

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ПЕТРОЗАВОДСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ИНСТИТУТ МАТЕМАТИКИ И ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

ОТЧЕТ О ПРОХОЖДЕНИИ ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ

Выполнила студентка 4 курса группы 22407
Екатерина Сергеевна Жданович

Направление подготовки:
09.03.04 — "Программная инженерия"

Место прохождения практики:
ООО "ГЕА РУ Си-Ай-Эс"

Период прохождения практики: 11.01.21–31.01.21

Руководитель практики:
О. Ю. Богоявленская, доцент

Оценка

оценка

Дата

дата

Петрозаводск — 2021

Содержание

1	Характеристика предприятия	3
2	Используемые технологии	5
3	Задачи, решаемые в ходе практики	7
3.1	Ознакомление с теоретической частью	7
3.2	Анализ веб-сайтов	8
3.3	Извлечение данных с сайта	8
3.4	Работа с данными, преобразование их в необходимый формат	8
3.5	Геокодирование	8
3.6	Визуализация и анализ результатов	9
3.7	Создание итоговой презентации	10
4	Дневник производственной практики	11
5	Заключение	12
	Список использованных источников	13

1 Характеристика предприятия

ООО "ГЕА РУ Си-Ай-Эс в котором я проходила производственную практику, является российским подразделением компании HERE Technologies. HERE Technologies — это международная технологическая компания, которая занимается разработкой картографического контента, предоставляет данные о местоположении, а также сопутствующие услуги частным лицам и компаниям. Компания признается лидером по применению навигационных систем.

Комплексная платформа (HERE Platform) решает проблему фрагментированных и разрозненных источников данных. Сервисы для транспортных средств используют и передают данные с датчиков в режиме реального времени в транспортные средства и обратно. Используется в многочисленных пакетах программного обеспечения, включая системы навигации. Карты HERE включают такие функции, как поиск по карте, спутниковые карты, построение маршрутов, 3D-карты, отражение пробок в реальном времени.

В портфеле компании более 40 картографических продуктов. В спектр их возможностей входят:

- обеспечение автомобильного сервиса, оснащение машинных бортовых систем;
- организация мобильности и отслеживания деятельности выездного персонала;
- возможность отслеживания геоаналитики, геовизуализация и обработка других геолокационных данных;
- мобильная помощь в управлении транспортным парком, включая обеспечение поддержкой техники и персонала;
- умные города — система управления общественным транспортом и другой инфраструктурой.

HERE Technologies - ведущая мировая платформа геоинформационных данных. Ядро решений HERE — это постоянно развивающаяся и обновляющаяся карта. Платформы HERE Technologies принимают информацию из 80 тыс. источников. В них входят: глобальная карта автомобильных средств, информация GPS-зондов, датчики на автотранспорте, а также открытые базы местных органов власти.

200 стран покрыты картами HERE, в 63 странах предоставляется информация о пробках, в 33 странах — о движении в режиме реального времени. 4 из 5 машин с навигационной системой в Европе и Северной Америке используют карты HERE. Карты HERE на борту 150 миллионов транспортных средств.

В компании работают более 9000 сотрудников в 56 странах мира.

Подразделение HERE Technologies в России осуществляет сбор, частичную обработку и контроль за обработкой данных, а также участвует в подготовке карты и продуктов к релизу. Выпуск продуктов делается не в России, а в Германии и США.

Уже более 7 лет компания HERE Technologies активно сотрудничает с ведущими высшими учебными заведениями нашей страны. В рамках такого сотрудничества специалисты компании HERE, руководители университетов и студенты участвуют в совместных проектах и организуют производственные и учебные практики. Концепция подготовки молодых специалистов была разработана весной 2013 года, а сегодня ее реализуют множество университетов в России и странах СНГ - студенты занимаются созданием геолокационной основы в картографическом веб-сервисе Map Creator, принимают участие в проектах по сбору данных на местности с последующей камеральной обработкой, участвуют в тренингах и мапатонах, посвященных разработке новых продуктов и скоростному картированию.

