



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**Институт филологии, иностранных языков и медиакоммуникации**

**Факультет теоретической и прикладной филологии**

**Кафедра русского языка и общего языкознания**



**Рабочая программа дисциплины**

**Б1.В.15 Введение в системный анализ и машинное обучение**

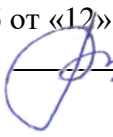
Направление подготовки **45.04.01 Филология**

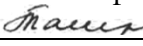
Направленность (профиль) подготовки

**«Современная филология: аналитика текстовых данных»**

Квалификация (степень) выпускника – **магистр**

Форма обучения – **очная**

Согласовано с УМК ИФИЯМ  
Протокол № 6 от «12» февраля 2024 г.  
Председатель  О.Л. Михалёва

Рекомендовано кафедрой  
русского языка и общего языкознания  
Протокол № 5 от «08» февраля 2024 г.  
Зав. кафедрой  М.Б. Ташлыкова

Иркутск 2024 г.

## Содержание

I.		Цели и задачи дисциплины	3
II.		Место дисциплины в структуре ОПОП	3
III.		Требования к результатам освоения дисциплины	3
IV.		Содержание и структура дисциплины	5
	4.1.	Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов	
	4.2.	План внеаудиторной самостоятельной работы обучающихся по дисциплине	6
	4.3.	Содержание учебного материала	7
	4.3.1.	Перечень семинарских, практических занятий и лабораторных работ	10
	4.3.2.	Перечень тем (вопросов), выносимых на самостоятельное изучение в рамках самостоятельной работы студентов	11
	4.4.	Методические указания по организации самостоятельной работы студентов	12
V.		Учебно-методическое и информационное обеспечение дисциплины	12
	а)	перечень литературы	
	б)	периодические издания	
	в)	список авторских методических разработок	
	г)	базы данных, поисково-справочные и информационные системы	
VI.		Материально-техническое обеспечение дисциплины	14
	6.1.	Учебно-лабораторное оборудование	
	6.2.	Программное обеспечение	
	6.3.	Технические и электронные средства обучения	
VII.		Образовательные технологии	15
VIII.		Оценочные материалы для текущего контроля и промежуточной аттестации	16

## **I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ**

**Цель** дисциплины – обучение теории и практике автоматической обработки текстов на естественном языке, формирование представлений об основных методах машинного обучения, применяемых в филологии.

**К задачам** дисциплины относится

- формирование у студентов представлений о принципах и приемах предобработки текстов, извлечения признаков для анализа текстовых данных;
- знакомство с основными методами валидации данных и оценки качества алгоритмов;
- получение навыков работы с библиотеками для машинного обучения и обработки текстов, применения современных методов машинного обучения для решения задач в области филологии.

## **II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП**

**2.1.** Учебная дисциплина «Введение в системный анализ и машинное обучение» относится к базовой части учебного плана.

Место курса в системе подготовки магистров определяется, с одной стороны, его прикладной направленностью, с другой – тесной связью с фундаментальными лингвистическими дисциплинами, знание которых обеспечивает студента необходимым инструментарием для решения задач практического характера.

**2.2.** Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами:

Б1.О.06 Теория текста

Б1.В.08 Устная и письменная речь как система и структура

Б1.В.16 Языки программирования и анализ данных

Б1.В.17 Автоматическая обработка текстов на естественном языке

Б1.В.13 Межличностная коммуникация: способы речевого взаимодействия

## **III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ**

Процесс изучения дисциплины направлен на формирование следующих компетенций:

- ПК-1 – способен находить, отбирать и систематизировать, творчески и критически осмысливать информацию для решения научно-исследовательских и практических задач в сфере профессиональной деятельности; способен ставить и решать на основе имеющихся данных прикладные задачи, связанные с использованием языка и литературы в процессах образования, устной и письменной коммуникации.

