

Министерство науки и высшего образования Российской Федерации  
НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ (НИ ТГУ)

Филологический факультет

УТВЕРЖДЕНО:  
Декан  
И. В. Тубалова

Рабочая программа дисциплины

**Введение в анализ естественного языка (NLP)**

по направлению подготовки

**45.04.03 Фундаментальная и прикладная лингвистика**

Направленность (профиль) подготовки:  
**Компьютерная и когнитивная лингвистика**

Форма обучения  
**Очная**

Квалификация  
**Магистр**

Год приема  
**2024**

СОГЛАСОВАНО:  
Руководитель ОП  
З.И. Резанова

Председатель УМК  
Ю.А. Тихомирова

Томск – 2025

## **1. Цель и планируемые результаты освоения дисциплины**

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-3 Способен выбирать оптимальные подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.

ОПК-4 Способен расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук.

ОПК-6 Способен осуществлять эффективное управление разработкой программных средств информационных проектов в сфере своей профессиональной деятельности.

ПК-1 Способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований.

ПК-4 Способен разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.

Результатами освоения дисциплины являются следующие индикаторы достижения компетенций:

ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий

ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов

ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования

ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере

ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта

ИПК-4.2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта

## **2. Задачи освоения дисциплины**

– получение студентом знаний об основных теоретических и прикладных направлениях науки в области компьютерной лингвистики, основные понятий в области компьютерной обработки естественного языка и методов машинного обучения, применяемых в данной области

– получение студентом навыков критического сопоставления существующих подходов и методов решения конкретных научных и прикладных задач, выбора математических и лингвистических методов решения профессиональных задач с применением языков программирования

– получение студентом умений формулирования целей проекта в области компьютерной лингвистики и разработки соответствующей программы действий.

## **3. Место дисциплины в структуре образовательной программы**

Дисциплина относится к Блоку 1 «Дисциплины (модули)».

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений, предлагается обучающимся на выбор.

#### **4. Семестр(ы) освоения и форма(ы) промежуточной аттестации по дисциплине**

Первый семестр, экзамен

#### **5. Входные требования для освоения дисциплины**

Для успешного освоения дисциплины требуются компетенции, сформированные в ходе освоения образовательных программ предшествующего уровня образования.

#### **6. Язык реализации**

Русский

#### **7. Объем дисциплины**

Общая трудоемкость дисциплины составляет 4 з.е., 144 часов, из которых:

-лекции: 10 ч.

-практические занятия: 32 ч.

в том числе практическая подготовка: 32 ч.

Объем самостоятельной работы студента определен учебным планом.

#### **8. Содержание дисциплины, структурированное по темам**

Тема 1. Введение в проблематику обработки языка. Теоретические проблемы, соотношение NLP и лингвистики, основные источники. N-граммы и моделирование языка. Векторная семантика

Тема 2. Современное состояние рынка вакансий в области NLP. Вакансии в области обработки языковых данных: требования и зарплаты

Тема 3. Машинное обучение в применении к обработке естественного языка. Naive Bayes - наивный байесовский классификатор. k-nearest neighbor, k-means clustering - k ближайших соседей, кластеризация по k-средним. Hidden Markov Models - скрытые марковские модели. Decision trees, Random Forest - дерево решений, случайный лес. Support Vector Machine - метод опорных векторов. Principal Component Analysis - метод главных компонент. Artificial Neural Networks - нейронные сети. Logistic Regression, можно больше о регрессии. Gradient Boosting и подобные

Тема 4. Сентимент-анализ. Современное состояние поля, реализации Bag-of-Words и более продвинутых подходов. Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных

Тема 5. Виртуальные ассистенты и чат-боты

Тема 6. Машинный перевод: от истоков к трансформерам, не забывая про конечного пользователя

Тема 7. Speech Recognition: фонетическая структура языка, акустика речи, восприятие и производство речи. Машинное обучение и распознавание речи.

Тема 8. Этика искусственного интеллекта. Этика искусственного интеллекта: проблемы, подходы, решения. Скандалный случай из области больших данных: краткий доклад

## 9. Текущий контроль по дисциплине

Текущий контроль по дисциплине проводится путем контроля посещаемости, проведения контрольных работ, выполнения домашних заданий и фиксируется в форме контрольной точки не менее одного раза в семестр.

