Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Государственный институт русского языка им. А.С. Пушкина»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Цифровая лингвистика

наименование дисциплины (модуля)

Уровень высшего образования:

магистратура

указывается: бакалавриат, магистратура

Направление подготовки:

45.04.01 – «Филология»

(код и название направления подготовки)

Направленность (профиль) образования:

Прикладная филология

Форма обучения:

очная

очная, очно-заочная, заочная

Рабочая программа дисциплины (модуля) разработана в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки 45.04.01 – «Филология», утвержденным приказом Минобрнауки России № 980 от 12.08.2020.

1. Место дисциплины (модуля) в структуре ОПОП

В соответствии с учебным планом по направлению подготовки 45.04.01 – «Филология», дисциплина «Цифровая лингвистика» относится к части, формируемой участниками образовательного процесса программы бакалавриата Б1.В.ДВ.06.01

2. Цель и задачи освоения дисциплины (модуля)

Целью дисциплины является сформировать у обучающихся целостное представление о теоретических основах, методах и технологиях цифровой лингвистики как области применения цифровых и корпусных технологий к анализу естественного языка, а также развить практические умения и навыки использования цифровых инструментов и корпусных ресурсов для работы с языковыми данными и решения прикладных лингвистических задач в профессиональной деятельности.

3. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	**Индикаторы достижения компетенций**	**Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций**
УПК-3. Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели	УК-3.1 - Вырабатывает стратегию сотрудничества и на ее основе организует работу команды для достижения поставленной цели	Знать: - теоретические основы организации и руководства командной работой в цифровых лингвистических проектах: принципы формирования команды, распределения ролей и задач, выработки командной стратегии, а также методы планирования и координации совместной деятельности при выполнении проектов по цифровой обработке текстов и корпусных исследований. Уметь: - организовывать и руководить работой команды при реализации цифровых лингвистических проектов: формулировать общую цель и командную стратегию, распределять задачи между участниками в соответствии

		с их компетенциями, координировать совместную работу над лингвистическим проектом и оценивать результаты командной деятельности. Владеть: - приёмами и навыками организации и руководства командой в условиях цифрового лингвистического проекта: постановки командных задач, выбора инструментов и методов цифровой лингвистики для совместной работы, обеспечения эффективной коммуникации и взаимодействия участников, а также достижения поставленной цели в установленные сроки.
ПК-1. Способен проводить разноаспектное описание лексических единиц для решения теоретических и прикладных задач	ПК-1.2. Применяет цифровые и корпусные технологии для работы с лексическими данными	Знать: - основные типы цифровых корпусов и лексикографических ресурсов (национальные корпусы, специализированные корпусы, корпусные менеджеры), а также базовые методы корпусного анализа лексики — частотный анализ, извлечение коллокаций, анализ ключевых слов, используемые в цифровой лингвистике для исследования лексических данных. Уметь: - применять цифровые и корпусные технологии для поиска, отбора и анализа лексических единиц в корпусных ресурсах: формулировать корпусные запросы, извлекать контексты употребления слов, выявлять статистические характеристики лексики и устанавливать системные связи между лексическими единицами на основе корпусных данных. Владеть: - навыками практической работы с цифровыми корпусными инструментами для обработки лексических данных: использования корпусных менеджеров и языков запросов, построения частотных словарей и лексических профилей, интерпретации количественных

		данных о лексике при решении учебных и исследовательских задач в области цифровой лингвистики.
ПК-2. Способен осуществлять деятельность по созданию и редактированию текстового контента в соответствии с задачами профессиональной деятельности	ПК-2.1. Редактирует, трансформирует, реферирует различные типы текстов в зависимости от задач профессиональной деятельности ПК-2.2. Владеет приемами и навыками создания эффективных текстов в зависимости от задач профессиональной деятельности	Знать - цифровые средства обработки текстов для решения различных профессиональных задач Уметь -создавать и оптимизировать тексты различных типов с использованием методов и инструментов цифровой лингвистики: анализировать лексико-статистические и структурные характеристики текста, оценивать его эффективность и корректировать содержание и форму для достижения целей профессиональной коммуникации Владеть - приёмами и навыками создания эффективных текстов с применением цифровых лингвистических инструментов (частотный анализ, извлечение ключевых слов, оценка читаемости, тематическое моделирование) в зависимости от задач профессиональной деятельности.

