

Создание парсера архива структурированных почасовых климатических параметров засухи NOAA Global Hourly

Вихляев Дмитрий Романович

Приамурский государственный университет имени Шолом-Алейхема

Студент

Аннотация

В данной статье рассматривается создание парсера для сбора данных метеостанций с интернет архива National Oceanic and Atmospheric Administration. В исследовании описаны методы сбора, обработки, и хранения метеорологических данных. Приведены программные реализации и структура базы данных.

Ключевые слова: Python, База данных, парсер, метеоданные.

Creating a parser for the archive of structured hourly climatic parameters of the NOAA Global Hourly drought

Vikhlyayev Dmitry Romanovich

Sholom-Aleichem Priamursky State University

Student

Abstract

This article discusses the creation of a parser for collecting weather station data from the National Oceanic and Atmospheric Administration Internet archive. The study describes methods for collecting, processing, and storing meteorological data. The software implementations and the database structure are given.

Keywords: Python, Database, parser, weather data.

1 Введение

1.1 Актуальность

Создания системы сбора данных метеостанций обусловлена необходимостью в оперативной, точной и структурированной информации о метеорологических показателях. Изменения климата, частота экстремальных погодных явлений и их влияние на экономику, и общественную безопасность требуют более глубокого анализа и прогнозирования, основанного на больших объемах данных. Использование данных от NOAA позволяет охватить широкие географические и временные масштабы, что особенно важно для долгосрочных климатических исследований и прикладных задач. Особенности данного исследования являются интеграция с онлайн-ресурсом NOAA для автоматизированного сбора данных, реализация методов их обработки и эффективного хранения, что обеспечивает непрерывный доступ к актуальной информации. Эта система может быть полезна не только

для климатологов, но и для организаций, которым необходимы точные прогнозы и исторические данные для анализа и планирования: это могут быть предприятия в сферах сельского хозяйства, транспорта, энергетики, а также учреждения, занимающиеся чрезвычайными ситуациями.

1.2 Обзор исследований

А.И.Честнов, И.А.Сикарев, В.М.Абрамов создали геоинформационную систему по визуализации данных сервиса noaa climate data online на территории России [1]. А.С.Халиманенков, В.А.Глаголев документировали особенности расшифровки показателей метеорологических данных сети NOAA [2]. И.Ю.Савин, А.Вриелинг провели анализ многолетней динамики растительных ресурсов на территории России по данным NOAA AVHRR [3]. Е.В.Варламова, В.С.Соловьев провели исследование вариаций вегетационного индекса NDVI арктической зоны Якутии [4].

1.1 Цель исследования

Цель исследования – описать структурированность действий и методы системы для сбора, обработки, и хранения метеоданных предоставляющие ресурсом NOAA.

2 Материалы и методы

Для реализации используется язык программирования Python и СУБД MySQL.

3 Результаты и обсуждения

Сервис NOAA (National Oceanic and Atmospheric Administration) предоставляет файловое хранилище собранных многолетних данных метеостанций по всему миру [5]. Данные разбиты на директории индексированные по годам. В каждой директории хранятся файлы в формате csv. Название файлов обусловлено слиянием индекса метеостанции VMO и номера NCDC WBAN.

Для сбора метеоданных написан скрипт на языке Python с методами загрузки файлов с удалённого источника и записи в структуру данных «Pandas.DataFrame» (рис.1).