**Перечень планируемых результатов обучения,  
соотнесенных с индикаторами достижения компетенций**

<b>Компетенция</b>	<b>Индикаторы компетенций</b>	<b>Результаты обучения</b>
<p><i>ПК-1</i> способен находить, отбирать и систематизировать, творчески и критически осмысливать информацию для решения научно-исследовательских и практических задач в сфере профессиональной деятельности; способен ставить и решать на основе имеющихся данных прикладные задачи, связанные с использованием языка и литературы в процессах образования, устной и письменной коммуникации.</p>	<p><i>ИДК ПК 1.1.</i> Отбирает и систематизирует филологическую информацию в рамках изучаемых дисциплин, демонстрирует ее творческое и критическое осмысление для решения теоретических (научно-исследовательских) и практических (проектных) задач в сфере профессиональной деятельности.</p>	<p><b>Знать:</b></p> <ul style="list-style-type: none"> <li>• принципы и приёмы автоматической обработки текстов на естественном языке.</li> </ul> <p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>• осуществлять предобработку языковых данных.</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>• навыками использования обработки языка в задачах информационного поиска;</li> <li>• навыками обработки языкового материала с помощью библиотек, предназначенных для машинного обучения.</li> </ul>
	<p><i>ИДК ПК 1.2.</i> Решает поставленные на основе имеющихся данных прикладные задачи, связанные с использованием языка и литературы в процессах образования, устной и письменной коммуникации.</p>	<p><b>Знать:</b></p> <ul style="list-style-type: none"> <li>• основные методы валидации данных и оценки качества алгоритмов.</li> </ul> <p><b>Уметь:</b></p> <ul style="list-style-type: none"> <li>• применять методы современных методов машинного обучения в рамках собственных исследований.</li> </ul> <p><b>Владеть:</b></p> <ul style="list-style-type: none"> <li>• навыками интерпретации и представления полученных данных.</li> </ul>

#### IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов. Форма промежуточной аттестации – экзамен.

##### 4.1. Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

№ п/п	Раздел дисциплины/темы	курс	Всего часов	Из них – практическая подготовка обучающихся	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)			Самостоятельная работа	Формы текущего контроля успеваемости
					Контактная работа преподавателя с обучающимися				
					Лекции	Семинарские (практические занятия)	Консультации		
1	Предобработка и визуализация данных	2	9		2	2		5	проверочная работа
2	Библиотеки для машинного обучения. Выделение признаков.	2	17		6	4		7	устный опрос, проверочная работа
3	Векторное представление слов в обработке текстов	2	17		4	6		7	устный опрос, проверочная работа
4	Методы снижения размерности	2	15		4	4		7	устный опрос, проверочная работа
	<b>Итого часов (контроль 45)</b>		<b>58</b>		<b>16</b>	<b>16</b>		<b>26</b>	<b>5</b>

#### 4.2. План внеаудиторной самостоятельной работы обучающихся по дисциплине

Курс	Название раздела / темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно-методическое обеспечение СР
		Вид СР	Сроки	Трудо-емкость (в часах)		
2	Задачи обработки текстов. Проблемы формализации естественных языков. Предобработка текстов для выделения признаков. Лемматизация. Токенизация. Стоп-слова. TF-IDF. «Мешок слов». Регулярные выражения.	Работа с основной и дополнительной литературой и свободной средой разработки. Подготовка к проверочной работе.	2 неделя семестра	2	проверочная работа	Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с.
2	Библиотеки NLTK, gensim, SpaCy, Yargy.	Работа с основной и дополнительной литературой и свободной средой разработки. Подготовка к проверочной работе.	3 неделя семестра	3	проверочная работа	Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с.
2	Задачи машинного обучения. Базовые библиотеки для машинного обучения: Pandas, Scikit-learn, Matplotlib.	Работа с основной и дополнительной литературой. Подготовка к устному опросу.	4 неделя семестра	2	устный опрос	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Изда-

