

Документ подписан Министерством науки и высшего образования Российской Федерации
Информация о владельце: ФГБОУ ВО «Тверской государственный университет»
ФИО: Смирнов Сергей Николаевич
Должность: врио ректора
Дата подписания: 25.10.2023 09:52:26
Уникальный программный ключ:
69e375c64f7e975d4e8830e7b4fcc2ad1bf35f08

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «Тверской государственный университет»

Утверждаю:
Руководитель ООП
М.Л. Логунов



Рабочая программа дисциплины (с аннотацией)

Инструменты искусственного интеллекта для
анализа текста

Направление подготовки
45.03.01 Филология

Направленность (профиль)
Отечественная филология

Для студентов 1 курса
очной формы обучения

Составитель:
Старший преподаватель
Е.В. Тишина

Тверь, 2023

I. Аннотация

1. Цель и задачи дисциплины

Целью данного курса является знакомство с технологиями автоматической обработки естественного языка с целью 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. Результатом занятий должно стать приобретение студентами навыков работы с приложениями и веб-сервисами для обработки естественного языка, системами автоматического перевода, корпусами текста и электронными словарями, а также получение представлений о возможном применении данных средств и ресурсов при проведении исследований. Изучение дисциплины «Инструменты искусственного интеллекта для анализа текста» способствует развитию у будущих специалистов склонности и способности к творческому мышлению, выработке системного подхода к исследуемым явлениям.

2. Место дисциплины в структуре ООП

Дисциплина Б1.О.52 Инструменты искусственного интеллекта для анализа текста относится к обязательной части и ОПОП ВО. Дисциплина изучается на 1 курсе (2 семестр) и заканчивается зачетом.

Требования к уровню подготовки студентов до обучения: знания теоретических основ информатики и вычислительной техники в объеме программы общеобразовательной школы, а также иметь базовые знания школьной программы по математике.

По окончании курса студент должен уметь работать с технологиями искусственного интеллекта в области обработки естественного языка и грамотно применять их в соответствии с типом задачи, знать различные типы электронных лингвистических ресурсов, иметь навыки работы с электронными словарями, корпусами текстов, лингвистическими базами данных, приложениями для автоматического анализа морфологии и синтаксиса, системами автоматизированного перевода.

3. Объем дисциплины: 2 зачетные единицы, 72 академических часа, в том числе:

контактная аудиторная работа: лекции 18 часов, практические занятия 18 часов;

самостоятельная работа: 36 часов

4. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы

Планируемые результаты освоения образовательной программы (формируемые компетенции)	Планируемые результаты обучения по дисциплине
УК-1.1 Анализирует задачу, выделяя ее базовые составляющие	Знает принципы сбора, отбора и обобщения информации.
УК-1.2 Определяет, интерпретирует и	Умеет соотносить разнородные явления и

ранжирует информацию, требуемую для решения поставленной задачи;	систематизировать их в рамках избранных видов профессиональной деятельности; работать с интернет источниками, систематизировать материал в виде таблиц и схем, подбирать и модифицировать материал для профессиональной деятельности
УК-1.3 Осуществляет поиск информации для решения поставленной задачи по различным типам запросов;	Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт библиографического разыскания, создания научных текстов.
ОПК-7.1 Применяет современные информационные технологии для решения задач профессиональной (педагогической) деятельности;	Знает методы доступа к информационным ресурсам. Умеет пользоваться базовыми функциями персонального компьютера; пользоваться стандартными офисными приложениями (текстовый процессор, редактор таблиц, редактор презентаций); пользоваться общепринятыми ресурсами сети Интернет.
ОПК-7.2 Применяет современные информационные технологии для решения задач профессиональной (научно-исследовательской) деятельности;	Имеет практический опыт работы с поисковыми машинами, справочными и библиотечными системами, системами дистанционного образования
ОПК-6.1 Применяет современные информационные технологии сбора, хранения, обработки, анализа и передачи филологической информации; основы защиты информации;	Знает различные типы электронных лингвистических ресурсов Владеет: навыками информационно-библиографического поиска, в том числе в сети Интернет, в том числе с использованием профессиональных электронных ресурсов; навыками защиты конфиденциальности данных; навыками защиты от компьютерных вирусов и других вредоносных программ;
ОПК-6.3 Самостоятельно осуществляет поиск, накопление и расширение объёма профессионально значимых знаний при посредстве информационно-коммуникативных технологий и использовать их для решения научно-исследовательских и производственно-технологических задач профессиональной деятельности	Имеет навыки работы с электронными словарями, корпусами текстов, лингвистическими базами данных, приложениями для автоматического анализа морфологии и синтаксиса, системами автоматизированного перевода

5. Форма промежуточной аттестации и семестр прохождения

Зачет, 2 семестр.

6. Язык преподавания русский.

II. Содержание дисциплины, структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

Учебная программа – наименование разделов и тем	Всего (час.)	Контактная работа (час.)				Самостоятельная работа, в том числе Контроль (час.)
		Лекции		Практические занятия		
		всего	в т.ч.	всего	в т.ч.	

