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

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Информационные ресурсы в филологической деятельности

Б1.О.02

Уровень высшего образования:

магистратура

Направление подготовки:

45.04.01 Филология

Направленность (профиль) образования:

Лингвистическая аналитика и экспертиза текстов

Форма обучения:

очная

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 магистратуры 45.04.01 «Филология» по специальности «Лингвистическая аналитика и экспертиза текстов» дисциплина Б1.О.02 «Информационные ресурсы в филологической деятельности» относится к обязательной части и изучается в 1 семестре.

2. Цель и задачи освоения дисциплины (модуля)

Курс «Информационные ресурсы в филологической деятельности» является обязательной частью профессионального цикла и направлен на комплексную интеграцию классической филологической науки с современными цифровыми технологиями. Главной целью курса является формирование информационно-технологической компетентности как неотъемлемой составляющей профессиональной подготовки магистра-филолога. Для реализации этой цели предполагается освоение инструментальных средств и компьютерных технологий, которые служат научно-практической базой для осуществления профессиональной деятельности в области филологии, лингвистики и межкультурной коммуникации.

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**
УК-1	УК-1.3 Критически оценивает надежность источников информации, работает с противоречивой информацией из разных источников	Знать: Системную структуру современного информационного пространства в гуманитарных

		науках: классификацию баз данных, корпусов текстов, цифровых архивов и электронных библиотек, а также их взаимосвязи. Уметь: самостоятельно проектировать базы данных для лингвистических исследований, проводить корпусный анализ, критически оценивать достоверность сетевых источников, а также применять методы автоматической обработки текста для решения конкретных научных и прикладных задач. Владеть: Техникой критического аудита информационных ресурсов
УК-4	УК 4.5 Демонстрирует интегративные умения выполнять разные типы перевода академического текста с иностранного (-ых) на государственный язык в профессиональных целях. Умеет использовать сеть интернет и социальные сети в процессе учебной и академической профессиональной коммуникации	Знать: Методологию критического анализа информационных ресурсов: критерии релевантности, репрезентативности, верифицируемости данных, а также способы выявления лакунарности в источниках. Уметь: Критически сопоставлять

		данные из различных информационных систем для выявления противоречий и верификации гипотез Владеть: Навыком разработки и защиты исследовательской стратегии: способностью обосновать, почему для решения конкретной проблемы выбран именно данный набор информационных инструментов и данная последовательность действий.
ОПК-2	ОПК-2.2. Обладает навыками чтения и интерпретации научных трудов в избранной области филологии	Знать: Принципы стратегического планирования информационного поиска и исследования: от постановки научной проблемы до выбора оптимального комплекса инструментов и прогнозирования результатов. Уметь: осуществлять декомпозицию филологической проблемной ситуации на информационные составляющие (выделить корпус источников, историографию

ОПК-3	ОПК-3.2 Использует навыки работы с текстом в научной, педагогической, журналистской, прикладной и других видах деятельности ОПК-3.3 Корректно применяет приемы лингвистического и литературоведческого анализа	вопроса, фактологическую базу, недостающие данные). Владеть: Навыком системного информационного мониторинга для комплексного решения филологической задачи Знать: Методологические ограничения цифровых инструментов (статистическую погрешность при работе с малыми корпусами, алгоритмические искажения поисковых систем) и способы их критической оценки. Уметь: Выстраивать иерархию информационных ресурсов по степени их значимости для конкретной исследовательской стратегии Владеть: Навыком интеграции результатов системного анализа в итоговый научный продукт с наглядной демонстрацией логики принятых решений.

	текста в избранной области филологии	Знать: Типологию проблемных ситуаций в филологической деятельности, требующих системного анализа. Уметь: Формулировать многоуровневую стратегию действий: определять приоритетные направления поиска, распределять ресурсы, выбирать методы обработки данных (количественные, качественные, текстологические) и предусматривать альтернативные сценарии при отсутствии данных. Владеть: Практикой применения специализированного программного обеспечения для анализа данных в рамках единой стратегии исследования.

