

МИНОБРНАУКИ РОССИИ

**Федеральное государственное автономное образовательное
учреждение высшего образования "Пермский
государственный национальный исследовательский
университет"**

Авторы-составители: **Исаева Екатерина Владимировна**

Программа учебной практики

ПРАКТИКА ПО АВТОМАТИЗИРОВАННОМУ АНАЛИЗУ ТЕКСТА

Код УМК 98507

Утверждено
Протокол №9
от «12» мая 2023 г.

Пермь, 2023

1. Вид практики, способ и форма проведения практики

Вид практики **учебная**

Тип практики **учебно-ознакомительная практика**

Способ проведения практики **стационарная, выездная**

Форма (формы) проведения практики **дискретная**

2. Место практики в структуре образовательной программы

Учебная практика « Практика по автоматизированному анализу текста » входит в базовую часть Блока « М.2 » образовательной программы по направлениям подготовки (специальностям):

Направление: **45.04.02** Лингвистика

направленность Цифровая лингвистика и перевод

Цель практики :

Сформировать компетенцию по парсингу и обработке данных на естественном языке для решения лингвистических задач.

Задачи практики :

1. Закрепить навык парсинга текстовых данных;
2. Закрепить навык автоматической обработки текстовых данных;
3. Научить студентов готовить научный отчет, документирование и визуализацию полученных результатов.

3. Перечень планируемых результатов обучения

В результате прохождения практики **Практика по автоматизированному анализу текста** у обучающегося должны быть сформированы следующие компетенции:

45.04.02 Лингвистика (направленность : Цифровая лингвистика и перевод)

ОПК.7 Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации

Индикаторы

ОПК.7.2 Проводит лингвистический анализ дискурса с использованием современных экспертных систем и систем управления базами данных

ПК.3 Способен проводить самостоятельные научные исследования по актуальной проблематике в области лингвистики, переводоведения и межкультурной коммуникации

Индикаторы

ПК.3.2 Применяет современные методики поиска, анализа и обработки материала исследования на основе системного представления о динамике развития переводоведения, лингвистики и межкультурной коммуникации

4. Содержание и объем практики, формы отчетности

Данная практика является естественным продолжением дисциплин: "Computer Linguistics for International Professional Communication", "Основы программирования для решения профессиональных задач" и практик: "Программирование для решения профессиональных задач (ознакомительная практика)", "Data Processing in Linguistics". Во время этой практики студенты будут применять свои навыки в использовании методов Low-Coding для обработки текстовых данных, в т.ч. больших данных. Студенты закрепят навыки собирать и анализировать текстовые данные, находить в них лингвистические закономерности. Будет выполнен финальный проект, который потребует от студентов проведения исследования текстовых данных и языка, подготовки отчета с визуализацией данных, выступления на конференции и написания научной статьи.

Направление подготовки	45.04.02 Лингвистика (направленность: Цифровая лингвистика и перевод)
форма обучения	очная
№№ триместров, выделенных для прохождения практики	3
Объем практики (з.е.)	3
Объем практики (ак.час.)	108
Форма отчетности	Экзамен (3 триместр)

Примерный график прохождения практики

Количество часов	Содержание работ	Место проведения
1. Парсинг и очистка данных на естественном языке		
36	Этот этап стажировки посвящен парсингу языковых данных для последующего проведения лингвистического исследования. Этап предполагает работу с платформами данных, такими как kaggle, очистку данных, избавление от стоп-слов, выбросов или избыточных данных, разметку, классификацию и т.д.	Кафедра английского языка профессиональной коммуникации
2. Автоматизированная обработка данных на естественном языке		
36	На данном этапе практики студенты применяют навыки программирования, полученные на дисциплине "Основы программирования для решения профессиональных задач" и практике "Программирование для решения профессиональных задач (ознакомительная практика)" для обработки полученных языковых данных и решения научных лингвистических задач.	Кафедра английского языка профессиональной коммуникации
3. Подготовка научного отчета и публикация материалов		
36	Этот этап посвящен документированию результатов, визуализации данных, написанию научной работы (научной статьи) и представлению результатов на научной конференции. Условия организации и проведения практики для лиц с ОВЗ	Кафедра английского языка профессиональной коммуникации