**Таблица 1. Критерии и уровни освоения компетенции.**

**Компетенция** Способен выбирать оптимальные подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	допороговый	<i>Не может критически сопоставлять и оценивать существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.</i>	2
	пороговый	<i>Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий на минимальном уровне</i>	3
	достаточный	<i>Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий на достаточном уровне</i>	4
	продвинутый	<i>Полно, быстро и эффективно критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.</i>	5

**Компетенция** Способен расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания

Инд-р Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.	1. допорогово ый	<i>Не демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</i>	2
	порогово й	<i>Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов в минимальном объеме</i>	3
	достаточ ный	<i>Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов в достаточном объеме</i>	4
	продвину тый	<i>Полно, быстро и эффективно демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</i>	5

**Компетенция** Способен осуществлять эффективное управление разработкой программных средств и информационных проектов в сфере своей профессиональной деятельности

Индикатор	Уро вни освоения	Описание уровня освоения	Шкала оценивания
Инд-р Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	1. допорогово ый	<i>Не выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.</i>	2
	порогово й	<i>Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования. Аргументация осуществляется в минимальном объеме.</i>	3

	достаточный	<i>Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования. Аргументация осуществляется в достаточном объеме.</i>	4
	продвинутый	<i>Полно, быстро и эффективно аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.</i>	5

**Компетенция** Способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	допороговый	<i>Не обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.</i>	2
	пороговый	<i>Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере в минимальном объеме.</i>	3
	достаточный	<i>Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере в достаточном объеме</i>	4
	продвинутый	<i>Полно, быстро и эффективно обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.</i>	5

**Компетенция** Способность разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных

технических средств и информационных технологий, в том числе в области искусственного интеллекта

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.	допороговый	<i>Не формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</i>	2
	пороговый	<i>Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта в минимальном объеме.</i>	3
	достаточный	<i>Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта в достаточном объеме.</i>	4
	продвинутый	<i>Полно, быстро и эффективно формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</i>	5
Инд-р 2 Разрабатывает программу действий по решению задач проекта в	допороговый	<i>Не разрабатывает программу действий по решению задач проекта в области когнитивной</i>	2

области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта		<i>и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта</i>	
	пороговый	<i>Минимально разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта</i>	3
	достаточный	<i>В достаточном объеме разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта</i>	4

**Таблица 2. Индикаторы формирования компетенции и оценочные средства.**  
**Компетенция** Способен выбирать оптимальные подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий

Содержание	Индикатор компетенции	Оценочные средства
Тема 1. Введение в проблематику обработки языка. Теоретические проблемы, соотношение NLP и лингвистики, основные источники. N-граммы и моделирование языка. Векторная семантика	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Сообщение Контрольная работа Эссе

Тема 2. Современное состояние рынка вакансий в области NLP. Вакансии в области обработки языковых данных: требования и зарплаты	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Письменная работа
Тема 3. Машинное обучение в применении к обработке естественного языка. Naive Bayes - наивный байесовский классификатор. k-nearest neighbor, k-means clustering - k ближайших соседей, кластеризация по k-средним. Hidden Markov Models - скрытые марковские модели. Decision trees, Random Forest - дерево решений, случайный лес. Support Vector Machine - метод опорных векторов. Principal Component Analysis - метод главных компонент. Artificial Neural Networks - нейронные сети. Logistic Regression, можно больше о регрессии. Gradient Boosting и подобные	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Сообщение Контрольная работа Эссе
Тема 4. Сентимент-анализ. Современное состояние поля, реализации Bag-of-Words и более продвинутых подходов. Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Письменная работа
Тема 5. Виртуальные ассистенты и чат-боты	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Сообщение Контрольная работа Эссе