4. Объем дисциплины (модуля) _2_ з.е., в том числе __32__академических часа на контактную работу обучающихся с преподавателем; ___38__академических часов на самостоятельную работу обучающихся.

(В конце данного пункта следует отметить, если дисциплина или часть ее реализуется с применением электронного обучения и дистанционных образовательных технологий)

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). **Форма промежуточной аттестации по дисциплине (модулю)**	**Всего (часы)**	В том числе							
		Контактная работа (работа во взаимодействии с преподавателем) ***Виды контактной работы, часы***					**Самостоятельная работа обучающегося** ***Виды самостоятельной работы, часы***		
		Занятия лекционного типа	Занятия семинарского типа (практические/ лабораторные)	Групповые консультации	Индивидуальная работа с обучающимися	**Всего**	**Доклад**	**Тест**	**Всего**
Тема 1. Архитектура цифрового гуманитарного знания (Digital Humanities)		2	2			4		4	8

Тема 2 Компьютерная лексикография и лексикология: от бумажного словаря к интерактивной базе данных		2	2			4		4	8
Тема 3 Информационное обеспечение корпусных и квантитативных исследований		2	2			4		4	8
Тема 4 Построение частотных словарей, конкордансов и коллокационных сетей. Визуализация данных.		2	2			4		4	8
Тема 5 Информационный поиск, библиометрия и управление научным аппаратом:		2	2			4		4	8

стратегии поиска.									
Тема 6 Информационные ресурсы в прикладной и педагогической деятельности лингвиста		2	2			4		6	10
Тема 7 Базы тестовых заданий и языковые корпуса для создания учебных материалов. Системы управления обучением (LMS) и инструменты для дистанционного преподавания языка. Справочные порталы и их интеграция в методику преподавания и		2	2			4		6	10

научную деятельность.									
Тема 8 Информационная безопасность, этика данных и цифровые гуманитарные проекты		2	2			4		6	10
Промежуточная аттестация *(указывается форма проведения)*	*зачет*								2
Итого		16	16			32		38	72

6. Фонд оценочных материалов по дисциплине (модулю)

Типовые задания и иные материалы, необходимые для оценки результатов обучения

Тест по курсу «Информационные ресурсы в филологической деятельности».

Блок 1: Основные понятия

Вопросы на знание терминологии

1. Что такое информационный ресурс?

2. Дайте определение понятию «информация».
3. Чем отличаются первичные и вторичные информационные ресурсы? Приведите примеры.
4. Что такое лингвистический корпус? Для чего он используется?
5. Назовите три основных вида информационных ресурсов.
6. В чем разница между библиографическими и полнотекстовыми ресурсами?
7. Что такое метаданные?
8. Каковы функции электронных библиотек?
9. Что такое цифровой архив?
10. Объясните понятие «открытый доступ к информации».

Блок 2: Типология и классификация ресурсов

11. Какие виды информационных ресурсов используются в филологических исследованиях?
12. Классифицируйте электронные словари по типу доступа (онлайн/офлайн).
13. Перечислите основные типы баз данных, применяемых в лингвистике.
14. Что такое гипертекст? Приведите пример.
15. В чем отличие справочных систем от поисковых машин?
16. Какие ресурсы относятся к мультимедийным информационным ресурсам?
17. Назовите критерии оценки качества информационного ресурса.

18. Что такое репозиторий научных публикаций?
19. Какие существуют способы организации информации в цифровых коллекциях?
20. Опишите особенности национальных корпусов текстов.

Блок 3: Работа с информационными ресурсами

21. Опишите алгоритм поиска научной литературы в электронной библиотеке.
22. Как использовать расширенный поиск в Google Scholar?
23. Назовите основные операторы языка запросов в поисковых системах.
24. Как осуществляется фильтрация результатов поиска по дате публикации?
25. Что такое цитирование и как оно связано с наукометрией?
26. Как проверить достоверность найденной информации?
27. Опишите этапы создания аннотированной библиографии.
28. Как работать с электронными таблицами при анализе частотности языковых единиц?
29. Назовите программы для автоматического извлечения ключевых слов из текста.
30. Как организовать личную электронную библиотеку?

