

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ
Заведующий кафедрой
теоретической и прикладной лингвистики



Шилихина К.М.
30.05.2025 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.05 Анализ данных для лингвиста

1. Код и наименование направления подготовки/специальности:

45.03.03 Фундаментальная и прикладная лингвистика

2. Профиль подготовки/специализация:

Экспертно-аналитическая деятельность

3. Квалификация выпускника: бакалавр

4. Форма обучения: очная

5. Кафедра, отвечающая за реализацию дисциплины: кафедра теоретической и прикладной лингвистики

6. Составители программы: Дони́на О́льга Вале́рьевна, кандидат филол. наук, доцент кафедры теоретической и прикладной лингвистики

7. Рекомендована: Научно-методическим советом факультета РГФ, протокол № 8 от 18.04.2025 г.

8. Учебный год: 2028/2029

Семестр: 7

9. Цели и задачи учебной дисциплины

Целью освоения учебной дисциплины является формирование представления о различных методах анализа данных.

Задачи учебной дисциплины:

- формировать представление об основных понятиях и методах анализа данных;
- развивать умение ставить и понимать исследовательские задачи, которые могут быть решены с помощью методов анализа данных;
- формировать навыки применения методов анализа данных при самостоятельном решении исследовательских задач.

10. Место учебной дисциплины в структуре ООП: дисциплина Б1.В.05 Анализ данных для лингвиста входит в Блок Б1 учебного плана, часть, формируемая участниками образовательных отношений, вариативная дисциплина. Для ее успешного освоения необходимы базовые знания, умения и навыки, сформированные в процессе изучения дисциплин Б1.О.14 Математическая статистика, Б1.В.01 Проектирование баз данных, Б1.В.04 Технологии корпусной лингвистики, Б1.В.10 Формальные модели в лингвистике. Изучение данной дисциплины предшествует освоению дисциплин Б1.В.ДВ.05.01 Общая и компьютерная лексикография, ФТД.02 Основы лексико-семантической прогностики.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Коды	Индикаторы	Планируемые результаты обучения
ПК-3	Владеет навыками оформления и представления результатов научного исследования	ПК-3.1	Выбирает адекватный способ представления результатов исследований в соответствующей области знаний	Знать: общие представления о способе подачи научных материалов; методы оформления результатов научного исследования в виде статей, квалификационных работ, постеров, презентаций, докладов на конференциях и семинарах. Уметь: грамотно и в соответствии с требованиями оформить научную работу, раздаточный материал или презентацию к докладу; сделать устный доклад в соответствии с регламентом. Владеть: практическим опытом выступлений перед аудиторией, а также создания письменных работ для представления результатов научного исследования.
ПК-4	Владеет методами сбора и документации лингвистических данных	ПК-4.2	Собирает, обрабатывает, анализирует и обобщает результаты собственных исследований	Знать: основные методы документирования естественного языка; принципы полевой работы с носителями естественных языков; технологии проведения сбора языковых данных; базовые принципы передачи и хранения собранной информации. Уметь: производить запись речевого сигнала на цифровой носитель; аннотировать, обрабатывать и анализировать

				собранные данные; собирать метаданные, в частности, социолингвистические данные о носителе языка. Владеть: практическим опытом документирования лингвистических данных: записи речевого сигнала с последующей обработкой и анализом; обработки, структурирования и хранения метаинформации о собранных данных.
ПК-5	Способен пользоваться лингвистически ориентированными программными продуктами	ПК-5.1	Осуществляет постановку задачи на технологические исследования	Знать: основные типы систем, использующих модули лингвистического анализа; основные принципы и методы компьютерного моделирования лингвистических задач. Уметь: анализировать работу различных систем обработки текста и звучащей речи для выявления основных лингвистических компонентов и основных типов обработки текста, используемых в данных системах; подбирать необходимые лингвистические ресурсы для различных задач лингвистического обеспечения систем (например, лексикографических, задач морфологического анализа и т.п.). Владеть: навыком работы с различными системами автоматической и экспертной обработки текста и звучащей речи.
		ПК-5.2	Анализирует результаты технологических исследований	
ПК-77	Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умеет пользоваться такими ресурсами.	ПК-7.1	Разрабатывает и документирует программные интерфейсы	Знать: основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных. Уметь: пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами Владеть: навыком разработки электронных языковых ресурсов; опытом применения основных методов, способов и средств получения, хранения, переработки информации
		ПК-7.2	Пользуется электронными языковыми ресурсами для решения прикладных задач	
		ПК-7.3	Анализирует требования к программному обеспечению	

12. Объем дисциплины в зачетных единицах/час. — 3 з.е. /108 ч.

Форма промежуточной аттестации: зачет

13. Трудоемкость по видам учебной работы

Вид учебной работы		Трудоемкость	
		Всего	По семестрам
			7 семестр
Аудиторные занятия		36	36
в том числе:	лекции	18	18
	практические	-	-
	лабораторные	18	18
Самостоятельная работа		72	72
в том числе: курсовая работа (проект)		-	-
Форма промежуточной аттестации (<i>зачет</i>)		-	Зачет
Итого:		108	108

13.1. Содержание дисциплины

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК
1. Лекции			
1.1	Получение данных	Роль данных в интеллектуальном анализе текста. Типы данных. Источники данных. Как отобрать материал для исследования. Выборка. Теоретическое формирование выборки. Онлайн источники данных. Проблемы и ограничения использования онлайн-данных. Big Data.	https://edu.vsu.ru/course/view.php?id=3923
1.2	Этика исследования	Этические принципы исследовательского проекта. Этические принципы Ассоциации интернет-исследователей (AoIR). Конфиденциальность. Общедоступные и частные данные. Информированное согласие. Издательская этика.	https://edu.vsu.ru/course/view.php?id=3923
1.3	Философия и логика интеллектуального анализа текста	Онтологические и эпистемологические позиции. Метатеория. Индуктивная логика, дедуктивная логика, абдуктивная логика.	https://edu.vsu.ru/course/view.php?id=3923
1.4	Разработка исследовательского проекта	Стратегия создания проекта исследования. Дизайн исследования (тип исследования, уровень анализа, режим анализа, выбор данных, логика вывода). Идиографические и номотетические подходы к исследованию. Уровни анализа: текстовый, контекстуальный, социологический. Качественные, количественные и смешанные методы исследования. Отбор данных; формирование выборки.	https://edu.vsu.ru/course/view.php?id=3923
1.5	Веб-скрейпинг и краулинг	Этапы веб-краулинга. Методы веб-краулинга. Программное обеспечение для автоматического сбора текстовой информации из веб-страниц. Сравнение веб-краулинга и веб-скрейпинга. Сравнение инструментов для веб-краулинга и веб-скрейпинга.	https://edu.vsu.ru/course/view.php?id=3923
1.6	Лексические ресурсы	Типы лексических ресурсов. Роль лексических	https://edu.vsu.ru/course/view.php?id=3923

