

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.
10.06.2022 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.ДВ.06.02 Квантитативная лингвистика

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. Составитель программы: Донина Ольга Валерьевна, кандидат филол. наук, преподаватель кафедры теоретической и прикладной лингвистики

7. Рекомендована: Научно-методическим советом факультета РГФ, протокол № 8 от 23.05.2022 г.

8. Учебный год: 2025/2026

Семестр: 7

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются:

- формирование способностей пользоваться лингвистически ориентированными программными продуктами (ПКВ-3);
- овладение навыками оформления и представления результатов научного исследования (ПКВо-6).

Задачи учебной дисциплины.

Фиксация индикаторов компетенции ПК-3. Обучающийся, прошедший курс обучения дисциплине «Б1.8.07 Квантитативная лингвистика», в результате обучения:

ПК-3.1 - Осуществляет постановку задачи на технологические исследования

ПК-3.2 - Анализирует результаты технологических исследований.

Фиксация индикаторов компетенции ПК-7.

ПК-7.1 – Выбирает адекватный способ представления результатов исследований в соответствующей области знаний

ПК-7.2 – Оформляет результаты научно-исследовательских работ в соответствии с существующими требованиями

10. Место учебной дисциплины в структуре ООП: дисциплина Б1.В.ДВ.06.02

Квантитативная лингвистика входит в часть, формируемую участниками образовательных отношений, Блока Б1 учебного плана, дисциплина по выбору. Для ее успешного освоения необходимы базовые знания, умения и навыки, сформированные в процессе изучения дисциплин Б1.О.03 Иностранный язык (английский), Б1.О.11 Алгебра и начала анализа, Б1.О.17 Введение в теорию языка, Б1.О.25 Информатика и основы программирования, Б1.В.01 Проектирование баз данных, Б1.В.03 Введение в прикладную лингвистику, Б1.В.04 Технологии корпусной лингвистики, Б1.В.06 Символьные вычисления, Б1.В.09 Семантический WEB, Б1.В.10 Формальные модели в лингвистике. Изучение данной дисциплины предшествует освоению дисциплин Б1.О.27 Основные проблемы современной лингвистики, 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 | Способен пользоваться лингвистически ориентированными программными продуктами | ПК-3.1 | 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, |
| | | ПК-3.2 | Анализирует результаты технологических исследований | |

| | | | | |
|------|---|--------|--|--|
| | | | | лексикографических, задач морфологического анализа и т.п.). Владеть: навыком работы с различными системами автоматической и экспертной обработки текста и звучащей речи. |
| ПК-7 | Владеет навыками оформления и представления результатов научного исследования | ПК-7.1 | 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: навыком выступлений перед аудиторией, а также создания письменных работ для представления результатов научного исследования. |
| | | ПК-7.2 | Оформляет результаты научно-исследовательских работ в соответствии с существующими требованиями | |