Блок 4: Современные технологии и цифровизация

31. Что такое искусственный интеллект в обработке естественного языка (*NLP*)?

32. Как применяются нейросети в автоматической обработке текстов?
33. Что такое *Big Data* и как она применяется в филологии?
34. Опишите возможности *семантического веба*.
35. Что такое облачные сервисы хранения данных? Приведите примеры.
36. Как работают системы автоматического перевода?
37. Что такое цифровая гуманитаристика (*Digital Humanities*)?
38. Как используются социальные сети как источник лингвистических данных?
39. Что такое краудсорсинг в создании лингвистических ресурсов?
40. Как обеспечивается сохранность цифровых информационных ресурсов?

Блок 5: Правовое регулирование и этика

41. Что такое авторское право в цифровой среде?
42. Как регулируется использование чужих текстов в научных работах?
43. Что такое лицензия Creative Commons?
44. Как защитить свои персональные данные при работе с онлайн-ресурсами?
45. Что такое плагиат и как его обнаружить?
46. Каковы правила цитирования источников?
47. Что такое конфиденциальность информации?

48. Каковы требования к хранению и передаче цифровых данных?
49. Что такое информационная безопасность?
50. Какова роль этики в использовании информационных ресурсов?

Рекомендации по использованию:

- Можно комбинировать вопросы разных типов: выбор одного или нескольких правильных ответов, открытые вопросы, задания на сопоставление.
- К каждому блоку можно добавить практические кейсы — например, анализ корпуса текстов или составление запроса к базе данных.

Задания по курсу для самостоятельной работы студентов

1. **Аудит и картографирование научного профиля.** *Задача:* Составить подробную карту информационных ресурсов (баз данных, журналов, репозиториев, новостных агрегаторов) для узкой лингвистической специализации (например, «корпусные исследования дискурсивных маркеров в русском языке» или «диалектология северных говоров»). *Инструменты:* Zotero/Mendeley для сбора метаданных, Miro или XMind для визуализации. *Результат:* Интерактивная карта с разделением на первичные (корпусы, словари), вторичные (статьи, монографии) и третичные (рецензии, обзоры) источники, а также оценка их полноты и доступности.

2. **Проектирование сложного поискового запроса.** *Задача:* Для заданной спорной лингвистической проблемы (например, статус «звательного падежа» в современном русском или природа эвиденциальности) составить и обосновать поисковый запрос для библиографической базы (eLibrary, Google Scholar, Scopus). *Критерии:* Использование булевых операторов, фильтров по типу документа,

временных срезов и цепочек синонимичных терминов. *Результат:* Сравнительная таблица выдачи по трем разным базам данных с анализом причин расхождений (разница в индексации, полнотекстовом покрытии, языке интерфейса).

3. **Критическая оценка цифрового словаря.** *Задача:* Провести сравнительный анализ двух лексикографических ресурсов (например, «Академос» и «Грамота.ру» для орфоэпии или «Национальный корпус русского языка» и «Wiktionary» для семантики). *Параметры:* Актуальность нормативной базы, глубина описания (наличие примеров, этимологии, стилистических помет), скорость обновления, API-доступность, наличие пользовательских правок. *Результат:* Паспорт ресурса (1 страница) и экспертное заключение о применимости каждого словаря в лексикографической практике и при подготовке академической статьи.

4. **Сбор и очистка пользовательских данных.** *Задача:* С помощью парсера (например, на Python с библиотеками Beautiful Soup или Selenium, либо без кода через ParseHub) собрать массив пользовательских текстов по заданной теме (отзывы на маркетплейсе, ветки форума, комментарии в Telegram-канале). *Ограничение:* Объем — не менее 5000 токенов. *Результат:* Очищенный датасет в формате CSV/JSON с удалением HTML-тегов, служебных символов и дублей, готовый для загрузки в инструмент частотного анализа (например, AntConc).