		ресурсов в интеллектуальном анализе текста. Словари, тезаурусы, семантические сети, конкордансы. Корпуса.	.ru/course/view.php?id=3923
1.7	Базовая обработка текста	Токенизация, удаление стоп-слов, стемминг и лемматизация. Текстовая статистика и законы, регулирующие распределение слов в тексте (закон Ципфа, закон Хипса). Языковые модели и их применение. Продвинутое этапы обработки текстов (частеречная разметка, идентификация коллокаций, синтаксический анализ, распознавание именованных сущностей, дизамбигуация, близость слов).	https://edu.vsu.ru/course/view.php?id=3923
1.8	Методы текстового анализа	История текстового анализа. Методы анализа текстов. Контент-анализ. Обоснованная теория. Этнографические методы. Конверсационный анализ. Анализ способа категоризации участия. Анализ нарративов. Нарративная семиотика. Структурная семантика. Symlog. Анализ тем. Анализ метафор. Критический дискурс анализ. Функциональная прагматика. Теория различий. Объективная герменевтика.	https://edu.vsu.ru/course/view.php?id=3923
1.9	Методы интеллектуального анализа текстов	Data Science. Data Mining. Text Mining. Процесс интеллектуального анализа текста. Машинное обучение. Машинное обучение с учителем (деревья решений, метод опорных векторов, регрессия, метод k ближайших соседей, нейронные сети). Классификация текстов. Подходы к классификации текстов (наивный байесовский классификатор, классификатор Роккио, бутстрэппинг). Применение классификации текстов. Кластеризация текстов. Анализ тональности. Извлечение информации, извлечение именованных сущностей, извлечение отношений. Тематическое моделирование (LDA, LSA).	https://edu.vsu.ru/course/view.php?id=3923
2. Практические занятия			
2.1			
2.2			
3. Лабораторные занятия			
3.1	Получение данных	Сбор данных. Парсинг. ParseHub, OctoParse. Источники данных для интеллектуального анализа текста: The American Presidency Project, arXiv Bulk Data Access, Category:Dataset, Cmu Movie Summary Corpus, Congressional and Federal Government Web Harvests, Congressional Record, Consumer Complaint Database, Corpus of Contemporary American English, Documentcloud, Ebsco Newspaper Source, Corpus of Global Web-based English, Hathitrust, Internet Archive, Jstor for Research, Lexisnexis Academic, Observatory on Social Media, Openlibrary, Public.Resource.Org, Pubmed, Robots Reading Vogue, Text Creation Partnership, The @Unitedstated Project, University of Oxford Test Archive, Yahoo Webscope Program	https://edu.vsu.ru/course/view.php?id=3923
3.2	Лексические ресурсы	Лексические ресурсы: WordNet, Global WordNet, WordNet Domains, WordNet-Affect, тезаурус Роже, Linguistic Inquiry and Word Count (LIWC), General Inquirer, SentiWordNet, OpinionFinder, Википедия, Викисловарь, BabelNet. Конкордансеры: Adelaide Text Analysis Tool, AntConc, Simple Concordance Program, TextSTAT, Wmatrix, Wordsmith. Выравнивание битекстов: ABBYY Aligner. Базы данных: SQL.	https://edu.vsu.ru/course/view.php?id=3923

3.3	Базовая обработка текста	Программное обеспечение для подготовки и очистки текста: регулярные выражения, Adobe Acrobat, BBEEdit, OpenRefine, TextCleanr, TextPipe, TextSoap, Trifacta Wrangler, UltraEdit. Программное обеспечение для обработки текста: NLTK, Stanford CoreNLP toolkit, LingPipe, стеммер Портера и др., WordNet Similarity, word2vec.	https://edu.vsu.ru/course/view.php?id=3923
3.4	Анализ и обработка данных	Морфологическая разметка: MyStem. Обработка естественного языка с помощью Python: Библиотеки Python для анализа данных. Программное обеспечение для анализа текста: Leximancer, Linguistic Inquiry and Word Count (LIWC), RapidMiner, TextAnalyst, WordStat. Тематическое моделирование: Topic Modeling Tool. Анализ тональности текста: Lexicoder, OpinionFinder, RapidMiner Sentiment Analysis, SAS Sentiment Analysis Studio. Качественный анализ данных: ATLAS.ti, Dedoose, f4analyse, HyperRESEARCH, Kwalitan, MAXQDA, NVivo, QDA Miner, Qualrus, Quirkos, AQUAD, Cassandre, Coding Analysis Toolkit, CATMA, Compendium, FreeQDA, libreQDA, Open Code, QDA Miner Lite, RQDA, Saturate, Text Analysis Markup System, Text Analysis Markup System Analyzer. Статистические инструменты: коэффициенты надежности, дисперсионный анализ, критерий хи-квадрат, регрессия; STATA, SPSS, SAS, R.	https://edu.vsu.ru/course/view.php?id=3923
3.5	Методы текстового анализа	Программное обеспечение для нарративного анализа (SAS, PC-ACE, UCINET, NetDraw, NVivo). Программное обеспечение для анализа метафор (ATLAS.ti, MAXQDA, NVivo, TACT, TextAnalyst, Wmatrix, WordSmith).	https://edu.vsu.ru/course/view.php?id=3923
3.6	Методы интеллектуального анализа текстов	Программное обеспечение для машинного обучения с учителем: Weka, Scikit, Caret, Theano, SVM-Light, Decision trees implementations. Программное обеспечение для классификации текстов: MALLET, AI::Categorizer, NLTK, Weka, Scikit. Программное обеспечение для анализа тональности: OpinionFinder, Weka, Scikit. Программное обеспечение для извлечения информации: GATE, ReVerb, DeepDive, MITIE. Программное обеспечение для тематического моделирования: MALLET, Alceste.	https://edu.vsu.ru/course/view.php?id=3923
3.7	Представление результатов лингвистических исследований	Организация данных в программе EXCEL (сортировка, статистическая обработка языковых данных). Представление информации в виде диаграмм, гистограмм, таблиц. Визуализация текстовых данных: Облака слов (Wordle, TagCrowd, QDAS, Sparklounds), деревья слов и сети фраз (IBM Many Eyes, NVivo, QDAS), матрицы и карты (QDAS, ATLAS.ti, MAXQDA, NVivo). Программное обеспечение для анализа тем (NVivo, Dedoose). Gephi. Tableau.	https://edu.vsu.ru/course/view.php?id=3923