			практическая подготовка		практическая подготовка	
Новые информационные технологии и Искусственный интеллект (ИИ). Основные направления исследований в области искусственного интеллекта. История создания искусственного интеллекта. Natural Language Processing (NLP). Описание исследовательских и практических задач, в решение которых вовлекаются технологии ИИ	4	2				2
Филологические и лингвистические ресурсы и программы. Электронные лингвистические ресурсы. Цифровые библиотеки. Информационные системы в филологических задачах. Задачи информационного поиска с точки зрения филолога.	11	2		4		5
Анализ текстов с использованием открытого программного обеспечения. Voyant tools, Natasha, kartaslov.ru, Яндекс-Томица, MyStem, AntConc и др	12	2		4		6
Инструменты автоматической обработки языка. Приложения для автоматического анализа морфологии и синтаксиса. Регулярные выражения	10	2		2		6
Аспекты интеллектуального анализа текстов. Текст как объект интеллектуального анализа. Распознавание и извлечение информации из текста. Анализ тональности. Автоматическое аннотирование и реферирование текста.	9	2		2		5
ИИ в переводческой деятельности. История работы над программами для автоматического перевода. Автоматизированный перевод и машинный перевод. Системы автоматизированного перевода. Модели машинного перевода. Программные средства помощи переводчику	12	4		2		6
Национальные и иные корпуса. Национальный корпус русского языка и другие русскоязычные	14	4		4		6

корпусы. British National Corpus как первый пример национального корпуса. Корпусы иных языков. . Google books Ngram Viewer и поиск словосочетаний. Сервисы исследования биграмм и n-грамм. Обработка корпусных данных. Структура и назначение параллельных корпусов в решении задач ИИ. Примеры исследований с использованием корпусных данных.						
ИТОГО	72	18		18		36

III. Образовательные технологии

Учебная программа – наименование разделов и тем (в строгом соответствии с разделом II РПД)	Вид занятия	Образовательные технологии
Новые информационные технологии и Искусственный интеллект (ИИ). Основные направления исследований в области искусственного интеллекта. История создания искусственного интеллекта. Natural Language Processing (NLP). Описание исследовательских и практических задач, в решение которых вовлекаются технологии ИИ	Лекция	традиционная
Филологические и лингвистические ресурсы и программы. Электронные лингвистические ресурсы. Цифровые библиотеки. Информационные системы в филологических задачах. Задачи информационного поиска с точки зрения филолога.	Лекция	лекция-визуализация
	Практические занятия	Информационные (цифровые)
Анализ текстов с использованием открытого программного обеспечения. Voyant tools, Natasha, kartaslov.ru, Яндекс-Томита, MyStem, AntConc и др	Лекция	лекция-визуализация
	Практические занятия	Проектная технология
Инструменты автоматической обработки языка. Приложения для автоматического анализа морфологии и синтаксиса. Регулярные выражения	Лекция	лекция-визуализация
	Практические занятия	Информационные (цифровые)
Аспекты интеллектуального анализа текстов. Текст как объект интеллектуального анализа. Распознавание и извлечение информации из текста. Анализ тональности. Автоматическое аннотирование и реферирование текста.	Лекция	лекция-визуализация
	Практические занятия	Информационные (цифровые)
ИИ в переводческой деятельности. История работы над программами для автоматического перевода. Автоматизированный перевод и машинный перевод. Системы автоматизированного перевода. Модели машинного перевода. Программные средства помощи переводчику	Лекция	Активное слушание
	Практические занятия	Информационные (цифровые)
Национальные и иные корпуса. Национальный корпус русского языка и другие русскоязычные корпуса. Корпусы иных языков. British National Corpus как	Лекция	лекция-визуализация

<p>первый пример национального корпуса. Google books Ngram Viewer и поиск словосочетаний. Сервисы исследования биграмм и n-грамм. Обработка корпусных данных. Структура и назначение параллельных корпусов в решении задач ИИ. Примеры исследований с использованием корпусных данных.</p>	<p>Практические занятия</p>	<p>Информационные (цифровые)</p>
--	-----------------------------	----------------------------------

IV. Оценочные материалы для проведения текущей и промежуточной аттестации

Оценочные материалы для проведения текущей аттестации

1. контрольные вопросы

1. История ИИ как научного направления. Пути развития.
2. Основные разделы теории и приложений ИИ.
3. Сферы использования ИИ в филологических и лингвистических исследованиях.
4. Современные информационно-поисковые системы (Google, Яндекс и др.). Возможности расширенного поиска Синтаксис запросов
5. Основные приложения ИИ в филологии и языкознании.
6. Взаимодействие искусственного интеллекта с другими научными направлениями
7. Системы русского морфологического анализа (Mystem).
8. Синтаксический анализ текста.
9. Семантический анализ целого текста.
10. Системы машинного перевода.
11. Существующие системы машинного перевода.
12. Семантические сети и графы.
13. Какие подкорпуса входят в состав НКРЯ?
14. Найдите в сети Интернет электронные лингвистические словари.
15. Опишите синтаксис языка поискового запроса Яндекс (Google). Приведите примеры поиска.
16. Филологические ресурсы Интернета как главный источник информации для филолога
17. Филологическая и лингвистическая информация в электронных энциклопедиях. Проблема поиска, доступности и авторитетности информации
18. Поисковые системы и поисковая машина. Отечественные поисковые системы
19. Цифровое обучение и онлайн-образование. Массовые открытые онлайн курсы (MOOK)
20. Понятие компьютерного вируса. Технология антивирусной защиты.
21. Вопросы компьютерной и информационной безопасности. Потенциальные угрозы сохранности информации.