5. **Создание аннотированного мини-корпуса.** *Задача:* Собрать, разметить и зааннотировать корпус из 20 текстов (объемом 15–20 тысяч словоупотреблений) по заданному социолингвистическому критерию (например, «молодежный сленг геймеров 2020–2025 годов» или «официальные ответы ведомств в соцсетях»). *Разметка:* Обязательное тегирование метаданных (автор, возраст, платформа, дата) и минимум трех уровней лингвистической разметки (часть речи, лексико-семантическая группа, прагматическая функция). *Инструменты:* Конкорданс, построение частотных списков и списка

коллокаций. *Результат:* Корпуса в формате TXT/XML и отчет с 5 наиболее показательными примерами контекстного употребления ключевых единиц.

6. **Верификация данных в Википедии и языковых разделах.** *Задача:* Выбрать сложную лингвистическую статью (например, «Падеж», «Синтаксис» или «Этимология») и проследить историю правок за последний год. *Задача:* Выявить три спорных утверждения, проверить их по профильным академическим источникам (монографиям, индексируемым статьям) и оценить авторитетность редакторов. *Результат:* Сводная таблица «Утверждение — Первичный источник — Статус в статье» и эссе о рисках использования Википедии как стартовой точки для филологического исследования.

7. **Работа с API научного репозитория.** *Задача:* Написать скрипт (или пошагово выполнить запросы через интерфейс API, например, Crossref, OpenAlex или PubMed) для автоматического извлечения метаданных по заданному DOI-листу (20–30 статей). *Цель:* Выгрузить данные о цитированиях, аффилиациях авторов и динамике публикаций по теме за 10 лет. *Результат:* Сводная таблица и визуализация (график роста числа публикаций или карта соавторства), выполненная в Excel, Python (Matplotlib/Plotly) или инструментах Google.

8. **Фиксация и архивирование эфемерных данных.** *Задача:* Исследовать лингвистический феномен, существующий только в цифровой среде (например, эволюция использования стикеров как синтаксических единиц в Telegram или мемы в студенческих чатах). *Задача:* Разработать протокол фиксации: выбор инструментов архивирования (Wayback Machine, StreamYard, локальное сохранение), форматов (MP4, PNG, JSON) и структуры хранения. *Результат:* Архив объемом не менее 50 единиц, его техническое описание и аналитическая записка о проблеме «исчезающих данных» в интернет-лингвистике.

9. **Сравнительный анализ инструментов машинного перевода.** *Задача:* Взять сложный филологический текст (фрагмент орнаментальной прозы, философский трактат или диалектную запись) и пропустить его

через три системы (DeepL, Яндекс Переводчик, ChatGPT). *Оценка:* Провести постредактирование каждого варианта, зафиксировав типы системных ошибок (потеря культурных реалий, разрушение синтаксических связей, стилистическая нейтрализация). *Результат:* Сводная таблица ошибок и стратегия интеграции МП в работу филолога (на каких этапах он ускоряет работу, а на каких — вводит в заблуждение).

10. **Проектирование информационной стратегии исследования.** *Задача:* Разработать дорожную карту годового исследования (например, магистерской диссертации) от формулировки гипотезы до защиты. *Карта:* Должна включать календарный план мониторинга новых поступлений (алерты в Google Scholar, RSS-ленты профильных журналов), график работы с закрытыми архивами, план управления библиографией и стратегию самопродвижения (профили в ORCID, ResearchGate, Academia.edu). *Результат:* Интерактивный план в Trello/Notion с настроенными шаблонами еженедельных отчетов и системой оповещений.