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				
		Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1	Получение данных	2	-	2	8	12
2	Этика исследования	2	-	-	4	6
3	Философия и логика	2	-	-	4	6

	интеллектуального анализа текста					
4	Разработка исследовательского проекта	2	-	-	4	6
5	Веб-скрейпинг и краулинг	2	-	-	4	6
6	Лексические ресурсы	2	-	2	8	12
7	Базовая обработка текста	2	-	2	8	12
8	Анализ и обработка данных	-		4	8	12
9	Методы текстового анализа	2	-	2	8	12
10	Методы интеллектуального анализа текстов	2	-	4	8	14
11	Представление результатов лингвистических исследований	-	-	2	8	10
	Итого:	18	-	18	72	108

14. Методические указания для обучающихся по освоению дисциплины

Необходимо регулярное посещение лекционных и лабораторных занятий, работа с литературой по дисциплине, выполнение индивидуальных лабораторных работ. Самостоятельная работа обучающихся предусматривает подготовку к аудиторным и лабораторным занятиям; выполнение домашних заданий; подготовку презентаций.

При изучении материала учебной дисциплины по учебнику нужно, прежде всего, уяснить существо каждого излагаемого там вопроса. Главное - это понять изложенное в учебнике, а не «заучить».

Изучать материал рекомендуется по темам конспекта лекций и по главам (параграфам) учебника (учебного пособия). Сначала следует прочитать весь материал темы (параграфа), особенно не задерживаясь на том, что показалось не совсем понятным: часто это становится понятным из последующего. Затем надо вернуться к местам, вызвавшим затруднения и внимательно разобраться в том, что было неясно.

Особое внимание при повторном чтении необходимо обратить на формулировки соответствующих определений, формулы и т.п. (они обычно бывают набраны в учебнике курсивом); в точных формулировках, как правило, существенно каждое слово и очень полезно понять, почему данное положение сформулировано именно так. Однако не следует стараться заучивать формулировки; важно понять их смысл и уметь изложить результат своими словами.

Закончив изучение раздела, полезно составить краткий конспект, по возможности, не заглядывая в учебник (учебное пособие).

При изучении учебной дисциплины особое внимание следует уделить приобретению навыков решения профессионально-ориентированных задач. Для этого, изучив материал данной темы, надо сначала обязательно разобраться в решениях соответствующих задач, которые рассматривались на практических занятиях, приведены в учебно-методических материалах, пособиях, учебниках, ресурсах Интернета, обратив особое внимание на методические указания по их решению.

Закончив изучение раздела, нужно проверить умение ответить на все вопросы программы курса по этой теме (осуществить самопроверку).

Все вопросы, которые должны быть изучены и усвоены, в программе перечислены достаточно подробно. Однако очень полезно составить перечень таких вопросов самостоятельно (в отдельной тетради) следующим образом:

- начав изучение очередной темы программы, выписать сначала в тетради последовательно все перечисленные в программе вопросы этой темы, оставив справа широкую колонку;

- по мере изучения материала раздела (чтения учебника, учебно-методических пособий, конспекта лекций) следует в правой колонке указать страницу учебного издания (конспекта лекции), на которой излагается соответствующий вопрос.

В результате в этой тетради будет полный перечень вопросов для самопроверки, который можно использовать и при подготовке к экзамену. Кроме того, ответив на вопрос или написав соответствующую формулу, можете по учебнику (конспекту лекций) быстро проверить, правильно ли это сделано, если в правильности своего ответа Вы сомневаетесь. Наконец, по тетради с такими вопросами Вы можете установить, весь ли материал, предусмотренный программой, Вами изучен.

Следует иметь в виду, что в различных учебных изданиях материал может излагаться в разной последовательности. Поэтому ответ на какой-нибудь вопрос программы может оказаться в другой главе, но на изучении курса в целом это, конечно, никак не скажется.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины

а) основная литература:

№ п/п	Источник
1	Интеллектуальные информационные системы и технологии : учебное пособие / Ю. Ю. Громов, О. Г. Иванова, В. В. Алексеев и др. – Тамбов : Тамбовский государственный технический университет (ТГТУ), 2013. – 244 с. – URL: https://biblioclub.ru/index.php?page=book&id=277713

б) дополнительная литература:

№ п/п	Источник
2	Ignatow G. Text Mining: A Guidebook for the Social Sciences / G. Ignatow, R. F. Mihalcea. - SAGE Publications, Inc, 2016 – 208 p.
3	Самойленко, А.П. Информационные технологии статистической обработки данных : учебное пособие / А.П. Самойленко, О.А. Усенко ; Южный федеральный университет, Инженерно-технологическая академия. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2017. – 127 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=500042
4	Жуковский, О.И. Информационные технологии и анализ данных : учебное пособие / О.И. Жуковский ; Томский Государственный университет систем управления и радиоэлектроники (ТУСУР). – Томск : Эль Контент, 2014. – 130 с. : схем., ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=480500
5	Лазницас, Е.А. Базы данных и системы управления базами данных : учебное пособие : [12+] / Е.А. Лазницас, И.Н. Загумёникова, П.Г. Гилевский. – Минск : РИПО, 2016. – 267 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=463305
6	Технологии обработки информации : учебное пособие / авт.-сост. Н.В. Кандаурова, В.С. Чеканов ; Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2014. – 175 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=457753
7	Серегин, М.Ю. Интеллектуальные информационные системы : учебное пособие / М.Ю. Серегин, М.А. Ивановский, А.В. Яковлев ; Тамбовский государственный технический университет. – Тамбов : Тамбовский государственный технический университет (ТГТУ), 2012. – 205 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=277790
8	Щелоков, С.А. Базы данных : учебное пособие / С.А. Щелоков ; Оренбургский государственный университет, Кафедра программного обеспечения вычислительной техники и автоматизированных систем. – Оренбург : Оренбургский государственный университет, 2014. – 298 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=260752
9	Хныкина, А.Г. Информационные технологии : учебное пособие / А.Г. Хныкина, Т.В. Минкина ; Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2017. – 126 с. : схем., ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=494703
10	Чубукова И.А. Data Mining / И.А. Чубукова. - Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Бином. Лаборатория знаний, 2008 - 383 с. Режим доступа: http://www.intuit.ru/departament/database/datamining/
11	Грас Дж. Data Science. Наука о данных с нуля / Дж. Грас. – пб.: БХВ-Петербург, 2017. – 336 с.
12	Скиена, Стивен С. Наука о данных: учебный курс. / Стивен С. Скиена. - СПб. : ООО "Диалектика", 2020. - 544 с.
13	Силен Д. Основы Data Science и Big Data. Python и наука о данных / Д. Силен, А. Мейсман, М. Али. – СПб.: Питер, 2017. – 336 с.
14	Андерсон, К. Аналитическая культура. От сбора данных до бизнес-результатов / К. Андерсон. — М. : Манн, Иванов и Фербер, 2017. — 336 с.
15	Пучковская А.А. Введение в цифровые гуманитарные исследования / А.А. Пучковская, Л.В. Зимина, Д.А. Волков – СПб: Университет ИТМО, 2021. – 61 с.

