

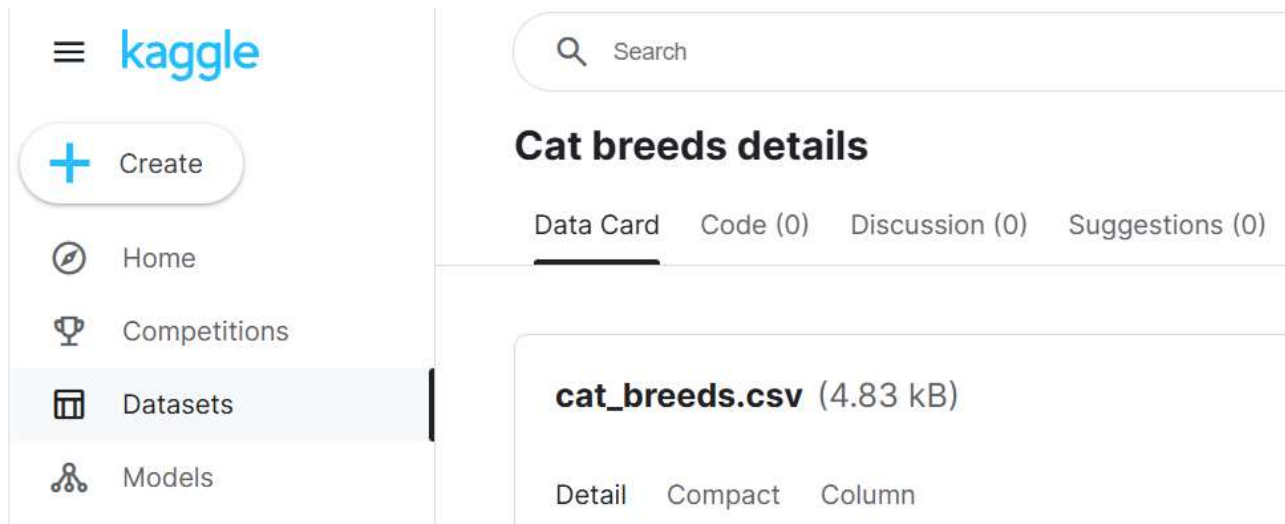
# ІТМО

## База данных хранения пород кошек



Федоров И.А., R4240, ISU 370113

# Данные 1



В качестве начальных данных выбран датасет:

<https://www.kaggle.com/datasets/warcoder/cat-breeds-details>

# Данные 2

	name	length	min_life_expectancy	max_life_expectancy	min_weight	max_weight	family_friendly	shedding	general_health	playfulness	children_friendly	groom
0	Abyssinian	12 to 16 inches	9.0	15.0	6.0	10.0	3	3	2	5	5	
1	Aegean	Medium	9.0	10.0	7.0	10.0	5	3	4	4	5	
2	American Bobtail	Medium	11.0	15.0	8.0	13.0	4	4	4	4	4	
3	American Shorthair	12 to 15 inches	15.0	20.0	7.0	12.0	3	3	4	2	4	
4	American Wirehair	Medium to large	14.0	18.0	8.0	15.0	4	3	5	3	4	
...	...	...	...	...	...	...	...	...	...	...	...	...
57	Toyger	Medium	10.0	15.0	7.0	15.0	5	1	3	4	4	
58	Turkish Angora	Medium	12.0	18.0	5.0	9.0	5	2	3	5	3	
59	Turkish Van	14 to 17 inches	12.0	17.0	10.0	18.0	3	2	4	4	3	
60	Ukrainian Levkoy	Medium	9.0	15.0	8.0	12.0	5	1	3	4	5	
61	York Chocolate	Large	13.0	15.0	10.0	18.0	4	4	3	4	4	

Внешний вид данных

# Данные 3

**DOGELL**

Search for dog breed...

### Cat Breeds A-Z List

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Showing 1-20 of 79 items.

