



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.17 Автоматическая обработка текстов на естественном языке

Направление подготовки **45.04.01 Филология**

Направленность (профиль) подготовки
«Современная филология: практики работы с текстами»

Квалификация (степень) выпускника – **магистр**
Форма обучения – **очная**

Согласовано с УМК ИФИЯМ
Протокол № 6 от «21» февраля 2023 г.
Председатель О.Л. Михалёва

Рекомендовано кафедрой
русского языка и общего языкознания
Протокол № 5 от «9» февраля 2023 г.
Зав. кафедрой М.Б. Ташлыкова

Иркутск 2023 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	5
	4.2.	План внеаудиторной самостоятельной работы обучающихся по дисциплине	6
	4.3.	Содержание учебного материала	10
	4.3.1.	Перечень семинарских, практических занятий и лабораторных работ	11
	4.3.2.	Перечень тем (вопросов), выносимых на самостоятельное изучение в рамках самостоятельной работы студентов	12
	4.4.	Методические указания по организации самостоятельной работы студентов	13
V.		Учебно-методическое и информационное обеспечение дисциплины	13
	а)	перечень литературы	
	б)	периодические издания	
	в)	список авторских методических разработок	
	г)	базы данных, поисково-справочные и информационные системы	
VI.		Материально-техническое обеспечение дисциплины	15
	6.1.	Учебно-лабораторное оборудование	
	6.2.	Программное обеспечение	
	6.3.	Технические и электронные средства обучения	
VII.		Образовательные технологии	16
VIII.		Оценочные материалы для текущего контроля и промежуточной аттестации	17

I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель дисциплины – обучение теории и практике автоматической обработки текстов на естественном языке.

К **задачам** дисциплины относятся

- формирование у студентов представлений о принципах и приемах автоматической обработки текстов на естественном языке;
- знакомство со спектром задач и проблем, решаемых в этой области;
- получение навыков предобработки и фильтрации текстовых данных, автоматического морфологического анализа, извлечения информации, анализа тональности текстов.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина «Автоматическая обработка текстов на естественном языке» 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.04 Методология научного познания в гуманитарных науках

Б1.О.06 Теория текста

Б1.О.07 Художественный текст как знаковая система

Б1.В.07 Функциональное разнообразие русской речи

Б1.В.13 Межличностная коммуникация: способы речевого взаимодействия

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной:

Б1.О.05 Введение в системный анализ и машинное обучение

Б1.В.15 Технология анализа спорных текстов в рамках лингвистической экспертизы

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**

Компетенция	Индикаторы компетенций	Результаты обучения
<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>Знать:</p> <ul style="list-style-type: none"> • принципы и приёмы автоматической обработки текстов на естественном языке. <p>Уметь:</p> <ul style="list-style-type: none"> • осуществлять предобработку языковых данных. <p>Владеть:</p> <ul style="list-style-type: none"> • навыками автоматического морфологического анализа; • навыками использования обработки языка в задачах информационного поиска; • навыками обработки языкового материала с помощью современных языков программирования.
	<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>Знать:</p> <ul style="list-style-type: none"> • основные сферы применения автоматической обработки текстов на естественном языке. <p>Уметь:</p> <ul style="list-style-type: none"> • применять методы автоматической обработки текстов для оптимизации работы в рамках собственных исследований. <p>Владеть:</p> <ul style="list-style-type: none"> • навыками анализа тональности текстов; • навыками интерпретации и представления полученных данных.

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)				Формы текущего контроля успеваемости
					Контактная работа преподавателя с обучающимися			Самостоятельная работа	
					Лекции	Семинарские (практические занятия)	Консультации		
1	Предобработка и фильтрация данных	2	14		2	4		8	проверочная работа
2	Автоматический морфологический анализ	2	10		4	2		4	устный опрос, проверочная работа
3	Обработка языка в задачах информационного поиска	2	15		2	4		9	устный опрос, проверочная работа
4	Анализ тональности	2	15		4	2		9	устный опрос, проверочная работа
	Итого часов (контроль 49)		54		12	12		30	5