16	Базалева, О. И. Мастерство визуализации данных. / О.И. Базалева - СПб.: 000 "Диалектика", 2018.-192 с.
17	Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP : учебное пособие по специальности 071900 "Информационные системы и технологии" направления 654700 "Информационные системы" / А.А. Барсегян [и др.] — 2-е изд. — СПб : БХВ-Петербург, 2007 .— 375 с.
18	Макленнен, Дж. Microsoft SQL Server 2008 Data mining. Интеллектуальный анализ данных : [пер. с англ.] / Дж. Макленнен, Ч. Танг, Б. Криват .— СПб : БХВ-Петербург, 2009 .— 700 с.
19	Jiawei H., Micheline K., Jian P. "Data Mining: Concepts and Techniques", Third Edition – July 6, 2011. Режим доступа: http://web.engr.illinois.edu/~hanj/bk2/toc.pdf
20	Бенгфорт Б., Билбро Р., Охеда Т. Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. — СПб.: Питер, 2019. — 368 с.
21	Мастичкий С.Э., Шитиков В.К. Статистический анализ и визуализация данных с помощью R. — 2014. Электронная книга, адрес доступа: http://r-analytics.blogspot.com
22	Litosseliti L. Research Methods in Linguistics. - Bloomsbury Publishing. 2018. — 377 p.
23	Walkenbach J., Alexander M. Excel VBA Programming For Dummies. - John Wiley & Sons. 2019. — 415 p.
24	Harvey G. Excel Timesaving Techniques For Dummies, - Wiley Publishing, 2005. — 410 p.
25	Bluttman K. Excel Formulas and Functions For Dummies. - John Wiley & Sons. 2013. — 410 p.
26	Harvey G. Excel 2019 All-in-One For Dummies. - John Wiley & Sons. 2019. — 819 p.
27	Айзек М.П., Финков М.В. Графики, анализ данных в Excel: пошаговые примеры. — СПб.: Наука и Техника, 2019. — 384 с.
28	Комарова З. И. Методология, метод, методика и технология научных исследований в лингвистике: учебное пособие. — Екатеринбург: Изд-во УрФУ, 2012. — 818 с.
29	North M. Data Mining for the Masses. - A Global Text Project Book, 2012. — 264 p.
30	Frye C. D. Microsoft Excel 2013: Step by Step. - O'Reilly Media, 2013. — 505 p.
31	Каптерев А. Мастерство презентации. Как создавать презентации, которые могут изменить мир / Алексей Каптерев ; пер. с англ. С. Кировой. — М. : Манн, Иванов и Фербер, Эксмо, 2012. — 336 с.
32	Thomas J. Discovering English with Sketch Engine: Workbook & Glossary. — VERSATILE, 2016. — 141 p.
33	Harvey G. Excel Workbook For Dummies. - Wiley Publishing, 2006. — 371 p.
34	Ismay C., Kim A. Y. An Introduction to Statistical and Data Sciences via R. - Chapman and Hall/CRC, 2018. — 216 p.
35	Даниленко В.П. Методы лингвистического анализа : курс лекций / В.П. Даниленко. — М. : ФЛИНТА : Наука, 2011. — 280 с.
36	Mirkin B. Core Concepts in Data Analysis: Summarization, Correlation and Visualization. — Springer, 2011. — 402 p.
37	Арнольд И.В. Основы научных исследований в лингвистике: Учеб. пособие. — М.: Высш. шк., 1991. — 140 с.
38	Шипунов А.Б. Наглядная статистика. Используем R / А. Б. Шипунов, Е.М. Балдин, П.А. Волкова, А.И. Коробейников, С.А. Назарова, С.В. Петров, В.Г. Суфиянов. — 2014. — 296 с.

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
39	ЭБС Лань. — Режим доступа: по подписке. — URL: ЭБС Лань (lanbook.com)
40	ЭБС «Университетская библиотека онлайн». — Режим доступа: по подписке. — URL: ЭБС "Университетская библиотека онлайн" читать электронные книги (biblioclub.ru)
41	ЭБС ЮРАЙТ.— Режим доступа: по подписке. — URL: Образовательная платформа Юрайт. Для вузов и ссузов. (urait.ru)

16. Перечень учебно-методического обеспечения для самостоятельной работы

№ п/п	Источник
1	Донина, О.В. Введение в анализ текстовых данных: учебное пособие / О.В. Донина., К.А. Сидоров, Н.С. Горбунов – Воронеж : Издательский дом ВГУ, 2022. — 105 с.
2	Донина О.В. Автоматизация лингвистических исследований: учебное пособие / О.В. Донина. – Воронеж : Издательский дом ВГУ, 2022. — 125 с.
3	Донина О.В. Введение в анализ данных для лингвиста: учебное пособие / О.В. Донина – Воронеж : Издательский дом ВГУ, 2024. — 115 с.