Тема 6. Машинный перевод: от истоков к трансформерам, не забывая про конечного пользователя	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Письменная работа
Тема 7. Speech Recognition: фонетическая структура языка, акустика речи, восприятие и производство речи. Машинное обучение и распознавание речи.	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Сообщение Контрольная работа Эссе
Тема 8. Этика искусственного интеллекта. Этика искусственного интеллекта: проблемы, подходы, решения. Скандалный случай из области больших данных: краткий доклад	ИОПК-3.2 Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Письменная работа

**Компетенция** Способен расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук

Содержание	Индикатор компетенции	Оценочные средства
Тема 1. Введение в проблематику обработки языка. Теоретические проблемы, соотношение NLP и лингвистики, основные источники. N-граммы и моделирование языка. Векторная семантика	ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.	Сообщение Контрольная работа Эссе
Тема 2. Современное состояние рынка вакансий в области NLP. Вакансии в области обработки языковых данных: требования и зарплаты	ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.	Письменная работа

<p>Тема 3. Машинное обучение в применении к обработке естественного языка. Naive Bayes - наивный байесовский классификатор. k-nearest neighbor, k-means clustering - k ближайших соседей, кластеризация по k-средним. Hidden Markov Models - скрытые марковские модели. Decision trees, Random Forest - дерево решений, случайный лес. Support Vector Machine - метод опорных векторов. Principal Component Analysis - метод главных компонент. Artificial Neural Networks - нейронные сети. Logistic Regression, можно больше о регрессии. Gradient Boosting и подобные</p>	<p>ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</p>	<p>Сообщение Контрольная работа Эссе</p>
<p>Тема 4. Сентимент-анализ. Современное состояние поля, реализации Bag-of-Words и более продвинутых подходов. Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных</p>	<p>ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</p>	<p>Письменная работа</p>
<p>Тема 5. Виртуальные ассистенты и чат-боты</p>	<p>ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</p>	<p>Сообщение Контрольная работа Эссе</p>
<p>Тема 6. Машинный перевод: от истоков к трансформерам, не забывая про конечного пользователя</p>	<p>ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</p>	<p>Письменная работа</p>
<p>Тема 7. Speech Recognition: фонетическая структура языка, акустика речи, восприятие и производство речи. Машинное обучение и распознавание речи.</p>	<p>ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</p>	<p>Сообщение Контрольная работа Эссе</p>

Тема 8. Этика искусственного интеллекта. Этика искусственного интеллекта: проблемы, подходы, решения. Скандалный случай из области больших данных: краткий доклад	ИОПК-4.1 Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.	Письменная работа

**Компетенция** Способен осуществлять эффективное управление разработкой программных средств и информационных проектов в сфере своей профессиональной деятельности

Содержание	Индикатор компетенции	Оценочные средства
Тема 1. Введение в проблематику обработки языка. Теоретические проблемы, соотношение NLP и лингвистики, основные источники. N-граммы и моделирование языка. Векторная семантика	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Сообщение Контрольная работа Эссе
Тема 2. Современное состояние рынка вакансий в области NLP. Вакансии в области обработки языковых данных: требования и зарплаты	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Письменная работа
Тема 3. Машинное обучение в применении к обработке естественного языка. Naive Bayes - наивный байесовский классификатор. k-nearest neighbor, k-means clustering - k ближайших соседей, кластеризация по k-средним. Hidden Markov Models - скрытые марковские модели. Decision trees, Random Forest - дерево решений, случайный лес. Support Vector Machine - метод опорных векторов. Principal Component Analysis - метод главных компонент. Artificial Neural Networks - нейронные сети. Logistic Regression, можно больше о регрессии. Gradient Boosting и подобные	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Сообщение Контрольная работа Эссе
Тема 4. Сентимент-анализ.	ИОПК-6.1	Письменная работа

Современное состояние поля, реализации Bag-of-Words и более продвинутых подходов. Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных	Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	нная работа
Тема 5. Виртуальные ассистенты и чат-боты	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Сообщение Контрольная работа Эссе
Тема 6. Машинный перевод: от истоков к трансформерам, не забывая про конечного пользователя	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Письменная работа
Тема 7. Speech Recognition: фонетическая структура языка, акустика речи, восприятие и производство речи. Машинное обучение и распознавание речи.	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Сообщение Контрольная работа Эссе
Тема 8. Этика искусственного интеллекта. Этика искусственного интеллекта: проблемы, подходы, решения. Скандалный случай из области больших данных: краткий доклад	ИОПК-6.1 Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Письменная работа

