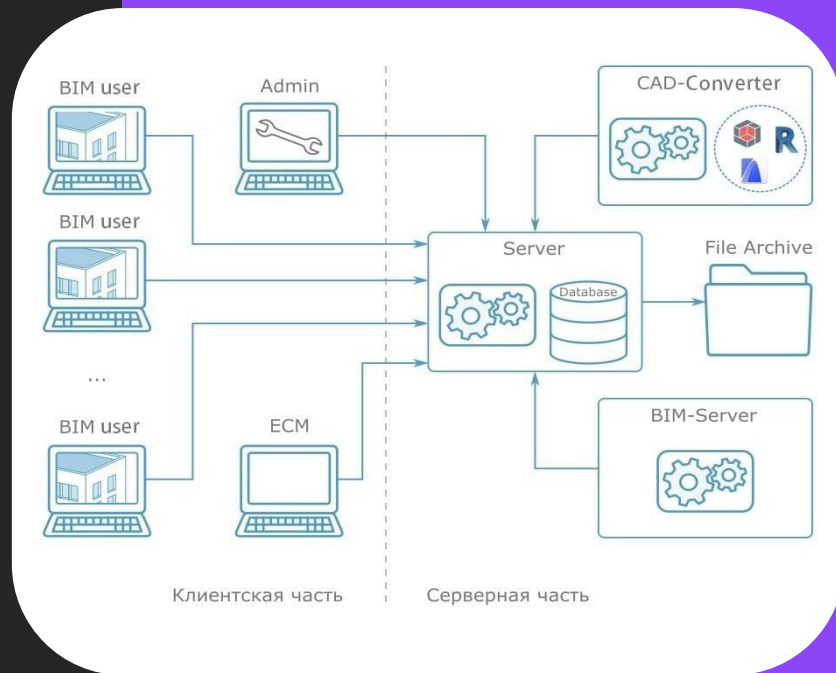


# Формирование базы данных

Сбор, фильтрация данных) и синхронизация баз





## Марина Федорова

Аналитик, год поступления в GeekBrains 2021

Немного о себе. Краткое описание  
в несколько строк:

- Живу в Санкт-Петербурге
- Люблю цифры , Клайва Льюиса, фэнтези и учиться.
- Кандидат экономических наук. Не боюсь нового!
- 7 лет трудилась художником-иконописцем, 3 года обучала детей английскому. Вырастила своих троих детей.



# Поставленная задача

## Описание задачи

Для приложения, которое позволяет отсканировать штрихкод продукта и получить его состав необходимо расширить базу данных.

## Цель проекта

Сделать общую базу, которую можно будет использовать для Приложения из двух доступных баз - из каталога продуктов магазина “Призма” и открытой базы данных OpenFoodFacts

## План работы:

Изучить структуру базы <https://world.openfoodfacts.org/data>

Развернуть базу локально, чтобы иметь возможность с ней работать.

Преобразовать и отфильтровать данные из базы.

Влить данные из базы в существующую на данный момент базу.

## Проблема, которую будет решать проект

Увеличение количества распознаваемых приложением продуктов.



## Решение задачи / План работы

1. Развернула базу на сервере
  2. Вид базы после запуска клиента MongoDB Compass
- ### запуска клиента MongoDB Compass

```
C:\WINDOWS\system32\cmd.exe - mongorestore.exe -d opff -c opff D:\mongodb\bin\dump\opff\products.bson
```

```
Microsoft Windows [Version 10.0.18363.1556]  
(c) Корпорация Майкрософт (Microsoft Corporation), 2019. Все права защищены.
```

```
C:\Users\Сепрей>D:
```

```
D:\>cd D:\Program Files\MongoDB\Tools\100\bin
```

```
D:\Program Files\MongoDB\Tools\100\bin>dir
```

```
Том в устройстве D не имеет метки.  
Серийный номер тома: 7523-8BF0
```