2. задания для практических занятий,

22. В словаре трудностей произношения и ударения в современном русском языке указано что правильная форма: танцОвщик, танцОвщица (не рекомендуется танцовщИк, танцовщИца). Пользуясь акцентологическом корпусом определите встречаются или нет тексты, где ударение стоит на последнем слоге. И если «да» приведите этот пример (эти примеры).
23. Books Ngram Viewer — это сервис от Google, который строит график частотности по упоминаниям в книгах начиная с 16 века. Данная визуализация повторений может быть полезна для какого-либо исследования, как помощник в реализации проекта, для выполнения филологического задания и в других видах деятельности:
 - a. Исследуйте с какими вместе словами встречается слово «счастье» в разных языках (русский, английский, испанский, китайский)
 - b. Найдите 10 самых частотных продолжений фразы ____ ("Самое лучшее место на земле..."). В инструкции по Advanced Usage данный тип поиска называется Wildcard search.
24. С помощью системы автоматизированного перевода выполните перевод заданного фрагмента. При переводе используйте параллельный корпус.
25. Выполните анализ текста с помощью программы AntConc

26. Используя программу AntConc, сравните текст А.С. Пушкина «Капитанская дочка» и произведение любого современного автора. Для этих двух текстов:
 - a. Отсортируйте слова по частоте (Word List - Sort by Freq). Какие слова оказались в 10-ке наиболее частотных?
 - b. Какие слова оказались самыми редкими
 - c. Выберите 3 слова и проанализируйте контексты, в которых они используются (Concordance)
 - d. выберите по одному слову из каждого текста и составьте для них список N-грамм
 - e. сделайте вывод об индивидуальных особенностях авторских стилей.
27. Используя примеры изменений, произошедших в лексике русского языка за последние сто лет (взять из <https://yandex.ru/company/researches/2021/old-words>) проанализировать частоту употребления словосочетаний с помощью НКРЯ.
28. Используя Books Ngram Viewer выясните какое из заклинаний чаще всего использовалось в серии книг о Гарри Поттере. (<https://phylodasha.github.io/>)
29. Корпусное исследование: говорят «в Украине» или «на Украине»?
30. Работу выполнить с помощью сервиса Google Books Ngram Viewer. Пользуясь знаком «*», исследовать употребительность сочетаний с переменной частью или несколькими переменными частями
31. Работу выполнить с помощью сервиса Google Books Ngram Viewer. Используя тэг `_END/_START_`, найти относительные частоты встречаемости классов слов в конечной / начальной позиции предложения в английском и русском языках.
32. Работу выполнить с помощью сервиса Google Books Ngram Viewer. Определить частоты встречаемости слова «робот». Проанализировать, почему это слово встречается до 1951 года.
33. Работу выполнить с помощью сервисов Google Books Ngram Viewer и НКРЯ. Сочетаемость слов. Например: 'старался', 'пытался', 'делал', 'пробовал' и др.
34. Работу выполнить с помощью сервиса Google Books Ngram Viewer. Динамика лексикона языка.
35. Много ли встречается в русской речи слов с сочетанием «оо» в середине слева? Проверьте это средствами корпуса.
36. Возможен ли в МУРКО семантический поиск? Можно ли, например, найти звуковые фрагменты с упоминанием животных?
37. Проверить: между домов - это устаревшая норма, между домами – современная.
38. Проверить: когда в Москве начали с конца 20-х годов планировать и строить метро, оно первое время называлось в мужском роде, но к тому времени, как его построили и открыли, оно уже закрепилось в среднем.
39. Какие прилагательные сочетаются с существительным _____ (например: "судьба", "House", «образование», «любовь»)?
40. Частотность словосочетаний типа: Adj + N, Adj + Adj + N.
41. Примеры применения корпусов текстов для лингвистических исследований

3. темы рефератов

1. Тест Тьюринга и критерии "интеллектуальности" информационных систем.
2. Определение и классификация систем искусственного интеллекта, цели и пути их создания.
3. Основные направления исследований в области Искусственного Интеллекта (ИИ)
4. История корпусной лингвистики в мире и в России.
5. Информационно-поисковые языки.
6. Лингвистические электронные библиотеки.
7. Автоматический графематический анализ текстов.
8. Автоматическая морфологическая разметка текстов.
9. Методы снятия морфологической омонимии.
10. Способы использования корпусов в лексикографии.
11. Система Multitran.
12. Использование корпусов в лингвистических исследованиях.

4. творческие задания

1. Считается, что Уильям Вордсворт воспевал и романтизировал одиночество, поскольку в его произведениях часто встречаются слова «одинок», «сам», «один». Конечно, Вордсворт использовал их очень часто. Но чаще ли, чем другие авторы? Можно ли это считать характерной особенностью его творчества?
(ист. sysblok.ru/nlp/ne-takie-uzh-i-liriki-kompjuternye-tehnologii-dlja-gumanitariiev/)
2. Преподаватели школы магии Хогвартс. Речевой портрет, описание деталей внешности (какие эпитеты, метафоры, и др словосочетания), манеры говорения, манеры поведения, сопутствующие эмоции и чувства, методов обучения, поощрения и наказания учеников, взаимоотношения с другими учителями. (ист. <https://phylodasha.github.io/>)