4	Пушкарёва, Т.П. Основы компьютерной обработки информации : учебное пособие / Т.П. Пушкарёва ; Сибирский федеральный университет. – Красноярск : Сибирский федеральный университет (СФУ), 2016. – 180 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=497475
5	Исакова, А.И. Информационные технологии : учебное пособие / А.И. Исакова ; Томский Государственный университет систем управления и радиоэлектроники (ТУСУР), Кафедра автоматизированных систем управления (АСУ). – Томск : ТУСУР, 2013. – 207 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=480610
6	Гасанов, Э. Э. Интеллектуальные системы. Теория хранения и поиска информации : учебник для вузов / Э. Э. Гасанов, В. Б. Кудрявцев. — 2-е изд., испр. и доп. — Москва : Издательство Юрайт, 2020. — 271 с. — (Высшее образование). — ISBN 978-5-534-08684-3. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: https://urait.ru/bcode/452220

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение):

При реализации дисциплины могут проводиться различные типы лекций (вводная, обзорная и т.д.). При проведении лабораторных работ предпочтение отдается применению классических технологий: обсуждение со студентами заранее подготовленных ими тем и разбор практических задач и лабораторных работ.

18. Материально-техническое обеспечение дисциплины:

/ауд. 12/ - компьютерный класс: Компьютер Arbyte Tempo/AOC (12 шт.), Проектор Benq MW523 (1 шт.), Сканер Canon Canoscan LiDE 120 (5 шт.) Экран проекционный (1 шт.)

19. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция	Индикаторы достижения компетенции	Оценочные средства
1	1 Получение данных	ПК-3	Выбирает адекватный способ представления результатов исследований в соответствующей области знаний (ПК-3.1)	Выполнение индивидуальных лабораторных работ, тестовые задания
	2 Этика исследования			
	3 Философия и логика интеллектуального анализа текста	ПК-4	Собирает, обрабатывает, анализирует и обобщает результаты собственных исследований (ПК-4.2)	
	4 Разработка исследовательского проекта			
	5 Веб-скрейпинг и краулинг	ПК-5	Осуществляет постановку задачи на технологические исследования (ПК-5.1)	
	6 Лексические ресурсы		Анализирует результаты технологических исследований (ПК-5.2)	
	7 Базовая обработка текста	ПК-7	Разрабатывает и документирует программные интерфейсы (ПК-7.1)	
	8 Анализ и обработка данных		Пользуется электронными языковыми ресурсами для решения прикладных задач (ПК-7.2)	
	9 Методы текстового анализа		Анализирует требования к программному обеспечению (ПК-	

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция	Индикаторы достижения компетенции	Оценочные средства
	10 Методы интеллектуального анализа текстов		7.3)	
	11 Представление результатов лингвистических исследований			
Промежуточная аттестация форма контроля - зачет				Практические задания

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств:

Лабораторные работы, в том числе домашние задания, тестовые задания,

Тестовые задания

- Семантический словарь, т. е. словарь, в котором представлены смысловые связи слов — синонимические, отношения Род-Вид (иногда называемые отношением Выше-Ниже), Часть-Целое, ассоциации, пр.
 - База данных
 - Тезаурус
 - Корпус
 - Онтология
- Набор понятий, сущностей определенной области знаний, ориентированный на многократное использование для различных задач.
 - Онтология
 - Корпус
 - Лингвистическая разметка
 - Словарь
- Что из перечисленного является примером онтологии?
 - GloWbE
 - PyТез
 - НКРЯ
 - WordNet
- Какой ресурс Вы будете использовать для морфологического анализа русского языка:
 - Sublime
 - Gephi
 - Skuuper
 - MyStem
- Процедура снятия неоднозначности – это _____ .
- _____ -слова – слова, знаки и символы, которые не обладают значением сами по себе, например, предлоги, частицы и местоимения.
- _____ выражения – формальный язык поиска, основанный на использовании специальных метасимволов.
- N-_____ – последовательности из N слов
- Метод, создающий базу для последующего морфологического и синтаксического анализа на основе выделения слов, цифровых комплексов, формул и т. д.

- a. Семантический анализ
- b. Классификация
- c. Графематический анализ
- d. Кластеризация

10. _____ – этап, в рамках которого происходит разделение текста на более мелкие единицы — на предложения и слова.
11. Какую программу Вы будете использовать для визуализации данных:
- a. Tableau
 - b. MAMP
 - c. MyStem
 - d. ParseHub
12. Тематическое _____ – построение модели коллекции текстовых документов, которая определяет, к каким темам относится каждый из документов.
13. _____ – это список всех употреблений заданного языкового выражения в контексте.
14. _____ данных – совокупность данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными.
15. Text _____ – извлечения информации из неструктурированных текстовых данных.
16. Корпусный _____ – поисковая система по текстовым данным, которая предоставляет статистику о языковых единицах и приводит информацию в удобный для анализа вид.
17. Какая технология компьютерной лингвистики поможет сосчитать в тексте количество глаголов и прилагательных?
- a. морфологический анализатор
 - b. распознавание символов
 - c. анализ тональности
 - d. извлечение ключевых слов
18. Какая технология компьютерной лингвистики поможет сократить текст, убрав из него все причастные и деепричастные обороты?
- a. синтаксический анализ
 - b. извлечение именованных сущностей
 - c. анализ жанров
 - d. поиск плагиата
19. Какая технология компьютерной лингвистики поможет проанализировать комментарии к одной из нашумевших новостей и оценить отношение к событию?
- a. морфологический анализ
 - b. синтаксический анализ
 - c. реферирование текстов
 - d. анализ тональности
20. Какая технология компьютерной лингвистики поможет проанализировать массив текстов и понять, о чем они?
- a. поиск плагиата
 - b. извлечение именованных сущностей
 - c. распознавание символов
 - d. извлечение ключевых слов
21. Облачный сервис для работы с кодом и библиотеками Python:
- a. R Studio
 - b. RapidMiner
 - c. GATE
 - d. Google Colaboratory
22. Библиотека для получения векторных представлений слов на основе их совместной встречаемости в текстах.
- a. TEI
 - b. word2vec