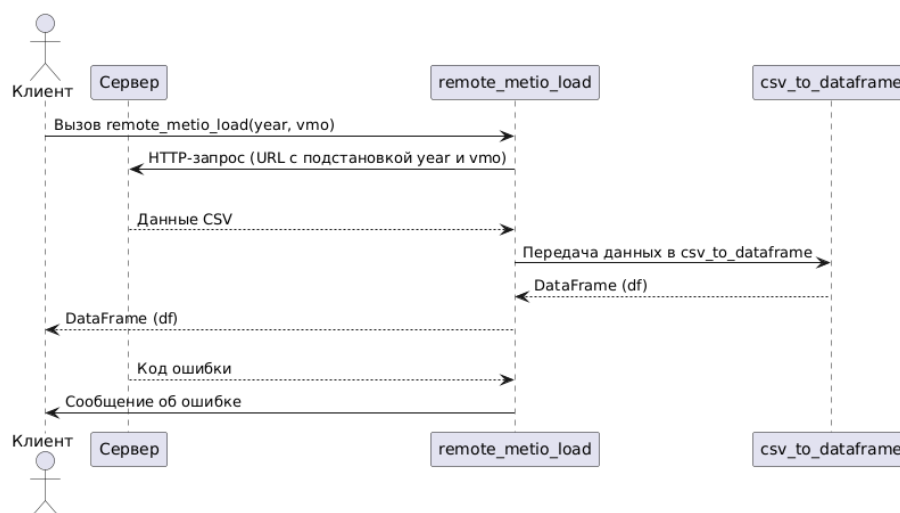


Рис. 1. Функции загрузка метеоданных с файлового хранилища NOAA

В документе csv содержатся информационные данные соответственной метеостанции и метеорологические показатели, разделённые по времени, рисунок 2. Метеостанции делают записи каждые 3 часа. NOAA оставляет дату и время по UTC. Помимо этого информация по показателям закодирована. Для перевода показателя в соответствующий вид нужно воспользоваться документацией [7].

STATION	DATE	SOUF	LATITUDE	LONGITU	ELEVATI	NAME	REPORT_	CALL_SIC	QUALIT	WND	CIG	VIS	TMP	DEW	SLP
31713099999	2003-01-01T00:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	99999.9.N	010000.1.N.1	-0292.1	-0315.1	10298.1
31713099999	2003-01-01T03:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	360.1.N.0020.1	99999.9.N	020000.1.N.1	-0234.1	-0266.1	10290.1
31713099999	2003-01-01T06:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	99999.9.N	020000.1.N.1	-0212.1	-0252.1	10288.1
31713099999	2003-01-01T12:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	99999.9.N	999999.9.N.1	-0278.1	-0303.1	10310.1
31713099999	2003-01-01T15:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0296.1	-0319.1	10316.1
31713099999	2003-01-01T18:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0315.1	-0338.1	10329.1
31713099999	2003-01-01T21:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0329.1	-0351.1	10333.1
31713099999	2003-01-02T00:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	050000.1.N.1	-0341.1	-0358.1	10348.1
31713099999	2003-01-02T03:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	020000.1.N.1	-0231.1	-0255.1	10345.1
31713099999	2003-01-02T06:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	020000.1.N.1	-0206.1	-0246.1	10337.1
31713099999	2003-01-02T09:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0253.1	-0284.1	10327.1
31713099999	2003-01-02T12:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0276.1	-0306.1	10328.1
31713099999	2003-01-02T15:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0298.1	-0321.1	10323.1
31713099999	2003-01-02T18:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	22000.1.9.N	999999.9.N.1	-0309.1	-0332.1	10309.1
31713099999	2003-01-03T00:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	99999.9.N	004000.1.N.1	-0262.1	-0287.1	10295.1
31713099999	2003-01-03T03:00:00...	4	48.7333333	132.95	80.0	BIROBIDZHAN...	FM-12	99999	V020	999.9.C.0000.1	99999.9.N	004000.1.N.1	-0246.1	-0270.1	10284.1

Рис. 2. Файл метеостанции Биробиджан с данными на 2003 год

NOAA также предоставляет список индексируемых метеостанций с дополнительной информацией [6]. Документ с подробным описанием имеет формат txt, но в этой же директории хранится одноимённый файл в формате csv, с которого удобно получить список необходимых станций по стране или географическому расположению (рис.3). Чтобы получить все метеостанции на территории Российской Федерации, нужно получить записи, где поле «CTRY» имеет значение «RS». Описываемая система загружает данные csv в структуру данных «Pandas.DataFrame» и собирает нужные значения.