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

Форма промежуточной аттестации: экзамен

13. Трудоемкость по видам учебной работы

| Вид учебной работы | | Трудоемкость | |
|--|--------------|--------------|--------------|
| | | Всего | По семестрам |
| | | | |
| Аудиторные занятия | | 36 | 36 |
| в том числе: | лекции | 18 | 18 |
| | практические | 18 | 18 |
| | лабораторные | - | - |
| Самостоятельная работа | | 36 | 36 |
| в том числе: курсовая работа (проект) | | - | - |
| Форма промежуточной аттестации (экзамен – 36 ч.) | | 36 | Экзамен |
| 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 | Количественные данные в лингвистических исследованиях | Естественный язык и язык математики как семиотические системы. Статистические закономерности в языковых структурах. Основные статистические категории: выборка и совокупности, типы переменных, их классификация. Меры центральной тенденции и рассеяния. Требования к выборочной лингвистической совокупности – репрезентативность и рациональный объем выборки. Репрезентативность; приемы, обеспечивающие надежную репрезентативность тематических выборок. Определение минимально достаточного объема выборки в грамматических, фонетико-фонологических и лексикологических исследованиях. Относительная ошибка и надежность. Корреляционный анализ в лингвистических исследованиях. Дисперсионный анализ. Кластерный анализ. Факторный анализ. | |
| 1.2 | Закон Ципфа — Мандельброта и его следствия | Частота как характеристика употребительности слова в тексте. Закон Ципфа и его уточнение. Понятие о частотных словарях и их применении в лингвистических исследованиях. Лексикография и частотные словари. Частота слова и продуктивность. Частота слова и многозначность. Частота слова и его возраст. | |
| 1.3 | Применение статистических методов в основных разделах лингвистики | Фоностатистика. Квантитативная типология Дж. Гринберга. Математическое моделирование диахронических процессов. Метод радиоуглеродного датирования. Постулаты классической глоттохронологии. Современная глоттохронология. Лексикостатистика. Статистические методы в анализе дискурса. Социолингвистические переменные, их оценивание. Психометрические методы. Контент-анализ, его методика и сущность с лингвистической точки зрения. Этапы подготовки и проведения контентанализа: выбор корпуса языковых данных; выбор концептуальной переменной и определение ее значений; выбор единицы кодирования; отбор кодировщиков и формулировка инструкций по кодированию; кодировка данных; подсчет данных и интерпретация результатов. Виды контент-анализа. жесткий и мягкий, содержательный и структурный. | |
| 1.4 | Опыт квантитативной лексикологии | Полевая модель организации лексического пространства. Критерии отбора лексики языка (функциональный, синтагматический, эпидигматический, парадигматический). Понятие лексико-семантического ядра. Параметрический анализ лексики как метод лексико-семантической типологии. Моделирование лексико-семантического пространства языка. | |
| 1.5 | Квантитативные методы в корпусной лингвистике | Исходные понятия корпусной лингвистики: проблемная область, корпус данных, корпус текстов, единица хранения, типы корпусов данных, способы представления и хранения корпуса данных. Требования к корпусу текстов как к выборочной лингвистической совокупности, отражающей проблемную область. Корпусные статистические методы анализа лексики. Квантитативные методы автоматического | |

| | | | |
|--------------------------------|---|---|--|
| | | выделения ключевых слов и терминов, устойчивых словосочетаний, синонимов. | |
| 1.6 | Квантитативная лингвистика: дешифровка текста, авторизация текста | Позиционная статистика как метод дешифровки. Задачи атрибуции текста и стилеметрия. Примеры применения дешифровки и авторизации. | |
| 2. Практические занятия | | | |
| 2.1 | Количественные данные в лингвистических исследованиях | Естественный язык и язык математики как семиотические системы. Статистические закономерности в языковых структурах. Основные статистические категории: выборка и совокупности, типы переменных, их классификация. Меры центральной тенденции и рассеяния. Требования к выборочной лингвистической совокупности – репрезентативность и рациональный объем выборки. Репрезентативность; приемы, обеспечивающие надежную репрезентативность тематических выборок. Определение минимально достаточного объема выборки в грамматических, фонетико-фонологических и лексикологических исследованиях. Относительная ошибка и надежность. Корреляционный анализ в лингвистических исследованиях. Дисперсионный анализ. Кластерный анализ. Факторный анализ. | |
| 2.2 | Закон Ципфа — Мандельброта и его следствия | Частота как характеристика употребительности слова в тексте. Закон Ципфа и его уточнение. Понятие о частотных словарях и их применении в лингвистических исследованиях. Лексикография и частотные словари. Частота слова и продуктивность. Частота слова и многозначность. Частота слова и его возраст. | |
| 2.3 | Применение статистических методов в основных разделах лингвистики | Фоностатистика. Квантитативная типология Дж. Гринберга. Математическое моделирование диахронических процессов. Метод радиоуглеродного датирования. Постулаты классической глоттохронологии. Современная глоттохронология. Лексикостатистика. Статистические методы в анализе дискурса. Социолингвистические переменные, их оценивание. Психометрические методы. Контент-анализ, его методика и сущность с лингвистической точки зрения. Этапы подготовки и проведения контентанализа: выбор корпуса языковых данных; выбор концептуальной переменной и определение ее значений; выбор единицы кодирования; отбор кодировщиков и формулировка инструкций по кодированию; кодировка данных; подсчет данных и интерпретация результатов. Виды контент-анализа. жесткий и мягкий, содержательный и структурный. | |
| 2.4 | Опыт квантитативной лексикологии | Полевая модель организации лексического пространства. Критерии отбора лексики языка (функциональный, синтагматический, эпидигматический, парадигматический). Понятие лексико-семантического ядра. Параметрический анализ лексики как метод лексико-семантической типологии. Моделирование лексико-семантического пространства языка. | |

