические сети и графы. Фреймы
9. Приобретение и формализация знаний. Трудности построения баз знаний
10. Методы моделирования и обучения нейронных сетей
11. Семантический анализ целого текста. Анализ тональности

#### Вопросы для подготовки к зачету

1. Сферы использования ИИ в филологических и лингвистических исследованиях.
2. Основные приложения ИИ в филологии и языкознании. Автоматический анализ речевого сообщения. Системы распознавания речи.
3. Основные приложения ИИ в филологии и языкознании. Приложения для автоматического анализа морфологии и синтаксиса.

4. Основные приложения ИИ в филологии и языкознании. Средства представления лексических знаний; онтологии.
5. Текст как объект интеллектуального анализа. Распознавание и извлечение информации из текста. Вопросно-ответные системы.
6. Текст как объект интеллектуального анализа. Анализ тональности.
7. Текст как объект интеллектуального анализа. Автоматическое аннотирование и реферирование текста.
8. История работы над программами для автоматического перевода.
9. Автоматизированный перевод и машинный перевод. Принципы работы современных систем автоматического перевода.
10. Системы автоматизированного перевода. Модели машинного перевода. Программные средства помощи переводчику.

Определение авторства как филологическая задача. Признаки, используемые в стилеметрии.

### Критерии оценивания рефератов / презентаций

Оценка	Критерии оценки
«Отлично»	<p>1. Доклад четко выстроен, сопровождается презентацией, отражающей проблематику доклада, имеет четкие выводы, полностью характеризующие работу. Автор отлично ориентируется в материале и отвечает на все вопросы, владеет специальной терминологией.</p> <p>2. Студент активно участвует в обсуждении докладов других студентов из группы.</p>
«Хорошо»	<p>1. Доклад структурирован, презентация не отражает в полной мере содержание доклада, допущены неточности, выводы нечеткие. Автор затрудняется ответить на вопросы, слабо использует специальную терминологию.</p> <p>2. Студент участвует в обсуждении докладов других студентов из группы.</p>
«Удовлетворительно»	<p>1. Доклад структурирован, содержание не полностью раскрывает цель доклада, отсутствует презентация.</p> <p>2. Студент пассивен при обсуждении докладов других студентов из группы.</p>
«Неудовлетворительно»	<p>1. Доклад не структурирован/отсутствует, презентация отсутствует. Автор не владеет материалом.</p> <p>2. Студент не участвует в обсуждении докладов.</p>

## 6. Учебно-методическое и информационное обеспечение дисциплины

### 6.1. Список источников и литературы

## Основная литература

- Автоматическая обработка текстов на естественном языке и анализ данных: учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017.
- Бовтенко М.А. Компьютерная лингводидактика: учеб. пособие / М. А. Бовтенко. - М.: Флинта: Наука, 2005.
- Введение в электронные лингвистические ресурсы [Электронный ресурс] / сост. В. Е. Гольдин, О. Ю. Крючкова. Саратов: 2011.
- Зубов А.В., Зубова И.И. Основы искусственного интеллекта для лингвистов: учебное пособие для вузов. Москва: Логос, 2007.
- Леонтьева Н.Н. Автоматическое понимание текстов: системы, модели, ресурсы. М., 2006

## Дополнительная литература

- Андреев В.С. Моделирование индивидуального стиля (на основе лингвистических характеристик): монография. М.: Флинта – Наука, 2012.
- Герд А.С. Прикладная лингвистика. – СПб: Изд-во Санкт-Петербургского ун-та, 2005
- Добров Б.В., Иванов В.В., Лукашевич Н.В., Соловьев В.Д. Онтологии и тезаурусы: модели, инструменты, приложения. Изд-во ИНТУИТ, 2009.
- Мартыненко Г.Я. Основы стилеметрии. Л.: Изд-во Ленинградского ун-та, 1988.
- Марусенко М.А. Атрибуция анонимных и псевдонимных литературных произведений методами теории распознавания образов. Л.: Изд-во Ленинградского ун-та, 1990.
- Минин А. Я. Информационные технологии в образовании: Учебное пособие / Москва: Московский педагогический государственный университет, 2016.
- Осипов Г. С. Методы искусственного интеллекта. М.: Физматлит, 2011.
- Паттерсон, Д. Глубокое обучение с точки зрения практика / Д. Паттерсон, А. Гибсон. — Москва: ДМК Пресс, 2018.
- Прикладная и компьютерная лингвистика / Под ред. И. С. Николаева, О. В. Митрениной, Т. М. Ландо. Изд. 2-е. М.: ЛЕНАНД, 2017.
- Lucci S., Kopec D. Artificial intelligence in the 21st century. Stylus Publishing, LLC, 2015.