Количество часов	Содержание работ	Место проведения
	<p>и инвалидностью.</p> <p>Для лиц с ОВЗ и инвалидностью программой предусмотрены альтернативные места прохождения практики. Индивидуальная программа практики составляется с учетом рекомендаций психолого-медико-педагогической комиссии или МСЭ.</p> <p>Возможно изменение временных рамок прохождения текущей промежуточной аттестации, а именно, предусмотрено увеличение времени на подготовку и сдачу отчетности по практике в пределах 1 академического часа.</p> <p>Формы проведения практики определяются с учетом особенностей психофизиологического развития, индивидуальных возможностей и состояния здоровья обучающихся.</p>	

5. Перечень учебной литературы, необходимой для проведения практики

Основная

1. Кознов, Д. В. Основы визуального моделирования : учебное пособие / Д. В. Кознов. — 3-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 246 с. — ISBN 978-5-4497-0674-4. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <http://www.iprbookshop.ru/97561.html>

2. Онтологии и тезаурусы: модели, инструменты, приложения : учебное пособие / Б. В. Добров, В. В. Иванов, Н. В. Лукашевич, В. Д. Соловьев. — 3-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 172 с. — ISBN 978-5-4497-0668-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/97555.html>

Дополнительная

1. Valentina Janev. Knowledge Graphs and Big Data Processing / Valentina Janev, Damien Graux, Hajira Jabeen, Emanuel Sallinger // — 2020. — 209 p. — ISBN978-3-030-53199-7. [Электронный ресурс]. <https://link.springer.com/book/10.1007/978-3-030-53199-7>

2. Протодяконов, А. В. Алгоритмы Data Science и их практическая реализация на Python : учебное пособие / А. В. Протодяконов, П. А. Пылов, В. Е. Садовников. — Москва, Вологда : Инфра-Инженерия, 2022. — 392 с. — ISBN 978-5-9729-1006-9. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/124000>

6. Перечень ресурсов сети «Интернет», требуемых для проведения практики

При прохождении практики требуется использование следующих ресурсов сети «Интернет» :

<https://colab.research.google.com/> Google Colab

<https://www.kaggle.com/> Kaggle

<https://huggingface.co/> Hugging Face

7. Перечень информационных технологий, используемых при проведении практики

Образовательный процесс по практике **Практика по автоматизированному анализу текста** предполагает использование следующего программного обеспечения и информационных справочных систем:

- 1) презентационные материалы (слайды по темам лекционных и практических занятий);
- 2) доступ в режиме on-line в Электронную библиотечную систему (ЭБС)
- 3) доступ в электронную информационно-образовательную среду университета.

Перечень необходимого лицензионного и (или) свободно распространяемого программного обеспечения:

1. Приложение, позволяющее просматривать и воспроизводить медиаконтент PDF-файлов «Adobe Acrobat Reader DC».
2. Программы, демонстрации видео материалов (проигрыватель) «WindowsMediaPlayer».
3. Программа просмотра интернет контента (браузер) «Google Chrome».
4. Офисный пакет приложений «LibreOffice» или Microsoft Office 2003 Professional.
5. Справочно-правовая программа КонсультантПлюс

Дополнительно может использоваться:

1. Лицензионное программное обеспечение Sanako Lab 250;
2. Электронный словарь Abby Lingvo;
3. Программное обеспечение свободного доступа Hot Potatoes, Quizlet.