| | | | |
|--------------------------------|---|---|--|
| 2.5 | Квантитативные методы в корпусной лингвистике | Исходные понятия корпусной лингвистики: проблемная область, корпус данных, корпус текстов, единица хранения, типы корпусов данных, способы представления и хранения корпуса данных. Требования к корпусу текстов как к выборочной лингвистической совокупности, отражающей проблемную область. Корпусные статистические методы анализа лексики. Квантитативные методы автоматического выделения ключевых слов и терминов, устойчивых словосочетаний, синонимов. | |
| 2.6 | Квантитативная лингвистика: дешифровка текста, авторизация текста | Позиционная статистика как метод дешифровки. Задачи атрибуции текста и стилеметрия. Примеры применения дешифровки и авторизации. | |
| 3. Лабораторные занятия | | | |
| 3.1 | | | |
| 3.2 | | | |

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 | Количественные данные в лингвистических исследованиях | 2 | 2 | - | 10 | 16 |
| 2 | Закон Ципфа — Мандельброта и его следствия | 2 | 2 | - | 10 | 16 |
| 3 | Применение статистических методов в основных разделах лингвистики | 4 | 4 | - | 14 | 20 |
| 4 | Опыт квантитативной лексикологии | 2 | 2 | - | 10 | 16 |
| 5 | Квантитативные методы в корпусной лингвистике | 4 | 4 | - | 14 | 20 |
| 6 | Квантитативная лингвистика: дешифровка текста, авторизация текста | 4 | 4 | - | 14 | 20 |
| | Итого: | 18 | 18 | - | 72 | 108 |

14. Методические указания для обучающихся по освоению дисциплины

Для изучения разделов данной учебной дисциплины необходимо вспомнить и систематизировать знания, полученные ранее по лингвистике.

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. Затем необходимо самостоятельно решить несколько аналогичных задач из сборников задач, приводимых в разделах рабочей программы, и после этого решать соответствующие задачи из сборников тестовых заданий и контрольных работ. 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, а также номер формулы, которые выражают ответ на данный 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 (уравнение), 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.

Указания по выполнению тестовых заданий и контрольных работ приводятся в учебно-методической литературе, в которых к каждой задаче даются конкретные методические указания по ее решению и приводится пример решения.

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

а) основная литература:

| № п/п | Источник |
|-------|--|
| 1 | Марчук Ю. Н. Компьютерная лингвистика : учебное пособие для студентов вузов, специализирующихся по направлению и специальности "Филология" / Ю.Н. Марчук. - М. : Восток-Запад, 2007. - 317 с. |
| 2 | Романская квантитативная лексикология : (материалы к спецкурсу) : [учебнометодическое пособие для студентов 4 курса дневного и вечернего отделений] / В.Т. Титов ; Воронеж. гос. ун-т. — Воронеж : ЛОП ВГУ, 2006. — 43 с. : ил. — Библиогр. в конце кажд. главы, с.: 41-42. |
| 3 | Моисеева И. Ю. Квантитативная лингвистика и новые информационные технологии : учебное пособие / И. Ю. Моисеева. - Оренбург : Оренбургский государственный университет, 2017. - 103 с. - URL: http://biblioclub.ru/index.php?page=book&id=481797 |