## 6.2. Рекомендуемые интернет-ресурсы

1. International Computer Archive of Modern and Medieval English <http://icame.uib.no/>
2. Атлас языков мира. <http://wals.info/>
3. База данных «Языки мира» [www.dblang.ru](http://www.dblang.ru)
4. Интернет-портал «Historia linguisticae»
5. <http://histling.nw.ru/>
6. Коллекция словарей Института русского языка им. В.В. Виноградова РАН <http://slovari.ru>
7. Корпуса английского языка <https://www.english-corpora.org/>
8. Лингвистические данные Linguistic data consortium: <https://www ldc.upenn.edu/language-resources>
9. Лингвистический процессор «ЭТАП-3» <http://proling.iitp.ru/ru/etap>
10. Машинный фонд русского языка <http://cfri.ru>
11. Национальный корпус русского языка. <http://ruscorpora.ru/>
12. Новый частотный словарь русской лексики. <http://dict.ruslang.ru/freq.php>
13. Программное обеспечение в области автоматической обработки текста. [www.aot.ru](http://www.aot.ru)
14. Проект «Вавилонская башня» <http://starling.rinet.ru>
15. Русский ворднет <http://wordnet.ru/>

16. Системы автоматического аннотирования текстов [www.copernic.com](http://www.copernic.com)
17. Словари, созданные на основе Национального корпуса русского языка.  
<http://dict.ruslang.ru/>
18. Тезаурус английского языка WordNet <http://wordnet.princeton.edu>
19. Технологии Академии «Яндекс» <http://company.yandex.ru/technologies/search>
20. Упсальский корпус русского языка. <http://www.slaviska.uu.se/korpus.htm>
21. Хельсинский аннотированный корпус русских текстов: [slav.helsinki.fi/hanco](http://slav.helsinki.fi/hanco)
22. Чешский национальный корпус. <http://ucnk.ff.cuni.cz/english/index.php>
23. Каталог лингвистических ресурсов CLARIN <https://www.clarin.eu/>
24. Каталог лингвистических ресурсов ELRA <http://www.elra.info/en/>.

### **6.3. Профессиональные базы данных и информационно-справочные системы**

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Консультант Плюс
2. Гарант

### **7. Материально-техническое обеспечение дисциплины**

Лекционный материал для студентов содержит статистические данные, а также сравнительные характеристики изменения динамики экономических показателей. Для проведения аудиторных занятий по дисциплине «Институциональная экономика» необходима аудитория, оснащенная ПК и мультимедиа-проектором:

Персональный компьютер, или ноутбук с процессором семейств Intel, или AMD, с тактовой частотой не менее 1 ГГц;

1. Операционная система: Microsoft Windows 2000, Microsoft Windows XP, Microsoft Windows Vista;  
- Не менее 256 МБ оперативной памяти, рекомендуемый объем - 512 МБ; -  
Видеокарта и монитор с разрешением не менее 1024x768 точек;
2. Интернет-браузер только Internet Explorer не ниже версии 5.5.
3. Мультимедиа-проектор

Для обучающихся с нарушениями опорно-двигательного аппарата обеспечить:

- безбарьерную среду, в кабинете предусмотрены регулируемые компьютерные кресла индивидуальные столы обеспечивающие реализацию эргономических принципов: техническими средствами обучения для обучающихся с различными видами ограничения здоровья.