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	Предобработка текстов для подсчета статистики. Перевод текста в таблицу. Лемматизация. Токенизация. Изменение регистра. Стоп-слова. Подсчет частотности. Распределение слов. Ngramms. TF-IDF. Udrpipe	Работа с основной и дополнительной литературой и свободной средой разработки. Подготовка к проверочной работе.	2 неделя семестра	4	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	Фильтрация данных. Соединение, разъединение, замена компонентов строк. Поиск по подстроке. Сортировка значений. Отбор данных по комбинации условий. Запись в файл.	Работа с основной и дополнительной литературой и свободной средой разработки. Подготовка к проверочной работе.	3 неделя семестра	4	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	Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов. Словарь словоформ.	Работа с основной и дополнительной литературой. Подготовка к устному опросу.	4 неделя семестра	1	устный опрос	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	Проблемы морфологического аннотирования русского языка.	Работа с основной и дополнительной литературой и средой разработки. Подготовка к устному опросу	5 неделя семестра	1	устный опрос	Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лу-кашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с..
2	Mystem. Pymorphy2. DeepMorphy.	Работа с основной и дополнительной литературой и средой разработки. Подготовка к проверочной работе.	6 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	Специфика задач, подходы к решению, извлекаемая информация.	Работа с основной и дополнительной литературой. Подготовка к проверочной работе.	7 неделя семестра	3	устный опрос	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. Бессмертный И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Изда-

						тельство Юрайт, 2023. — 243 с.
2	Лингвистические шаблоны и правила. Инструментальные системы для извлечения информации.	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	8 неделя семестра	3	устный опрос	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. Бессмертный И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с.
2	Именованные сущности и особенности их извлечения. Особенности извлечения атрибутов, отношений и фактов. Извлечение терминологической информации.	Работа с основной и дополнительной литературой, конспектирование и средой разработки. Подготовка к устному опросу.	9 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. Бессмертный И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В.

						Платонов. — Москва : Издательство Юрайт, 2023. — 243 с.
2	Определения тональности текста как задача классификации. Сложности анализа тональности текстов. Словарные ресурсы для анализа тональности.	Работа с основной и дополнительной литературой, конспектирование. Подготовка к проверочной работе.	10 неделя семестра	3	устный опрос	Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лу-кашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. — 269 с. Бессмертный И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с.
2	Метрические методы классификации. Логические методы. Линейные методы. Профильные методы. Ансамбли решений. Нейросетевые методы.	Работа с основной и дополнительной литературой и свободной средой разработки. Подготовка к проверочной работе.	11 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. Бессмертный И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный,

						А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с.
2	Анализ тональности документов в целом. Анализ тональности по аспектам.	Работа с основной и дополнительной литературой и средой разработки. Подготовка к проверочной работе.	12 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. Бессмертный И. А. Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с.

Общий объем самостоятельной работы по дисциплине (час) – 30

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

4.3. Содержание учебного материала

Раздел 1. Предобработка и фильтрация данных

1.1. Предобработка текстов для подсчета статистики. Перевод текста в таблицу. Лемматизация. Токенизация. Изменение регистра. Стоп-слова. Подсчет частотности. Распределение слов. Ngramms. TF-IDF. Udpire.

1.2. Фильтрация данных. Соединение, разъединение, замена компонентов строк. Поиск по подстроке. Сортировка значений. Отбор данных по комбинации условий. Запись в файл.

Раздел 2. Автоматический морфологический анализ

2.1. Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов. Словарь словоформ.

2.2. Проблемы морфологического аннотирования русского языка.

2.3. Mystem. Rymorphy2. DeepMorphy.

Раздел 3. Обработка языка в задачах информационного поиска

3.1. Проблемы формализации естественных языков. Специфика задач, подходы к решению, извлекаемая информация.

3.2. Лингвистические шаблоны и правила. Инструментальные системы для извлечения информации.

3.3. Именованные сущности и особенности их извлечения. Особенности извлечения атрибутов, отношений и фактов. Извлечение терминологической информации.

Раздел 4. Анализ тональности

4.1. Определения тональности текста как задача классификации. Автоматическая категоризация текстов. Сложности анализа тональности текстов. Словарные ресурсы для анализа тональности.

4.2. Метрические методы классификации. Логические методы. Линейные методы. Профильные методы. Ансамбли решений. Нейросетевые методы.

4.3. Анализ тональности документов в целом. Анализ тональности по аспектам.