Sort by Popularity Rank ▾

**Exotic Shorthair** Show Details Compare

Origin: United States Crossbreed

Size: Medium Hypoallergenic: No ✖

Lap cat: **Yes** ✓

Lifespan: 12-15 years

**Ragdoll** Show Details Compare

Origin: United States Crossbreed

Size: **Large** Hypoallergenic: No ✖

Lap cat: **Yes** ✓

Lifespan: 12-17 years

**Size**

- ☐ Large
- ☐ Medium
- ☐ Small

**Lap Cat**

- ☐ No
- ☐ Yes

**Hypoallergenic**

- ☐ No
- ☐ Yes

**Shedding**

- ☐ High
- ☐ Minimal
- ☐ Moderate

Сайт для дополнительных данных

Был использован сайт для дополнительных данных  
<https://dogell.com/en/cat-breeds>

# Данные 4



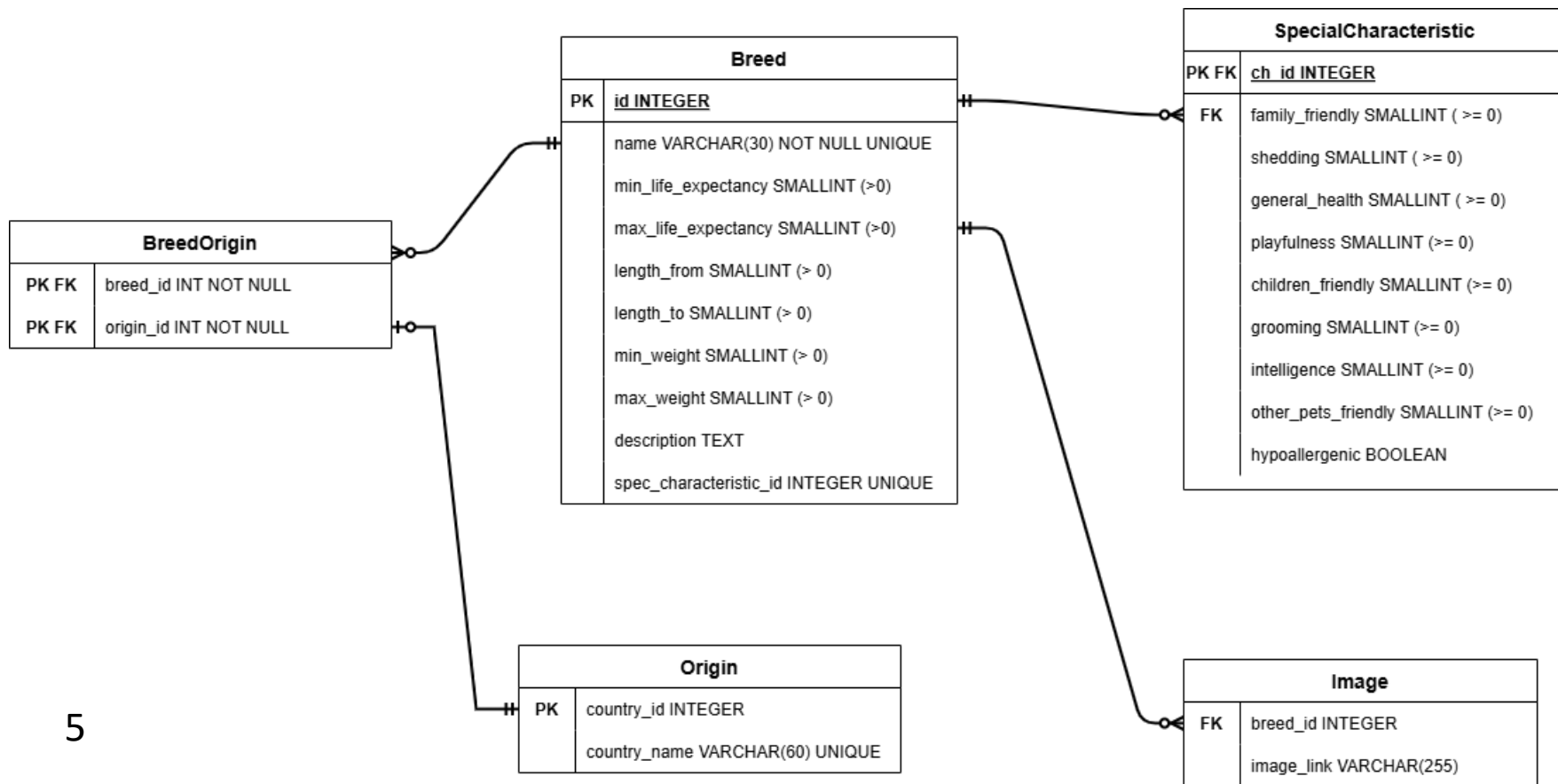
## Подготовка датасета

- ☐ Использованы pandas и numpy
- ☐ Импорт данных, удаления дубликатов, сохранение данных в csv-файлы
- ☐ Нормализация данных для дальнейшего сохранения (столбец длины, тип значений и т.д.)
- ☐ Объединение датасета и данных парсинга
- ☐ Разбиение на разные dataframe-ы для удобного импорта в БД

## Парсинг данных

- ☐ Использованы requests и BeautifulSoup4
- ☐ Получение страниц и запрос к серверу
- ☐ Обработка пагинации для анализа всех пород с сайта
- ☐ Парсинг ссылок на картинки, страну (origin) и свойство «гипоаллергенность»
- ☐ Формирование и сохранение dataframe в csv файл для объединения

# ER диаграмма БД (postgresql)



# Создание таблиц 1

```
CREATE TABLE Breed (  
  id INTEGER PRIMARY KEY,  
  name VARCHAR(30) UNIQUE NOT NULL,  
  min_life_expectancy SMALLINT CHECK(min_life_expectancy > 0),  