### **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

для слепых и слабовидящих:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
- письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
- обеспечивается индивидуальное равномерное освещение не менее 300 люкс; -

для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств; - письменные задания оформляются увеличенным шрифтом; - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

для глухих и слабослышащих:

- лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
- письменные задания выполняются на компьютере в письменной форме;
- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

для лиц с нарушениями опорно-двигательного аппарата:

- лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
- письменные задания выполняются на компьютере со специализированным программным обеспечением;
- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа. Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

для слепых и слабовидящих:

- в печатной форме увеличенным шрифтом; -
- в форме электронного документа;
- в форме аудиофайла.

для глухих и слабослышащих: -

- в печатной форме;
- в форме электронного документа.

для обучающихся с нарушениями опорно-двигательного аппарата: -

- в печатной форме;
- в форме электронного документа; -
- в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

для слепых и слабовидящих:

- устройством для сканирования и чтения с камерой SARA CE; -
- дисплеем Брайля PAC Mate 20;
- принтером Брайля EmBraille ViewPlus;

для глухих и слабослышащих:

- автоматизированным рабочим местом для людей с нарушением слуха и

слабослышащих;

- акустический усилитель и колонки;

для обучающихся с нарушениями опорно-двигательного аппарата:

- передвижными, регулируемые эргономическими партами СИ-1;

- компьютерной техникой со специальным программным обеспечением.

## **9. Методические материалы**

### **9.1. Планы проведения семинарских и практических занятий**

Тема 1. «Искусственный интеллект в гуманитарной сфере»

Необходимое материальное обеспечение: учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: программный пакет Microsoft Office

Рекомендуемые интернет-ресурсы: Вестник Digital Humanities <http://vdigital.me/>. Изучаем Digital Humanities <https://dhumanities.ru/>. Российская ассоциация цифровых гуманитарных наук <http://dhrussia.ru/#o-nas>. European Association for Digital Humanities <https://eadh.org//>

Виды учебных занятий: лекции, практические занятия, самостоятельная работа

Примерные задания для практических занятий: дебаты «Digital Humanities: приверженцы и критики»; обсуждение. Дискуссия «Этические аспекты функционирования ИИ».

Примерные задания для самостоятельной работы студентов: подготовка докладов «Проблемы искусственного интеллекта и типы решаемых задач»

«История ИИ как научного направления»

«Основные разделы теории и приложений ИИ»

«Взаимодействие искусственного интеллекта с другими научными направлениями».

«Введение в цифровую филологию»

«Направления цифровой филологии»

«История развития направления исследований Digital Humanities»

«Гуманитарная экспертиза технологических изменений»

«Этические аспекты функционирования ИИ»

«Сферы использования ИИ в филологических и лингвистических исследованиях».

«Основные приложения ИИ в филологии и языкознании»

Рекомендуемые формы отчетности: реферат, презентация

## Тема 2 «Лингвистические и филологические ресурсы и программы. Big Data»

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: Microsoft Office

Рекомендуемые интернет-ресурсы:

Филологический портал [www.philology.ru](http://www.philology.ru),

Международный лингвистический портал The LinguistList. URL: [www.linguistlist.org](http://www.linguistlist.org),

База данных языков мира SIL International. URL: <http://www.sil.org/>,

Новые возможности лингвистических исследований по исторической семантике с применением электронных ресурсов (<http://textualheritage.org/content/view/74/68/lang.ru..>),

Языковые ресурсы: традиции и инновации (<http://elib.grsu.by/katalog/161659-346552.pdf>),

Контроль использования интернет-ресурсов (<http://alexott.net/ru/writings/cf/JI200502.pdf>),

Атлас языков мира. <http://wals.info/>,

База данных «Языки мира» [www.dblang.ru](http://www.dblang.ru),

Лингвистические данные Linguistic data consortium: <https://www ldc.upenn.edu/language-resources>,

Лингвистический процессор «ЭТАП-3» <http://proling.iitp.ru/ru/etap>,

Интернет-портал «Historia linguisticae» <http://histling.nw.ru/>,

Проект «Вавилонская башня» <http://starling.rinet.ru>,

Русский ворднет <http://wordnet.ru/>,

Технологии Академии «Яндекс» <http://company.yandex.ru/technologies/search>,

Каталог лингвистических ресурсов CLARIN <https://www.clarin.eu/>,

Каталог лингвистических ресурсов ELRA <http://www.elra.info/en/>.