**Компетенция** Способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований

Содержание	Индикатор компетенции	Оценочные средства
Тема 1. Введение в проблематику обработки языка. Теоретические проблемы, соотношение NLP и лингвистики, основные источники. N-граммы и моделирование языка. Векторная семантика	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Сообщение Контрольная работа Эссе

Тема 2. Современное состояние рынка вакансий в области NLP. Вакансии в области обработки языковых данных: требования и зарплаты	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Письменная работа
Тема 3. Машинное обучение в применении к обработке естественного языка. Naive Bayes - наивный байесовский классификатор. k-nearest neighbor, k-means clustering - k ближайших соседей, кластеризация по k-средним. Hidden Markov Models - скрытые марковские модели. Decision trees, Random Forest - дерево решений, случайный лес. Support Vector Machine - метод опорных векторов. Principal Component Analysis - метод главных компонент. Artificial Neural Networks - нейронные сети. Logistic Regression, можно больше о регрессии. Gradient Boosting и подобные	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Сообщение Контрольная работа Эссе
Тема 4. Сентимент-анализ. Современное состояние поля, реализации Bag-of-Words и более продвинутых подходов. Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Письменная работа
Тема 5. Виртуальные ассистенты и чат-боты	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Сообщение Контрольная работа Эссе
Тема 6. Машинный	ИПК-1.1 Обнаруживает	Письменная работа

перевод: от истоков к трансформерам, не забывая про конечного пользователя	знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	нная работа
Тема 7. Speech Recognition: фонетическая структура языка, акустика речи, восприятие и производство речи. Машинное обучение и распознавание речи.	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Сообщение Контрольная работа Эссе
Тема 8. Этика искусственного интеллекта. Этика искусственного интеллекта: проблемы, подходы, решения. Скандалный случай из области больших данных: краткий доклад	ИПК-1.1 Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Письменная работа

**Компетенция** Способность разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта

Содержание	Индикатор компетенции	Оценочные средства
Тема 1. Введение в проблематику обработки языка. Теоретические проблемы, соотношение NLP и лингвистики, основные источники. N-граммы и моделирование языка. Векторная семантика	ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.	Сообщение Контрольная работа Эссе
Тема 2. Современное состояние рынка вакансий в области NLP. Вакансии в области обработки языковых данных: требования и зарплаты	ИПК-4.2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта.	Письменная работа
Тема 3. Машинное обучение в применении к обработке естественного языка. Naive Bayes - наивный байесовский классификатор. k-nearest neighbor, k-means clustering	ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения	Сообщение Контрольная работа Эссе

<p>- k ближайших соседей, кластеризация по k-средним. Hidden Markov Models - скрытые марковские модели. Decision trees, Random Forest - дерево решений, случайный лес. Support Vector Machine - метод опорных векторов. Principal Component Analysis - метод главных компонент. Artificial Neural Networks - нейронные сети. Logistic Regression, можно больше о регрессии. Gradient Boosting и подобные</p>	<p>современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</p>	
<p>Тема 4. Сентимент-анализ. Современное состояние поля, реализации Bag-of-Words и более продвинутых подходов. Решения в области, проприетарные и открытые; продукты для мониторинга соцсетей; использование метаданных</p>	<p>ИПК-4.2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта.</p>	<p>Письменная работа</p>
<p>Тема 5. Виртуальные ассистенты и чат-боты</p>	<p>ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</p>	<p>Сообщение Контрольная работа Эссе</p>
<p>Тема 6. Машинный перевод: от истоков к трансформерам, не забывая про конечного пользователя</p>	<p>ИПК-4.2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта.</p>	<p>Письменная работа</p>
<p>Тема 7. Speech Recognition: фонетическая структура языка, акустика речи, восприятие и производство речи. Машинное обучение и распознавание речи.</p>	<p>ИПК-4.1 Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</p>	<p>Сообщение Контрольная работа Эссе</p>

<p>Тема 8. Этика искусственного интеллекта. Этика искусственного интеллекта: проблемы, подходы, решения. Скандалный случай из области больших данных: краткий доклад</p>	<p>ИПК-4.2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта.</p>	<p>Письменная работа</p>
--	--	--------------------------

### Приложение к Таблице 2.

Примерные темы сообщений.