- c. tf-idf
 - d. regex
23. Модель, часто используемая при обработке текстов, представляющая собой неупорядоченный набор слов, входящих в обрабатываемый текст.
- a. word2vec
 - b. tf-idf
 - c. doc2vec
 - d. Bag of Words
24. Какое онлайн-приложение вы выберете для визуализации текста в виде облаков слов?
- a. AntConc
 - b. ELAN
 - c. Wordle
 - d. Praat
25. Какой ресурс вы будете использовать для построения карт?
- a. Topic Modelling Tool
 - b. ABBYY Aligner
 - c. Voyant tools
 - d. DataWrapper
26. Автоматическое нахождение основы слова:
- a. Лемматизация
 - b. Стемминг
 - c. Кластеризация
 - d. Анализ тональности
27. Автоматическое приведение словоформ к начальной форме:
- a. Стемминг
 - b. Разметка
 - c. Лемматизация
 - d. Tf-idf
28. Поисковая функция в Excel, которая находит значения в одной таблице и переносит их в другую.
- a. ВПР
 - b. ПСТР
 - c. ЗНАЧЕН
 - d. ПОИСКПОЗ
29. Функция в Excel, которая возвращает указанное число знаков из текстовой строки, начиная с указанной позиции; возвращает часть текста из середины строки.
- a. ЛЕВСИМВ
 - b. ПСТР
 - c. ПРАВСИМВ
 - d. ИНДЕКС
30. Какая функция в Excel позволит заменить все вхождения «ОАО» на «ЗАО»?
- a. ВПР
 - b. ПОИСК
 - c. ПОДСТАВИТЬ
 - d. ЗНАЧЕН
31. Какой из признаков не может быть измерен номинальной шкалой?
- a. Район города, в котором живет респондент
 - b. Образование респондента
 - c. Доход респондента
 - d. Номинальной шкалой можно измерить любой признак
32. Какие меры будут оценивать центр и вариативность распределения более точно?
- a. Мода и размах
 - b. Медиана и межквартильный размах

- c. Среднее арифметическое и СКО
 - d. Усеченное среднее и дисперсия
33. Закон _____ гласит следующее: Если все слова языка (или просто относительно длинного текста) упорядочить по убыванию частоты их использования, то частота n-го слова в таком списке окажется приблизительно обратно пропорциональной его порядковому номеру n (так называемому рангу этого слова).
34. _____ — статистическая мера, используемая для оценки важности слова в контексте документа, являющегося частью коллекции документов.
35. Мера _____ – метрика, показывающая все ли правильные примеры нашел классификатор и определяющаяся как количество найденных правильных примеров к общему количеству релевантных примеров в корпусе.
36. Мера _____ – метрика, отражающая количество релевантных категории примеров к общему количеству классифицированных как правильные примеров.
37. Метрика в задачах машинного обучения, являющаяся средним гармоническим между полнотой и точностью.
- a. F-Score
 - b. Precision
 - c. Recall
 - d. Accuracy
38. Пакет библиотек и программ для символьной и статистической обработки естественного языка, написанных на языке программирования Python:
- a. NLTK
 - b. Matplotlib
 - c. Pandas
 - d. SciPy
39. _____ текста – определение подлинности или подложности рукописного текста и установление его автора.
40. _____ – исследование и измерение стилевых характеристик текста с целью установления авторства или получения каких-либо сведений об авторе и условиях создания текстового документа.
41. Natural Language _____ – обработка естественного языка с помощью компьютерных технологий.

Лабораторная работа 1

Используя текстовые формулы в программе Excel, выполните следующие задания (таблицы с данными представлены в курсе <https://edu.vsu.ru/course/view.php?id=3923>):

- 1. Вывести инициалы по схеме "И.О."
- 2. Определить пол сотрудника
- 3. Вывести фамилии без пробелов
- 4. Вывести фамилии с буквами "Е", вместо "Ё"
- 5. Составить описание по схеме Фамилия Имя: пол, г.р. – 0
- 6. Определить длину текста Описания
- 7. Вычислить позицию первого пробела в Описании
- 8. Выделить Фамилию из Описания
- 9. Вычислить позицию двоеточия в Описании
- 10. Выделить Имя из Описания
- 11. Объединить данные столбцов Фамилия, Имя и Отчество в одну текстовую строку
- 12. Вывести ФИО без лишних пробелов
- 13. Вывести ФИО так, чтобы:
 - a. все буквы были прописными
 - b. все буквы были строчными

- с. первая буква в каждом слове была прописной, а последующие – строчными
14. Разбить ФИО на составляющие с помощью команды "Текст по столбцам"
 15. Вывести фамилию из ФИО
 16. Вывести индекс из Адреса
 17. Вывести город из Адреса
 18. Выделить три основные цифры Номера авто
 19. Выделить код региона (последний цифры Номера авто)
 20. В столбце [Инициал] (диапазон E5:E22) вывести первую букву имени
 21. В столбце [Пол] (диапазон F5:F22) определить пол по фамилии - если фамилия заканчивается на [а] вывести "Жен." и для всех остальных вывести "Муж."
 22. В столбце [Фамилия без пробелов] (диапазон G5:G22) из столбца [А] вывести фамилии без лишних пробелов
 23. В столбце [Описание] (диапазон H5:H22) сцепить из частей текст по схеме: [Фамилия без пробелов], [Инициал], [точка], [пробел], [Пол (в скобках)], [запятая], [слово возраст], [пробел], [значение возраста] - Например, Яковлева Е. (Жен.), возраст 24
 24. В столбце [Длина описания] (диапазон I5:I22) определить длину описания (кол-во символов)
 25. В столбце [Где 1-ая точка в Описании] (диапазон J5:J22) определить номер позиции первой точки в тексте Описания
 26. В столбце [Где открывающаяся скобка] (диапазон K5:K22) определить номер позиции открывающейся скобки в тексте Описания
 27. В столбце [Где закрывающаяся скобка] (диапазон L5:L22) определить номер позиции закрывающейся скобки в тексте Описания
 28. В столбце [Фамилия из Описания] (диапазон M5:M22) вывести фамилию из текста Описания
 29. В столбце [Пол из Описания] (диапазон N5:N22) вывести пол из текста Описания
 30. В столбце [Фамилии без "ё"] (диапазон O5:O22) заменить букву "ё" на букву "е" в фамилиях из столбца [Фамилия без пробелов]

Лабораторная работа 2

По аналогии с приведенным на занятии примером, используя ParseHub, извлеките текстовые данные с информационного ресурса по вашему выбору: СМИ, блоги, форумы, социальные сети, новостные источники отдельных сфер (кино, сериалы, мода, музыка, медицина и т.п.). В качестве дополнительной информации к текстовым данным совершите загрузку: заголовка, раздела, даты публикации, автора и тегов (при наличии). Извлекаемые тексты не должны повторяться. Минимальный объем текстовых данных: 25 000 словоформ. В качестве отчетности необходимо предоставить документ в формате XLSX.