б) дополнительная литература:

| № п/п | Источник |
|-------|---|
| 4 | Гребенщикова, А. В. Основы количественной лингвистики и новых информационных технологий : учебное пособие / А. В. Гребенщикова. — 2-е изд., стер. — Москва : ФЛИНТА, 2015. — 152 с. — ISBN 978-5-9765-2137-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/70334 |
| 5 | Грудева, Е.В. Корпусная лингвистика : учебное пособие / Е.В. Грудева ; науч. ред. Л.Н. Чурилина. — 3-е изд., стер. — Москва : ФЛИНТА, 2017. — 166 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=364207 |
| 7 | Ляшевская, О.Н. Корпусные инструменты в грамматических исследованиях русского языка / О.Н. Ляшевская. — Москва : Языки славянской культуры (ЯСК) : Рукописные памятники Древней Руси, 2016. — 520 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=473302 |
| 8 | Хроленко, А.Т. Современные информационные технологии для гуманитария : [16+] / А.Т. Хроленко, А.В. Денисов. — 5-е изд., стер. — Москва : ФЛИНТА, 2018. — 129 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=363413 |
| 9 | Щипицина, Л.Ю. Информационные технологии в лингвистике : учебное пособие : [16+] / Л.Ю. Щипицина. — Москва : ФЛИНТА, 2013. — 127 с. : табл. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=375745 |
| 10 | Гуслякова, А. В. Информационные технологии и лингвистика XXI века : учебное пособие / А. В. Гуслякова. — Москва : МПГУ, 2016. — 96 с. — ISBN 978-5-4263-0398-0. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/106069 |
| 11 | Формализация исследовательских процедур анализа семантики языковых единиц / М.В. Каменский, Т.Н. Ломтева, Н.С. Кабылкина и др. ; под общ. ред. М.В. Каменского ; Северо-Кавказский федеральный университет. — Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2016. — 170 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=466913 |
| 12 | Клячкин, В.Н. Статистические методы анализа данных : учебное пособие / В.Н. Клячкин, Ю.Е. Кувайскова, В.А. Алексеева. — Москва : Финансы и статистика, 2016. — 240 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=447697 |
| 13 | Кокорина, И.В. Основы математической обработки информации в филологии: комбинаторика, теория вероятностей и математическая статистика / И.В. Кокорина ; Северный (Арктический) федеральный университет им. М. В. Ломоносова. — Архангельск : Северный (Арктический) федеральный университет (САФУ), 2014. — 115 с. : ил. — Режим доступа: по подписке. — URL: https://biblioclub.ru/index.php?page=book&id=312317 |

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

| № п/п | Ресурс |
|-------|--|
| 14 | ЭБС Лань. — Режим доступа: по подписке. — URL: ЭБС Лань (lanbook.com) |
| 15 | ЭБС «Университетская библиотека онлайн». — Режим доступа: по подписке. — URL: ЭБС "Университетская библиотека онлайн" читать электронные книги (biblioclub.ru) |
| 16 | ЭБС ЮРАИТ.— Режим доступа: по подписке. — URL: Образовательная платформа Юрайт. Для вузов и ссузов. (urait.ru) |
| 17 | ЭБС IPRbooks. — Режим доступа: по подписке. — URL: Электронно-библиотечная система IPRbooks / Результат поиска (iprbookshop.ru) |
| 18 | СинТагРус http://ruscorpora.ru/search-syntax.html |
| 19 | Тестовый корпус с параллельной синтаксической разметкой http://otipl.philol.msu.ru/~soiza/testsynt/ |
| 20 | Rus-Treebank http://otipl.philol.msu.ru/~soiza/rtb/res01/rtb.php |
| 21 | Единая коллекция цифровых образовательных ресурсов - http://school-collection.edu.ru |
| 22 | Ресурсный центр «Информационные технологии в обучении языку» - http://www.itlt.edu.nstu.ru/cld_literature.php |
| 23 | Федеральный центр информационно-образовательных ресурсов - http://fcior.edu.ru |
| 24 | Единое окно доступа к образовательным ресурсам - http://window.edu.ru |
| 25 | Информатизация, Linux и СПО в российском образовании - http://www.linformatika.ru/SPO |
| 26 | Каталог лингвистических программ и ресурсов в сети - http://www.rvb.ru/soft/catalogue/index.html |
| 27 | Портал информационной и технической поддержки ПО образовательных учреждений РФ - http://www.spohelp.ru |