Оценочные материалы для проведения промежуточной аттестации

Планируемый образовательный результат (компетенция, индикатор)	Типовые контрольные задания	Критерии оценивания и шкала оценивания
<p>УК-1 Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач</p> <p>УК-1.1 Анализирует задачу, выделяя ее базовые составляющие</p> <p>УК-1.2 Определяет, интерпретирует и ранжирует информацию, требуемую для решения поставленной задачи</p> <p>УК-1.3 Осуществляет поиск информации для решения поставленной задачи по различным типам запросов</p>	<p>1. Какие подкорпуса входят в состав НКРЯ?</p> <p>2. Найдите в сети Интернет электронные лингвистические словари.</p> <p>3. Опишите синтаксис языка поискового запроса Яндекс (Google). Приведите примеры поиска.</p> <p>Устный опрос</p> <hr/> <p>В словаре трудностей произношения и ударения в современном русском языке указано что правильная форма: танцОвщик, танцОвщица (не рекомендуется танцовщИк, танцовщИца). Пользуясь акцентологическом корпусом определите встречаются или нет тексты, где ударение стоит на последнем слоге. И если «да» приведите этот пример (эти примеры).</p> <p>практическое задание</p>	<ul style="list-style-type: none"> • Полно и правильно дан ответ на все поставленные вопросы, приведены необходимые примеры; Ответ демонстрирует знание и корректное использование терминологии; ответ не содержит фактических ошибок. – 12 – 14 баллов • Полно и правильно дан ответ на все поставленные вопросы, приведены примеры, однако имеются неточности; в целом студент показывает понимание изученного материала – 8 – 11 балла • Ответ дан в основном правильно, но недостаточно аргументированы выводы, приведены не все необходимые примеры – 4 – 7 баллов • Даны неверные ответы на поставленные вопросы – 0 - 3 баллов <hr/> <p>Полно и правильно даны ответы на все поставленные вопросы, приведены необходимые примеры; студент показывает понимание излагаемого материала – 12 – 14 баллов</p> <p>Полно и правильно даны ответы на все поставленные вопросы, приведены примеры, однако имеются неточности; в целом студент показывает понимание изученного материала – 8 – 11 балла</p> <p>Ответ дан в основном правильно, но недостаточно аргументированы выводы, приведены не все необходимые примеры – 4 – 7 баллов</p> <p>Даны неверные ответы на поставленные вопросы – 0 - 3 баллов</p>
<p>ОПК-7 Способен понимать принципы работы современных информационных технологий и использовать их для решения задач</p>	<p>Books Ngram Viewer — это сервис от Google, который строит график частотности по упоминаниям в книгах начиная с 16 века. Данная визуализация повторений может быть полезна для какого-либо исследования, как помощник в реализации проекта, для</p>	<p>Полно и правильно дан ответ на все поставленные вопросы, приведены необходимые примеры; Ответ демонстрирует знание современных информационных технологий и корректное использование</p>

<p>профессиональной деятельности</p> <p>ОПК-7.1 Применяет современные информационные технологии для решения задач профессиональной (педагогической) деятельности</p> <p>ОПК-7.2 Применяет современные информационные технологии для решения задач профессиональной (научно-исследовательской) деятельности</p>	<p>выполнения филологического задания и в других видах деятельности:</p> <p>1. Исследуйте с какими вместе словами встречается слово «счастье» в разных языках (русский, английский, испанский, китайский) счастье>* _VERB, счастье>* _NOUN</p> <p>2. Найдите 10 самых частотных продолжений фразы "Где я могу найти... ". В инструкции по Advanced Usage данный тип поиска называется Wildcard search.</p>	<p>терминологии; ответ не содержит фактических ошибок. – 12 – 14 баллов</p> <p>Полно и правильно дан ответ на все поставленные вопросы, приведены примеры, однако имеются неточности; в целом студент показывает понимание изученного материала – 8 – 11 балла</p> <p>Ответ дан в основном правильно, но недостаточно аргументированы выводы, приведены не все необходимые примеры – 4 – 7 баллов</p> <p>Даны неверные ответы на поставленные вопросы – 0 - 3 баллов</p>
<p>ОПК-6 Способен решать стандартные задачи по организационному и документационному обеспечению профессиональной деятельности с применением современных технических средств, информационно-коммуникационных технологий с учетом требований информационной безопасности</p> <p>ОПК-6.1 Применяет современные информационные технологии сбора, хранения, обработки, анализа и передачи филологической информации; основы</p>	<p>Выполните анализ текста с помощью программы AntConc</p> <p>Используя программу AntConc, сравните текст А.С. Пушкина «Капитанская дочка» и произведение любого современного автора.</p> <p>Для обоих текстов:</p> <p>а) Отсортируйте слова по частоте (Word List - Sort by Freq). Какие слова оказались в 10-ке наиболее частотных? б) Какие слова оказались самыми редкими в) выберите 3 слова и проанализируйте контексты, в которых они используются (Concordance) г) постройте для них Concordance Plot для каждого слова, проанализируйте функционирование этих слов в разных частях текста д) выберите по одному слову из каждого текста и составьте для них список N-грамм е) сделайте вывод об индивидуальных особенностях авторских стилей.</p>	<p>Имеется полный анализ задачи, предложены верные шаги решения, приведено полное верное решение задачи, включающее правильный ответ – 12 – 14 баллов</p> <p>Имеется полный анализ задачи, предложен верный алгоритм решения задачи, приведено частичное решение задачи / решение недостаточно обосновано / в решении имеются лишние или неверные выводы – 8 - 11 балла</p> <p>Неполный анализ задачи, предложен верный алгоритм решения задачи, имеется верное решение лишь части задачи из-за ошибки – 4 – 7 баллов</p> <p>Не соответствует требованиям, изложенным выше – 0 - 3 баллов</p>

<p>защиты информации</p> <p>ОПК-6.3 Самостоятельно осуществляет поиск, накопление и расширение объёма профессионально значимых знаний при посредстве информационно-коммуникативных технологий и использовать их для решения научно-исследовательских и производственно-технологических задач профессиональной деятельности</p>	<p>3. Используя примеры изменений, произошедших в лексике русского языка за последние сто лет (взять из https://yandex.ru/company/researches/2021/old-words) проанализировать частоту употребления словосочетаний с помощью НКРЯ.</p> <p>4. Используя Books Ngram Viewer выясните какое из заклинаний чаще всего использовалось в серии книг о Гарри Поттере. (https://phylodasha.github.io/)</p> <p>5. Корпусное исследование: говорят «в Украине» или «на Украине»?</p>	<p>Полно и правильно дан ответ на все поставленные вопросы, приведены необходимые примеры; Ответ демонстрирует знание и корректное использование терминологии; ответ не содержит фактических ошибок. – 12 – 14 баллов</p> <p>Полно и правильно дан ответ на все поставленные вопросы, приведены примеры, однако имеются неточности; в целом студент показывает понимание изученного материала – 8 – 11 балла</p> <p>Ответ дан в основном правильно, но недостаточно аргументированы выводы, приведены не все необходимые примеры – 4 – 7 баллов</p> <p>Даны неверные ответы на поставленные вопросы – 0 - 3 баллов</p>
---	---	--