USAF	WBAN	STATIONNAME	CTRY	STATE	ICAO	LAT	LON	ELEV	BEGIN	END
134910	99999	SKOPJE	RS			+42.517	+022.283	+1176.0	19940107	19960215
200260	99999	VIKTORIYA ISLAND	RS			+80.167	+036.750	+0009.0	19591102	20020320
200340	99999	NAGURSKOYE	RS			+80.817	+047.417	+0018.0	19550305	20030314
200460	99999	POLARGMO IM. E. T. KRENKELJA	RS			+80.617	+058.050	+0022.0	19570703	20241030
200470	99999	BUKHTA YURIYA	RS			+80.317	+052.800	+0006.0	19550305	19591215
200490	99999	TIKHAYA BAY	RS			+80.367	+052.917	+0046.0	19320202	19971015
200660	99999	USHAKOVA ISLAND	RS			+80.833	+079.700	+0047.0	19550305	19971215
200690	99999	OSTROV VIZE	RS			+79.483	+076.983	+0010.0	19550305	20241030
200850	99999	OSTROV PIONEER	RS			+79.883	+091.233	+0030.0	19591101	19591225
200870	99999	MGMS IM. G.F. USHAKOVA	RS			+79.550	+090.567	+0008.0	19320101	20241030
200970	99999	CAPE PESCHANIY	RS			+79.383	+102.417	+0006.0	19590901	20040925
201460	99999	ZALIV GULDA	RS			+79.917	+058.917	+0099.0	19620201	19621230
201860	99999	KRASNOFLOTSKIYE ISL	RS			+78.617	+098.700	+0008.0	19550305	19871130
201990	99999	MALYY TAYMIR ISLAND	RS			+78.133	+107.200	+0012.0	19550305	19940815
202740	99999	OSTROV UEDINENIYA	RS			+77.500	+082.200	+0023.0	19350202	19961121
202770	99999	ISACHENKO ISLAND	RS			+77.283	+089.667	+0005.0	19550305	19970330
202890	99999	RUSSKIY ISLAND	RS			+77.183	+096.583	+0009.0	19550305	19960425
202910	99999	SOLNECHNAYA BAY	RS			+78.233	+103.117	+0030.0	19550305	19971025

Рис. 3. Данные отправляемые через POST запрос

Для хранения данных выбрана база данных Mysql. В первую очередь нужны 2 таблицы для содержания метеостанций и метеорологических показателей. На территории России удобно разделить расположение метеостанций на регионы. Это позволит правильно расставлять часовой пояс, который как описывалось ранее, содержится в UTC. На рисунке 4,5 представлены структура и sql схема базы данных, соответственно, состоящая из 3 таблиц, которая используется в системе. В качестве показателей выбраны температура воздуха, точка росы и количество осадков.