  max_life_expectancy SMALLINT CHECK(max_life_expectancy > 0),  
  length_from SMALLINT CHECK(length_from > 0),  
  length_to SMALLINT CHECK(length_to > 0),  
  min_weight SMALLINT CHECK(min_weight > 0),  
  max_weight SMALLINT CHECK(max_weight > 0),  
  description TEXT,  
  spec_characteristic_id INTEGER UNIQUE,  
  CHECK (max_weight >= min_weight),  
  CHECK (length_from <= length_to),  
  CHECK (max_life_expectancy >= min_life_expectancy)  
);
```

```
CREATE TABLE BreedOrigin (  
  breed_id INT NOT NULL,  
  origin_id INT NOT NULL,  
  PRIMARY KEY (breed_id, origin_id),  
  FOREIGN KEY (breed_id) REFERENCES Breed(id) ON DELETE CASCADE,  
  FOREIGN KEY (origin_id) REFERENCES Origin(country_id) ON DELETE CASCADE  
);
```

## Создание таблиц 2

```
CREATE TABLE SpecialCharacteristic (  
    ch_id INTEGER PRIMARY KEY,  
    family_friendly SMALLINT CHECK (family_friendly >= 0),  
    shedding SMALLINT CHECK (shedding >= 0),  
    general_health SMALLINT CHECK (general_health >= 0),  
    playfulness SMALLINT CHECK (playfulness >= 0),  
    children_friendly SMALLINT CHECK (children_friendly >= 0),  
    grooming SMALLINT CHECK (grooming >= 0),  
    intelligence SMALLINT CHECK (intelligence >= 0),  
    other_pets_friendly SMALLINT CHECK (other_pets_friendly >= 0),  
    hypoallergenic BOOLEAN,  
    FOREIGN KEY (ch_id) REFERENCES Breed (spec_characteristic_id) ON DELETE CASCADE  
);
```

```
CREATE TABLE Origin (  
    country_id SERIAL PRIMARY KEY,  
    country_name VARCHAR(60) UNIQUE  
);
```