Промежуточный контроль качества усвоения студентами содержания дисциплины проводится в виде зачета, на котором определяется: уровень усвоения студентами основного учебного материала;

- уровень понимания студентами изученного материала;
- способности студентов использовать полученные знания для решения конкретных задач.

V. Учебно-методическое и информационное обеспечение дисциплины

1) Рекомендуемая литература

а) Основная литература

1. Захаров, В. П. Корпусная лингвистика : учебник / В. П. Захаров, С. Ю. Богданова. - 3-е изд., перераб. - Санкт-Петербург : СПбГУ, 2020. - 234 с. - ISBN 978-5-288-05997-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1244746> (дата обращения: 19.12.2021). – Режим доступа: по подписке.
2. Надеина, Т. М. Основы прикладной и математической лингвистики : учебное пособие / Т. М. Надеина. — Москва : Норма : ИНФРА-М, 2020. — 152 с. - ISBN 978-5-00156-072-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1092460> (дата обращения: 19.12.2021). – Режим доступа: по подписке.
3. Голицына, О. Л. Информационные системы и технологии : учебное пособие / О.Л. Голицына, Н.В. Максимов, И.И. Попов. — Москва : ФОРУМ : ИНФРА-М, 2021. — 400 с. — (Среднее профессиональное образование). - ISBN 978-5-00091-592-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1138895> (дата обращения: 19.12.2021). – Режим доступа: по подписке.

б) Дополнительная литература

4. Моделирование в корпусной лингвистике: специализированные корпуса русского языка : монография / В. П. Захаров, И. В. Азаров, О. А. Митрофанова [и др.]. - СПб : Изд-во С.-Петербург. ун-та, 2019. - 208 с. - ISBN 978-5-288-05902-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1080953> (дата обращения: 19.12.2021). – Режим доступа: по подписке.
5. Баррат, Д. Последнее изобретение человечества: Искусственный интеллект и конец эры Homo sapiens: Научно-популярное / Баррат Д., Лисова Н. - Москва : Альпина нон-фикшн, 2016. - 304 с. (Искусственный интеллект) ISBN 978-5-91671-436-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/916060> (дата обращения: 19.12.2021). – Режим доступа: по подписке.

6. Павлов, С.И. Системы искусственного интеллекта : учебное пособие / С.И. Павлов. - Томск : Томский государственный университет систем управления и радиоэлектроники, 2011. - Ч. 2. - 194 с. - ISBN 978-5-4332-0014-2 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=208939>
7. Большакова Е. И., Воронцов К. В., Ефремова Н. Э., Клышинский Э. С., Лукашевич Н. В., Сапин А. С. 2017. Автоматическая обработка текстов на естественном языке и анализ данных: учеб. пособие. М.: Изд-во НИУ ВШЭ, 2017.

2) Программное обеспечение

а) Лицензионное программное обеспечение

1. Kaspersky Endpoint Security 10 для Windows – Акт на передачу прав №2129 от 25 октября 2016 г.
2. Microsoft Office профессиональный плюс 2013 – Акт приема-передачи № 369 от 21 июля 2017
3. Microsoft Windows 10 Enterprise – Акт приема-передачи № 369 от 21 июля 2017

б) Свободно распространяемое программное обеспечение

4. Adobe Acrobat Reader DC – бесплатно
5. GIMP 2.6.12-2 – бесплатно
6. Open Source Edition 4.1.1 – бесплатно
7. Google Chrome – бесплатно
8. SmartGit – бесплатно
9. WinDjView 2.0.2 - бесплатно

3) Современные профессиональные базы данных и информационные справочные системы

1. ЭБС «ZnaniUM.COM» www.znanium.com
2. ЭБС «ЮРАЙТ» www.biblio-online.ru
3. ЭБС «Университетская библиотека онлайн» <https://biblioclub.ru/>
4. ЭБС IPRbooks <http://www.iprbookshop.ru/>
5. ЭБС «Лань» <http://e.lanbook.com>
6. ЭБС BOOK.ru <https://www.book.ru/>
7. ЭБС ТвГУ <http://megapro.tversu.ru/megapro/Web>

4) Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины:

1. Сайт ТвГУ <http://university.tversu.ru/structure/>
2. Репозиторий ТвГУ <http://eprints.tversu.ru>
3. www.office.com Microsoft Office 365
4. Википедия общедоступная многоязычная универсальная интернет-энциклопедия со свободным контентом ru.wikipedia.org
5. Веб-сайт компьютерных словарей// www.dictionary.com
6. Коллекция словарей Института русского языка им. В.В. Виноградова РАН slovari.ru
7. Словари и энциклопедии//www.academic.ru
8. Национальный корпус русского языка// www.ruscorpora.ru
9. Антивирусы и программы безопасности comss.ru
10. Ресурсы МООС (Массовый открытый онлайн-курс) Coursera, edX, Stanford Online, Udacity, Российская национальная платформа открытого образования postnauka.ru/themes/linguistics, postnauka.ru/themes/language
11. Справочно-информационный портал ГРАМОТА.РУ – русский язык для всех
12. Учеб. пособие//www.informika.ru/text/inftech/edu/desing
13. Академия Google <https://scholar.google.com/>
14. Исследования Яндекс <https://yandex.ru/company/researches>
15. «Системный Блокъ» — издание о цифровых технологиях в гуманитарных науках, искусстве, образовании и не только <https://sysblok.ru/>
16. Автоматическая обработка текста (материалы сайта). URL: <http://www.aot.ru>.
17. Национальный корпус русского языка. URL: <http://www.ruscorpora.ru>

