



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Министерство науки и высшего образования Российской Федерации

ИВАНОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Кафедра зарубежной филологии

ОДОБРЕНО:

Руководитель ОП

С.А. Маник

(подпись)

« 29 » _____ сентября 2024 г.

Рабочая программа дисциплины
Применение Big Data в гуманитарной сфере

Уровень высшего образования:	магистратура
Квалификация выпускника:	магистр
Направление подготовки:	45.04.01 Филология
Направленность (профиль) образовательной программы:	Прикладная цифровая филология

Иваново 2024



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

1. Цели освоения дисциплины

Целью освоения дисциплины является обучение магистрантов навыкам использования больших (лингвистических) данных и управления ими для последующей разработки стратегий продвижения лингвистических продуктов и услуг в цифровом пространстве, а также создания цифрового продукта в своей профессиональной деятельности. Обучающиеся изучают ландшафт больших данных, включая характеристики больших данных и процессы работы с ними, знакомятся с примерами применения больших данных в гуманитарной сфере. Данная дисциплина призвана объяснить основные причины, условия и возможности применения компьютерных технологий в научных филологических исследованиях, прикладной, в том числе образовательной деятельности филолога; познакомить магистров с существующими разработками ИИ и доступными программными продуктами для осуществления профессиональной деятельности филолога.

В ходе занятий предполагается теоретическое знакомство с общими базовыми понятиями и принципами автоматизации работы филолога, приобретение базовых навыков использования компьютерных языков для работы с текстовыми и лингвистическими данными, а также осуществление практической подготовки обучающихся посредством выполнения определенных видов работ, связанных с будущей профессиональной деятельностью и направленных на формирование, закрепление, развитие практических навыков и компетенций по профилю образовательной программы.

Курс ориентирован на магистров ОП «Прикладная цифровая филология», примеры и задачи подбираются с учетом профессиональных интересов магистров. Особое внимание уделяется практическому освоению терминологии и определениям базовых терминов.

2. Место дисциплины в структуре ОП

Дисциплина относится к части образовательной программы, формируемой участниками образовательных отношений Б1.В.07.

Успешное освоение данной дисциплины способствует готовности студентов к профессиональной деятельности и предваряет научно-исследовательскую работу по направлению «Филология», направленность «Прикладная цифровая филология».

Данная учебная дисциплина опирается на материал предшествующих теоретических курсов «Филология в системе гуманитарного знания», «Информационные технологии в науке и образовании», «Основы программирования для филологов», «Управление цифровыми проектами», «Терминоведение и базы данных», «Цифровые инструменты в переводе», «Машинное обучение», «Программирование и анализ данных».

Для освоения данной дисциплины обучающийся должен:

Знать: основные аспекты цифровизации образования в мире и в России в частности; процессы цифровизации в гуманитарных областях; цифровые продукты (типа Алиса, цифровые музеи, и т.п.);

Уметь: находить информацию в сети, четко формулировать поисковый запрос;

Иметь: базовые цифровые навыки

3. Планируемые результаты обучения по дисциплине

3.1. Компетенции, формированию которых способствует дисциплина

При освоении дисциплины формируются следующие компетенции в соответствии с ФГОС ВО по данному направлению подготовки:



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

УК-4 - Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия

ПК 2 - Способен выявлять актуальные научные проблемы в области филологии, в том числе находящиеся на стыке различных областей наук, разрабатывать подходы к их решению и проводить научные исследования самостоятельно и в составе исследовательских коллективов с использованием информационных технологий

ПК 3 - Способен публично представлять результаты научных исследований и научно-исследовательских проектов в современной форме, включая результаты собственной научной деятельности, в том числе на иностранном(ых) языке(ах)

ПК-5 - Способен использовать современные информационные технологии, управлять данными с использованием прикладных программ, баз данных и корпусов текстов, выполнять перевод текстов на иностранный язык с их помощью

3.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения формируемых компетенций

В результате освоения дисциплины обучающийся должен:

Знать:

- теоретические, методологические и практические основы использования компьютерных технологий в лингвистике, мета язык цифровой гуманитаристики (ПК-2, ПК-5);
- общие принципы устройства и работы современной вычислительной техники (ПК-5);
- базовые понятия технологии Big Data для гуманитарной сферы (УК-4);
- подходы и этапы обработки текста на естественного языка (ПК-2, ПК-5);
- алгоритмы и этапы компьютерного (автоматического) анализа текста (ПК-2, ПК-5);
- основные виды корпусов на русском, английском и немецком языках, принципы работы с корпусом (ПК-2, ПК-5);