16. Перечень учебно-методического обеспечения для самостоятельной работы

| № п/п | Источник |
|-------|---|
| 1 | Коноплева, И.А. Информационные технологии : учебное пособие / И.А. Коноплева, О.А. Хохлова, А.В. Денисов ; под ред. И.А. Коноплевой. – 2-е изд., перераб. и доп. – Москва : Проспект, 2014. – 328 с. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=251652 |
| 2 | Тимофеева, М.К. Язык с позиций философии, психологии, математики : учебное пособие : [16+] / М.К. Тимофеева. – 2-е изд., стер. – Москва : ФЛИНТА, 2013. – 176 с. : ил. – Режим доступа: по подписке. – URL: https://biblioclub.ru/index.php?page=book&id=375760 |

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.). Для освоения материала на практических занятиях используются задания, предназначенные как для индивидуального решения задач по обработке текста, так и для коллективного обсуждения стратегии решения той или иной задачи. Особое внимание уделяется организации самостоятельной работы студентов с программным обеспечением, рассматриваемым в рамках курса.

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 текущей и промежуточной аттестаций

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. Количественные данные в лингвистических исследованиях. 2. Закон Ципфа — Мандельброта и его следствия. 3. Применение статистических методов в основных разделах лингвистики. 4. Опыт квантитативной лексикологии. 5. Квантитативные методы в корпусной лингвистике. 6. Квантитативная лингвистика: дешифровка текста, авторизация текста. | ПК-3 | Осуществляет постановку задачи на технологические исследования (ПК-3.1) Анализирует результаты технологических исследований (ПК-3.2) | Тест № 1 |
| | | ПК-7 | Выбирает адекватный способ представления результатов исследований в соответствующей области знаний (ПК-7.1) | |
| | | | Оформляет результаты научно-исследовательских работ в соответствии с существующими требованиями (ПК-7.2) | |
| Промежуточная аттестация форма контроля – экзамен | | | | |

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:

- практические задания, в т.ч. домашние
 - тестовые задания
-

Тест № 1

1. Кто впервые заявил о математической лингвистике как о научном направлении?

- а) Джошуа Уотмоу
- б) Ноам Холмский
- в) А.Н. Баранов
- г) Р.Г. Пиотровский

2. Какая выборка считается малой?

- а) менее 100 единиц;
- б) менее 50 единиц;
- в) менее 30 единиц;
- г) менее 10 единиц.

3. Как называется величина, которая характеризует ширину доверительного интервала, в который попадает относительная частота исследуемого свойства лингвистической единицы?

- а) относительная ошибка
- б) коэффициент достоверности
- в) среднее квадратичное отклонение
- г) относительное квадратичное отклонение

4. Раздел прикладной лингвистики, занимающийся разработкой и использованием электронных словарей, называется

- а) компьютерная лингвистика
- б) компьютерная лексикография
- в) корпусная лингвистика
- г) машинный перевод

5. Словарь, в котором словарные единицы сгруппированы по семантическим полям и понятийным группам с указанием семантических отношений между единицами, называется _____.