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

Оценка качества учебных достижений обучающихся по дисциплине осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Темы докладов

1. Цифровые корпуса текстов как инструмент лексикографического анализа (на примере Национального корпуса русского языка). Специфика использования подкорпусов для исследования исторической семантики.
2. Автоматизированные методы стилистического анализа текста: применение стилометров для атрибуции анонимных произведений и выявления авторского идиостиля.

3. Цифровая эдиция (Digital Scholarly Editing): принципы подготовки, визуализации и комментирования рукописного наследия в академических онлайн-проектах.
4. Технологии Text Mining и дистрибутивная семантика в исследовании эволюции поэтических образов (на материале русской лирики XIX–XXI веков).
5. Виртуальные реконструкции историко-культурных ландшафтов в филологической деятельности: от создания ГИС-карт литературных путешествий до геопоэтики.
6. Специфика работы с фольклорными информационными ресурсами: базы данных сказочных сюжетов (типа СУС) и методы их количественного анализа.
7. Машинное обучение в решении задач межтекстовых связей: алгоритмы автоматического поиска цитат, аллюзий и реминисценций в больших массивах прозы.
8. Репрезентация диалектной лексики в современных лексикографических базах данных: проблемы полевого сбора, метаданных и пользовательских интерфейсов.
9. Информационное обеспечение переводоведения: CAT-инструменты, переводческая память и терминологические базы данных в работе с художественным текстом.
10. Социальные сети и блоги как новый тип филологического источника: методы сбора, верификации и этические проблемы анализа пользовательского контента.
11. Создание и анализ мультимодальных литературных ресурсов: интеграция текста, аудио (начитки), видео и гипертекста в едином исследовательском пространстве.
12. Филологическая критика данных (Data Criticism): оценка репрезентативности, полноты и предвзятости (bias) литературных баз данных и цифровых библиотек.
13. Использование технологий дополненной и виртуальной реальности (AR/VR) для иммерсивного анализа пространства в литературе (на примере романов Достоевского или Булгакова).
14. Информационные ресурсы в преподавании родного и иностранного языков: проектирование адаптивных обучающих сред на основе анализа данных об успеваемости.

15. Библиометрические и альтметрические методы в оценке научного вклада филологов и динамики развития научных направлений (анализ цитируемости, упоминаний в медиа).
16. Архитектура и наполнение тезаурусов и семантических сетей для автоматической обработки текстов (на материале русского языка).
17. Онтологическое моделирование художественного мира писателя: формализация персонажей, хронотопа и сюжетных линий для семантического поиска.
18. Информационная поддержка сравнительно-исторического языкознания: базы данных родства языков, этимологические онлайн-словари и филогенетические алгоритмы.
19. Анализ тональности и сентимент-анализ в литературоведении: автоматическое выявление эмоциональных доминант в драматургии и эпистолярных жанрах.
20. Цифровое сохранение исчезающих малых языков России: создание электронных архивов, словарей и корпусов (на примере финно-угорских или тюркских языков).

Рекомендации по выбору и формулировке темы:

- Актуализация: любую из предложенных тем можно сузить, ограничив материал конкретным автором (например, «Цифровые методы анализа синтаксиса в романах Андрея Белого»), историческим периодом или типом ресурса.
- Методологический баланс: для магистерского реферата важно соблюсти баланс между филологической проблемой (что исследуется) и информационным ресурсом или технологией (как исследуется).

Практическая часть: большинство тем предполагает не только обзор литературы, но и описание работы с конкретным инструментом (например, выгрузка данных из НКРЯ, построение облака слов, анализ метаданных в электронной библиотеке).

Требования к структуре и содержанию доклада:

1) регламент выступления - 10 минут; наличие текста доклада и компьютерной презентации (презентация и текст доклада не должны дублировать друг друга);

2) использование самостоятельно подобранного текстового материала и его анализа в аспекте, определяемом темой доклада;

3) наличие учебного задания для студентов группы по теме доклада.