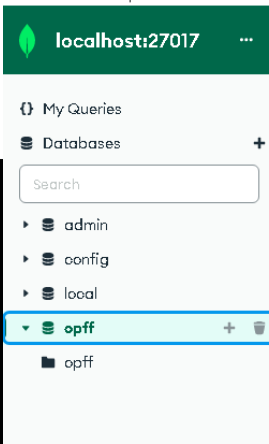
```
Содержимое папки D:\Program Files\MongoDB\Tools\100\bin
```

```
26.12.2022  19:14  <DIR>      .  
26.12.2022  19:14  <DIR>      ..  
03.11.2022  15:25      20 067 687 bsondump.exe  
03.11.2022  15:26      114 688 1lbsasl.dll  
03.11.2022  15:25      23 792 184 mongodump.exe  
03.11.2022  15:25      23 364 765 mongoexport.exe  
03.11.2022  15:25      24 645 984 mongofiles.exe  
03.11.2022  15:25      23 718 892 mongoimport.exe  
03.11.2022  15:25      24 304 747 mongorestore.exe  
03.11.2022  15:25      22 986 864 mongostat.exe  
03.11.2022  15:25      22 408 684 mongotop.exe  
9 файлов    185 388 495 байт  
2 папок     50 296 254 464 байт свободно
```

```
D:\Program Files\MongoDB\Tools\100\bin>mongorestore.exe -d opff -c opff D:\mongodb\bin\dump\opff\products.bson  
2022-12-26T19:20:24.385+0300 checking for collection data in D:\mongodb\bin\dump\opff\products.bson  
2022-12-26T19:20:24.683+0300 reading metadata for opff.opff from D:\mongodb\bin\dump\opff\products.metadata.json  
2022-12-26T19:20:24.688+0300 restoring to existing collection opff.opff without dropping  
2022-12-26T19:20:24.688+0300 restoring opff.opff from D:\mongodb\bin\dump\opff\products.bson  
2022-12-26T19:20:27.399+0300 [.....] opff.opff 92.5MB/38.7GB (0.2%)  
2022-12-26T19:20:30.384+0300 [.....] opff.opff 213MB/38.7GB (0.5%)  
2022-12-26T19:20:33.384+0300 [.....] opff.opff 273MB/38.7GB (0.7%)  
2022-12-26T19:20:36.384+0300 [.....] opff.opff 411MB/38.7GB (1.0%)
```

MongoDB Compass - localhost:27017/opff

Connect View Help



Collections

Create collection

View

Uncompressed data size: 41.60 GB

Storage size:  
14.83 GB

Documents:  
2.7 M

Avg. document  
15.27 kB



## □ Приведение данных к единому формату

### Подготовка данных

Создаем коллекцию с нужным нам фильтром – с наличием ингредиентов на русском языке

The screenshot shows the MongoDB Compass application window titled 'MongoDB Compass - localhost:27017/opff.opff'. The interface includes a sidebar on the left with 'My Queries' and 'Databases' sections. Under 'Databases', the 'opff' database is selected, showing a list of collections: 'admin', 'config', 'local', and 'opff'. The main panel displays the 'opff.opff' collection details. At the top right, it shows '2.7m' documents and '64' indexes. Below this, there are tabs for 'Documents', 'Aggregations', 'Schema', 'Explain Plan', 'Indexes', and 'Validation'. The 'Documents' tab is active, showing a filter query: `{ ingredients_text_ru: { $exists: true } }`. Below the filter, the 'Project' field is set to `{ field: 0 }`. The 'Sort' field is set to `{ field: -1 } or [[ 'field', -1 ]]`. The 'Collection' field is set to `{ locale: 'simple' }`. On the right side of the filter section, there are buttons for 'Reset', 'Find', and 'Less Options'. Below the filter section, there are buttons for 'ADD DATA' and 'EXPORT COLLECTION'. At the bottom of the window, there is a status bar showing '1 of N/A' and a 'Fetching Documents' progress indicator with a 'Stop' button.