						тельство Юрайт, 2022. — 140 с.
2	Признаки и работа с ними. Валидация данных и оценка качества алгоритмов. Функции потерь, метрики качества. Разделение набора данных. Кросс-валидация.	Работа с основной и дополнительной литературой и средой разработки. Подготовка к устному опросу	5 неделя семестра	3	устный опрос	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Издательство Юрайт, 2022. — 140 с.
2	Простые векторные представления слов: One-hot векторы, SVD, Hashing Trick	Работа с основной и дополнительной литературой и средой разработки. Подготовка к проверочной работе.	6 неделя семестра	2	проверочная работа	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Издательство Юрайт, 2022. — 140 с.
2	Обучаемые векторные представления слов: word2vec (CBOW, skipGram), GloVe, fastText.	Работа с основной и дополнительной литературой. Подготовка к проверочной работе.	7 неделя семестра	2	устный опрос	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Изда-

						тельство Юрайт, 2022. — 140 с.
2	Методы обучения и донастройки неглубоких эмбедингов.	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	8 неделя семестра	3	устный опрос	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Издательство Юрайт, 2022. — 140 с.
2	Использование неглубоких эмбедингов для классификации.	Работа с основной и дополнительной литературой, конспектирование и средой разработки. Подготовка к устному опросу.	9 неделя семестра	2	проверочная работа	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Издательство Юрайт, 2022. — 140 с.
2	Извлечение ключевых слов. Создание краткого содержания текстов.	Работа с основной и дополнительной литературой, конспектирование. Подготовка к проверочной работе.	10 неделя семестра	1	устный опрос	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Изда-



						тельство Юрайт, 2022. — 140 с.
2	Тематическое моделирование. PLSA (вероятностная тематическая модель представления текста на естественном языке), LDA (метод тематического моделирования, позволяющий объяснять результаты наблюдений с помощью неявных групп).	Работа с основной и дополнительной литературой и свободной средой разработки. Подготовка к проверочной работе.	11 неделя семестра	3	проверочная работа	Автоматическая обработка <i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Издательство Юрайт, 2022. — 140 с.
2	Тематическое моделирование. ARTM (многокритериальный подход к построению вероятностных тематических моделей коллекций текстовых документов).	Работа с основной и дополнительной литературой и средой разработки. Подготовка к проверочной работе.	12 неделя семестра	3	проверочная работа	<i>Платонов, А. В.</i> Машинное обучение : учебное пособие для вузов / А. В. Платонов. — М.: Издательство Юрайт, 2023. — 85 с <i>Горохов, А. В.</i> Основы системного анализа: учебное пособие для вузов / А. В. Горохов. — М.: Издательство Юрайт, 2022. — 140 с.

Общий объем самостоятельной работы по дисциплине (час) – 26

**Из них объем самостоятельной работы с использованием электронного обучения и дистанционных образовательных технологий (час)**

### **4.3. Содержание учебного материала**

#### **Раздел 1. Предобработка и визуализация данных**

1.1. Задачи обработки текстов. Проблемы формализации естественных языков. Предобработка текстов для выделения признаков. Лемматизация. Токенизация. Стоп-слова. TF-IDF. «Мешок слов». Регулярные выражения.

1.2. Библиотеки NLTK, gensim, SpaCy, Yargy.

#### **Раздел 2 Библиотеки для машинного обучения. Выделение признаков.**

2.1. Задачи машинного обучения. Базовые библиотеки для машинного обучения: Pandas, Scikit-learn, Matplotlib.

2.2. Признаки и работа с ними. Валидация данных и оценка качества алгоритмов. Функции потерь, метрики качества. Разделение набора данных. Кросс-валидация.

2.3. Простые векторные представления слов: One-hot векторы, SVD, Hashing Trick

#### **Раздел 3. Векторное представление слов в обработке текстов**

3.1. Обучаемые векторные представления слов: word2vec (CBOW, skipGram), GloVe, fastText.

3.2. Методы обучения и донастройки неглубоких эмбеддингов.

3.3. Использование неглубоких эмбеддингов для классификации.

#### **Раздел 4. Методы снижения размерности**

4.1. Извлечение ключевых слов. Создание краткого содержания текстов.

4.2. Тематическое моделирование. PLSA (вероятностная тематическая модель представления текста на естественном языке), LDA (метод тематического моделирования, позволяющий объяснять результаты наблюдений с помощью неявных групп).

4.3. Тематическое моделирование. ARTM (многокритериальный подход к построению вероятностных тематических моделей коллекций текстовых документов).