- а) конкордансом
- б) тезаурусом
- в) энциклопедией
- г) толковым

6. Файлы-изображения и звуковые файлы в табличных базах данных содержатся в _____.

- а) поле «гиперссылка»
- б) поле «логические данные»
- в) поле «объект ole»

г) поле «текстовые данные»

7. Формат kwic при построении конкорданса означает, что _____.

- а) список примеров выровнен по ключевому слову в центре строки
- б) список примеров представлен в алфавитном порядке
- в) представлены только самые частотные сочетания с ключевым словом

8. Программа, осуществляющая морфологическую или синтаксическую разметку текста в корпусе, называется _____.

- а) парсер
- б) корпусный менеджер
- в) конкордансер
- г) токен

9. Достаточно высокая синтаксическая и морфологическая точность, а также стабильность и предсказуемость результата являются преимуществами _____.

- а) систем машинного перевода, основанного на правилах
- б) систем статистического машинного перевода

10. Системы Trados и Wordfast относятся к системам _____.

- а) машинного перевода
- б) информационного поиска
- в) памяти переводов
- г) автоматического синтеза текста

11. В «квантитативной лингвистике» в качестве основного инструмента изучения языка и речи используются:

- а) качественные и количественные методы анализа
- б) качественные и статистические методы анализа
- в) количественные или статистические методы анализа

12. Комбинаторная лингвистика не использует в исследованиях

- а) теорию множеств
- б) стохастические процессы
- в) фрактальные модели
- г) конечный набор правил

13. Метод лингвистических переменных это

- а) метод построения терм-множеств
- б) формульное представление количественных подсчетов
- в) метод анализа деятельности нейронов при компьютерной томографии

14. Количественные методы в лингвистике включают

- а) когнитивное моделирование
- б) статистический анализ
- в) дефиниционный анализ
- г) компонентный анализ

15. Компьютерная лингвистика – направление в прикладной лингвистике, ориентированное на

а) использование программ, компьютерных технологий организации и обработки данных для моделирования функционирования языка в тех или иных условиях, ситуациях, проблемных сферах

б) применение компьютерных моделей языка в лингвистике и смежных дисциплинах

в) решение практических задач, связанных с изучением языка.

16. Объяснение фактов языка является задачей

- а) теоретической лингвистики
- б) прикладной лингвистики
- в) математической лингвистики
- г) компьютерной лингвистики

17. Данные – это

- а) факты, отражающие объекты, процессы и явления предметной области
- б) закономерности, установленные в результате практической деятельности и накопления профессионального опыта в некоторой проблемной области и позволяющие специалистам ставить и решать задачи в этой области
- в) сведения, рассматриваемые в каком-либо контексте и позволяющие пользователю составить собственное мнение

18. Информация соответствует _____ аспекту отражения действительности.

- а) синтаксическому
- б) семантическому
- в) прагматическому

19. При оценке информации различают такие ее аспекты как

- а) логический
- б) синтаксический
- в) математический
- г) физический
- д) прагматический
- е) семантический

20. Объем данных – это _____ мера оценки информации.

- а) логическая
- б) семантическая
- в) алгоритмическая
- г) физическая
- д) прагматическая
- е) синтаксическая
- ж) абсолютная

21. Семантический аспект передает _____ содержание информации.

- а) структурное
- б) полезное
- в) смысловое
- г) качественное
- д) формальное

22. Прагматический аспект отражает _____ свойства информации.

- а) потребительские
- б) структурные
- в) количественные
- г) качественные
- д) смысловые
- е) динамические

23. Знания – это

а) факты, характеризующие объекты, процессы и явления предметной области, а также их свойства

б) закономерности, установленные в результате практической деятельности и накопления профессионального опыта в некоторой проблемной области и позволяющие специалистам ставить и решать задачи в этой области

в) сведения, рассматриваемые в каком-либо контексте и позволяющие пользователю составить собственное мнение

24. Факты, отображающие объекты, процессы и явления предметной области, а также их свойства – это

а) данные

б) знания

в) информация

г) описания

25. Сведения, рассматриваемые в каком-либо контексте, которые имеют значение для пользователя, -

а) данные

б) знания

в) описания

г) результаты

д) информация

Описание технологии проведения

Тест-задание выдается студенту на электронном или бумажном носителе. Время выполнения задания – 20 мин. Максимальное количество баллов за все правильно выполненные задания – 25.