Уметь:

- определять массивы больших данных (УК-4, ПК-5);
- составлять собственный исследовательский корпус (ПК-4, ПК-5);
- осуществлять кодировку текста и сортировку элементов списка (ПК-5);
- использовать возможности и инструменты ИИ и компьютерных технологий при решении собственных задач в исследовательской и практической работе (ПК-4, ПК-5);
- анализировать кластеры больших данных в гуманитарной сфере (ПК-5);
- самостоятельно изучать, подключать и использовать компьютерные программы и модули, необходимые для решения конкретных задач компьютерной лингвистики (ПК-4, ПК-5);

Иметь практический опыт/Иметь навыки:

- использовать инструменты для работы с большими данными естественного языка, составления собственного исследовательского корпуса (ПК-4; ПК-5);
- внедрять современные технологии создания и обслуживания больших данных (УК-4, ПК-5);
- применять наиболее популярные компьютерные лингвистические инструменты, библиотеки и приложения для работы с лингвистическими данными (ПК-4, ПК-5).

4. Объем и содержание дисциплины

Объем дисциплины составляет 4 зачетных единиц (144 академических часов), в т.ч.: практическая подготовка (ПП) – 32 академических часов в очной форме



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

4.1. Содержание дисциплины по разделам (темам), соотнесенное с видами и трудоемкостью занятий лекционно-семинарского типа

Объем иной контактной работы и самостоятельной работы обучающегося по дисциплине указан в учебном плане образовательной программы.

№ п/п	Разделы (темы) дисциплины	Семестр	Виды занятий, их объем (в ак. часах, по очной форме обучения)		Формы текущего контроля успеваемости (по очной форме обучения)
			Занятия лекционного типа	Занятия семинарского типа	Формы промежуточной аттестации
1.	Вводный. Введение в проблематику дисциплины, представление рабочей программы, осмысление требований к организации процесса обучения, самостоятельной работы и форм аттестации. Понимание ландшафта больших данных	4		2 практ. занятие	Входная диагностика: тест с последующим обсуждением результатов. Список вопросов, интересующих студента по содержанию дисциплины (сдается в письменном виде)
2.	Примеры работы с большими данными	4		16 практ. занятие	Практическое задание, презентация, контрольная работа, опрос
3.	Подходы и этапы обработки текста на ЕЯ. Компьютерный (автоматический) анализ текста	4		4 практ. занятие	Устный опрос, практическое задание
4.	Корпусная лингвистика: истоки, определение и текущее состояние. Перспективы корпусной лингвистики. Понятийный аппарат	4		2 практ. занятие	Контрольная работа
5.	Инструменты корпусной лингвистики. Принципы работы с корпусом.	4		4 практ. занятие	Отчет
6.	Составление собственного исследовательского корпуса. Заключительный. Подведение и анализ результатов освоения дисциплины	4		4 практ. занятие	практическое задание
Итого за семестр:				32	Экзамен
Итого по дисциплине:				32	

4.2. Развернутое описание содержания дисциплины по разделам (темам)

1. Вводный. Введение в проблематику дисциплины, представление рабочей программы, осмысление требований к организации процесса обучения, самостоятельной работы и форм аттестации. Понимание ландшафта больших данных

Определение, цель работы с большими данными. Представление данных и динамика. Процесс. Вычислительная мощность. Характеристики больших данных

2. Примеры работы с большими данными

Предикативная аналитика. Big data в промышленности: предсказание аварий и оптимизация производства. Big data в финансах: оценка платежеспособности и повышение качества сервиса.



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Big data в HR: наем сотрудников и предупреждение увольнений. Big data в образовании: помощь в выборе курсов и предотвращение отчислений. Перспективы big data. Использование больших данных в лингвистических исследованиях. Интеллектуальный анализ.

3. Подходы и этапы обработки текста на ЕЯ. Компьютерный (автоматический) анализ текста

Графемный анализ текста, морфологический анализ, синтаксический анализ, семантический анализ

4. Корпусная лингвистика: истоки, определение и текущее состояние. Перспективы корпусной лингвистики. Понятийный аппарат

Корпусная лингвистика как новое направление в филологии. Понятие перформативности, проявление перформативности в языке, отражение метонимии в словарях, референциальные свойства слова, типы референции. Корпус. Конкорданс. Репрезентативность. Полнота корпуса.