#### 4.3.1. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы дисциплины	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)		Оценочные средства	Формируемые компетенции и индикаторы
			Всего часов	Из них – практическая подготовка		
1	2	3	4	5	6	7
1	1.1	Задачи обработки текстов. Проблемы формализации естественных языков. Предобработка текстов для выделения признаков. Лемматизация. Токенизация. Стоп-слова. TF-IDF. «Мешок слов». Регулярные выражения.	1		проверочная работа	ПК-1.1, ПК-1.2
2	1.2	Библиотеки NLTK, gensim, SpaCy, Yargy.	1		проверочная работа	ПК-1.1, ПК-1.2
	2.1	Задачи машинного обучения. Базовые библиотеки для машинного обучения: Pandas, Scikit-learn, Matplotlib.	2		устный опрос	ПК-1.1, ПК-1.2
	2.2	Признаки и работа с ними. Валидация данных и оценка качества алгоритмов. Функции потерь, метрики качества. Разделение набора данных. Кросс-валидация.	1		устный опрос	ПК-1.1, ПК-1.2
3	2.3	Простые векторные представления слов: One-hot векторы, SVD, Hashing Trick	1		проверочная работа	ПК-1.1, ПК-1.2
	3.1	Обучаемые векторные представления слов: word2vec (CBOW, skipGram), GloVe, fastText.	2		устный опрос	ПК-1.1, ПК-1.2
4	3.2	Методы обучения и донастройки неглубоких эмбедингов.	2		устный опрос	ПК-1.1, ПК-1.2
5	3.3	Использование неглубоких эмбедингов для	2		проверочная работа	ПК-1.1,

		классификации.				ПК-1.2
	4.1	Извлечение ключевых слов. Создание краткого содержания текстов.	1		устный опрос	ПК-1.1, ПК-1.2
	4.2	Тематическое моделирование. PLSA (вероятностная тематическая модель представления текста на естественном языке), LDA (метод тематического моделирования, позволяющий объяснять результаты наблюдений с помощью неявных групп).	1		проверочная работа	ПК-1.1, ПК-1.2
6	4.3	Тематическое моделирование. ARTM (многокритериальный подход к построению вероятностных тематических моделей коллекций текстовых документов).	2		проверочная работа	ПК-1.1, ПК-1.2

**4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение студентами в рамках самостоятельной работы (СРС)**

№ п/п	Тема	Задание	Формируемая компетенция	ИДК
1.	Проблемы формализации естественных языков.	Прочитайте главу 7 «Интеллектуальные системы: учебник и практикум для вузов». Законспектируйте параграф 7.2. Выполните задание 2 в разделе «Практикум».	<i>ПК-1</i>	ПК-1.1, ПК-1.2

#### 4.4. Методические указания по организации самостоятельной работы студентов

**Раздел 1.** В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Задачи обработки текстов. Проблемы формализации естественных языков. Предобработка текстов для выделения признаков. Лемматизация. Токенизация. Стоп-слова. TF-IDF. «Мешок слов». Регулярные выражения. 2. Библиотеки NLTK, gensim, SpaCy, Yargy.

**Раздел 2.** В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Задачи машинного обучения. Базовые библиотеки для машинного обучения: Pandas, Scikit-learn, Matplotlib. 2. Признаки и работа с ними. Валидация данных и оценка качества алгоритмов. Функции потерь, метрики качества. Разделение набора данных. Кросс-валидация. 3. Простые векторные представления слов: One-hot векторы, SVD, Hashing Trick.

**Раздел 3.** В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Обучаемые векторные представления слов: word2vec (CBOW, skipGram), GloVe, fastText. 2. Методы обучения и донастройки неглубоких эмбедингов. 3. Использование неглубоких эмбедингов для классификации.

**Раздел 4.** В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Извлечение ключевых слов. Создание краткого содержания текстов. 2. Тематическое моделирование. PLSA (вероятностная тематическая модель представления текста на естественном языке), LDA (метод тематического моделирования, позволяющий объяснять результаты наблюдений с помощью неявных групп). 3. Тематическое моделирование. ARTM (многокритериальный подход к построению вероятностных тематических моделей коллекций текстовых документов).