18. British National Corpus. <http://www.natcorp.ox.ac.uk/>
 19. International Computer Archive of Modern and Medieval English <http://icame.uib.no/>
 20. Атлас языков мира. <http://wals.info/>
 21. База данных «Языки мира» www.dblang.ru
 22. Корпуса английского языка <https://www.english-corpora.org/>
 23. Лингвистические данные Lingistic data consortium:
<https://www ldc.upenn.edu/languageresources>
 24. Машинный фонд русского языка <http://cfri.ru>
 25. Новый частотный словарь русской лексики. <http://dict.ruslang.ru/freq.php>
 26. Программное обеспечение в области автоматической обработки текста. www.aot.ru
 27. Проект «Вавилонская башня» <http://starling.rinet.ru>
 28. Русский ворднет <http://wordnet.ru/>
 29. Системы автоматического аннотирования текстов www.copernic.com
 30. Словари, созданные на основе Национального корпуса русского языка. dict.ruslang.ru/
 31. Тезаурус английского языка WordNet <http://wordnet.princeton.edu>
 32. Технологии Академии «Яндекс» company.yandex.ru/technologies/search
 33. Упсальский корпус русского языка. snd.gu.se/en/catalogue/study/ext0071
 34. Хельсинский аннотированный корпус русских текстов: h248.it.helsinki.fi/hanco/
 35. Чешский национальный корпус. <https://ucnk.ff.cuni.cz/cs/>
 36. Каталог лингвистических ресурсов CLARIN <https://www.clarin.eu/>
 37. Каталог лингвистических ресурсов ELRA <http://www.elra.info/en/>.
 38. Каталог сетевых ресурсов по русской филологии и фольклору <http://feb-web.ru/feb/feb/sites.htm>
- Ресурсы по грамматике русского языка
39. «Русская грамматика» <http://www.rusgram.narod.ru/>
 40. Интернет-портал «Русская корпусная грамматика» <http://rusgram.ru/>
- Лексикографические ресурсы
41. Орфографический академический ресурс «Академос» <http://orfo.ruslang.ru/>
 42. Орфографическое комментирование русского словаря <http://oross.ruslang.ru/>
 43. С. И. Ожегов, Н. Ю. Шведова. Толковый словарь русского языка
<http://ozhegov.info/slovar/>
 44. Словарь русского языка в 4-х томах («МАС», Малый академический словарь) <http://feb-web.ru/feb/mas/mas-abc/default.asp>
 45. Новый объяснительный словарь синонимов русского языка
http://ruslang.ru/text_noss2_title
 46. Словари и Энциклопедии <https://gufo.me/>
 47. Проект «Лексикограф» <http://lexicograph.ruslang.ru/>
 48. М. Фасмер. Этимологический словарь русского языка
<http://etymolog.ruslang.ru/index.php?act=contents&book=vasmer>
 49. База частот значений слов для многозначных существительных, глаголов, прилагательных <http://sensefreq.ruslang.ru/>
 50. База AdaGram (кластеры контекстов, соответствующие употреблению слова в корпусе)
<http://adagram.ll-cl.org/>
 51. Проект «Большой толковый словарь русских прилагательных и наречий»
<http://www.ruslang.ru/node/1638>
 52. Информационно-поисковая система «Образный инструментарий русской лирики»
<http://lexrus.ru/default.aspx?p=3118>
 53. Интерактивный словарь компаративных тропов русской художественной литературы XIX-XXI вв. https://www.ruslang.ru/slovar_tropov
- Ресурсы по истории русского языка
54. Рукописные памятники древней Руси <http://www.lrc-lib.ru/>
 55. Этимология и история слов русского языка <http://etymolog.ruslang.ru/>
 56. Сводный словник древнерусских памятников для дополнения базы данных «Словаря русского языка XI–XVII вв.» Логин: user. Пароль: useslovník
<http://ruslang.ru:8080/slovník/web/login.jsp>

57. Электронная информационно-поисковая система с реализацией доступа на базе «Словаря древнерусского языка (XI–XIV вв.)» <http://www.slovari.ru/default.aspx?p=2641>
58. Партнерство «Рукописные памятники Древней Руси» (древнерусские берестяные грамоты, русские летописи, издания славянских памятников) <http://www.lrc-lib.ru/>
59. Славяно-русский Пролог по древнейшим рукописям. Тексты, источники, комментарии <http://prolog-manuscript.org/?fbclid=IwAR14oJJJ5ftPDkGC2LleruswJ9FB6y2nqo5bqA2WOX3thysmomce7DpYe4s>
- Ресурсы по русской диалектологии
60. Каталог сетевых ресурсов по русской филологии и фольклору <http://feb-web.ru/feb/feb/sites.htm>
61. Корпус говора села Роговатка <http://lingconlab.ru/rogovatka/>
62. Корпус опочечских говоров <http://lingconlab.ru/opochka/>
63. Корпус говора села Малинино <http://lingconlab.ru/malinino/>
64. Школьный диалектологический атлас. Язык русской деревни <http://gramota.ru/book/village/>
65. А. В. Тер-Аванесова, С. А. Крылов. Конкорданс словоформ к записям Харовских говоров Вологодского диалекта русского языка по материалам экспедиций 2001–2002 гг. <https://www.ruslang.ru/doc/krylov/harovskie11b.pdf>
66. А. В. Тер-Аванесова, С. А. Крылов. Лексико-грамматическая база данных по севернорусскому слободскому говору деревень Арзубиха, Захариха и Злобиха Харовского-р-на Вологодской области (скачиваемая в формате .rar база данных для программы Starling)
67. С. А. Крылов, А. В. Тер-Аванесова. Электронные базы данных по русским народным говорам https://www.ruslang.ru/krylov_dialect
- Русский язык в контексте других славянских языков
68. Общеславянский лингвистический атлас <http://www.slavatlas.org/>
69. Комиссия по лексикологии и лексикографии при Международном комитете славистов <https://www.ruslang.ru/lexcomission>
70. Национальный комитет славистов РФ <http://nkslav.ru/>