8. Описание материально-технической базы, необходимой для проведения практики

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕЛБС ПГНИУ (student.psu.ru). При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

- 1) Практические занятия проводятся в аудитории, оснащенной презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением и (или) компьютерном классе (аппаратное и программное оснащение определено в паспорте компьютерного класса).
- 2) система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).
- 2) система LMS Moodle (<http://e-team.psu.ru/>), которая поддерживает возможность использования подключения к сети «Интернет», обеспеченная доступом в электронную информационно-образовательную среду университета. Помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.
5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными

компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

9. Методические указания для обучающихся по прохождению практики

В соответствии с настоящей Программой и формируемыми компетенциями, в период переводческой практики студенты призваны осуществлять следующие виды практической деятельности:

1) предобработка данных:

подготовительная фаза аналитической деятельности (парсинг данных на естественном языке, очистка данных, восстановление пропущенных данных, устранение противоречий и т.д.); постановка исследовательских задач и прогнозирование основных результатов, экспертная обработка данных. Навыки работы с документацией: изучение литературы и Интернет источников; навыки составления технического задания; постоянное расширение общего кругозора и тематических знаний.

2) собственно аналитическая деятельность:

- подготовка алгоритма автоматизированной обработки данных (возможно использование готовых шаблонов с последующей доработкой адаптацией модели в соответствии с исследовательскими задачами);

- сбор аналитики, подтверждающих и валидационных материалов;

3) документирование результатов

- подготовка отчета по практике (техническая и научная составляющая, подготовка графической информации);

- подготовка научной публикации (студент обязан представить текст научной статьи по результатам реализации исследовательского проекта на рецензирование научному руководителю).

Студент-практикант обязан:

1. Ознакомиться с программой переводческой практики.

2. Своевременно, качественно и в полном объеме выполнить содержательную часть программы переводческой практики.

3. Соблюдать установленные в принимающей организации правила внутреннего распорядка и трудовой дисциплины. В период производственной практики на студента распространяются правила охраны труда и правила внутреннего распорядка, действующие в принимающей организации.

4. Соблюдать технику безопасности на рабочем месте.

5. Выполнять требования руководителей (от ПГНИУ и организации).

6. Соблюдать профессиональную переводческую этику

7. В установленный срок отчитаться о практике, сдав все отчетные документы.

Фонды оценочных средств для проведения промежуточной аттестации

Планируемые результаты обучения по практике для формирования компетенции. Индикаторы и критерии их оценивания

ОПК.7

Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ОПК.7.2 Проводит лингвистический анализ дискурса с использованием современных экспертных систем и систем управления базами данных</p>	<p>Знать современные инструменты для автоматизированного дискурсивного анализа; Уметь использовать программные средства и настраивать рабочие процессы для автоматизации дискурсивного анализа; Владеть современными инструментами для качественного дискурсивного анализа.</p>	<p style="text-align: center;">Неудовлетворительно</p> <p>Не сформированы знания, умения и навыки, предусмотренные компетенцией.</p> <p style="text-align: center;">Удовлетворительно</p> <p>Знает современные инструменты для автоматизированного дискурсивного анализа; Умеет использовать программные средства, но не умеет настраивать рабочие процессы для автоматизации дискурсивного анализа; Не владеет современными инструментами для качественного дискурсивного анализа.</p> <p style="text-align: center;">Хорошо</p> <p>Знает современные инструменты для автоматизированного дискурсивного анализа; Умеет использовать программные средства и настраивать рабочие процессы для автоматизации дискурсивного анализа; Не владеет современными инструментами для качественного дискурсивного анализа.</p> <p style="text-align: center;">Отлично</p> <p>Знает современные инструменты для автоматизированного дискурсивного анализа; Умеет использовать программные средства и настраивать рабочие процессы для автоматизации дискурсивного анализа; Владеет современными инструментами для качественного дискурсивного анализа.</p>

ПК.3

Способен проводить самостоятельные научные исследования по актуальной проблематике в области лингвистики, переводоведения и межкультурной коммуникации