#### 4.5. Примерная тематика курсовых работ (проектов)

Курсовые работы в рамках данной дисциплины не предусмотрены.

### V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ:

#### а) перечень литературы

##### основная литература

1. **Автоматическая обработка текстов на естественном языке и анализ данных** : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с. ISBN 978-5-9909752-1-7

2. **Платонов, А. В.** Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/520544> (дата обращения: 28.09.2023).

3. **Горохов, А. В.** Основы системного анализа : учебное пособие для вузов / А. В. Горохов. — Москва : Издательство Юрайт, 2022. — 140 с. — (Высшее образование). — ISBN 978-5-534-09459-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/492860> (дата обращения: 28.09.2023).

#### б) периодические издания

Вопросы языкознания  
Филологические науки  
Сибирский филологический журнал

#### **в) список авторских методических разработок**

Методические материалы по курсу размещены на образовательном портале [belca.isu.ru](https://belca.isu.ru): <https://belca.isu.ru>

#### **г) базы данных, поисково-справочные и информационные системы**

##### **Корпуса русского языка**

1. <http://www.ruscorpora.ru/>  
<http://www.ruscorpora.ru/corpora-other.html>  
Национальный корпус русского языка – информационно-справочная система, основанная на собрании русских текстов в электронной форме
2. <http://www.slaviska.uu.se/korpus.htm>  
Упсальский корпус русского языка
3. <http://www.sfb441.uni-tuebingen.de/b1/rus/korpora.html>  
Тюбингенский корпус русского языка
4. <http://cfri.ru/>  
Машинный фонд русского языка
5. <http://www.philol.msu.ru/~lex/corpus/>  
Компьютерный корпус текстов русских газет конца XX века

##### **Электронные библиотеки**

1. [Научная электронная библиотека eLIBRARY.RU](http://www.eLIBRARY.RU)
2. [Коллекция журналов издательства «НАУКА»](#)
3. [Журналы Института научной информации по общественным наукам \(ИНИОН РАН\)](#)
4. [Университетская информационная система РОССИЯ \(УИС РОССИЯ\)](#)
5. [Журнал Science издательства American Association for the Advancement of Science \(AAAS\)](#)
6. <http://www.philology.ru/>  
Библиотека филологических текстов (статей, монографий)
7. <http://philologos.narod.ru/>  
Материалы по теории языка и литературы
8. <http://www.scribd.com/>  
База письменных документов (научных статей, монографий, художественных текстов etc.) на разных языках
9. <http://ellib.library.isu.ru>  
Электронная библиотека «Труды ученых ИГУ»
10. <http://library.cjes.org>  
Библиотека Центра экстремальной журналистики

##### **Другие сайты**

10. <http://www.gramota.ru>  
Словари и справочники, представленные на портале
11. <http://www.grammar.ru>  
Словари и справочники, представленные на портале
12. <http://linguistlist.org/>

- «The world's largest online linguistic resource»: информация о конференциях, публикации, каталог ссылок на другие лингвистические ресурсы etc.
13. <http://www.ruthenia.ru/web/rusweb.html>  
«Русистика на Вебе»: ссылки на сайты филологических факультетов, отделений русистики и славистики
14. <http://educa.isu.ru/>; <http://belca.isu.ru/>  
Образовательный портал Иркутского государственного университета, образовательный портал ИФИЯМ

## VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 6.1. Учебно-лабораторное оборудование.

<p><b>Специальные помещения</b></p> <p>Учебная аудитория для проведения занятий лекционного и семинарского типа.</p>	<p><b>Оборудование аудитории</b></p> <p><b>Мебель:</b></p> <ul style="list-style-type: none"> <li>• специализированная учебная мебель на 30 посадочных мест;</li> <li>• стол-кафедра – 1 шт.;</li> <li>• стол преподавателя – 1 шт.; стул – 1 шт.</li> </ul> <p><b>Доска</b> настенная магнитно-меловая, одноэлементная – 1 шт.;</p> <p><b>Технические средства обучения</b>, служащие для представления учебной информации большой аудитории по дисциплине:</p> <ul style="list-style-type: none"> <li>• стационарный ПК – 1 шт.;</li> <li>• проектор Epson EB-X10;</li> <li>• экран Digis;</li> <li>• колонки Sven.</li> </ul> <p><b>Учебно-наглядные пособия:</b></p> <ul style="list-style-type: none"> <li>• презентации в формате Microsoft Power Point по каждой теме РПД дисциплины.</li> </ul> <p><b>Программное обеспечение:</b> офисный пакет приложений Microsoft Office (в т.ч. программа Microsoft Power Point для создания и демонстрации презентаций, иллюстраций и других учебных материалов по дисциплине).</p>
<p><b>Специальные помещения</b></p> <p>Компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, организации самостоятельной работы, в том числе научно-исследовательской.</p>	<p><b>Оборудование аудитории</b></p> <p><b>Мебель:</b></p> <ul style="list-style-type: none"> <li>• специализированная учебная мебель на 20 посадочных мест;</li> <li>• стол-кафедра – 1 шт.;</li> <li>• стол преподавателя – 1 шт.; стул – 1 шт.</li> </ul> <p><b>Технические средства обучения</b>, служащие для представления учебной информации большой аудитории по дисциплине:</p> <ul style="list-style-type: none"> <li>• стационарный ПК с неограниченным доступом к сети Интернет – 20 шт.;</li> <li>• колонки Sven.</li> </ul> <p><b>Программное обеспечение:</b> офисный пакет приложений Microsoft Office</p>

Специализированное учебное оборудование не используется.

### 6.2. Программное обеспечение



Университет обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства (подробная информация размещена на сайте ИГУ)

№ п/п	Наименование
1.	Adobe Reader DC 2019.008.20071
2.	IrfanView 4.42
3.	Foxit PDF Reader 8.0
4.	Google Chrome
5.	Java 8
6.	Kaspersky Endpoint Security для бизнеса- Стандартный Russian Edition. 250-499.
7.	Microsoft Office Enterprise 2007 Russian Academic OPEN No Level
8.	Mozilla Firefox
9.	Office 365 ProPlus for Students
10.	OpenOffice 4.1.3
11.	Opera 45
12.	PDF24Creator 8.0.2
13.	Skype 7.30.0
14.	VLC Player 2.2.4
15.	WinPro 10 RUS Upgrd OLP NL Acdmc.
16.	7zip 18.06

### 6.3. Технические и электронные средства

Авторские презентации к лекциям и практическим занятиям, фрагменты открытых лекций по проблематике.

С подробным перечнем можно ознакомиться по ссылке:

<https://belca.isu.ru/course/view.php?id=>

## VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Преподавание дисциплины осуществляется в форме практических и лекционных занятий, а также содержит разные формы самостоятельной работы студентов: подготовка к аудиторным занятиям (в соответствии с планами занятий); выполнение отдельных видов самостоятельной работы, предусмотренных рабочей программой учебной дисциплины, в том числе подготовка докладов, конспектов и иных письменных работ, выполнение практических заданий, связанных со сбором и анализом фактического и теоретического материала, поиском материалов в Интернете, разбор конкретных ситуаций, подготовка к выполнению тестовых и творческих заданий и др. Выполнение самостоятельной работы студентами контролируется в ходе практических занятий, на консультациях, при проверке выполненных заданий, в процессе самоконтроля, а также включает проверочные работы студентов по освоению теоретического и дидактического материала. В ходе проведения практических занятий используются языки программирования, интерактивные формы проведения занятий (деловые и ролевые игры, компьютерные симуляции, тренинги) с целью формирования и развития профессиональных навыков студентов. Широко применяются мультимедийные средства. Сетевые компьютерные технологии включают использование ресурсов Интернет и виртуального университета.

## **VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

### **Оценочные средства для входного контроля**

Для обеспечения входного контроля используется самостоятельная работа, ориентированная на проверку остаточных знаний, полученных во время обучения на первом курсе.