2 Используемые технологии

В ходе практики я выполняла задания на языке Python. Дополнительно использовала библиотеки:

BeautifulSoup — это библиотека Python для извлечения данных из файлов HTML и XML. Может преобразовать даже неправильную разметку в дерево синтаксического разбора. Он поддерживает простые и естественные способы навигации, поиска и модификации дерева синтаксического разбора. В большинстве случаев он поможет сэкономить часы и дни работы.

Библиотека использовалась мной для извлечения данных с сайтов и выделения определенных данных.

Pandas - программная библиотека для обработки и анализа данных. Основная область применения — обеспечение работы в рамках среды Python не только для сбора и очистки данных, но для задач анализа и моделирования данных, без переключения на более специфичные языки. Основные возможности библиотеки:

- Объект DataFrame для манипулирования индексированными массивами двумерных данных.
- Инструменты для обмена данными между структурами в памяти и файлами различных форматов.
- Встроенные средства совмещения данных и способы обработки отсутствующей информации.
- Переформатирование наборов данных, в том числе создание сводных таблиц.
- Слияние и объединение наборов данных.
- И т.д.

В рамках практики я использовала библиотеку pandas для обработки данных, полученных с сайтов.

Requests — библиотека, позволяющая взаимодействовать с веб-приложениями. Это необходимо для решения любых задач, связанных с передачей информации от пользо-

вателя к серверу и обратно. Я использовала эту библиотеку для отправки запросов с целью получения данных с сайтов.

Json. Модуль `json` из стандартной библиотеки языка программирования Python является достаточно эффективным средством для взаимодействия с форматом JavaScript Object Notation. Функции данного модуля позволяют легко кодировать и декодировать информацию, работая с разными JSON-объектами. Это значительно повышает удобство создания веб-приложений в Python. В ходе практики я использовала библиотеку `Json` для перевода данных в Json-формат для дальнейшей работы с ними.

Geojson. Модуль аналогичен модулю `json`, используется для работы с `geojson`-объектами. Библиотека `geojson` использовалась мной для перевода данных в `geojson`-формат, так как для дальнейшей визуализации требовался именно этот формат.

Для визуализации результатов я использовала `HERE Studio`. `HERE Studio` - это интерактивное визуальное веб-приложение для доступа к данным и создания карт без использования кода. Позволяет создавать интерактивные пользовательские карты без программирования или внутренней инфраструктуры. Можно загружать большие наборы данных в различных форматах и геокодировать для мгновенного просмотра и редактирования.

Внешний вид карт можно настроить с помощью интуитивно понятного графического интерфейса. Можно выбрать из различных стилей базовой карты и настроить цвета, текст надписи, линии, маркеры и другие стилистические элементы. Также существует возможность создать интерактивные карточки для описания объектов пользователю, когда он выбирает точку, линию или многоугольник на карте. Можно добавить на карту такие объекты, как точки, линии или полилинии.

В ходе практики я использовала `HERE Studio` для визуализации полученных данных. Данные в формате `geojson` я загружала в веб-приложение, и далее они отображались на карте. Таким образом я проверяла, нет ли никаких проблем с данными. В финальной части практики, `HERE Studio` потребовалось мне для того, чтобы визуально увидеть разницу между реальными координатами объекта и координатами, полученными в результате геокодирования адреса объекта.

3 Задачи, решаемые в ходе практики

Моими основными задачами на практику были:

1. Ознакомление с теоретической частью
2. Анализ веб-сайтов
3. Извлечение данных с сайта
4. Работа с данными, преобразование их в необходимый формат
5. Геокодирование
6. Визуализация и анализ результатов
7. Создание итоговой презентации

3.1 Ознакомление с теоретической частью

В начале практики была проведена серия вебинаров для ознакомления с теоретической частью. В ходе вебинаров я ознакомилась с такими вещами, как веб-скрейпинг, геокодирование и анализ его точности, визуализация геоданных, кластеризация и поиск дубликатов. На вебинарах подробно рассказывалось то, с чем мне предстоит работать далее.