5. Инструменты корпусной лингвистики. Принципы работы с корпусом.

6. Составление собственного исследовательского корпуса. Заключительный. Подведение и анализ результатов освоения дисциплины

5. Образовательные технологии

Рекомендуемые образовательные технологии: практические занятия с технологиями учебной дискуссии, проектная работа, представляемая в виде презентаций, опросы, конспекты, технологии смешанного обучения, индивидуальные задания, самостоятельная работа студентов, контрольные работы. Студенты знакомятся с технологиями работы и обработки больших лингвистических данных.

При проведении практических занятий рекомендуется помимо обсуждения теоретических вопросов использование проектных методик, выполнение практических заданий на основе Интернет-ресурсов, преимущественно англоязычных, работа с известными корпусами текстов на разных языках. Студенты представляют результаты своей самостоятельной и практической работы в виде презентаций с использованием программы PowerPoint и публикации материалов в ЭИОС.

6. Учебно-методическое обеспечение самостоятельной работы обучающихся

Студентам на ресурсе «Мой университет» выкладываются ссылки на записанные вебинары по разным темам, которые они могут посмотреть в удобное время; список площадок дистанционных курсов, на которые они должны подписаться и прислать отчет прохождения (скриншоты по неделям и успеваемость по курсу).

Полностью весь методический материал по обеспечению самостоятельной работы студентов приводится в Приложении 1 к РП.

7. Характеристика оценочных средств для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

Вопросы к обсуждению в ходе собеседования. Индивидуальные задания/работы с постановкой конкретной задачи и предписанным алгоритмом ее выполнения. Устный опрос. Подготовка проектов в форме презентации в PowerPoint.



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

Примерные вопросы к устному опросу.

- 1) Понятие «big data»
- 2) Применение больших данных в разных областях
- 3) Компьютерные методы лингвистических исследований.
- 4) Автоматический анализ текста.
- 5) Лингвистические модели.
- 6) Формализация языковой структуры.
- 7) Использование корпусов текстов в научных исследованиях.
- 8) Системы машинного перевода.
- 9) Локализация электронных ресурсов и программного обеспечения.
- 10) Лингвистические ресурсы и поиск в Интернет.
- 11) Лингвистические технологии информационного поиска.
- 12) Лингвистические аспекты искусственного интеллекта.

Критерии оценки опроса:

Оценка	Критерии
Отлично	Отвечает на все вопросы, продемонстрирована способность анализировать и обобщать информацию, сделаны обоснованные выводы, разъяснения
Хорошо	Отвечает не на все вопросы, но на большинство. Умеет в некоторой степени анализировать и обобщать информацию. Выводы умеет делать
Удовлетворительно	Отвечает не на все вопросы (меньше 50%). Демонстрирует некоторые способности анализировать и обобщать информацию. Выводы делать не может
Неудовлетворительно	Задание не выполнено

Опорный конспект – особый вид текста, в основе которого лежит аналитико-синтетическая переработка информации первоисточника (исходного текста).

Перечень тем для презентаций:

Перспективы автоматизации обработки данных

Big data в гуманитарных областях (по областям)

Виды корпусов

Национальный корпус русского языка: варианты поиска по корпусу

Также студенты выполняют индивидуальные задания по каждой теме, представляя результаты анализа и синтеза данных.

В конце магистранты сдают экзамен. Если студент присутствовал на 75% занятий и выполнил не менее 8 заданий на практических заданиях, то он получает «отлично» автоматически как накопительную оценку. Если данное требование не выполнено, то проходит собеседование с преподавателем.

Типовые варианты тестовой работы и вопросы к экзамену представлены в фонде оценочных средств (Приложение 2).