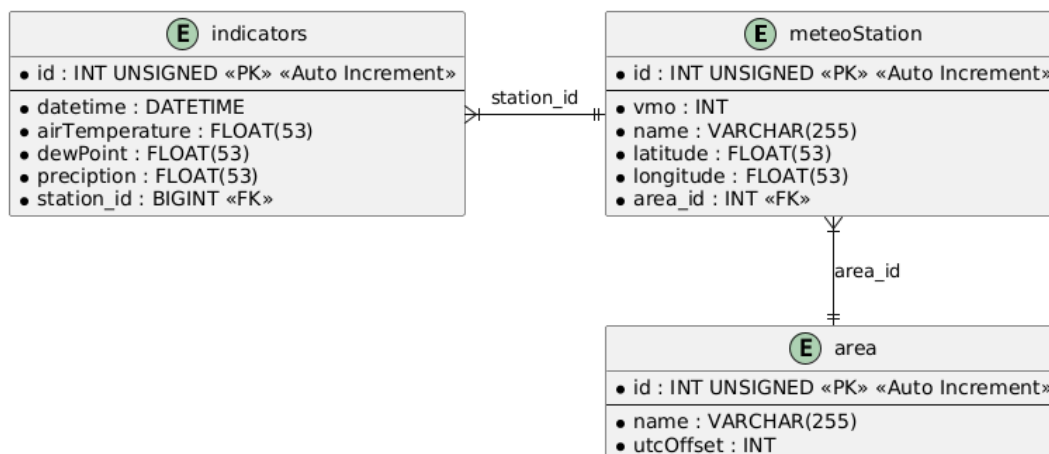


Рис. 4. Структура базы данных

```

CREATE TABLE `meteoStation` (
  `id` INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
  `vmo` INT NOT NULL,
  `name` VARCHAR(255) NOT NULL,
  `latitude` FLOAT(53) NOT NULL,
  `longitude` FLOAT(53) NOT NULL,
  `area_id` INT NOT NULL
);
CREATE TABLE `indicators` (
  `id` INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
  `datetime` DATETIME NOT NULL,
  `airTemperature` FLOAT(53) NOT NULL,
  `dewPoint` FLOAT(53) NOT NULL,
  `precipitation` FLOAT(53) NOT NULL,
  `station_id` BIGINT NOT NULL
);
CREATE TABLE `area` (
  `id` INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
  `name` VARCHAR(255) NOT NULL,
  `utcOffset` INT NOT NULL
);
ALTER TABLE
`meteoStation` ADD CONSTRAINT `meteoStation_area_id_foreign` FOREIGN KEY(`area_id`) REFERENCES `area`(`id`);
ALTER TABLE
`indicators` ADD CONSTRAINT `indicators_station_id_foreign` FOREIGN KEY(`station_id`) REFERENCES `meteoStation`(`id`);

```

Рис. 5. SQL схема таблиц

Прежде чем заполнять базу данных метеорологическими показателями, необходимо провести обработку данных. Нужно учитывать, что среди записей могут быть пропуски. Этот момент можно отследить, так как записи должны иметь разницу в 3 часа. Помимо целой записи, пропуск может присутствовать для конкретного показателя. Каждый показатель закодирован и хранит как минимум 2 значения разделённые запятой сами данные и код характеризующий значение показателя. На рисунке 6 представлена блок-схема функции для обработки 3 показателей. Обработка поля даты и времени не представлено, так как оно не закодировано, для него меняется только часовой пояс.

Показатели температуры воздуха и точки росы состоят из двух значений. Первое значение – это температура в градусах Цельсия умноженное в 10 раз. При расшифровке получится число с плавающей точкой с одной цифрой после запятой. Второе значение – код качества, подробнее о кодах и о том как кодируются показатели можно узнать из документации, о которой ранее уже упоминалось. В качестве кодов, которые были выбраны неприемлемыми к хранению, были выбраны 3, 7 и 9, которые означают ошибочные и пропущенные значения. Уже упоминалось о пропущенных значениях показателя, но в данном случае значение записано, но указано как пропущенное и скорее всего, будет иметь вид «9999,9».

Показатель количество осадков имеет 4 значения. Первый содержит количество осадков за период времени в миллиметрах, умноженный в 10 раз, также как с температурой воздуха. Однако осадки не замеряются каждые 3 часа, значения могут быть записаны в любом из 8 промежутков времени за день. К тому же сами значения при отсутствии осадков указаны как пропущенные, что может означать как отсутствие данных, так и отсутствие самих осадков на территории станции. Другими словами при отсутствии осадков данные не записываются, поэтому сложно отсудить действительно ли не было осадков или был допущен промах при фиксации. Четвёртое значения код качества уже был описан для температуры воздуха. Второе и третье значения не используются в данной системе.