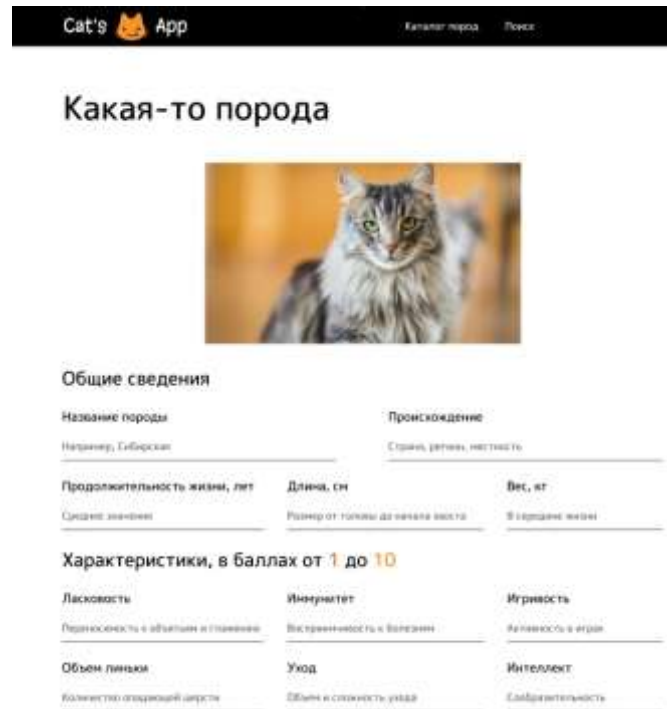
```
CREATE TABLE Image (  
    breed_id INTEGER,  
    image_link VARCHAR(255),  
    FOREIGN KEY (breed_id) REFERENCES Breed (id) ON DELETE CASCADE  
);
```



# Пример возможного интерфейса для понимания сценариев



«Сетка пород» с краткими хар-ками



Страница с конкретной породой

# Запрос кратких данных для «сетки» пород

SELECT

b.id, b.name, b.min\_life\_expectancy, b.max\_life\_expectancy, b.length\_from,  
b.length\_to, b.min\_weight, b.max\_weight, b.description, i.image\_link

FROM Breed b

LEFT JOIN (SELECT DISTINCT ON (breed\_id) breed\_id, image\_link

FROM Image ORDER BY breed\_id, image\_link) i

ON b.id = i.breed\_id;

#	id	name	min_lif...	max_lif...	length...	length_to	min_w...	max_w...	d...	image_link
1		Abyssinian	9	15	12	16	6	10	-	https://dogell.b-cdn.net/uploads/breed/thumb_abyssini...
2		Aegean	9	10	14	17	7	10	-	https://dogell.b-cdn.net/uploads/breed/thumb_aegean...
3		American ...	11	15	14	15	8	13	-	https://dogell.b-cdn.net/uploads/breed/thumb_america...
4		American ...	15	20	12	15	7	12	-	https://dogell.b-cdn.net/uploads/breed/thumb_america...
5		American ...	14	18	15	20	8	15	-	https://dogell.b-cdn.net/uploads/breed/thumb_america...
6		Arabian M...	12	14	15	16	8	16	-	https://dogell.b-cdn.net/uploads/breed/thumb_arabian...
7		Asian	15	15	14	15	10	10	-	https://dogell.b-cdn.net/uploads/breed/thumb_asian.png
8		Australian...	15	18	15	16	8	15	-	https://dogell.b-cdn.net/uploads/breed/thumb_australi...
9		Bambino	9	15	9	11	4	9	-	https://dogell.b-cdn.net/uploads/breed/thumb_bambin...

# Запрос данных для «сетки» пород (со странами)



```
SELECT b.id, b.name, b.min_life_expectancy, b.max_life_expectancy,  
       b.length_from, b.length_to, b.min_weight, b.max_weight, i.image_link,  
       array_agg(o.country_name) AS countries  
FROM Breed b  
LEFT JOIN (SELECT DISTINCT ON (breed_id) breed_id, image_link  
FROM Image ORDER BY breed_id, image_link) i ON b.id = i.breed_id  
LEFT JOIN BreedOrigin bo ON b.id = bo.breed_id  
LEFT JOIN Origin o ON bo.origin_id = o.country_id  
GROUP BY b.id, i.image_link;
```