Критерии оценки доклада

Критерии	Показатели
Наличие необходимых элементов	Есть письменный текст доклада, компьютерная презентация, не дублирующие друг друга, студент выступил с докладом
Соблюдение требований к структуре и содержанию доклада	В докладе используется самостоятельно подобранный текстовый материал, есть его анализ и учебное задание

Шкала оценивания доклада (в баллах):

- все требования соблюдены, необходимые элементы есть, выполнено качественно: 15 б.

- все требования соблюдены, необходимые элементы есть, выполнено качественно, но презентация и текст доклада дублируют друга друга – 10 б

- не все требования соблюдены, не все необходимые элементы в наличии (нет текста доклада/презентации/нет самостоятельно подобранного текстового материала/задания – 5-7 баллов).

- доклад не представлен – 0 баллов.

7. Методические материалы

7.1. Перечень основной и дополнительной учебной литературы и иные материалы

Основная литература:

1. Сысоев П.В. Информационные и коммуникационные технологии в лингвистическом образовании. М.: Книжный дом "Либроком", 2019. - 264 с. ISBN 978-5-397-06692-1 .

Дополнительная литература:

1. Старостина С. А. Информационные технологии в филологии: учебное пособие / С. А. Старостина; Тамбовский государственный университет имени Г. Р. Державина. — Тамбов: Издательский дом "Державинский", 2022

2. Гуслякова А. В. Информационные технологии и лингвистика XXI века [Текст] : учебное пособие / А. В. Гуслякова ; М-во образования и науки Российской Федерации, Федеральное гос. бюджетное образовательное учреждение высш. образования Московский пед. гос. ун-т. - Москва : МГПУ, 2016. - 95 с. : ил.; 20 см.; ISBN 978-5-4263-0398-0

3. Щипицина Л.Ю. Информационные технологии в лингвистике [Текст] : учебное пособие / Л. Ю. Щипицина. - Москва : Флинта : Наука, 2013.

Словари и справочники:

1. Энциклопедия «Русский язык»/Институт русского языка им. В.В. Виноградова РАН, Гл. ред. А.М. Молдован. – 3-е изд., переработанное и дополненное. – М.: АСТ-ПРЕСС ШКОЛА, 2020. – 904 с. – ISBN: 978-5-907126-47-3

Периодические издания:

1. Вопросы языкознания: Научный журнал. - https://vja.ruslang.ru/ru/archive (Архив журнала с открытым доступом в Интернет).

7.2. Перечень лицензионного программного обеспечения

8. Ресурсное обеспечение

8.1. Перечень профессиональных баз данных и информационных справочных систем:

Наименование	Ссылка на ресурс	Доступность (свободный доступ / ограниченный доступ)
1	***2***	***3***
Информационно-справочные системы		
Библиографические базы данных по общественным наукам ИНИОН	http://inion.ru/	Свободный доступ
Методология: словарь системы основных понятий. Сайт академика РАО Новикова А.М.	http://www.anovikov.ru/dict/met_sl.htm	Свободный доступ

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

Наименование	Ссылка на ресурс	Доступность (свободный доступ / ограниченный доступ)
1	***2***	***3***
Электронно-библиотечные системы		
ЭБС Юрайт	https://urait.ru	После регистрации с компьютера Института – доступ с любого компьютера, подключенного к сети Интернет.
Профессиональные базы данных		
Справочно-информационный портал ГРАМОТА.РУ – русский язык для всех	http://gramota.ru/	Свободный доступ
«ИНФРА–М»	http://znanium.com	Свободный доступ
BOOK.ru	https://www.book.ru	После регистрации с компьютера Института – доступ с любого компьютера, подключенного к сети

		Интернет
Научная электронная библиотека РИНЦ	https://elibrary.ru/	Свободный доступ
Научная электронная библиотека «КиберЛенинка»	https://cyberleninka.ru/	Свободный доступ

9. Язык преподавания: русский

10. Разработчики программы

Лунькова Е. С., д.филол. наук, профессор кафедры ОиРЯ ФГБОУ ВО «Государственный институт русского языка им. А. С. Пушкина»