### **Примерная программа, на основе которой формулируются вопросы для входного контроля**

1. Приведите примеры тем исследований, где могут использоваться следующие методы: функциональный анализ, сравнительный, метод моделирования, метод стилистического анализа, тезаурусный метод.
2. Понятие структурной связанности текста. Левосторонние и правосторонние семантико-синтаксические средства связи.
3. Понятие цельности текста. Ключевые слова. Виды повторной номинации.
4. Основные функции повторной номинации в разных текстах: текстообразующая, информационно-описательная, ситуативная, экспрессивно-оценочная, стилистически дифференцирующая, редуцирующая. Textoобразующие средства повторной номинации. Повторная номинация на уровне лексическом, стилистическом, морфологическом, синтаксическом.
5. Виды информации в тексте: фактуальная, концептуальная, методическая, эмоциональная.
6. Говорящий и текст.
7. Проблема текста в лингвистике, филологии и других гуманитарных науках (по М. М. Бахтину).
8. Проблема декодирования текстов.
9. Семиотика предмета: предмет и вещь, их соотношение в тексте.
10. Сопоставление анализаторов Mystem, Rymorphy2, DeepMorphy. Плюсы и минусы. Особенности использования.
11. Обработка языка в задачах информационного поиска. Проблемы формализации естественных языков. Специфика задач, подходы к решению, извлекаемая информация.
12. Лингвистические шаблоны и правила в задачах информационного поиска.
13. Инструментальные системы для извлечения информации.
14. Именованные сущности и особенности их извлечения.
15. Предобработка и фильтрация данных. Основные этапы. Условия для отбора данных.

## Оценочные средства текущего контроля

Текущий контроль за усвоением содержания дисциплины осуществляется в следующих формах:

- тематический опрос студентов на практических занятиях;
- выполнение заданий, ориентированных на усвоение навыков программирования;
- изучение и анализ текстов различных типов;
- выступления (с презентацией) творческого характера;
- обсуждение отдельных вопросов курса в консультационном режиме.

## Список вопросов (программа) к экзамену

1. Задачи обработки текстов. Проблемы формализации естественных языков. Предобработка текстов для выделения признаков.
2. Лемматизация. Токенизация. Стоп-слова. TF-IDF.
3. «Мешок слов». Регулярные выражения.
4. Библиотеки NLTK, gensim, SpaCy, Yargy.
5. Задачи машинного обучения.
6. Базовые библиотеки для машинного обучения: Pandas.
7. Базовые библиотеки для машинного обучения: Scikit-learn.
8. Базовые библиотеки для машинного обучения: Matplotlib.
9. Признаки и работа с ними. Валидация данных и оценка качества алгоритмов.
10. Функции потерь, метрики качества. Разделение набора данных. Кросс-валидация.
11. Простые векторные представления слов: One-hot векторы, SVD, Hashing Trick
12. Обучаемые векторные представления слов: word2vec (CBOW, skipGram).
13. Обучаемые векторные представления слов: GloVe.
14. Обучаемые векторные представления слов: fastText.
15. Методы обучения и донастройки неглубоких эмбеддингов.
16. Использование неглубоких эмбеддингов для классификации.
17. Извлечение ключевых слов.
18. Создание краткого содержания текстов.
19. Тематическое моделирование. PLSA (вероятностная тематическая модель представления текста на естественном языке).
20. Тематическое моделирование. LDA (метод тематического моделирования, позволяющий объяснять результаты наблюдений с помощью неявных групп).
21. Тематическое моделирование. ARTM (многокритериальный подход к построению вероятностных тематических моделей коллекций текстовых документов).

**Разработчик:**

У  
(подпись)

ст.преподаватель  
(занимаемая должность)

У.Э. Чекмез  
(инициалы, фамилия)

Программа рассмотрена на заседании кафедры русского языка и общего языкознания  
(наименование)

Программа составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 45.04.01 Филология.

«08» февраля 2024 г.

Протокол № 5 Зав.кафедрой Ташлыкова М.Б. Ташлыкова

*Настоящая программа не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.*