Искусственный и естественный интеллект.

Истоки науки об обработке естественного языка. Роль кибернетики и искусственного интеллекта в ее становлении.

Когнитивная наука и проблемы искусственного интеллекта.

Примерные темы письменных работ.

Модели семантики в компьютерном анализе языка.

Когнитивные основы искусственного интеллекта.

Компьютерные модели когнитивных структур.

Междисциплинарность – основа подхода к профессиональному интеллекту.

Примерные темы рефератов.

«Когнитивная лингвистика и искусственный интеллект»

«Автоматическое выявление образного языка»

«Неоднозначность как фундаментальная проблема языка»

Примерный тест.

1. К обработке естественного языка не относится следующее понятие:

А) сфера-источник

Б) вектор

В) униграммы

2. Термин «конверсационный агент лингвистика» появился в

А) 1960 году

Б) 1975 году

В) 1990 году

3. Какой из авторов являлся частым соавтором Д. Журафски?

А) Фабьен Гандон

Б) Джеймс Хендлер

В) Джеймс Дин

В полном объеме средства оценивания для текущей аттестации находятся в LMS IDO (<https://lms.tsu.ru/course/view.php?id=14711>).

Анализ и оценивание существующих исследований на предмет охвата материала, качества анализа и теоретических оснований. Представление критического анализа в устном выступлении на семинаре.

Написание эссе на темы, предложенные преподавателем

Планирование исследования: студенты получают практическое задание в области компьютерной лингвистики, например, «Чат-бот «Психолог» для студентов университета» и должны разработать план выполнения с конкретными источниками литературы, аналогичными проектами и шагами реализации.

Подготовка аннотированной биографии (10-15 источников) к одной из тем семинаров.

Индивидуальный/парный/групповой семинар: студентам предъявляется тема для мозгового штурма, например, «Какое приложение может иметь теория функциональной

грамматики в области семантических репрезентаций?». Даётся пять минут на индивидуальную подготовку, затем студенты объединяются в пары и выбирают три лучших идеи (следующие пять минут). После этого группа выбирает 3 лучших идеи и представляет их преподавателю.

Задания на определение ключевых понятий различных областей NLP.

Лабораторная работа по заполнению словаря, программированию словарных статей, корпусов различных типов, правке онтографии, извлечению знаний из интеллектуальной информационной системы, распознаванию речи с помощью доступных программных платформ.

Написанию отчета по сравнению результатов работы с различными системами, написанию заявки в техническую поддержку интеллектуальной информационной системы; написание предложений по улучшению и расширению функционала существующих систем.

Оценочные материалы текущего контроля размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

## **10. Порядок проведения и критерии оценивания промежуточной аттестации**

Экзамен в первом семестре проводится устной форме. Экзаменационное задание состоит из трех частей. Продолжительность экзамена 1,5 часа.

Первая часть работы проверяет компетенции ИПК-4.2, ИПК-4.1, ИПК-1.1. Вторая часть работы проверяет компетенции ИОПК-6.1, ИОПК-4.1, ИОПК-3.2.

**Таблица 1. Критерии и уровни освоения компетенции.**

**Компетенция** Способен выбирать оптимальные подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	допороговый	<i>Не может критически сопоставлять и оценивать существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.</i>	2
	пороговый	<i>Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий на минимальном уровне</i>	3
	достаточный	<i>Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области</i>	4

		<i>лингвистики и информационных технологий на достаточном уровне</i>	
	продвинутый	<i>Полно, быстро и эффективно критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.</i>	5

**Компетенция** Способен расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.	допороговый	<i>Не демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</i>	2
	пороговый	<i>Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов в минимальном объеме</i>	3
	достаточный	<i>Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов в достаточном объеме</i>	4
	продвинутый	<i>Полно, быстро и эффективно демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.</i>	5

**Компетенция** Способен осуществлять эффективное управление разработкой программных средств и информационных проектов в сфере своей профессиональной деятельности