#	id	name	min_lif...	max_lif...	length...	length_to	min_w...	max_w...	image_...	countries
8		Australia...	15	18	15	16	8	15	https://d...	{Australia}
25		Donskoy	12	15	15	16	6	15	https://d...	{Russia}
12		Bombay	12	20	13	20	8	15	https://d...	{"United State...
46		Sokoke	9	15	15	17	6	10	https://d...	{"United State...
32		Nebelung	11	18	16	20	7	15	https://d...	{"United State...
1		Abyssinian	9	15	12	16	6	10	https://d...	{Ethiopia,Iran}
48		Sphynx	8	14	13	15	6	12	https://d...	{Canada}

```
INSERT INTO BreedOrigin (breed_id, origin_id)  
SELECT b.id, o.country_id FROM Breed b, Origin o  
WHERE b.name = 'Abyssinian' AND o.country_name = 'Iran';
```

# Доп. запрос для страницы конкретной породы



```
SELECT
    b.name, s.family_friendly, s.shedding, s.general_health,
    s.playfulness, s.children_friendly, s.grooming,
    s.intelligence, s.other_pets_friendly, s.hypoallergenic
FROM Breed b
LEFT JOIN SpecialCharacteristic s
ON b.spec_characteristic_id = s.ch_id
WHERE b.id = 1;
```

```
1 SELECT
2     b.name, s.family_friendly, s.shedding, s.general_health,
3     s.playfulness, s.children_friendly, s.grooming,
4     s.intelligence, s.other_pets_friendly, s.hypoallergenic
5 FROM Breed b
6 LEFT JOIN SpecialCharacteristic s
```

name	family_...	shedding	genera...	playful...	childre...	groomi...	intellig...	other_...	hypoallerge...
Abyssinian	3	3	2	5	5	3	5	5	f

# Получение картинок породы (галерея для конкретной)

```
SELECT i.image_link FROM Breed b
JOIN Image i ON b.id = i.breed_id
WHERE b.id = 1;
```

```
1 SELECT i.image_link FROM Breed b
2 JOIN Image i ON b.id = i.breed_id

image_link

https://dogell.b-cdn.net/uploads/breed/thumb_abyssinian.png

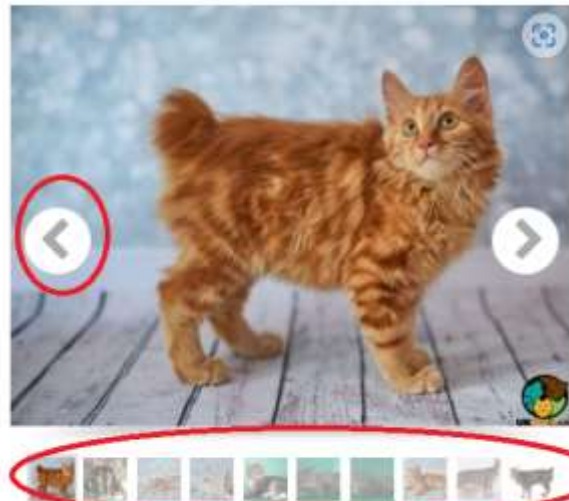
https://dogell.b-cdn.net/uploads/breed/thumb_abyssinian1.png
```

```
SELECT ARRAY_AGG(i.image_link) AS image_links
FROM Breed b
JOIN Image i ON b.id = i.breed_id
WHERE b.id = 1;
```

```
1 SELECT ARRAY_AGG(i.image_link) AS image_links
2 FROM Breed b

image_links

(https://dogell.b-cdn.net/uploads/breed/thumb_abyssinian.png,https://dogell.b-cdn.net/uploads/breed/thumb_abyssinian1.png)
```



## ИТМО

Size: ★★★★★  
Grooming: ★★★★★  
Exercise Level: ★★★★★  
Trainability: ★★★★★  
Good with Children: ★★★★★  
Good with other pets: ★★★★★  
Affectionate: ★★★★★  
Height: 3 - 10cm M | 3 - 10cm F  
Weight: 3 - 6kg M | 3 - 6kg F  
Life Expectancy: 13 - 15 Years