4. Объем дисциплины 2 з.е., в том числе 32 академических часа на контактную работу обучающихся с преподавателем; 38 академических часов на самостоятельную работу обучающихся.

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов

Наименование и краткое содержание разделов и тем дисциплины (модуля).	**Всего (часы)**	В том числе	
		Контактная работа (работа во взаимодействии с преподавателем) ***Виды контактной работы, часы***	**Самостоятельная работа обучающегося** ***Виды самостоятельной работы, часы***

Форма промежуточной аттестации по дисциплине (модулю)									
		Занятия лекционного типа	Занятия семинарского типа (практические/ лабораторные)	Групповые консультации	Индивидуальная работа с обучающимися	**Всего**	**УО, Д.**	**Тест**	**Всего**
Тема 1. Введение в цифровую лингвистику: предмет, цели, направления.	10	2	2			4	4	2	6
Тема 2. История развития компьютерной и цифровой лингвистики.	10	2	2			4	4	2	6
Тема 3. Корпуса текстов: принципы создания и использования.	8	2	2			4	4		4
Тема 4. Электронные лексикографические ресурсы, словари и языковые базы данных	8	2	2			4	4		4
Тема 5. Основы автоматической обработки текста	10	2	2			4	4	2	6
Тема 6. Морфологический и синтаксический анализ в цифровой среде	8	2	2			4	4		4
Тема 7. Квантитативные	8	2	2			4	4		4

методы и визуализация языковых данных									
Тема 8. Цифровая лингвистика в прикладных задачах: ИИ, перевод, чат-боты, поисковые системы.	8	2	2			4	4		4
Промежуточная аттестация *(указывается форма проведения)*	*2*								
Итого	72								

6. Фонд оценочных материалов по дисциплине (модулю)

Типовые задания и иные материалы, необходимые для оценки результатов обучения

6.1. **Тематика и требования к структуре и содержанию докладов**

Примерная тематика докладов:

1. Цифровые технологии в переводоведении.
2. Машинный перевод как прикладная лингвистическая технология.
3. Языковые технологии в разработке чат-ботов и голосовых ассистентов.
4. Морфологический анализ в цифровой лингвистике.
5. Синтаксический анализ и его прикладное значение.
6. Токенизация, лемматизация и частеречная разметка.
7. Автоматическая и ручная аннотация языковых данных.
8. Ошибки и ограничения систем автоматической обработки естественного языка.
9. Методы оценки качества лингвистических инструментов.
10. Прикладные направления цифровой лингвистики.
11. Инструменты автоматической обработки текста.
12. Применение цифровых инструментов в переводческой деятельности.

13. Лингвистические основы чат-ботов и голосовых помощников.
14. Визуализация языковых данных и интерпретация результатов анализа.
15. Работа с большими массивами языковых данных.

Требования к структуре и содержанию доклада:

Доклад – это продукт самостоятельной работы студента, публичное выступление по представлению полученных результатов решения определенной учебно-практической, учебно-исследовательской или научной темы

Критерии оценки доклада

Критерии	Показатели
Актуальность темы, корректность предлагаемых решений.	-умение формулировать проблему исследования; -выдвигать корректные гипотезы; -умение намечать пути исследования; -умение отбирать материал;
Качество доклада. Полнота раскрытия темы.	- изложение полученных знаний полное; -объем и полнота разработок, - самостоятельность, законченность, подготовленность к докладу, -уровень творчества: оригинальность раскрытия темы, подходов, предлагаемых решений. -эрудиция, наличие межпредметных (междисциплинарных) связей.
Критерий «Выбор и использование средств решения проблемы, адекватных целям»	- **Умение представлять подготовленный доклад.** - Качество презентации: оформление, соответствие стандартным требованиям, качество схем, рисунков. - Научный стиль речи: навыки публичного выступления; - аттрактивность, - чувство времени; - Культура речи.
Глубина проработки материала, умение держать внимание аудитории	Ответы на вопросы: полнота, аргументированность, убежденность, стремление раскрыть тему и сильные стороны работы.

Шкала оценивания доклада:

«отлично» – все критерии соответствуют показателям в полном объеме.