В ходе практики мне нужно было провести веб-скрейпинг точек с 5 сайтов, обработать полученные данные и визуализировать их. Для одного из сайтов нужно было произвести геокодирование, проанализировать результаты и также визуализировать их. Список сайтов, которые я выбрала для работы:

- sushiwok.ru
- tanuki.ru
- victoria-group.ru
- i-teka.kz
- aparking.kz

3.2 Анализ веб-сайтов

Первым этапом работы над каждым сайтом был его анализ. Нужно было понять, как брать данные с сайта — из HTML-кода или по API. Для этого я заходила на сайт, проверяла, нет ли необходимых данных в ответах от сервера при загрузке страницы. Иногда возникали такие ситуации, когда данные приходили не в одном ответе, а в нескольких. Далее я искала данные в Html-коде. Анализируя html-код, нужно было сначала найти нужные данные, а потом уже понять структуру — внутри каких тегов они находятся.

3.3 Извлечение данных с сайта

Для извлечения данных с сайта, я посылала запрос с помощью библиотеки requests. В ответ мне присылался либо ответ от сервера (в случае получения данных по API), либо весь html-код запрашиваемой страницы. В этом случае мне нужно было дополнительно работать с полученными данными.

3.4 Работа с данными, преобразование их в необходимый формат

В случае получения данных через html-код, нужно было дополнительно извлечь тот набор данных, который необходим. Для начала, с помощью библиотеки BeautifulSoup, я извлекала все скрипты из кода, находила тот, в котором лежат нужные данные. Далее, я работала с этими данными, переводила их в формат DataFrame для более удобного представления. И в конце преобразовывала данные в geojson-формат для дальнейшей визуализации.

3.5 Геокодирование

Для данных, полученных с сайта sushiwok.ru я провела геокодирование, то есть назначение географических координат точкам через их адрес. Для этого мне пришлось дополнительно поработать с данными, так как для геокодирования нужно было использовать их полный адрес. В результате, я получила 3 geojson-файла — с реальными координатами, указанными на сайте, с координатами, полученными в результате геокодирования, с линиями, показывающими расстояния между парами координат. Это необходимо для того, чтобы в дальнейшем понять, какие ошибки произошли в процессе геокодирования.

3.6 Визуализация и анализ результатов

Для каждого из сайтов я визуализировала полученные данные с помощью веб-приложения HERE Studio. Таким образом я смогла наглядно увидеть все полученные точки и понять, нет ли ошибок в работе.

