# МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

## ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ» Обнинский институт атомной энергетики Отделение интеллектуальных кибернетических систем

### Лабораторная работа №2

по курсу «Нереляционные базы данных» Вариант №4 — города мира.

Подписи:	
Исполнитель	
студент гр. ИС-Б17	 В. Ю. Петренко
Принял	
ктн	С В Грицюк

#### Задание:

> 1000, name, country, region, population, etc.

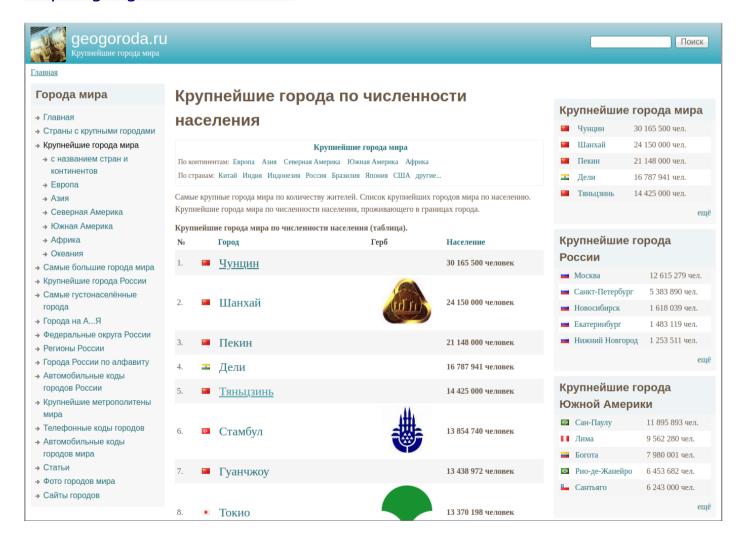
Make a research and find the source(s) of the data for your variant. As data, you should consider only

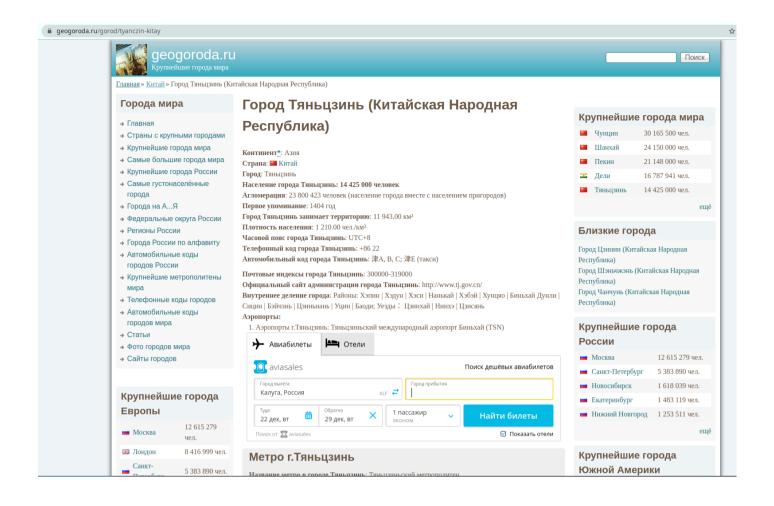
weakly structured data sources (web pages). Your report should contain:

- and data location (website with examples)
- and data structure, format, field types
- describe data extraction/parsing logic (algorithm)
- describe the planned procedures for data cleaning
- describe the output format (as JSON) with field types description
- $_{\neg \land}$  develop a list of queries (3 simple + 2 advanced) that could be executed for your dataset (statistics, samples)

#### Города мира:

https://geogoroda.ru/citizens





Структура данных (формат, типы полей):

Название — string

Континент — string

Страна — string

Haceление города - Int

Первое упоминание - Int

Занятая территория - Int

Плотность населения - Int

Часовой пояс - Int

Алгоритм извлечения:

Считываем веб-страницу;

Извлекаем необходимую информацию.

<div class="views-field views-field-nothing"> <span class="fieldcontent"><div><b>Континент<a class="tooltip" href="#">\*<span</pre>

```
class="custom info"><img src="/sites/default/files/info.png"
alt="Информация" height="48" width="48">
</span></a></b>: Азия</div>
<div itemscope=""
itemtype="http://schema.org/Country"><b><mark>Страна</mark></b>: <div
class="countryicon-sprite iconset-shiny countrycode-cn"></div> <span
itemprop="name"><a href="/strana/kitay" typeof="skos:Concept"
property="rdfs:label skos:prefLabel" datatype="">Китай</a></span></
div>
<div itemscope="" itemtype="http://schema.org/City"><b><u>Город</u></b>:
<span itemprop="name">Тяньцзинь</span></div>
<div></div></span> </div>
 <div class="views-field views-field-field-naselenie">
                                                        <strong
class="field-content">Население города Тяньцзинь: 14 425 000
<mark>человек</mark></strong> </div>
 </div>
 <div class="views-field views-field-field-osnovan">
                                                       <div
class="field-content"><b>Первое упоминание</b>: 1404 год</div>
</div>
 <div class="views-field views-field-field-ploshad-goroda">
                                                              <div
class="field-content"><b>Город Тяньцзинь занимает
территорию</b>: 11 943.00 км<sup>2</sup></div> </div>
 <div class="views-field views-field-field-plotnost">
                                                      <div
class="field-content"><b>Плотность населения</b>: 1 210.00
чел./км<sup>2</sup></div> </div>
 <div class="views-field views-field-field-time-zone">
                                                        <div
class="field-content"><b> Часовой пояс города Тяньцзинь </b>:
UTC+8 </div>
Выходной формат с описанием типа полей (JSON):
{
```

```
пате: 'Тяньцзинь',
continent: 'Азия',
country: 'Китай',
population: 14 425 000,
fmention: 1404,
area: 11 943,
density: 1 210,
timezone: 8
}

Список возможных запросов:
Вывести самый населенный город;
Вывести самый древний город;
Вывести все города Азии;
Вывести города с часовым поясом +3;
Вывести все города России;
```

Вывести города с плотностью населения от 800 до 1200 чел./км<sup>2</sup>.