Лабораторная работа 3

По аналогии с приведенным на занятии примером, проведите морфологический анализ текстовых данных (без метаданных), которые были получены в ходе выполнения Лабораторной работы 2.

Для извлечения необходимых материалов можете воспользоваться программой для работы с электронными таблицами (например, Microsoft Excel), скопировать столбец с текстовыми данными и разместить его в файле input.txt.

В качестве отчетности необходимо предоставить полученные результаты морфологической разметки и лемматизации в виде документа в формате TXT.

Лабораторная работа 4

1. Скачайте набор данных из 76 поваренных книг: <https://lib.msu.edu/feedingamericadata/>

2. Загрузите их и извлеките столбец «region». Выведите количество уникальных регионов в наборе данных.
3. Используя тот же столбец «region», отобразите частотное распределение регионов в данных.
4. Создайте гистограмму различных регионов, аннотированных в наборе данных.
5. Используйте функцию `plot_trend()`, чтобы создать график использования в рецептах трех ингредиентов на ваш выбор.

В качестве отчетности предоставьте полученные результаты в виде графиков и кода.

Лабораторная работа 5

По аналогии с приведенным на занятии примером проведите тематическое моделирование. В качестве материала исследования можете использовать художественное произведение (с разделением на главы или с делением на равные по объему части) или несколько новостных/научных статей по одной теме. Минимальное количество входных файлов (глав/статей/частей) равняется шести. Количество тем должно быть не менее семи, количество итераций: 5 000. Исходя из употребляемости тем и их ключевых слов, опишите приблизительное содержание как каждого файла, так и всей текстовой выборки. В качестве отчетности необходимо предоставить итоговый файл «all_topics.html» с темами, а также текстовый документ с интерпретацией полученных результатов.

Лабораторная работа 6

Импортируйте текстовые данные (без метаданных), полученные в ходе выполнения Лабораторной работы 2, в AntConc.

1. Определите 10 самых используемых словоформ в тексте, отсортировав их в порядке частотности;
2. Найдите контекстные употребления самого часто используемого слова;
3. Во вкладке Plot создайте позиционные графики слов, входящих в топ-10 по частоте употребления;
4. Создайте списки кластеров со словами, входящими в топ-5 по частоте употребления;
5. Найдите 10 коллокаций со словом на ваш выбор;
6. * Выполните пункты 1-5, предварительно проведя лемматизацию текста при помощи MyStem; сопоставьте полученные результаты с лемматизацией и без нее.

В качестве отчета предоставьте текстовый документ с анализом полученного материала и список слов.

Лабораторная работа 7

По аналогии с приведенным на занятии примером, воспользовавшись ABBYY Aligner и Notepad++, выровняйте параллельный текст.

1. Найдите текст нехудожественной литературы объемом не более 15 000 слов на иностранном языке и его перевод на русский язык в одном из поддерживаемых ABBYY Aligner форматов.
2. Импортируйте документы в ABBYY Aligner и запустите автоматическое выравнивание.
3. Найдите соседние строки, требующие слияния. Используйте инструмент «Объединить».
4. Перейдите к следующей ошибке и исправьте, используя соответствующий инструмент.
5. Найдите пустые строки и используйте инструмент «Удалить».
6. Исправьте все оставшиеся в проекте ошибки, используя доступные инструменты программы.
7. Экспортируйте выровненный текст в формате RTF.
8. Конвертируйте документ в формат TXT.
9. Откройте TXT документ в Notepad++ и удалите пустые строки.
10. Запишите макрос для создания пустой строки после каждых двух предложений.
11. Иницилируйте запуск макроса до конца документа.

В качестве отчетности предоставьте полученный результат (выровненный параллельный текст) в формате TXT.

Лабораторная работа 8

По аналогии с приведенным на занятии примером создайте терминологическую базу данных какой-либо сферы деятельности на ваш выбор. База должна включать в себя термин на русском языке, его перевод на английский, год возникновения, а также контекст использования на двух языках. Внесите в таблицу данные о 20 терминах.

В качестве отчетности предоставьте SQL-код выполненного задания.

Лабораторная работа 9

По аналогии с приведенным на занятии примером создайте граф с персонажами произведения на ваш выбор. Для выполнения задания воспользуйтесь материалами корпуса текстов драм и пьес DraCor.

В качестве отчетности представьте полученный граф в формате PNG, а также текстовый документ с краткой интерпретацией результатов:

1. Персонажи какого пола преобладают в произведении?
2. Кто обладает наибольшим и наименьшим количеством реплик?
3. У какого узла (персонажа) больше всего ребер (контактов с другими персонажами)?
4. Основываясь на ребрах графа, можно ли выявить неочевидные связи между персонажами?

Лабораторная работа 10

По аналогии с рассмотренным на занятии примером приведите данные вероятностной тематической модели, полученные в ходе выполнения Лабораторной работы 5, в вид графика с помощью программы Tableau.

В качестве отчетности предоставьте полученный график в формате PNG, дополнив его текстовым документом с описанием тем тематического моделирования (полным списком слов) и анализом результатов.

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: практические задания 1, 2

Практическое задание № 1

1) Проведите исследование, используя изученные в рамках курса методы. Опишите результаты исследования в виде научной статьи. Оформите статью в текстовом процессоре Microsoft Word.

2) Создайте титульный лист, содержание и введение в соответствии с темой вашей работы.

3) Текст работы должен соответствовать следующим требованиям:

- шрифт Times New Roman, размер 14, через 1,5 интервал, стиль "Обычный" (весь текст);
- абзац - 0,5 см; выравнивание по ширине;
- все инициалы в тексте набираются через неразрывный пробел ([Ctrl + Shift + Пробел] или команда "Вставка" - "Символ" - "Специальные символы" - "Неразрывный пробел");
- используйте список
- вставьте номера страниц, для этого: Вставка, номер страниц.

4) Список литературы дается по алфавиту в конце статьи с обязательной нумерацией и озаглавляется ЛИТЕРАТУРА (выравнивание по центру);

5) Ссылки на литературу оформляются следующим образом: Л.В. Щерба отмечал: "..."
[Щерба 1990, с. 1-5].