Требования к выполнению заданий (или шкалы и критерии оценивания)

Выполнение теста оценивается по двухбалльной шкале: зачтено или не зачтено. Оценка «зачтено» ставится, если студент набрал 15 и более баллов, т.е. выполнил правильно 60% -100% заданий. Оценка «не зачтено» ставится в том случае, если студент набрал менее 15 баллов, т.е. выполнил менее 60% заданий.

20.2 Промежуточная аттестация

Промежуточная аттестация по дисциплине осуществляется с помощью следующих оценочных средств: собеседование по экзаменационным билетам.

Перечень вопросов к экзамену:

1. Квантитативная лингвистика как наука.
2. Взаимодействие языка и математики.
3. Лингвистическое явление как математическая величина. Переменные и постоянные величины.
4. Дискретность и непрерывность в языке и речи.
5. Числовые функции в лингвистике.
6. Способы задания числовых лингвистических функций.
7. Знак в естественном языке и языке математики: сходства и отличия.

8. Структура языкового знака.
9. Речевой аспект языкового знака.
10. Квантитативные исследования языка на уровне фонетики. Количественное распределение слов с точки зрения начала и конца.
11. Количественное распределение слов с точки зрения длины слов.
12. Квантитативные исследования языка на уровне словообразования.
13. Синтетические и аналитические языки. Количественное измерение степени синтетичности языка.
14. Квантитативное изучение лексики.
15. Классификация лексики на основе грамматических признаков.
16. Квантитативное изучение омонимии и полисемии.
17. Вычисление частот различных фактов и объектов на всех уровнях языка.
18. Закон Ципфа-Мандельброта и его следствия.
19. Языкознание и теория вероятности. Лингвистические испытания и события.
20. Вероятность лингвистического события. Определения вероятности, применяемые в квантитативной лингвистике.
21. Классическое определение вероятности, его применение в лингвистических исследованиях.
22. Использование комбинаторики в лингвистических исследованиях. Размещения и размещения с повторениями.
23. Использование комбинаторики в лингвистических исследованиях. Перестановки и перестановки с повторениями.
24. Использование комбинаторики в лингвистических исследованиях. Сочетания и сочетания с повторениями.
25. Статистическое определение вероятности, его применение в лингвистических исследованиях.
26. Субъективная вероятность. Методика оценки субъективной вероятности и сравнения с объективной вероятностью.
27. Частотные словари как результат квантитативных исследований языка. Типы частотных словарей.
28. Методика отбора текстового материала для частотного словаря.
29. Общие принципы построения частотных словарей.
30. Правила анализа различных частей речи при построении частотного словаря.
31. Алгоритм построения частотного словаря.
32. Практическое использование данных частотного словаря.
33. Теория множеств и ее применение при изучении языка.
34. Классификация лингвистических множеств.
35. Множества на различных уровнях языка.
36. Основные операции над лингвистическими множествами.
37. Квантитативные методы, применяемые в лексикографии.
38. Использование квантитативных методов в компьютерной лингвистике.
39. Квантитативная лингвистика как теоретико-методологическая основа статистической обработки лингвистических данных. Статистические методы лингвистических исследований.
40. Регулярные выражения как методологическая основа автоматизации поиска и анализа текстовой информации на естественном языке.

Описание технологии проведения

Экзамен проводится по билетам, содержащим два теоретических вопроса. На подготовку ответа отводится 30 минут. Правильный ответ на каждый вопрос в билете оценивается в 10 баллов. Максимальное количество набранных баллов – 20.