VI. Методические материалы для обучающихся по освоению дисциплины

Преподавание дисциплины осуществляется в форме лекционных и практических занятий, которые сопровождаются демонстрацией наглядного материала – специализированных приложений, Интернет-ресурсов и презентаций. Во время занятий обучающиеся выполняют практические задания. Для закрепления пройденного материала предлагаются домашние задания по каждой из тем. Успешное овладение содержанием дисциплины «Инструменты искусственного интеллекта для анализа текста» предполагает работу обучающихся в группах в аудитории, а также их самостоятельную работу.

вопросы для подготовки к зачету;

1. История ИИ как научного направления.
2. Основные разделы теории и приложений ИИ
3. Интернет как источник лингвистической информации (хранилище электронных словарей, энциклопедий, глоссариев и разнообразных справочных материалов, корпус текстов, из которого можно черпать необходимую лингвистическую информацию; вместилище специализированных лингвистических ресурсов)
4. Взаимодействие искусственного интеллекта с другими научными направлениями
5. Обработка речевого сигнала в системах распознавания речи.
6. Синтаксический анализ текста.
7. Системы машинного перевода: история разработок, нерешенные проблемы, перспективы.
8. Требования, предъявляемые к системам представления и обработки знаний.
9. Семантические сети и графы. Фреймы.
10. Приобретение и формализация знаний. Трудности построения баз знаний.
11. Методы моделирования и обучения нейронных сетей.
12. Семантический анализ целого текста. Анализ тональности.

13. Дать определения терминов корпусной лингвистики (описание понятий): Разметка корпусов, Репрезентативность, Метаданные, Корпус-менеджер, Tree bank, Лемматизация, Конкорданс, Параллельный корпус
14. Сферы использования ИИ в филологических и лингвистических исследованиях.
15. Основные приложения ИИ в филологии и языкознании. Автоматический анализ речевого сообщения. Системы распознавания речи.
16. Основные приложения ИИ в филологии и языкознании. Приложения для автоматического анализа морфологии и синтаксиса.
17. Основные приложения ИИ в филологии и языкознании. Средства представления лексических знаний; онтологии.
18. Текст как объект интеллектуального анализа. Распознавание и извлечение информации из текста. Вопросно-ответные системы.
19. Текст как объект интеллектуального анализа. Анализ тональности.
20. Текст как объект интеллектуального анализа. Автоматическое аннотирование и реферирование текста.
21. История работы над программами для автоматического перевода.
22. Автоматизированный перевод и машинный перевод. Принципы работы современных систем автоматического перевода.
23. Системы автоматизированного перевода. Модели машинного перевода. Программные средства помощи переводчику.
24. Определение авторства как филологическая задача. Признаки, используемые в стилеметрии.
25. Методы и подходы в современной стилеметрии. Анализ стиля и искусственный интеллект.
26. Использование элементов ИИ в обучении языку. Компьютерные обучающие программы. Два направления в разработке искусственного интеллекта. Основные задачи искусственного интеллекта
27. выполнить поиск информации по тематике научного исследования; продемонстрировать корректность использования полученной научной информации;
28. представить научно-справочный аппарат со списком ссылок на научные ресурсы, найденные в сети Интернет;
29. выполнить сбор информации (предоставить не менее 15-20 фрагментов со ссылками на источник) для подготовки публикации о наиболее известных событиях, описанных в мировой литературе;
30. выбрать текст (фрагмент статьи любого компьютерного издания), содержащий сленговые выражения и слова, объемом не более одной страницы и преобразовать в текст, не содержащий сленговых выражений и слов. Должен быть составлен переводной словарь данного текста.
31. Трудности автоматизации обработки естественного языка в интеллектуальных системах.
32. Этапы анализа предложений на естественном языке.
33. Интеллектуальные системы, использующие естественный язык.
34. Задачи морфологического анализа, морфологический разбор, стемминг, лемматизация.
35. Понятия лексем, словоформы, леммы, морфемы, псевдо-основы и псевдо-окончания.
36. Грамматические модели русского языка в контексте автоматической обработки.
37. Понятие статистической языковой модели. Области применения. N-граммы.
38. История возникновения науки об искусственном интеллекте. Основоположники. Пути развития.

рекомендации по подготовке к разным видам учебных занятий;

Организуя свою учебную работу, студенты должны:

Во-первых, выявить рекомендуемый режим и характер учебной работы по изучению теоретического курса, практическому применению изученного материала, по выполнению заданий для самостоятельной работы, по использованию информационных технологий и т.д.