Виды учебных занятий: лекции, практические занятия, самостоятельная работа

Примерные задания для практических занятий: создание каталога лингвистических и филологических ресурсов и программ. Дискуссия «Верификация данных. Открытые данные и проблема их обработки»

Примерные задания для самостоятельной работы студентов: подготовка докладов

«Электронные лингвистические ресурсы»

«Лингвистические базы данных»

«Построение и применение лингвистических ресурсов»

«Цифровые библиотеки»

«Информационные системы в филологических задачах»

«Открытые данные и проблема их обработки»

«Верификация данных»

«Большие данные в филологии и языкознании».

Рекомендуемые формы отчетности: презентация, реферат

### Тема 3 «Компьютерная лексикография»

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Рекомендуемые интернет-ресурсы:

Электронные словари МультиТран: [www.multitran.ru](http://www.multitran.ru),

Подбор слов с помощью тезауруса: <https://support.office.com/ru-ru>,

Инструкция по использованию программ конкорданс:

<https://eflnotes.wordpress.com/2013/03/06/building-your-own-corpus-textstat-antconc/>,

Использование ABBY Lingvo: <http://www.ciao-italy.ru/materialy-i-sredstva/abbyy-lingvo-tutor-poshagovaya-instrukciya.html>,

Коллекция словарей Института русского языка им. В.В. Виноградова РАН <http://slovari.ru>,

Новый частотный словарь русской лексики. <http://dict.ruslang.ru/freq.php>,

Словари, созданные на основе Национального корпуса русского языка. <http://dict.ruslang.ru/>,

Тезаурус английского языка WordNet <http://wordnet.princeton.edu>,

Виды учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа

Примерные задания для практических занятий: дискуссия «Лексическая информация в системах ИИ».

Примерные задания для лабораторных занятий: сравнение электронных словарей.

Примерные задания для самостоятельной работы студентов: подготовка докладов «Прикладные аспекты лексикографии»

«Электронные словари: основные типы словарей и принципы их организации»

«Структура словарной статьи электронного словаря»

«Электронные словари, доступные в сети»

«Частотные словари»

«Иноязычные словари»

Рекомендуемые формы отчетности: презентация, реферат, лабораторная работа



#### Тема 4 «Корпусная лингвистика»

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: ABBYY Aligner, Skuuper Cleaner, YouAlign, Champollion, Hunalign, LEOBILINGUA, AntConc,

Рекомендуемые интернет-ресурсы:

Национальный корпус русского языка: [www.ruscorpora.ru](http://www.ruscorpora.ru),

The Corpus of Contemporary American English: <http://corpus.byu.edu>,

British National Corpus. <http://www.natcorp.ox.ac.uk/>,

Корпуса английского языка <https://www.english-corpora.org/>,

Упсальский корпус русского языка. <http://www.slaviska.uu.se/korpus.htm>,

Хельсинский аннотированный корпус русских текстов: [slav.helsinki.fi/hanco](http://slav.helsinki.fi/hanco),

Чешский национальный корпус. <http://ucnk.ff.cuni.cz/english/index.php>,

Виды учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа

Примерные задания для практических занятий: Характеристика наиболее известных корпусов английского языка. Case-study «Разработка собственного корпуса»

Дискуссия «Сложности при создании корпусов и возможности их преодоления».

Примерные задания для лабораторных занятий: Работа с параллельными корпусами (OPUS, Linguee, НКРЯ). Выравнивание (Skuuper Cleaner, YouAlign, Champollion, Hunalign, LEOBILINGUA, ABBYY Aligner). Корпусные приложения (Google books Ngram Viewer, SketchEngine), корпусные менеджеры (AntConc). Разметка TEI. Markdown. BaseX. Язык запросов XPath.

Примерные задания для самостоятельной работы студентов: подготовка докладов «Принципы организации корпуса»

«Основные свойства корпуса»

«Классификации корпусов»

«Современные корпуса текстов»

«Интернет как корпус»

«Национальный корпус русского языка и другие русскоязычные корпусы»

«Корпусы иных языков»

«Обработка корпусных данных»

«Работа с параллельными корпусами»

«Структура и назначение параллельных корпусов в решении задач ИИ».