Требования к выполнению заданий, шкалы и критерии оценивания

Для оценивания результатов обучения на экзамене используются следующие показатели:

1. знание основных типов систем, использующих модули лингвистического анализа; основных принципов и методов компьютерного моделирования лингвистических задач; способов подачи научных материалов; методов оформления результатов научного исследования в виде статей, квалификационных работ, постеров, презентаций, докладов на конференциях и семинарах;

2. 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.); грамотно и в соответствии с требованиями оформить научную работу, раздаточный материал или презентацию к докладу; сделать устный доклад в соответствии с регламентом;

3. владение навыком работы с различными системами автоматической и экспертной обработки текста и звучащей речи; выступлений перед аудиторией, а также создания письменных работ для представления результатов научного исследования.

Для оценивания результатов обучения на экзамене используется 4-х балльная шкала: «Отлично», «Хорошо», «Удовлетворительно», «Неудовлетворительно».

1. Оценка «Отлично» ставится в случае, если студент набрал 18-20 баллов.

2. Оценка «Хорошо» ставится в случае, если студент набрал 15-17 баллов.

3. Оценка «Удовлетворительно» ставится в случае, если студент набрал 12-14 баллов.

4. Оценка «Неудовлетворительно» ставится в случае, если студент набрал менее 12 баллов.

Соотношение показателей, критериев и шкалы оценивания результатов обучения.

| Критерии оценивания компетенций | Уровень сформированности компетенций | Шкала оценок |
|--|--------------------------------------|--------------|
| Полное соответствие ответа обучающегося всем перечисленным критериям. Продемонстрировано знание основных типов систем, использующих модули лингвистического анализа; основных принципов и методов компьютерного моделирования лингвистических задач; способов подачи научных материалов; методов | Повышенный уровень | Отлично |

| | | |
|--|------------------------|---------------|
| <p>оформления результатов научного исследования в виде статей, квалификационных работ, постеров, презентаций, докладов на конференциях и семинарах; 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.); грамотно и в соответствии с требованиями оформить научную работу, раздаточный материал или презентацию к докладу; сделать устный доклад в соответствии с регламентом; владение навыком работы с различными системами автоматической и экспертной обработки текста и звучащей речи; выступлений перед аудиторией, а также создания письменных работ для представления результатов научного исследования.</p> | | |
| <p>Ответ на контрольно-измерительный материал 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; способов подачи научных материалов; методов оформления результатов научного исследования в виде статей, квалификационных работ, постеров, презентаций, докладов на конференциях и семинарах; 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.); грамотно и в соответствии с требованиями оформить научную работу, раздаточный материал или презентацию к докладу; сделать устный доклад в соответствии с регламентом; владение навыком работы с различными системами автоматической и экспертной обработки текста и звучащей</p> | <p>Базовый уровень</p> | <p>Хорошо</p> |

| | | |
|---|-------------------|---------------------|
| речи; выступлений перед аудиторией, а также создания письменных работ для представления результатов научного исследования. | | |
| <p>Ответ на контрольно-измерительный материал 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; способов подачи научных материалов; методов оформления результатов научного исследования в виде статей, квалификационных работ, постеров, презентаций, докладов на конференциях и семинарах; 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.); грамотно и в соответствии с требованиями оформить научную работу, раздаточный материал или презентацию к докладу; сделать устный доклад в соответствии с регламентом; владение навыком работы с различными системами автоматической и экспертной обработки текста и звучащей речи; выступлений перед аудиторией, а также создания письменных работ для представления результатов научного исследования.</p> | Пороговый уровень | Удовлетворительно |
| <p>Ответ на контрольно-измерительный материал не соответствует любым трем (четырем) из перечисленных показателей. Обучающийся демонстрирует отрывочные, фрагментарные знания, допускает грубые ошибки при практическом применении приобретенных знаний.</p> | – | Неудовлетворительно |