# SQL база, с которой уже можно работать

MySQL Workbench

Local instance MySQL80

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

Filter objects

products

Tables

Views

Stored Procedures

Functions

sys

Administration Schemas

Information

Table: products

Columns:

id int UN AI PK bigint UN text UN components text double fat double

Output

Action Output

Time Action

Введите здесь текст для поиска

products-- MySQL dump 10.13 Distrib 8.0.29, for Linux (x86\_64)

```
--
-- Host: localhost    Database: simple_product
--
-- Server version: 8.0.29-0ubuntu0.20.04.3

/*!40101 SET @OLD_CHARACTER_SET_CLIENT=@@CHARACTER_SET_CLIENT */;
/*!40101 SET @OLD_CHARACTER_SET_RESULTS=@@CHARACTER_SET_RESULTS */;
/*!40101 SET @OLD_COLLATION_CONNECTION=@@COLLATION_CONNECTION */;
/*!50503 SET NAMES utf8mb4 */;
/*!40103 SET @OLD_TIME_ZONE=@@TIME_ZONE */;
/*!40103 SET TIME_ZONE='+00:00' */;
/*!40014 SET @OLD_UNIQUE_CHECKS=@@UNIQUE_CHECKS, UNIQUE_CHECKS=0 */;
/*!40014 SET @OLD_FOREIGN_KEY_CHECKS=@@FOREIGN_KEY_CHECKS, FOREIGN_KEY_CHECKS=0 */;
/*!40101 SET @OLD_SQL_MODE=@@SQL_MODE, SQL_MODE='NO_AUTO_VALUE_ON_ZERO' */;
/*!40111 SET @OLD_SQL_NOTES=@@SQL_NOTES, SQL_NOTES=0 */;

--
-- Table structure for table 'products'
--
DROP TABLE IF EXISTS `products`;
/*!40101 SET @saved_cs_client = @@character_set_client */;

```

Automatic context help is disabled. Use the toolbar to manually get help for the current caret position or to toggle automatic help.

SCHEMAS

Filter objects

products

Tables

Views

Stored Procedures

Functions

sys

Administration Schemas

Information

Table: products

Columns:

id int UN AI PK bigint UN text UN components text double fat double

1 SELECT \* FROM products.products;

Result Grid

Filter Rows:

Edit Export/Import Wrap Cell Content Fetch rows

id	barcode	name	components	energy	proteins	fat	carbo
1	4607191650210	Чёрная скородина замороженная, Hortex, 30...	NULL	NULL	NULL	NULL	NULL
2	7340191100732	Шоколад тёмный, X-tra, 100 г, Франция	какао тёртое, сахар, эмульгаторы (E322): ле...	509	6.9	28	53
3	4605246005817	Чай чёрный Golden Ceylon, Greenfield, 100 па...	NULL	NULL	NULL	NULL	NULL
4	4607191650319	Клюква замороженная, Hortex, 250 г	NULL	NULL	NULL	NULL	NULL
5	6415712600297	Сёмгла целая в маринаде, Kotimaista, 700 г, ...	сёмгла, вода, уксус, сахар, соль, ароматиза...	50	1.1	0.2	11
6	4601132998478	Рисет лососевый с каперсами и укропом, Мер...	NULL	NULL	NULL	NULL	NULL
7	2000345100005	Плюс с курой, 100 г	NULL	NULL	NULL	NULL	NULL
8	4600605024065	Кисло-молочный продукт со вкусом ягодного...	молоко обезжиренное, вода, сахар, сливки, ...	80	2.6	2.5	11.7
9	4101540758744	Кекс марципановый рождественский, Kuche...	NULL	NULL	NULL	NULL	NULL
10	3083681126686	Овощная смесь Фасульве, Bonduelle, 400 г	NULL	NULL	NULL	NULL	NULL
11	6420101443447	Кофе молотый светлой обжарки, Saludo, 450...	Натуральный кофе первой степени обжарки...	NULL	NULL	NULL	NULL
12	4640001731181	Сыр полутвёрдый Лёгкий, фасованный, 17...	NULL	NULL	NULL	NULL	NULL
13	4607024591222	Хлебцы хрустящие Лёпёшка с кунжутом и сы...	NULL	NULL	NULL	NULL	NULL
14	4605246002427	Кофе в зёрнах Классический, Жюкей, 500 г	NULL	NULL	NULL	NULL	NULL
15	37466019871	Шоколад молочный Swiss Classic с цельным ф...	NULL	NULL	NULL	NULL	NULL
16	4607030636665	Мороженое фисташковый пломбир в вафель...	Пломбир: молоко коровье цельное, молоко ц...	249	4.2	10.5	34.5
17	6415712133221	Джем малиновый терпостойкий, Rainbow, 30...	малина, сахар, вода, крахмал кукурузный м...	189	0.3	0.2	46