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	1. допороговый	<i>Не выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.</i>	2
	пороговый	<i>Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования. Аргументация осуществляется в минимальном объеме.</i>	3
	достаточный	<i>Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования. Аргументация осуществляется в достаточном объеме.</i>	4
	продвинутый	<i>Полно, быстро и эффективно аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.</i>	5

**Компетенция** Способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	допороговый	<i>Не обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.</i>	2
	пороговый	<i>Обнаруживает знания об актуальных направлениях междисциплинарных</i>	3

		<i>лингвистических исследований в избранной научной сфере в минимальном объеме.</i>	
	достаточный	<i>Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере в достаточном объеме</i>	4
	продвинутый	<i>Полно, быстро и эффективно обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.</i>	5

**Компетенция** Способность разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных технических средств и информационных технологий, в том числе в области искусственного интеллекта

Индикатор	Уровни освоения	Описание уровня освоения	Шкала оценивания
Инд-р 1. Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.	допороговый	<i>Не формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</i>	2
	пороговый	<i>Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта в минимальном объеме.</i>	3
	достаточный	<i>Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость</i>	4

		<i>применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта в достаточном объеме.</i>	
	продвинутый	<i>Полно, быстро и эффективно формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.</i>	5
Инд-р 2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта	допороговый	<i>Не разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта</i>	2
	пороговый	<i>Минимально разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта</i>	3
	достаточный	<i>В достаточном объеме разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта</i>	4
	продвинутый	<i>Полно, быстро и эффективно разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических</i>	5

		<i>средств и информационных технологий, в том числе в области искусственного интеллекта</i>	
--	--	---	--

**Таблица 2. Индикаторы формирования компетенции и оценочные средства.**

**Компетенция** Способен выбирать оптимальные подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий

Индикатор компетенции	Оценочные средства
Инд-р 1. Критически сопоставляет и оценивает существующие подходы и методы решения конкретных научных и прикладных задач в области лингвистики и информационных технологий.	Устный вопрос по докладу Экзаменационная презентация

**Компетенция** Способен расширять сферу научной деятельности, участвовать в междисциплинарных исследованиях на стыке наук

Индикатор компетенции	Оценочные средства
Инд-р 1. Демонстрирует знание новых теорий в сфере междисциплинарного взаимодействия лингвистики и наук гуманитарного, математического и естественно-научного циклов.	Устный вопрос по докладу Экзаменационная презентация

**Компетенция** Способен осуществлять эффективное управление разработкой программных средств и информационных проектов в сфере своей профессиональной деятельности

Индикатор компетенции	Оценочные средства
Инд-р 1. Аргументированно выбирает математические и лингвистические методы решения профессиональных задач с применением языков программирования.	Устный вопрос по докладу Экзаменационная презентация

**Компетенция** Способен проводить самостоятельные исследования и получать новые научные результаты в области междисциплинарных лингвистических исследований

Индикатор компетенции	Оценочные средства
Инд-р 1. Обнаруживает знания об актуальных направлениях междисциплинарных лингвистических исследований в избранной научной сфере.	Устный вопрос по докладу Экзаменационная презентация

**Компетенция** Способность разрабатывать проекты прикладной направленности в области когнитивной и компьютерной лингвистики с применением современных

технических средств и информационных технологий, в том числе в области искусственного интеллекта

Индикатор компетенции	Оценочные средства
Инд-р 1. Формулирует цель проекта прикладной направленности в области когнитивной и компьютерной лингвистики, обосновывает необходимость применения современных технических средств и информационных технологий, в том числе в области искусственного интеллекта.	Устный вопрос по докладу Экзаменационная презентация
Инд-р 2 Разрабатывает программу действий по решению задач проекта в области когнитивной и компьютерной лингвистики с учетом имеющихся технических средств и информационных технологий, в том числе в области искусственного интеллекта	Устный вопрос по докладу Экзаменационная презентация

## Приложение к Таблице 2

Примерные задания к экзамену.

Промежуточная аттестация и ее оценивание осуществляются следующим образом.