Индикатор	Планируемые результаты обучения	Критерии оценивания результатов обучения
<p>ПК.3.2</p>	<p>Знать современные</p>	<p style="text-align: center;">Неудовлетворительно</p>

<p>Применяет современные методики поиска, анализа и обработки материала исследования на основе системного представления о динамике развития переводоведения, лингвистики и межкультурной коммуникации</p>	<p>инструменты для обработки текстовых данных; Уметь использовать программные средства и настраивать рабочие процессы для автоматизации парсинга, подготовки и обработки текстовых данных; Владеть современными программными средствами для оценки результатов автоматизированной обработки данных.</p>	<p>Неудовлетворительно Не сформированы знания, умения и навыки, предусмотренные компетенцией.</p> <p>Удовлетворительно Знает современные инструменты для обработки текстовых данных; Не умеет использовать программные средства и настраивать рабочие процессы для решения двух из задач: автоматизация парсинга, подготовки и обработки текстовых данных; Либо не умеет использовать программные средства и настраивать рабочие процессы для решения одной из задач: автоматизация парсинга, подготовки и обработки текстовых данных и не владеет современными программными средствами для оценки результатов автоматизированной обработки данных .</p> <p>Хорошо Знает современные инструменты для обработки текстовых данных; Не умеет использовать программные средства и настраивать рабочие процессы для решения одной из задач: автоматизация парсинга, подготовки и обработки текстовых данных; Либо не владеет современными программными средствами для оценки результатов автоматизированной обработки данных.</p> <p>Отлично Знает современные инструменты для обработки текстовых данных; Умеет использовать программные средства и настраивать рабочие процессы для автоматизации парсинга, подготовки и обработки текстовых данных; Владеет современными программные средства для оценки результатов автоматизированной обработки данных.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Оценочные средства

Вид мероприятия промежуточной аттестации : Экзамен

Способ проведения мероприятия промежуточной аттестации : Защищаемое контрольное мероприятие

**Продолжительность проведения мероприятия промежуточной аттестации :
время отводимое на доклад 4**

Показатели оценивания

Не сформированы знания, умения и навыки, предусмотренные компетенцией, отчет по практике не подготовлен или не соответствует требованиям.	Неудовлетворительно
Знает инструменты онтологизации текстовых данных, принципы работы с открытым программным кодом. Не умеет применять готовый исходный программный код для обработки текстовых данных, способен адаптировать программный код для решения собственных исследовательских задач, умеет осуществлять систематизацию предметной области, разрабатывать для нее классификаторы, рубрикаторы, проводить типологизацию данных. Владеет инструментами формата Low-Coding для принятия датацентричных решений на основе больших данных в лингвистике, владеет инструментами автоматического моделирования предметной области и разработки онтологий. Отчет по практике составлен в соответствии с требованиями.	Удовлетворительно
Знает инструменты онтологизации текстовых данных, принципы работы с открытым программным кодом. Умеет применять готовый исходный программный код для обработки текстовых данных, способен адаптировать программный код для решения собственных исследовательских задач, не умеет осуществлять систематизацию предметной области, разрабатывать для нее классификаторы, рубрикаторы, проводить типологизацию данных. Владеет инструментами формата Low-Coding для принятия датацентричных решений на основе больших данных в лингвистике, не владеет инструментами автоматического моделирования предметной области и разработки онтологий. Отчет по практике составлен в соответствии с требованиями.	Хорошо
Знает инструменты онтологизации текстовых данных, принципы работы с открытым программным кодом. Умеет применять готовый исходный программный код для обработки текстовых данных, способен адаптировать программный код для решения собственных исследовательских задач, умеет осуществлять систематизацию предметной области, разрабатывать для нее классификаторы, рубрикаторы, проводить типологизацию данных. Владеет инструментами формата Low-Coding для принятия датацентричных решений на основе больших данных в лингвистике, владеет инструментами автоматического моделирования предметной области и разработки онтологий. Отчет по практике составлен в соответствии с требованиями.	Отлично