Рекомендуемые формы отчетности: презентация, реферат, лабораторная работа

Тема 5 «ИИ в переводческой деятельности. Автоматизированный перевод. Машинный перевод»

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: SmartCAT, PROMT.

Рекомендуемые интернет-ресурсы:

Интернет в работе переводчика ([http://samlib.ru/w/wagapow\\_a\\_s/tissen.shtml](http://samlib.ru/w/wagapow_a_s/tissen.shtml)).

Системы машинного перевода текстов и словари <https://compress.ru/article.aspx?id=11757>.

Какие инструменты могут ускорить работу переводчиков: обзор сервисов: <https://rb.ru/opinion/instrumenty-dlya-perevoda/>.

24 полезные программы для переводчиков <http://blog.tran.su/shkola/goods/tm/>.

Виды учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа.

Примерные задания для практических занятий: дискуссия «Анализ алгоритмов работы систем машинного перевода»

Дискуссия «Перспективы развития переводческой отрасли»

Составление каталога полезных программ для переводчиков.

Примерные задания для лабораторных занятий: Сравнение систем машинного и автоматизированного перевода (Google Translate, Bing Translator, Yandex.Translate, Promt, Сократ Персональный). CAT-системы.

Примерные задания для самостоятельной работы студентов: подготовка докладов «Автоматический перевод»

«История создания систем автоматического перевода»

«Основные концептуальные идеи и принципы машинного перевода»

«Современное состояние отрасли»

«Классификация систем машинного перевода»

«Машинный vs автоматизированный перевод»

«Методы оценки качества перевода»

«Перспективы развития машинного перевода»

«Сравнение систем машинного перевода»

«Программные средства помощи переводчику»

«Потребительские и профессиональные переводческие прикладные программы»

«История работы над программами для автоматического перевода»

«Принципы работы современных систем автоматического перевода»

«Системы автоматизированного перевода»

«Модели машинного перевода»

«Системы машинного перевода: история разработок, нерешенные проблемы, перспективы».

Рекомендуемые формы отчетности: презентация, реферат, лабораторная работа

Тема 6. «Визуализация текстовых данных. Стилеметрия»

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: VoyantTools, Microsoft Office, Gephi, R, R Studio.

Рекомендуемые интернет-ресурсы: Способы визуализации текстовой информации  
<https://cs.hse.ru/vitext/visualize>.

Виды учебных занятий: лабораторные занятия, самостоятельная работа

Примерные задания для практических занятий: сравнение способов визуализации текстовых данных.

Примерные задания для лабораторных занятий: Визуализация данных (диаграммы, гистограммы, графики; облака слов, облака Вена; графовые методы анализа текста, сетевой анализ). Стилеметрия.

Примерные задания для самостоятельной работы студентов: подготовка докладов «Атрибуция текста»

«Определение авторства как филологическая задача»

«Признаки, используемые в стилеметрии».

Рекомендуемые формы отчетности: презентация, лабораторная работа

Тема 7. «Системы автоматической обработки письменного текста и звучащей речи».

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: Microsoft Office, ABBYY FineReader Engine, ELAN, Praat.

Рекомендуемые интернет-ресурсы:

Программы лингвистического анализа и обработки текста

<http://asknet.ru/analytics/programms.htm>.

Программы анализа и лингвистической обработки текстов <https://rvb.ru/soft/catalogue/c01.html>.

Программное обеспечение в области автоматической обработки текста. [www.aot.ru](http://www.aot.ru).

Виды учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа

Примерные задания для лабораторных занятий: Распознавание и графематический анализ текста (OCR: Convertio, ABBYY Fine Reader Online, New OCR, Online OCR). Работа с видео и аудио (Shotcut, ELAN, Praat). Обработка звучащей речи: программы анализа речи, программы синтеза речи.

Примерные задания для практических занятий: создание каталога программ автоматической обработки письменного текста; анализ и сопоставление программ.