Экзамен состоит в выполнении практического задания по позиционированию своего магистерского исследования в рамках поля NLP. Работа выполняется письменно и оформляется согласно требованиям, предъявляемым к курсовым работам. Объем работы – до 15 стр. А4. Защита работы происходит на экзамене в форме презентации и ответа на вопросы экзаменатора и присутствующих.

Работа и презентация может быть оценена, если она соответствует следующим критериям.

Работа демонстрирует наличие у студента знаний о поле NLP, основных и специальных методах и приемах анализа естественного языка. Магистрант должен проявить способность позиционировать свою работу в контексте современной практики обработки естественного языка. Работа должна излагать суть планируемого проекта магистранта и его место в современной парадигме NLP. Приветствуется наличие детализированного плана разработки в соотнесении с необходимыми для его реализации технологиями.

Работа должна отвечать специфике научного стиля, быть ясно изложена. Задание, выполненное студентом, должно демонстрировать знание принципов создания и презентации научного текста в форме плана проекта, реферата, плана исследования, научного эссе, их структурирования и оформления. Особое внимание уделяется корректности оформления текста, иллюстраций и ссылок по существующему ГОСТу.

Требования к презентации:

1. успешное удержание внимания на речи и презентации докладчика;
2. адекватное оформление презентационных материалов по времени, дизайну и структуре;
3. разделение поданной информации на главную и второстепенную;
4. умение адаптировать презентацию к нуждам аудитории;
5. умение корректно (в рамках научной дискуссии) отвечать на поставленные вопросы аудитории и сделанные замечания.

Оценка «удовлетворительно» может быть поставлена при выполнении 60 процентов самостоятельной работы в семестре, выполнении проектной работы, фрагментарно описывающей магистерскую проектную работу, отражающей недостаточное знание поля NLP, технологий обработки языка и современного состояния субдисциплин. Презентация работы выполнена с нарушением критериев, ответы на вопросы даны некорректно или не по существу.

Оценка «хорошо» может быть поставлена при выполнении 80 процентов самостоятельной работы в семестре, выполнении проектной работы, в целом описывающей магистерскую проектную работу, отражающей знание поля NLP, основных технологий обработки языка и современного состояния субдисциплин. Презентация работы выполнена согласно критериям, возможны недочеты в изложении содержания работы, затруднения при ответах на вопросы.

Оценка «отлично» может быть поставлена при выполнении 95-100 процентов самостоятельной работы в семестре, выполнении проектной работы, исчерпывающе описывающей магистерскую проектную работу, отражающей полноценное знание поля NLP, технологий обработки языка и современного состояния субдисциплин. Презентация полностью отвечает всем критериям, ответы на вопросы экзаменатора и присутствующих демонстрируют знание специфики релевантной области NLP.

В полном объеме средства оценивания для промежуточной аттестации находятся в LMS IDO (<https://lms.tsu.ru/course/view.php?id=14711>).

Оценочные материалы для проведения промежуточной аттестации размещены на сайте ТГУ в разделе «Информация об образовательной программе» - <https://www.tsu.ru/sveden/education/eduop/>.

## **11. Учебно-методическое обеспечение**

а) Электронный учебный курс по дисциплине в электронном университете «LMS IDO» - <https://lms.tsu.ru/course/view.php?id=14711>

б) Оценочные материалы текущего контроля и промежуточной аттестации по дисциплине представлены в курсе «LMS IDO»

в) План практических занятий по дисциплине соответствует п. 8 (лекции не предусмотрены).

г) Методические указания по организации самостоятельной работы студентов.

Самостоятельная работа студентов включает чтение текстов из списка литературы, подготовку к лекциям по технологии перевернутого класса. Подробные методические указания представлены в курсе «LMS IDO» - <https://lms.tsu.ru/mod/page/view.php?id=199720>.