«хорошо» – 1-2 критерия не в полном объеме соответствуют показателям.
«удовлетворительно» – все критерии реализованы не в полном объеме.
«неудовлетворительно» – результаты работы не соответствует всем или нескольким критериям

6.2. Вопросы для подготовки к контрольным работам

1. Что понимается под лингвистической базой данных?
2. Какие требования предъявляются к структуре и наполнению лингвистических баз данных?
3. Чем отличаются словарь, тезаурус, корпус и база данных как цифровые ресурсы?
4. Какие виды цифровых лингвистических ресурсов существуют?
5. Каковы основные принципы организации электронных словарей?
6. Что такое корпус текстов и каковы его основные характеристики?
7. Какие типы корпусов используются в лингвистических исследованиях?
8. Что такое разметка корпуса?
9. В чем состоит роль метаданных в лингвистических базах и корпусах?
10. Какие форматы хранения языковых данных являются наиболее распространенными?
11. В чем заключаются преимущества и ограничения использования цифровых лингвистических ресурсов?
12. Какие проблемы возникают при создании и пополнении лингвистических баз данных?
13. Как обеспечивается достоверность и актуальность цифровых языковых ресурсов?
14. Какие существуют этические и правовые аспекты использования цифровых языковых ресурсов?
15. Как цифровые ресурсы влияют на развитие современных методов описания языка?
16. Что понимается под прикладными технологиями в цифровой лингвистике?
17. Какие основные инструменты используются для автоматической обработки текста?
18. В чем заключается морфологический анализ текста и где он применяется?
19. Что такое синтаксический анализ и какую роль он играет в цифровой лингвистике?
20. Какие задачи решает лемматизация и токенизация?
21. Чем отличаются ручная и автоматическая разметка текста?
22. Какие существуют виды лингвистической разметки?
23. Что такое частеречная разметка и как она используется в анализе текста?
24. Какие ошибки характерны для автоматических систем анализа естественного языка?

25. Как можно оценить качество работы инструментов автоматической обработки текста?
26. Как используются цифровые инструменты в переводческой деятельности?
27. Какие технологии лежат в основе машинного перевода?
28. Как цифровая лингвистика связана с искусственным интеллектом?
29. Какие лингвистические методы используются при создании чат-ботов и голосовых помощников?
30. В чем состоит специфика обработки устной и письменной речи в цифровой среде?
31. Как интерпретировать результаты количественного анализа текста?
32. Каковы преимущества использования цифровых инструментов в лингвистическом исследовании?
33. Какие ограничения имеют современные технологии автоматической обработки языка?
34. Какие требования предъявляются к качеству лингвистических данных при работе с цифровыми инструментами?
35. Каковы перспективы развития прикладных технологий в цифровой лингвистике?

6.3. Типовые задания для самостоятельной проверочной работы.

Задание № 1

1. Охарактеризуйте электронные словари как цифровые лингвистические ресурсы.
2. Функции корпусных технологий в прикладной лингвистике.
3. Сравните бумажный словарь, электронный словарь и лексическую базу данных по 3–4 параметрам.

Задание № 2

1. Какие существуют методы аннотирования текстовых корпусов?
2. Что такое электронная лексикография?
3. Определите, какие виды разметки целесообразно использовать для корпуса новостных текстов, и обоснуйте выбор.

Задание № 3

1. Охарактеризуйте основные инструменты автоматизированной обработки естественного языка.
2. Дайте определения: токенизация, лемматизация, частеречная разметка.

3. Проанализируйте любой короткий текст: выполните токенизацию, определите леммы и укажите части речи для 10 слов.

Задание № 4

1. Какие ограничения имеют автоматизированные системы обработки языковых данных?
2. Назовите критерии оценки качества лингвистических систем.
3. Проанализируйте результат автоматической разметки небольшого текста и укажите возможные источники ошибок.