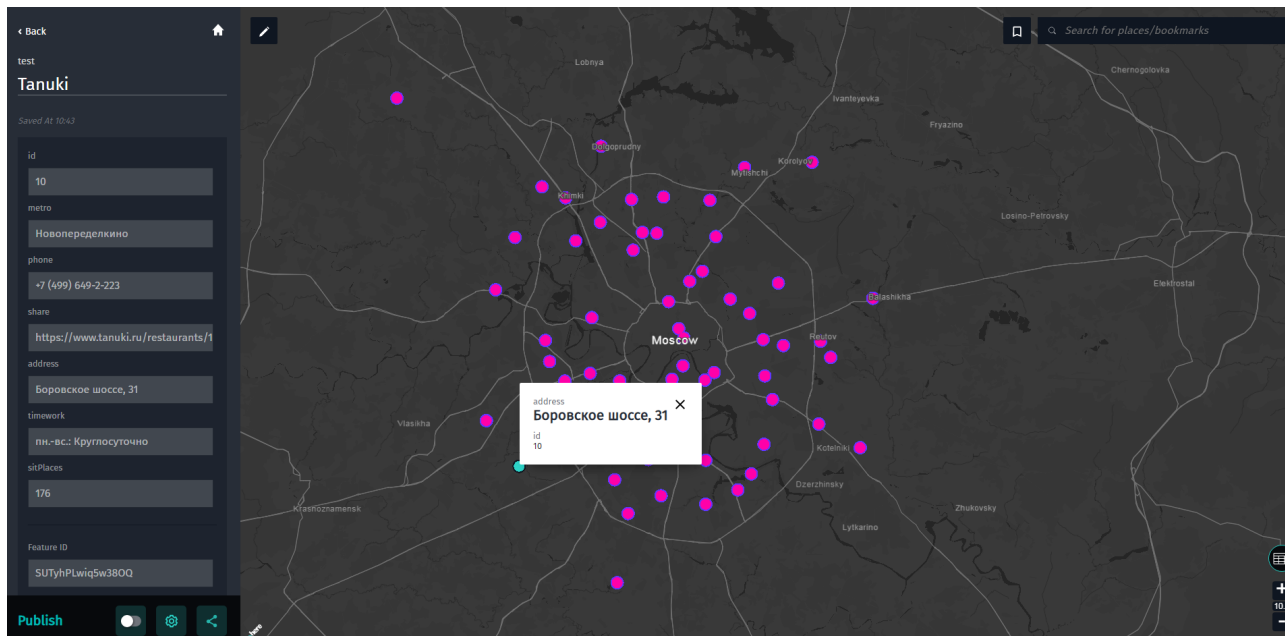
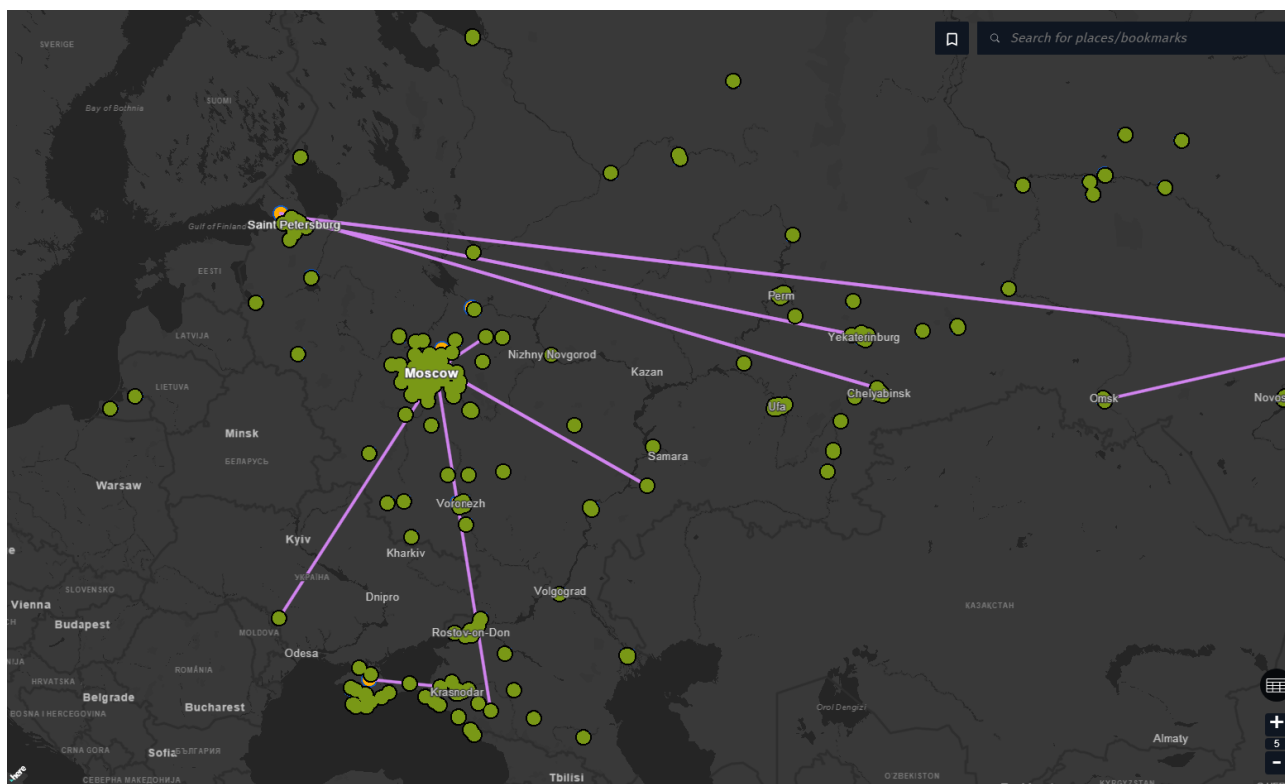