products 1

Output

Apply Revert



## ❑ Анализ базы MongoDB

Открываем Mongo. Нам нужно выбрать объекты, которые соответствуют критерию:

- наличие поля "ingredients\_text\_ru"

`{"ingredients_text_ru" : {$exists:true}}`

Выбрать штрихкоды, начинающиеся не с 2

. Используемое регулярное выражение:

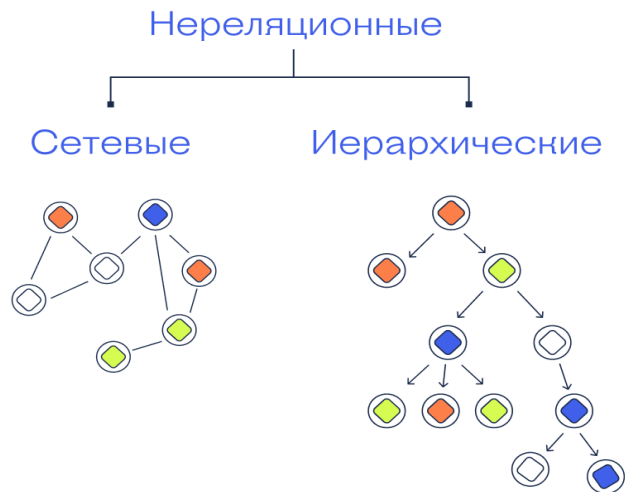
Нам нужно выбрать объекты, которые соответствуют критерию:

- наличие поля "ingredients\_text\_ru"

`{"ingredients_text_ru" : {$exists:true}}`

коды, начинающиеся не с 200

```
{
  $and : [
    "ingredients_text_ru" : {$exists:true},
    "code" : {$exists:true},
    "code" : {$not : {$regex : "^200"}}
  ]
}
```





- ❑ **Выбор технологий создания скрипта для выбора и записи данных из MomgoDB в SQL**

### Создание виртуального окружения

- ✓ Создаётся через модуль venv, который идёт в поставке Python 3.
- ✓ Используется команда "python -m venv" и название директории, в которой будет создано виртуальное окружение.  
`mkdir projectcd projectpython -m venv env`



У нас запущен демон MongoDB, пытаемся подключиться используя MongoClient

```
D:\Razrabitchik\Diploma\OFFparser>python -m venv env

D:\Razrabitchik\Diploma\OFFparser>dir
Том в устройстве D не имеет метки.
Серийный номер тома: 7523-8BF0

Содержимое папки D:\Razrabitchik\Diploma\OFFparser

05.02.2023  17:11    <DIR>          .
05.02.2023  17:11    <DIR>          ..
05.02.2023  17:11    <DIR>          env
                0 файлов             0 байт
                3 папок   52 985 532 416 байт свободно

D:\Razrabitchik\Diploma\OFFparser>
```

```
main.py - D:/Razrabitchik/Diploma/OFFparser/main.py (3.9.7)
File Edit Format Run Options Window Help

from pymongo import MongoClient
from pprint import pprint
client = MongoClient('localhost', 27017)
db = client.opff
collection = db.opff
pprint(collection.find_one())
```