6.4. Примерные вопросы для подготовки к зачету

1. Что изучает цифровая лингвистика как научная дисциплина?
2. В чем состоят ее предмет, цели и основные задачи?
3. Каковы прикладные аспекты цифровой лингвистики?
4. Чем цифровая лингвистика отличается от компьютерной и корпусной лингвистики?
5. Какие этапы развития прошла цифровая лингвистика?
6. Что понимается под автоматизированной обработкой естественного языка?
7. Какие инструменты используются для автоматической обработки текста?
8. Что такое морфологический анализ текстового материала?
9. В чем заключается синтаксический анализ в лингвистических исследованиях?
10. Что представляют собой токенизация, лемматизация и частеречная разметка?
11. Какие основные типы лингвистической разметки существуют?
12. В чем состоит смысл аннотирования текстовых корпусов?
13. Какие методы аннотирования текстовых корпусов используются в цифровой лингвистике?
14. Что такое корпус текстов и каковы его основные характеристики?
15. Каковы преимущества корпусных технологий в прикладной лингвистике?
16. Какие ограничения имеет автоматизированная обработка языковых данных?
17. Какие ошибки характерны для систем автоматической обработки естественного языка?
18. По каким критериям оценивается качество лингвистических систем?

19. Как оценивается качество результатов автоматической разметки?
20. Что такое электронная лексикография?
21. Что представляют собой лексические базы данных и тезаурусы?
22. В чем отличие электронного словаря от лексической базы данных?
23. Какие программные средства используются для анализа текстов и корпусов?
24. Какие возможности дают цифровые инструменты для исследования языка?
25. Как осуществляется обработка больших объемов языкового материала?
26. Какие методы визуализации языковых данных применяются в цифровой лингвистике?
27. Как используются цифровые ресурсы в лингвистических и прикладных исследованиях?
28. Как цифровые технологии применяются в переводе, лексикографии и языковом моделировании?
29. Какую роль играют цифровые лингвистические технологии в системах искусственного интеллекта?
30. Каковы возможности и ограничения чат-ботов и голосовых помощников с лингвистической точки зрения?
31. Каковы перспективы развития цифровых лингвистических инструментов?
32. Какие этические и правовые проблемы возникают при использовании цифровых языковых ресурсов?
33. Почему важно критически оценивать результаты автоматического анализа текста?

6.5. Шкала и критерии оценивания ***(шкала и критерии оценивания могут быть едиными (типовыми) для всех дисциплин (модулей), входящих в основную образовательную программу)***

Знания обучающего определяются следующими отметками: «отлично», «хорошо», «удовлетворительно», «неудовлетворительно». К промежуточной аттестации допускаются обучающиеся, принявшие участие во всех формах текущего контроля по дисциплине. В случае наличия задолженности по формам текущего контроля обучающийся отрабатывает пропущенные элементы текущего контроля по согласованию с преподавателем.

Промежуточная аттестация по дисциплине «Морфология» осуществляется в форме экзамена, а именно, в виде ответа на теоретические вопросы билета и выполнения практического задания.

Оценочные материалы для проведения промежуточной аттестации обучающихся приведены в приложении к рабочей программе дисциплины.

Критерии оценивания результатов обучения по дисциплине при проведении промежуточной аттестации:

<table>
<tr><td rowspan="3">Планируемые результа</td><td colspan="4">Критерии оценивания результатов обучения по дисциплине</td></tr>
<tr><td>«неудовлетворительно»</td><td>«удовлетворительно»</td><td>«хорошо»</td><td>«отлично»</td></tr>
<tr><td colspan="2">«не зачтено»</td><td colspan="2">«зачтено»</td></tr>
</table>

ты обучения по дисциплине				
Сформированность знаний (знать)	Уровень знаний ниже минимальных требований. Имели место грубые ошибки. Невозможно оценить полноту знаний вследствие отказа обучающегося от ответа.	Минимально допустимый уровень знаний. Допущено много негрубых ошибок.	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок.	Уровень знаний в объеме, соответствующем программе подготовки, или в объеме, превышающем программу подготовки, продемонстрирован без ошибок и погрешностей.
Сформированность умений (уметь)	Отсутствие минимальных умений. При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки. Невозможно оценить наличие умений вследствие отказа обучающегося от ответа.	Продемонстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме.	Продемонстрированы все основные умения. Решены все основные задачи, но некоторые с негрубыми ошибками. Выполнены все задания, в полном объеме, но некоторые с недочетами.	Продемонстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов.