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	Предобработка текстов для подсчета статистики. Перевод текста в таблицу. Лемматизация. Токенизация. Изменение регистра. Стоп-слова. Подсчет частотности. Распределение слов. Ngramms. TF-IDF. Udpire.	2		проверочная работа	ПК-1.1, ПК-1.2
2	1.2	Фильтрация данных. Соединение, разъединение, замена компонентов строк. Поиск по подстроке. Сортировка значений. Отбор данных по комбинации условий. Запись в файл.	2		проверочная работа	ПК-1.1, ПК-1.2
3	2.3	Mystem. Rymorphy2. DeepMorphy.	2		проверочная работа	ПК-1.1, ПК-1.2
4	3.2	Лингвистические шаблоны и правила. Инструментальные системы для извлечения информации.	2		проверочная работа	ПК-1.1, ПК-1.2
5	3.3	Именованные сущности и особенности их извлечения. Особенности извлечения атрибутов, отношений и фактов. Извлечение терминологической информации..	2		проверочная работа	ПК-1.1, ПК-1.2
6	4.3	Анализ тональности документов в целом. Анализ тональности по аспектам	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
2.	Автоматическая категоризация текстов.	Прочитайте главу 9 «Интеллектуальные системы: учебник и практикум для вузов». Выполните задание 3 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. Предобработка текстов для подсчета статистики. Перевод текста в таблицу. Лемматизация. Токенизация. Изменение регистра. Стоп-слова. Подсчет частотности. Распределение слов. Ngramms. TF-IDF. Udrpipe. 2. Фильтрация данных. Соединение, разъединение, замена компонентов строк. Поиск по подстроке. Сортировка значений. Отбор данных по комбинации условий. Запись в файл.

Раздел 2. В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов. Словарь словоформ. 2. Проблемы морфологического аннотирования русского языка. 3. Mystem. Pymorphy2. DeepMorphy.

Раздел 3. В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Проблемы формализации естественных языков. Специфика задач, подходы к решению, извлекаемая информация. 2. Лингвистические шаблоны и правила. Инструментальные системы для извлечения информации. 3. Именованные сущности и особенности их извлечения. Особенности извлечения атрибутов, отношений и фактов. Извлечение терминологической информации.

Раздел 4. В результате работы с основной и дополнительной литературой и свободной средой разработки студент должен уметь раскрыть содержание следующих тем: 1. Определения тональности текста как задача классификации. Автоматическая категоризация текстов. Сложности анализа тональности текстов. Словарные ресурсы для анализа тональности. 2. Метрические методы классификации. Логические методы. Линейные методы. Профильные методы. Ансамбли решений. Нейросетевые методы. 3. Анализ тональности документов в целом. Анализ тональности по аспектам.

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. **Бессмертный И. А.** Интеллектуальные системы : учебник и практикум для вузов / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. — Москва : Издательство Юрайт, 2023. — 243 с. — (Высшее образование). — ISBN 978-5-534-01042-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/511999> (дата обращения: 03.07.2023).

б) периодические издания

Вопросы языкознания
Филологические науки

в) список авторских методических разработок

Методические материалы по курсу размещены на образовательном портале **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://cfrl.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](#)
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>Специальные помещения</p> <p>Учебная аудитория для проведения занятий лекционного и семинарского типа.</p>	<p>Оборудование аудитории</p> <p>Мебель:</p> <ul style="list-style-type: none"> • специализированная учебная мебель на 30 посадочных мест; • стол-кафедра – 1 шт.; • стол преподавателя – 1 шт.; стул – 1 шт. <p>Доска настенная магнитно-меловая, одноэлементная – 1 шт.;</p> <p>Технические средства обучения, служащие для представления учебной информации большой аудитории по дисциплине:</p> <ul style="list-style-type: none"> • стационарный ПК – 1 шт.; • проектор Epson EB-X10; • экран Digis; • колонки Sven. <p>Учебно-наглядные пособия:</p> <ul style="list-style-type: none"> • презентации в формате Microsoft Power Point по каждой теме РПД дисциплины. <p>Программное обеспечение: 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>Специальные помещения</p> <p>Компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, организации самостоятельной работы, в том числе научно-исследовательской.</p>	<p>Оборудование аудитории</p> <p>Мебель:</p> <ul style="list-style-type: none"> • специализированная учебная мебель на 20 посадочных мест; • стол-кафедра – 1 шт.; • стол преподавателя – 1 шт.; стул – 1 шт. <p>Технические средства обучения, служащие для представления учебной информации большой аудитории по дисциплине:</p> <ul style="list-style-type: none"> • стационарный ПК с неограниченным доступом к сети Интернет – 20 шт.; • колонки Sven. <p>Программное обеспечение: 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 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. Дайте определение понятиям научный метод / прием, методология, методика научной работы.
2. Как разграничивают объект и предмет научного исследования? В чем специфика объекта лингвистики?
3. Приведите примеры тем исследований, где могут использоваться следующие методы: функциональный анализ, сравнительный, метод моделирования, метод стилистического анализа, тезаурусный метод.
4. Понятие структурной связанности текста. Левосторонние и правосторонние семантико-синтаксические средства связи.
5. Понятие цельности текста. Ключевые слова. Виды повторной номинации.
6. 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.
7. Виды информации в тексте: фактуальная, концептуальная, методическая, эмоциональная.
8. Говорящий и текст.
9. Проблема текста в лингвистике, филологии и других гуманитарных науках (по М. М. Бахтину).
10. Проблема декодирования текстов.
11. Семиотика предмета: предмет и вещь, их соотношение в тексте.
12. Стилиевые черты официально-делового стиля. Графические особенности текста документов. Языковые средства, реализующие основные стилиевые черты функции официально-делового стиля.
13. Стилиевые черты научного стиля. Языковые средства, реализующие основные стилиевые черты научного стиля.
Основные стилиевые черты публицистического стиля. Языковые средства, реализующие основные стилиевые черты публицистического стиля.
14. Языковые особенности разговорной речи: фонетический, лексический, морфологический, синтаксический уровень. Тенденции современной разговорной речи: тенденция к синкретизму и тенденция к расчленённости.
15. Языковые особенности просторечия: фонетический, лексический, морфологический, синтаксический уровень.