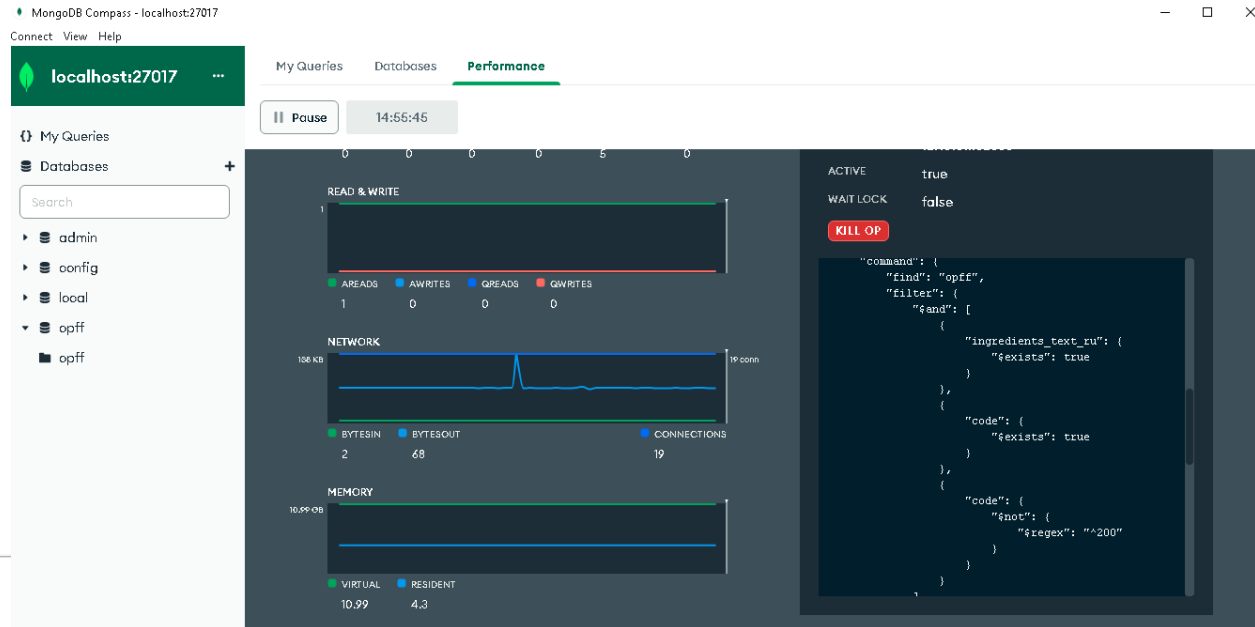
## ❑ Взаимодействие с базами данных

Выгрузка данных из базы  
OpenFoodFacts при  
помощи клиента MongoDB  
Compass

открыть - файл взаимодействия с сервером MySQL (с. 2-7)  
File Edit Format Run Options Window Help

```
from mysql import connector
from mysql.connector import errorcode

try:
    cnx = connector.connect(user='root', password='[REDACTED]',
                           host='127.0.0.1',
                           database='products')
except connector.Error as err:
    if err.errno == errorcode.ER_ACCESS_DENIED_ERROR:
        print("Something is wrong with your user name or password")
    elif err.errno == errorcode.ER_BAD_DB_ERROR:
        print("Database does not exist")
    else:
        print(err)
else:
    cnx.close()
```



Мы установили соединение между нашей  
программой и сервером MySQL.  
Обрабатываем исключения



## Достигнутые цели / Итоги в формате ДТП

В процессе работы над Дипломным проектом мне удалось:

- ✓ долить необходимые данные из базы MongoDB в существующую SQL базу
- ✓ В процессе работы я познакомилась с работой по SQL базы, развернула 2 базы на сервере, написала скрипт на python, который корректно отработал и дописал в базу необходимые мне значения



Самая большая трудность была в том, чтобы научиться правильно задавать вопрос. Оказалось, что все ответы есть, нужно только научиться спрашивать!