## **12. Перечень учебной литературы и ресурсов сети Интернет**

а) основная литература:

1. Jurafsky D., Martin J.H. Speech and Language Processing (3rd ed. draft). – URL: <https://web.stanford.edu/~jurafsky/slp3/>

б) дополнительная литература:

1. Molnar, Christoph. Interpretable machine learning. Lulu. com, 2020.
2. Coeckelbergh, Mark. AI ethics. MIT Press, 2020.
3. Franks, Bill. 97 Things About Ethics Everyone in Data Science Should Know. O'Reilly Media, 2020.
4. Handbook of Natural Language Processing. / Eds. Nitin Indurkhya, Fred J. Damerau. – 2nd ed. — Chapman & Hall/CRC, 2010. – 692 p.
5. The handbook of computational linguistics and natural language processing / Eds Alexander Clark, Chris Fox, Shalom Lappin. – Wiley-Blackwell, 2010. – 801 p.
6. Rogers, Simon, and Mark Girolami. A first course in machine learning. Chapman and Hall/CRC, 2016.
7. Гласснер Э. Глубокое обучение без математики. Т. 1: Основы / пер. с анг. В. А. Яроцкого. – М.: ДМК Пресс, 2019. – 584 с.: ил
8. Liu, Bing. Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge university press, 2020.

9. Danneman, Nathan, and Richard Heimann. Social media mining with R. Packt Publishing Ltd, 2014.
10. 97 Things Every Data Engineer Should Know. Tobias Macey 2021 O'Reilly Media, Inc.. 256 p.

б) ресурсы сети Интернет:

1. ACL Anthology. URL: <http://aclweb.org/anthology/>, <http://aclanthology.info/>
2. Behavior Research Methods. URL: <https://link.springer.com/journal/13428>
3. Computational Linguistics. URL: <http://www.mitpressjournals.org/loi/coli>
4. Computer Speech and Language. URL: <https://www.journals.elsevier.com/computer-speech-and-language>
5. International Journal of Corpus Linguistics. URL: <https://benjamins.com/#catalog/journals/ijcl/main>
6. Journal of Information Retrieval. URL: <http://www.springer.com/computer/database+management+%26+information+retrieval/journal/10791>
7. Journal of Machine Learning. URL: <http://www.springer.com/computer/ai/journal/10994>
8. Language and Linguistics Compass. URL: [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1749-818X](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1749-818X)
9. Language Resources and Evaluation. URL: <https://link.springer.com/journal/10579>
10. Machine Translation. URL: <https://link.springer.com/journal/10590>
11. Natural Language Semantics. URL: <https://link.springer.com/journal/11050>
12. Transactions of the Association for Computational Linguistics. URL: <https://www.transacl.org/ojs/index.php/tacl/issue/view/13>
13. Диалог. URL: <http://www.dialog-21.ru/digest/>

### **13. Перечень информационных технологий**

а) лицензионное и свободно распространяемое программное обеспечение:

- Microsoft Office 2019 Russian: пакет программ. Включает приложения: MS Office Word, MS Office Excel, MS Office PowerPoint, MS Office On-eNote, MS Office Publisher, MS Outlook, MS Office Web Apps (Word Excel MS PowerPoint Outlook);
- публично доступные облачные технологии (Google Docs, Яндекс диск и т.п.).

б) информационные справочные системы:

- Электронный каталог Научной библиотеки ТГУ –  
<http://chamo.lib.tsu.ru/search/query?locale=ru&theme=system>
- Электронная библиотека (репозиторий) ТГУ –  
<http://vital.lib.tsu.ru/vital/access/manager/Index>
- ЭБС Лань – <http://e.lanbook.com/>
- ЭБС Консультант студента – <http://www.studentlibrary.ru/>
- Образовательная платформа Юрайт – <https://urait.ru/>
- ЭБС ZNANIUM.com – <https://znanium.com/>
- ЭБС IPRbooks – <http://www.iprbookshop.ru/>

### **14. Материально-техническое обеспечение**

Аудитории для проведения занятий лекционного типа.

Аудитории для проведения занятий семинарского типа, индивидуальных и групповых консультаций, текущего контроля и промежуточной аттестации.

Помещения для самостоятельной работы, оснащенные компьютерной техникой и доступом к сети Интернет, в электронную информационно-образовательную среду и к информационным справочным системам.

## **15. Информация о разработчиках**

Шиляев Константин Сергеевич, к. филол. н., доцент, кафедра общей, компьютерной и когнитивной лингвистики