Сформированность навыков (владеть)	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки. Невозможно оценить наличие навыков вследствие отказа обучающегося от ответа.	Продемонстрирован минимальный набор навыков. Решены стандартные задачи с некоторыми недочетами.	Продемонстрированы базовые навыки. Решены стандартные задачи, но некоторые с недочетами.	Продемонстрировано всестороннее владение навыками. Задачи, в том числе нестандартные, решены без ошибок и недочетов.

7. Методические материалы

7.1. Перечень основной и дополнительной учебной литературы и иные материалы:

7.1.1. Основная литература

1. Компьютерная лингвистика и автоматическая обработка текстов : учебное пособие / Н.В. Лукашевич, А.А. Сорокин. – Москва : МАКС Пресс, 2025. – 608 с.

2. Мечковская, Н. Б. Практикум по лингвистической методологии и компьютерной лингвистике : учебное пособие / Н. Б. Мечковская. — Москва : ФЛИНТА, 2025. — 264 с. — ISBN 978-5-9765-5730-7.

3. Информатика для гуманитариев : учебник и практикум для вузов / под редакцией Г. Е. Кедровой. — 3-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2026. — 662 с. — (Высшее образование). — ISBN 978-5-534-16197-7.

4. Шунейко, А. А. Квантитативная лингвистика и новые информационные технологии : учебник для вузов / А. А. Шунейко, И. А. Авдеенко. — Москва : Издательство Юрайт, 2026. — 347 с. — (Высшее образование). — ISBN 978-5-534-15446-7.

7.1.2. Дополнительная литература

1. Башмаков, А. И. Интеллектуальные информационные технологии : учебное пособие / А. И. Башмаков, И. А. Башмаков. — Москва : МГТУ им. Н.Э. Баумана, 2005. — 304 с.

2. Потапова Р. К. Новые информационные технологии и лингвистика : учеб. пособие для вузов / Потапова Р. К. ; Московский гос. лингвистический ун-т. - 6-е изд. - М. : URSS : Ленанд, 2016. - 364 с.

3. Зубов А. В., Зубова И. И. Информационные технологии в лингвистике : учебник для высшего профессионального образования / Зубов А. В., Зубова И. И. - 2-е изд., испр. - М. : Академия, 2012. - 205 с.

4. Julia Lavid-López, Carmen Maíz-Arévalo, Juan Rafael Zamorano-Mansilla. Corpora in Translation and Contrastive Research in the Digital Age : Recent Advances and Explorations / Julia Lavid-López, Carmen Maíz-Arévalo, Juan Rafael Zamorano-Mansilla. - Amsterdam : John Benjamins Publishing Company, 2021. - ISBN 9789027259684.

7.2. Перечень лицензионного программного обеспечения

1. Microsoft Windows 10,
2. Microsoft Office 2016 стандартная,
3. Big Blue Button

8. Ресурсное обеспечение

8.1. Перечень профессиональных баз данных и информационных справочных систем:

Наименование	Ссылка на ресурс	Доступность (свободный доступ/ ограниченный доступ)
Информационно-справочные системы		
Интернет-портал «Русский язык»:	http://www.gramota.ru	свободный доступ
Портал «Образование на русском»	https://pushkininstitute.ru	ограниченный доступ, необходима регистрация
Электронно-библиотечные системы		
ЭБС Юрайт	https://urait.ru	ограниченный доступ, необходима регистрация
ЭБС IPR Smart	https://iprbookshop.ru	ограниченный доступ, необходима регистрация
Профессиональные базы данных		
Национальный корпус русского языка.	http://www/ruscorpora.ru/corpora-morph.html	свободный доступ
Научная электронная библиотека и Российский индекс научного цитирования	http://www.elibrary.ru	свободный доступ

8.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

Наименование	Ссылка на ресурс	Доступность

		(свободный доступ/ ограниченный доступ)
Электронно-библиотечные системы		
ЭБС Юрайт	https://urait.ru	ограниченный доступ, необходима регистрация
ЭБС IPR Smart	https://iprbookshop.ru	ограниченный доступ, необходима регистрация

9. Язык преподавания

Русский

10. Разработчики программы

Катышев Павел Алексеевич, доктор филологических наук, профессор, заведующий кафедрой общего и русского языкознания ФГБОУ ВО «Государственный институт русского языка им. А.С. Пушкина»