Во-вторых, ознакомиться с указанным в методическом материале по дисциплине (модулю) перечнем учебно-методических изданий, рекомендуемых студентам для подготовки к занятиям и выполнения самостоятельной работы, а также с методическими материалами на бумажных и электронных носителях, выпущенных кафедрой своими силами и предоставляемые студентам во время занятий.

Самостоятельная работа студентов, предусмотренная учебным планом должна соответствовать более глубокому усвоению изучаемого курса, формировать навыки исследовательской работы и ориентировать студентов на умение применять теоретические знания на практике.

1. Работа с учебными пособиями. Для полноценного усвоения курса студент должен, прежде всего, овладеть основными понятиями этой дисциплины. Необходимо усвоить определения и понятия, уметь приводить их точные формулировки, приводить примеры объектов, удовлетворяющих этому определению. Кроме того, необходимо знать круг фактов, связанных с данным понятием. Требуется также знать связи между понятиями, уметь устанавливать соотношения между классами объектов, описываемых различными понятиями.

2. Подготовка к самостоятельным занятиям. При подготовке к занятиям студентам рекомендуется следовать методическим рекомендациям по работе с учебными пособиями, приведенным выше. Самостоятельная работа студентов должна включать усвоение теоретического материала, подготовку к практическим занятиям, выполнение творческих заданий, работу с электронным учебно-методическим комплексом, подготовку к текущему контролю знаний, к промежуточной аттестации (зачету).

3. Подготовка к зачету. При подготовке к зачету студенты должны использовать как самостоятельно подготовленные конспекты, так и материалы, полученные в ходе лекций.

Для успешного освоения дисциплины важны следующие рекомендации: На первой лекции важно обратить внимание на конкретные требования к прохождению и сдаче курса. На данном курсе самостоятельные занятия являются самым важным компонентом обучающего процесса. На занятиях будет представлен необходимый теоретический материал по темам и представлены практические задания для самостоятельного выполнения.

требования к рейтинг-контролю с указанием баллов, распределенных между модулями и видами работы обучающихся

Процедура оценивания знаний, умений, владений и (или) опыта деятельности обучающихся по дисциплине производится в рамках балльно-рейтинговой системы, включая текущую и промежуточную аттестации.

Согласно подходам балльно-рейтинговой системы в рамках оценки знаний, умений, владений (умений применять) и (или) опыта деятельности дисциплины (модуля) установлены следующие аспекты:

Содержание учебной дисциплины в рамках одного семестра делится на два модуля (периода обучения). По окончании модуля (периода обучения) осуществляется рейтинговый контроль успеваемости знаний студентов.

Сроки проведения рейтингового контроля:

Рейтинговый контроль успеваемости проводится на 32-33 учебной неделе по графику учебного процесса, II рейтинговый контроль успеваемости - две последние недели фактического завершения семестра по графику учебного процесса;

Согласно Положению о рейтинговой системе оценки качества работы студентов ТвГУ, качество усвоения учебной дисциплины оценивается по 100-балльной шкале. Обучающемуся, набравшему 40 баллов и выше по итогам работы в семестре, в рейтинговой ведомости учета успеваемости и зачетной книжке выставляется отметка «зачтено». Обучающийся, набравший до 39 баллов включительно, сдает зачет.

I модуль – 50 баллов

II модуль – 50 баллов

Зачет

	Текущая работа студентов	Количество баллов
1-ая контрольная точка.		
1.	Выполнение самостоятельной работы	40 б.
2.	Тест 1 модуля	10 б.
	Всего:	50 б.
2-ая контрольная точка.		
1	Выполнение самостоятельной работы	40 б.

2	Письменный опрос	10 б.
	Всего:	50 б.

VII. Материально-техническое обеспечение

Лекционный зал, оборудованный современной презентационной техникой (проектор, экран, ноутбук). Компьютерный класс с локальной вычислительной сетью и возможностью выхода в Интернет.

Наименование специальных помещений	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
Помещение для самостоятельной работы обучающихся: Компьютерный класс. Учебная аудитория № 28(Л, П) (170002, Тверская область, Тверь, просп. Чайковского, д.70)	1. Компьютер Триолит. Монитор ЛОС1 Компьютер Norbel. Монитор BENQ 2. Компьютер Norbel. Монитор BENQ 3. Компьютер Norbel. Монитор BENQ 4. Компьютер Norbel. Монитор BENQ 5. Компьютер Norbel. Монитор BENQ 6. Компьютер Norbel. Монитор BENQ 7. Компьютер Norbel. Монитор BENQ 8. Компьютер Norbel. Монитор BENQ 9. Компьютер Norbel. Монитор BENQ 10. Компьютер Norbel. Монитор BENQ	Adobe Acrobat Reader DC – бесплатно GIMP 2.6.12-2 – бесплатно GlassFish Server Open Source Edition 4.1.1 – бесплатно Google Chrome – бесплатно Kaspersky Endpoint Security 10 для Windows – Акт на передачу прав №2129 от 25 октября 2016 г. Microsoft Office профессиональный плюс 2013 – Акт приема-передачи № 369 от 21 июля 2017 Microsoft Windows 10 Enterprise – Акт приема-передачи № 369 от 21 июля 2017 Net Beans IDE – бесплатно SmartGit – бесплатно WinDjView 2.0.2 - бесплатно
Учебная аудитория: 34 (170002, Тверская область, Тверь, просп. Чайковского, д.70)	Меловая доска, комплект учебной мебели, переносной ноутбук с возможностью подключиться к Интернет, проектор, сеть WiFi	

VIII. Сведения об обновлении рабочей программы дисциплины

№п.п.	Обновленный раздел рабочей программы дисциплины	Описание внесенных изменений	Реквизиты документа, утвердившего изменения
1.			
2.			