Рис. 1: Результаты для сайта tanuki.ru

Результаты геокодирования для сайта sushiwok.ru я также провизуализировала в HERE Studio. Для каждого объекта в результате визуализации отображается точка с реальными координатами объекта, точка с координатами, полученными в результате геокодирования, а также линия, показывающая расстояние между двумя этими точками.



Результаты геокодирования для сайта sushiwok.ru

3.7 Создание итоговой презентации

В конце практики был проведен еще один вебинар по навыкам презентации. На вебинаре рассказывалось, как правильно готовить и оформлять презентацию своего проекта, а также, как готовиться к выступлению, как выступать. Далее мной была подготовлена и представлена презентация, показывающая итоги выполненной работы.

4 Дневник производственной практики

Дата	Выполняемая работа
11.01.21	Вводный вебинар, ознакомление с планом работы.
12.01.21 — 15.01.21	Вебинары. Самостоятельное изучение дополнительных материалов.
15.01.21 — 22.01.21	Работа над извлечением данных с сайтов, приведением их к необходимому формату и визуализацией.
23.01.21 — 24.01.21	Геокодирование данных для сайта sushiwok.ru .
25.01.21	Вебинар по навыкам презентации.
25.01.21 — 27.01.21	Визуализация данных для сайта sushiwok.ru .
27.01.21 — 28.01.21	Подготовка презентации.
29.01.21	Презентация результатов работы.

5 Заключение

Производственная практика длилась в течение трех недель в соответствии с учебным планом подготовки студентов по направлению «Программная инженерия». Из-за коронавирусных ограничений практика проходила дистанционно.

В ходе прохождения практики я ознакомилась с компанией, задачами, которые она выполняет. Также я изучила много нового теоретического материала и выполнила практические задания. Практика позволила мне поработать в новой для меня сфере, научиться чему-то новому.

Список литературы

1. Статья "Как HERE Technologies задала стандарты геолокации" // hightech.fm URL: hightech.fm/2019/04/30/here-technologies
2. HERE Technologies — официальный сайт // www.here.com/ URL: www.here.com/
3. Here Technologies // [ru.wikipedia.org](https://ru.wikipedia.org/wiki/Here) URL: ru.wikipedia.org/wiki/Here
4. Компания HERE Technologies и студенты // [carrymap.com](https://carrymap.com/ru/cmblog/carrymap/here-technologies-and-students-from-leading-universities-use-carrymap-for-maps-update/) URL: carrymap.com/ru/cmblog/carrymap/here-technologies-and-students-from-leading-universities-use-carrymap-for-maps-update/
5. Документация по BeautifulSoup // [wiki.python.ru](https://wiki.python.ru/Документация/BeautifulSoup) URL: wiki.python.ru/Документация/BeautifulSoup
6. Библиотека Pandas // [ru.wikipedia.org](https://ru.wikipedia.org/wiki/Pandas) URL: ru.wikipedia.org/wiki/Pandas
7. Библиотека Requests // [python.ru](https://python.ru/post/97/) URL: python.ru/post/97/
8. Парсинг Json в Python 3 // [all-python.ru](https://all-python.ru/osnovy/json.html) URL: all-python.ru/osnovy/json.html
9. HERE Studio // [developer.here.com](https://developer.here.com/products/platform/studio) URL: developer.here.com/products/platform/studio