Сценарий где понадобится

# Фильтрационные запросы 1



Поиск длинных, но не толстых КОТОВ

```
SELECT id, name FROM Breed
WHERE length_to > 18 AND max_weight <= 15;
```

```
1 SELECT ID, NAME, avg_life_expectancy
2 FROM (SELECT ID, NAME,
```

id	name	avg_life_expectancy
22	Cornish Rex	13
23	Cypriot	13
25	Donskoy	13
29	Lykoi	13
35	Peterbald	13

```
1 SELECT ID, NAME FROM Breed
2 WHERE length_to > 18 AND max_weight <= 15;
```

id	name
5	American Wirehair
12	Bombay
18	California Spangled
32	Nebelung
42	Siamese Cat
45	Snowshoe

```
SELECT id, name, avg_life_expectancy
FROM (SELECT id, name,
      (min_life_expectancy + max_life_expectancy) / 2 AS avg_life_expectancy
      FROM Breed) AS bs
WHERE avg_life_expectancy > 12 AND avg_life_expectancy < 14;
```


«Маложивущих» не хотим, но и на пенсии долго «содержать» нежелательно

# Фильтрационные запросы 2



У нас аллергия, так что нужны гипоаллергенные

```
SELECT b.id, b.name FROM Breed b
JOIN SpecialCharacteristic sc
ON b.spec_characteristic_id = sc.ch_id
WHERE sc.hypoallergenic = TRUE;
```



#	id	name
10		Bengal Cats
16		Burmese
18		California Spangled
21		Colorpoint Shorthair
22		Cornish Rex
24		Devon Rex
25		Donskoy
27		Javanese

```
1 SELECT b.id, b.name, s.children_friendly, s.intelligence, s.hypoallergenic
2 FROM Breed b
```

#	id	name	children_friendly	intelligence	hypoallergenic
10		Bengal Cats	5	5	t
16		Burmese	5	5	t
42		Siamese Cat	4	5	t
53		Ukrainian Levkoy	5	5	t



```
SELECT b.id, b.name, s.children_friendly, s.intelligence, s.hypoallergenic
FROM Breed b
JOIN SpecialCharacteristic s ON b.spec_characteristic_id = s.ch_id
WHERE s.children_friendly >= 4 AND
s.intelligence > 4 AND s.hypoallergenic = TRUE;
```

Более полная фильтрация (по двум таблицам)



# Заключение

В результате работы были освоены навыки:

- Навыки программирования на ЯП Python
- Работы с библиотеками pandas и numpy для обработки данных;
- Работы с инструментами парсинга BeautifulSoup и requests;
- Работы с реляционными базами данных и языком запросов SQL;





# Спасибо за внимание!

**IT**'s *MO*re than a  
**UNIVERSITY**

[370113@edu.itmo.ru](mailto:370113@edu.itmo.ru)

[fan115-1@yandex.ru](mailto:fan115-1@yandex.ru)

[https://t.me/IF115\\_1](https://t.me/IF115_1)

# Дополнение

Тематика хранения пород позволяет использовать и нереляционные базы данных, как например **MongoDB**. Пример работы (сокращены некоторые поля) приведен ниже:

```
breed = {
  "name": "Сибирская",
  "origin": "Россия, Сибирь",
  "averageLifespan": 5,
  "length": {
    "from": 20,
    "to": 30
  },
  "weight": {
    "from": 3,
    "to": 15
  },
  "characteristics": {
    "gentleness": 5,
    "immunity": 5,
    "playfulness": 6,
    "molt": 3,
    "care": 8,
    "intelligence": 1,
    "childFriendliness": 10,
    "petFriendliness": 7
  },
  "description": "Очень крутая порода",
  "image": "https://www.google.com/url?sa
}

db.breeds.insertOne(breed)
```

```
db.breeds.find({
  "averageLifespan": {"$gte": 3, "$lte": 7},
  "length.from": 10,
  "length.to": 20,
  "weight.from": 10,
  "weight.to": 20,
  "characteristics.gentleness": {"$gte": 3, "$lte": 7},
  // ...
})
```

```
w = 18
db.breeds.find({'$and': [{'weight.from': {'$lte': w}}, {'weight.to': {'$gte': w}}]})
```



MongoDB может быть быстрее по скорости запросов, за счет структуры и хранения в одной «таблице»