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 и статистического анализа;
- изучение и анализ текстов различных типов;
- выступления (с презентацией) творческого характера;
- обсуждение отдельных вопросов курса в консультационном режиме.

Проверочная работа №1 (демонстрационный вариант)

Загрузите датасет с рассказами М. Зощенко (https://raw.githubusercontent.com/agricolanz/2020.11.01_appcogn_text_analysis/master/data/zoshenko.csv). Переведите текст в таблицу, лемматизируйте его, исключите стоп-слова, составьте рейтинг частотности из 100 пунктов.

Проверочная работа №2 (демонстрационный вариант)

С помощью пакета `gutenbergr` загрузите «Духовные оды» Г. Р. Даржавина. Используйте любой морфологический анализатор для автоматического морфологического анализа. Перечислите проблемные места для автоматического морфологического анализа. Какие элементы невозможно верно разметить без ручной обработки?

Проверочная работа №3 (демонстрационный вариант)

Создайте датасет из учебников по программированию. Извлеките терминологическую информацию.

Проверочная работа №4 (демонстрационный вариант)

Создайте датасет из отзывов к любому товару в любом интернет-магазине (не менее 500 отзывов). Осуществите общий анализ тональности.

Оценочные средства для промежуточной аттестации (в форме экзамена)

Список вопросов (программа) к экзамену

1. Предобработка и фильтрация данных. Основные этапы. Условия для отбора данных.
2. Предобработка текстовых данных как предварительный анализ текста. Лемматизация. Стоп-слова. Проблемы предобработки текстов на русском языке. Ngrams как источник информации о ключевых «персонажах» и синтагматических отношениях в тексте.
3. Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов. Словарь словоформ.
4. Проблемы морфологического аннотирования русского языка.
5. Сопоставление анализаторов `Mystem`, `Rymorphy2`, `DeepMorphy`. Плюсы и минусы. Особенности использования.
6. Обработка языка в задачах информационного поиска. Проблемы формализации естественных языков. Специфика задач, подходы к решению, извлекаемая информация.
7. Лингвистические шаблоны и правила в задачах информационного поиска.
8. Инструментальные системы для извлечения информации.
9. Именованные сущности и особенности их извлечения.
10. Особенности извлечения атрибутов, отношений и фактов.
11. Извлечение терминологической информации.

12. Определения тональности текста как задача классификации. Автоматическая категоризация текстов.
13. Сложности анализа тональности текстов.
14. Словарные ресурсы для анализа тональности текстов на русском языке.
15. Метрические методы классификации. Метод центроидов, k-ближайших соседей.
16. Логические методы. Метод деревьев решений. Нейросетевые методы.
17. Линейные методы. Метод опорных векторов, логистическая регрессия.
18. Профильные методы. Логический профиль, рассчитываемый профиль, экспертный профиль.
19. Ансамбли решений. Алгоритмы bagging, boosting.
20. Анализ тональности документов в целом. Анализ тональности по аспектам.