8. Учебно-методическое и информационное обеспечение дисциплины

Основная литература:



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

1. Моисеева, И. Ю. Квантитативная лингвистика и новые информационные технологии : учебное пособие / И. Ю. Моисеева ; Оренбургский государственный университет. – Оренбург : Оренбургский государственный университет, 2017. – 103 с. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=481797> (дата обращения: 06.09.2024). – Библиогр. в кн. – ISBN 978-5-7410-1713-5. – Текст : электронный.
2. Калугян, К. Х. Информационные технологии в лингвистике : учебное пособие : [16+] / К. Х. Калугян ; Ростовский государственный экономический университет (РИНХ). – Ростов-на-Дону : Издательско-полиграфический комплекс РГЭУ (РИНХ), 2021. – 52 с. : схем., табл. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=686245> (дата обращения: 06.09.2024). – Библиогр. в кн. – ISBN 978-5-7972-2901-8. – Текст : электронный
3. Грудева, Е. В. Корпусная лингвистика : учебное пособие : [16+] / 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> (дата обращения: 06.09.2024). – Библиогр. в кн. – ISBN 978-5-9765-1497-3. – Текст : электронный.
4. Долинский, В. А. Квантитативная лингвистика и новые информационные технологии : [16+] / В. А. Долинский ; Московский государственный лингвистический университет (МГЛУ). – Москва : Московский государственный лингвистический университет (МГЛУ), 2022. – 352 с. : ил., табл., схем. – Режим доступа: по подписке. – URL: <https://biblioclub.ru/index.php?page=book&id=710346> (дата обращения: 06.09.2024). – Библиогр. в кн. – ISBN 978-5-00120-343-8. – Текст : электронный.

Дополнительная литература:

1. [Электронный ресурс biblioclub: <http://biblioclub.ru/index.php?page=book&id=362986>] Моделирование систем: Подходы и методы: учебное пособие / В.Н. Волкова, Г.В. Горелова, В.Н. Козлов и др.; Министерство образования и науки Российской Федерации, Санкт-Петербургский государственный политехнический университет. – Санкт-Петербург: Издательство Политехнического университета, 2013. – 568 с.
2. 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> (дата обращения: 06.09.2024). – Библиогр.: с. 480-513. – ISBN 978-5-9907947-8-8. – Текст : электронный.
3. [Электронный ресурс biblioclub: <http://biblioclub.ru/index.php?page=book&id=480500>] Жуковский О. И. Информационные технологии и анализ данных: учебное пособие / О.И. Жуковский - Томск: Эль Контент, 2014. - 130 с.
4. Электронный ресурс: <http://www.sfs.uni-tuebingen.de/~hbaayen/publications/baayenCUPstats.pdf> Harald Baayen (2008). Analyzing Linguistic Data. A Practical Introduction to Linguistics using R, Cambridge University Press: Cambridge

Ресурсы информационно-телекоммуникационной сети «Интернет»:

Система электронной поддержки образовательного процесса «Мой университет» <https://uni.ivanovo.ac.ru>
Профессиональные базы данных и информационно-справочные системы:



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

ЭБС «Университетская библиотека онлайн» www.biblioclub.ru;
<http://lib.ivanovo.ac.ru/index.php/polnotekstovye-resursy/ebs-universitetskaya-biblioteka>
Электронная библиотека ИвГУ <http://lib.ivanovo.ac.ru/index.php/polnotekstovye-resursy/elibnew>
Электронный каталог НБ ИвГУ <http://lib.ivanovo.ac.ru/index.php/ek>

Программное обеспечение: операционная система Microsoft Windows, пакет офисных программ Microsoft Office и(или) LibreOffice, интернет-браузер Microsoft Edge и(или) Yandex Browser.

9. Материально-техническое обеспечение дисциплины

Учебные аудитории:

- для проведения занятий семинарского типа, консультаций, текущего контроля и промежуточной аттестации с комплектом специализированной учебной мебели и техническими средствами обучения;

Лаборатория, оснащенная лабораторным оборудованием, комплектом специализированной учебной мебели и техническими средствами обучения.

Помещение для самостоятельной работы, оснащенное комплектом специализированной учебной мебели, компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в ЭИОС.

Демонстрационное оборудование и учебно-наглядные пособия для занятий лекционного типа, обеспечивающие тематические иллюстрации: презентации PowerPoint, электронные словари и справочно-информационные порталы, видеоматериалы и т.п

Автор(ы) рабочей программы дисциплины: доцент кафедры английской филологии, д.ф.н., доц. С.А. Маник

Программа рассмотрена и утверждена на заседании кафедры зарубежной филологии
«29» августа 2024 г., протокол № 13.

Программа обновлена
протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)

Программа обновлена
протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)

Программа обновлена



Основная профессиональная образовательная программа
45.04.01 Филология
(Прикладная цифровая филология)

протокол заседания кафедры № _____ от «_____» _____ 20__ г.

Согласовано:

Руководитель ОП _____ / _____

(подпись)