Примерные задания для самостоятельной работы студентов: подготовка докладов

«Автоматический анализ естественного языка»

«Общие проблемы автоматической обработки естественного языка»

«Автоматическая обработка письменного текста»

«Автоматическое распознавание текста»

«Распознавание и извлечение информации из текста»

«Приложения для автоматического анализа морфологии и синтаксиса»

«Средства представления лексических знаний»

«Приобретение и формализация знаний»

«Трудности построения баз знаний»

«Онтологии»

«Автоматический анализ речевого сообщения»

«Системы распознавания речи»

«Чат-боты и диалоговые системы»

«Обработка речевого сигнала в системах распознавания речи»

«Синтаксический анализ текста».

Рекомендуемые формы отчетности: презентация, реферат, лабораторная работа

## Тема 8. «Аспекты интеллектуального анализа текста. Text Mining»

Необходимое материальное обеспечение: Учебная аудитория с мультимедийным комплексом.

Необходимое программное обеспечение: GATE, KNIME, Orange, RapidMiner, LPU

Рекомендуемые интернет-ресурсы:

ACE – Automatic Content Extraction  
<https://web.archive.org/web/20060308054306/http://www.itl.nist.gov/iad/894.01/tests/ace/>.

Интеллектуальный анализ текстов PROMT <https://www.promt.ru/technology/text-analysis/>.

ТОП-5 инструментов для Text Mining <http://datareview.info/article/top-5-instrumentov-dlya-text-mining/>.

Виды учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа.

Примерные задания для практических занятий: создание каталога программ интеллектуального анализа текста. Дискуссия «Машинное обучение в лингвистике: сильные и слабые стороны».

Примерные задания для лабораторных занятий: сравнение инструментов для Text Mining (GATE, KNIME, Orange, RapidMiner, LPU).

1. Примерные задания для самостоятельной работы студентов: подготовка докладов «Текст как объект интеллектуального анализа»
2. «Классификация и кластеризация текстов»
3. «Самообучающиеся системы: нейронные сети»
4. «Машинное обучение в лингвистике»
5. «Текст как объект интеллектуального анализа»
6. «Вопросно-ответные системы»
7. «Анализ тональности»
8. «Автоматическое аннотирование и реферирование текста»
9. «Методы моделирования и обучения нейронных сетей»
10. «Семантический анализ целого текста»
11. «Требования, предъявляемые к системам представления и обработки знаний»
12. «Семантические сети и графы»
13. «Фреймы и сценарии»
- 14.

Рекомендуемые формы отчетности: презентация, реферат, лабораторная работа

## АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

**Целью** курса «Инструменты искусственного интеллекта для анализа языка и текста» является знакомство с технологиями автоматической обработки естественного языка, обучение применению таких технологий в различных сферах филологической и лингвистической деятельности. В рамках курса предусматривается ознакомление учащихся с основными областями и задачами применения технологий искусственного интеллекта, а также с современными программными средствами для решения базовых филологических и лингвистических прикладных задач.

**Задачами** курса является приобретение студентами навыков работы с приложениями для обработки естественного языка, системами автоматического перевода, анализа тональности, а также получение представлений о возможном применении данных средств и ресурсов при проведении исследований.

Дисциплина направлена на формирование следующих компетенций:

**ОПК-6.1** Умеет отбирать наиболее подходящие для решения задачи информационные технологии и продукты

**ОПК-6.2** Умеет применять информационные технологии и продукты в зависимости от текущей задачи

**ПК-3.1** – Владеет основами методов научного исследования, информационной и библиографической культурой

**ПК-3.2** – Владеет стандартными методиками поиска, анализа и обработки материала исследования.

В результате изучения дисциплины студент должен:

**знать** различные типы электронных лингвистических ресурсов,

**уметь** работать с технологиями искусственного интеллекта в области обработки естественного языка и грамотно применять их в соответствии с типом задачи,

**владеть** навыками работы с электронными словарями, лингвистическими базами данных, приложениями для автоматического анализа морфологии и синтаксиса, системами автоматизированного перевода.

По дисциплине предусмотрена промежуточная аттестация в форме зачёта.

Общая трудоёмкость освоения дисциплины составляет 2 зачетные единицы.