Практическое задание № 2

Проведите исследование, используя изученные в рамках курса методы. Оформите результаты исследования в виде презентации в программе Power Point. Следуйте нижеизложенным советам.

1. Титульная страница необходима, чтобы представить аудитории Вас и тему Вашего доклада.

2. Оптимальное число строк на слайде — от 3 до 6. Перегруженность и мелкий шрифт тяжелы для восприятия.

3. Пункты перечней должны быть короткими фразами, тезисами. Чтение длинной фразы отвлекает внимание от речи. Короткая фраза легче запоминается визуально.

4. Распространённая ошибка — читать слайд дословно. Лучше всего, если на слайде будет написана подробная информация (определения, теоремы, формулы), а словами будет рассказываться их содержательный смысл. Информация на слайде может быть более формальной и строго изложенной, чем в речи.

5. Речь и слайды не должны совпадать, тогда презентация станет «объёмной». Речь должна быть более популярна и образна. Слайды должны содержать больше технических подробностей: формулы, схемы, таблицы, графики.

6. Над каждой фразой надо критически подумать: поймут ли её слушатели; достаточно ли у них специальных знаний, чтобы её понять? Непонятные фразы следует безжалостно изымать из презентации.

7. Любая фраза должна говорить за чем-то. Не просто потому, что Вы этим занимались в процессе работы. Каждая фраза должна логично подводить к следующим фразам, быть для них посылкой, и в конечном итоге всё выступление должно быть подчинено главной цели — донести до аудитории две-три по-настоящему ценных мысли. Тогда выступление будет цельным и оставит хорошее впечатление.

8. Имеет смысл быть аккуратным. Неряшливо сделанные слайды (разной в шрифтах и отступах, опечатки, типографические ошибки в формулах) вызывают подозрение, что и к содержательным вопросам докладчик подошёл спустя рукава.

9. Последний слайд – Спасибо за внимание.

10. Анимация в презентации допустима в редких случаях, если её цель информативная, а не развлекательная.

Описание технологии проведения

Студенты представляют доклады-презентации в рамках специально организованной учебной мини-конференции. Регламент выступления и последующего обсуждения доклада-презентации – не более 15 мин.

Требования к выполнению заданий (или шкалы и критерии оценивания)

Качество доклада оценивается по двухбалльной шкале: зачтено или не зачтено. Критериями оценки качества доклада являются 1. содержание (полнота, соответствие содержания теме доклада, раскрытие основных аспектов темы); 2. презентабельность (качество изложения материала, ораторские навыки, качество слайд-презентации). Каждый критерий оценивается 15-ю баллами. Оценка «зачтено» ставится, если качество доклада получает оценку 18-30 баллов. Оценка «не зачтено» ставится в том случае, если качество доклада получает оценку ниже 18 баллов.

Для оценивания результатов обучения на зачете используется 2-х балльная шкала: «Зачтено», «Не зачтено».

Соотношение показателей, критериев и шкалы оценивания результатов обучения.

Критерии оценивания компетенций	Уровень сформированности компетенций	Шкала оценок
---------------------------------	--------------------------------------	--------------

Полное соответствие ответа обучающегося всем перечисленным критериям. Продemonстрировано знание основных типов систем, использующих модули лингвистического анализа; основных принципов и методов компьютерного моделирования лингвистических задач, а также основных принципов обработки информации; базовых принципов корпусной лингвистики, лексикографии, математической статистики; базовых представлений о языковом разнообразии; наиболее полных и значимых лингвистических корпусов, электронных словарей и баз данных; умение анализировать работу различных систем обработки текста и звучащей речи для выявления основных лингвистических компонентов и основных типов обработки текста, используемых в данных системах; подбирать необходимые лингвистические ресурсы для различных задач лингвистического обеспечения систем (например, лексикографических, задач морфологического анализа и т.п.); пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами; владение навыками работы с различными системами автоматической и экспертной обработки текста и звучащей речи, а также разработки электронных языковых ресурсов; опытом применения основных методов, способов и средств получения, хранения, переработки информации.	Повышенный уровень	Зачтено
Ответ не соответствует одному (двум) из перечисленных показателей, но обучающийся дает правильные ответы на дополнительные вопросы. Недостаточно продемонстрировано знание основных типов систем, использующих модули лингвистического анализа; основных принципов и методов компьютерного моделирования лингвистических задач, а также основных принципов обработки информации; базовых принципов корпусной лингвистики, лексикографии, математической статистики; базовых представлений о языковом разнообразии; наиболее полных и значимых лингвистических корпусов, электронных словарей и баз данных; умение анализировать работу различных систем обработки текста и звучащей речи для выявления основных лингвистических компонентов и основных типов обработки текста, используемых в данных системах; подбирать необходимые лингвистические ресурсы для различных задач лингвистического обеспечения систем (например, лексикографических, задач морфологического анализа и т.п.); пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами; владение навыками работы с различными системами автоматической и	Базовый уровень	Зачтено

экспертной обработки текста и звучащей речи, а также разработки электронных языковых ресурсов; опытом применения основных методов, способов и средств получения, хранения, переработки информации.		
Ответ не соответствует любым двум(трем) из перечисленных показателей, обучающийся дает неполные ответы на дополнительные вопросы. Демонстрирует частичные знание основных типов систем, использующих модули лингвистического анализа; основных принципов и методов компьютерного моделирования лингвистических задач, а также основных принципов обработки информации; базовых принципов корпусной лингвистики, лексикографии, математической статистики; базовых представлений о языковом разнообразии; наиболее полных и значимых лингвистических корпусов, электронных словарей и баз данных; умение анализировать работу различных систем обработки текста и звучащей речи для выявления основных лингвистических компонентов и основных типов обработки текста, используемых в данных системах; подбирать необходимые лингвистические ресурсы для различных задач лингвистического обеспечения систем (например, лексикографических, задач морфологического анализа и т.п.); пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами; владение навыками работы с различными системами автоматической и экспертной обработки текста и звучащей речи, а также разработки электронных языковых ресурсов; опытом применения основных методов, способов и средств получения, хранения, переработки информации.	Пороговый уровень	Зачтено
Ответ не соответствует любым трем (четырем) из перечисленных показателей. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки при практическом применении приобретенных знаний; не может использовать программные средства для проведения анализа данных.	—	Не зачтено