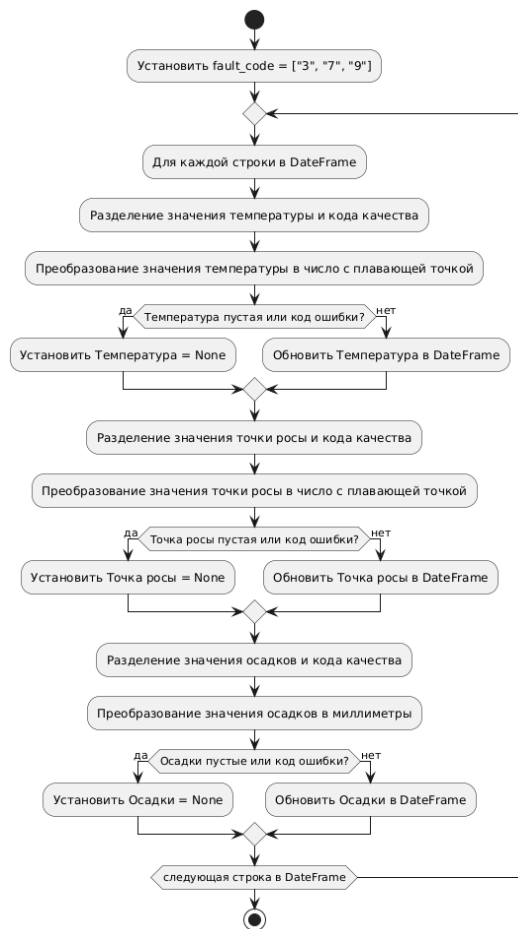


Рис. 6. Блок схема метода обработки значений показателей

Для примера была заполнена таблица в СУБД Mysql, данными метеостанции Биробиджан. И приведён пример запроса на вывод 10 самых высоких температур за весь период наблюдения (рис.7).

`SELECT `datetime`, `airtemperature` FROM `climat` ORDER BY `airtemperature` DESC LIMIT 10`

☐ Профилирование [\[Построчное ред.\]](#)

+ Параметры

	datetime	airtemperature
☐	2010-06-26 16:00:00	36.9
☐	1999-07-14 16:00:00	36.7
☐	2010-06-27 16:00:00	36.6
☐	1977-07-29 13:00:00	36
☐	2017-07-06 16:00:00	35.6
☐	1982-07-08 16:00:00	35.4
☐	2010-06-25 16:00:00	35.2
☐	2008-06-17 16:00:00	35.2
☐	2017-07-05 16:00:00	35
☐	2017-07-06 13:00:00	35

☐ Отметить все С отмеченными:

Рис. 7. Результат sql запроса на выбор 10 самых высоких температур

4 Выводы

В результате исследования были рассмотрены программные методы сбора метеоданных с сервиса NOAA. Описан способ раскодирования данных показателей. Представлен пример структуры базы данных для хранения информации.

Библиографический список

1. Честнов А.И., Сикарев И.А., Абрамов В.М. Создание геоинформационной системы по визуализации данных сервиса noaa climate data online на территории России // Информационные технологии и системы: управление, экономика, транспорт, право. 2023. Т. 1. № S2. С. 151-156.
2. Халиманенков А.С., Глаголев В.А. Формирование и хранение метеорологических данных сети NOAA в цифровом виде // Постулат. 2022. № 1 (75).
3. Савин И.Ю., Вриелинг А. Анализ многолетней динамики растительных ресурсов на территории России по данным NOAA AVHRR // Исследование Земли из космоса. 2008. № 5. С. 74-82.
4. Варламова Е.В., Соловьев В.С. Исследование вариаций вегетационного индекса NDVI арктической зоны Якутии // Современные проблемы дистанционного зондирования Земли из космоса. 2010. Т. 7. № 3. С. 226-230.
5. Файловое хранилище метеорологических данных NOAA URL: <https://www.ncei.noaa.gov/data/global-hourly> (дата обращения: 25.10.2024).
6. Список метеостанций предоставляющих NOAA и географическое расположение URL: <https://www.ncei.noaa.gov/pub/data/noaa/isd-history.txt> (дата обращения: 25.10.2024).
7. Документация данных для получения интегрированных данных метеорологических показателей предоставляемых NOAA URL: <https://www.ncei.noaa.gov/pub/data/noaa/ISH-DVD2012/ish-format-document.pdf> (дата обращения: 25.06.2025).