



ПРОГРАММНОЕ СРЕДСТВО ОПРЕДЕЛЕНИЯ И УНИФИКАЦИИ НАИМЕНОВАНИЙ НАУЧНЫХ ОРГАНИЗАЦИЙ

Студент: Чупрыгин Семён Сергеевич

Научный руководитель: Улизко Михаил Сергеевич

Дубна, 2024 г.



АКТУАЛЬНОСТЬ

- Разное указание аффилиаций авторами научных статей

Ресурсы

- Scopus
- Web of Science
- ArXiv
- ResearchGate
- ScienceDirect

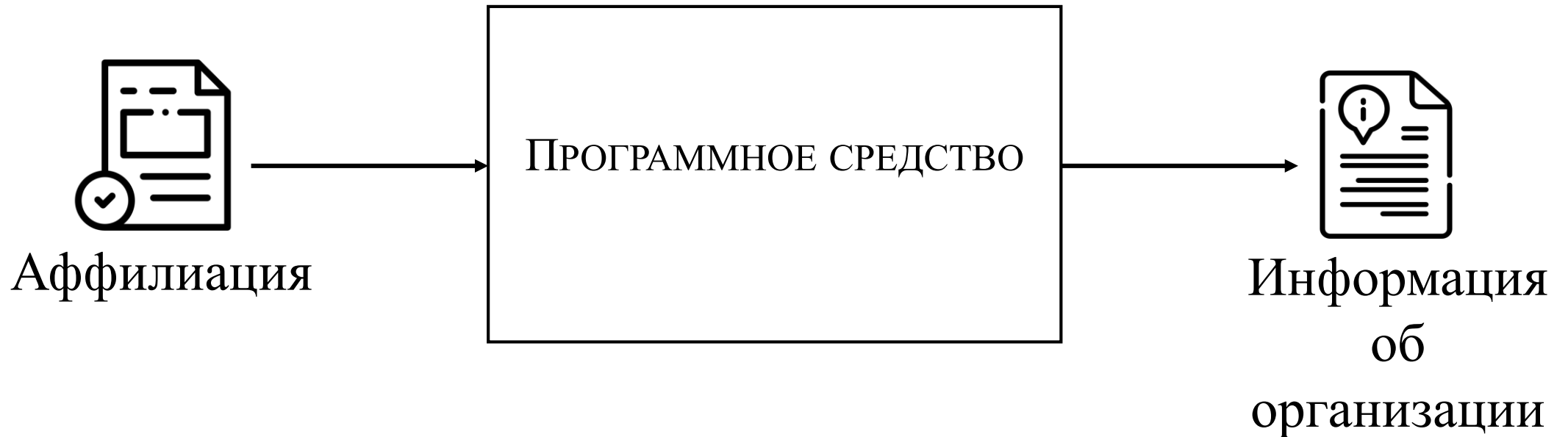
Аффилиации

- JINR
- Joint Institute for Nuclear Research
- Obyedinyennyi institut yadernykh isslyedovaniï
- Nuclear Research Joined Institute
- United Nuclear Research Institute
- Joint Inst for Nucl Research

Объединённый институт
ядерных исследований
(ОИЯИ)

Цель работы

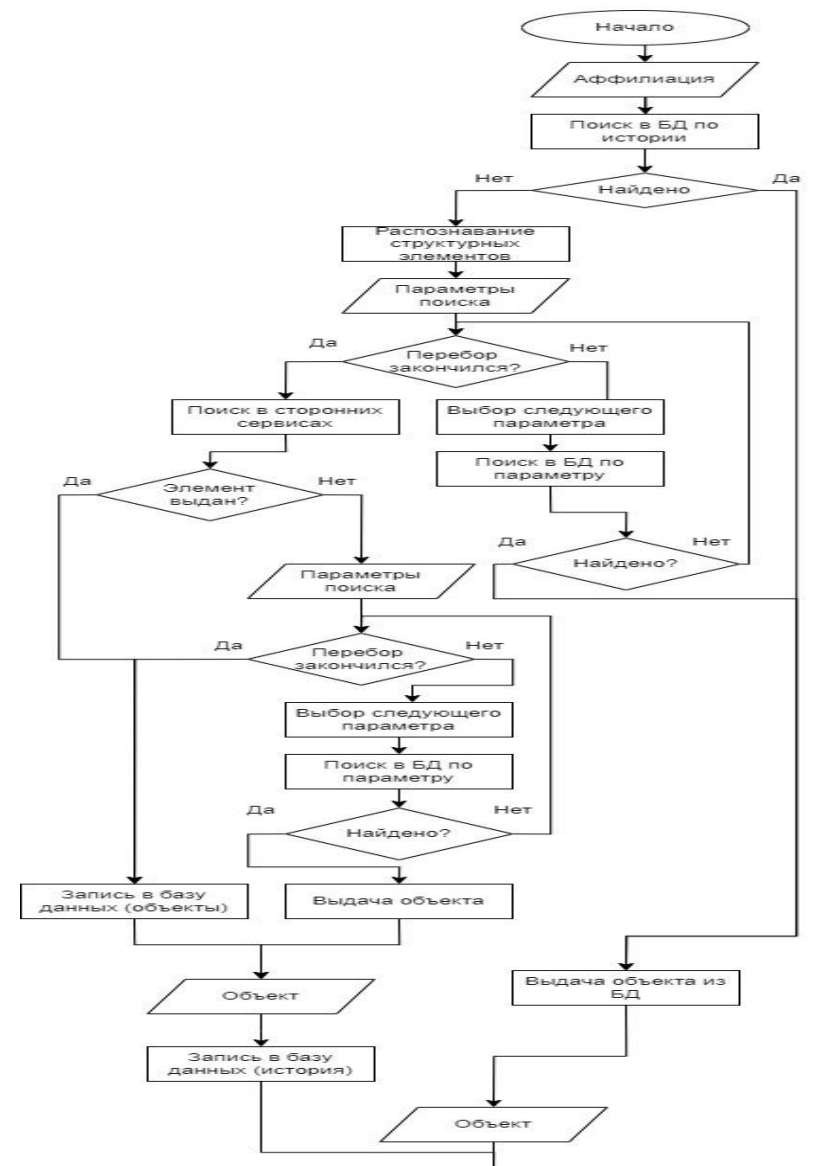
- Разработать программное средство определения и унификации наименований научных организаций



Целевое представление аффилиации

NAME_RU	Объединённый институт ядерных исследований (ОИЯИ)
NAME_ENG	Joint Institute for Nuclear Research (JINR)
NAME_NATIONAL	Объединённый институт ядерных исследований (ОИЯИ)
ZIP_CODE	141980
COUNTRY(ISO-2)	RU
LOCATION	Московская обл., Дубна, Жолио-Кюри, 6
COORDINATES	{ “lat”:56.7, “lon”:37.2 }

Алгоритм

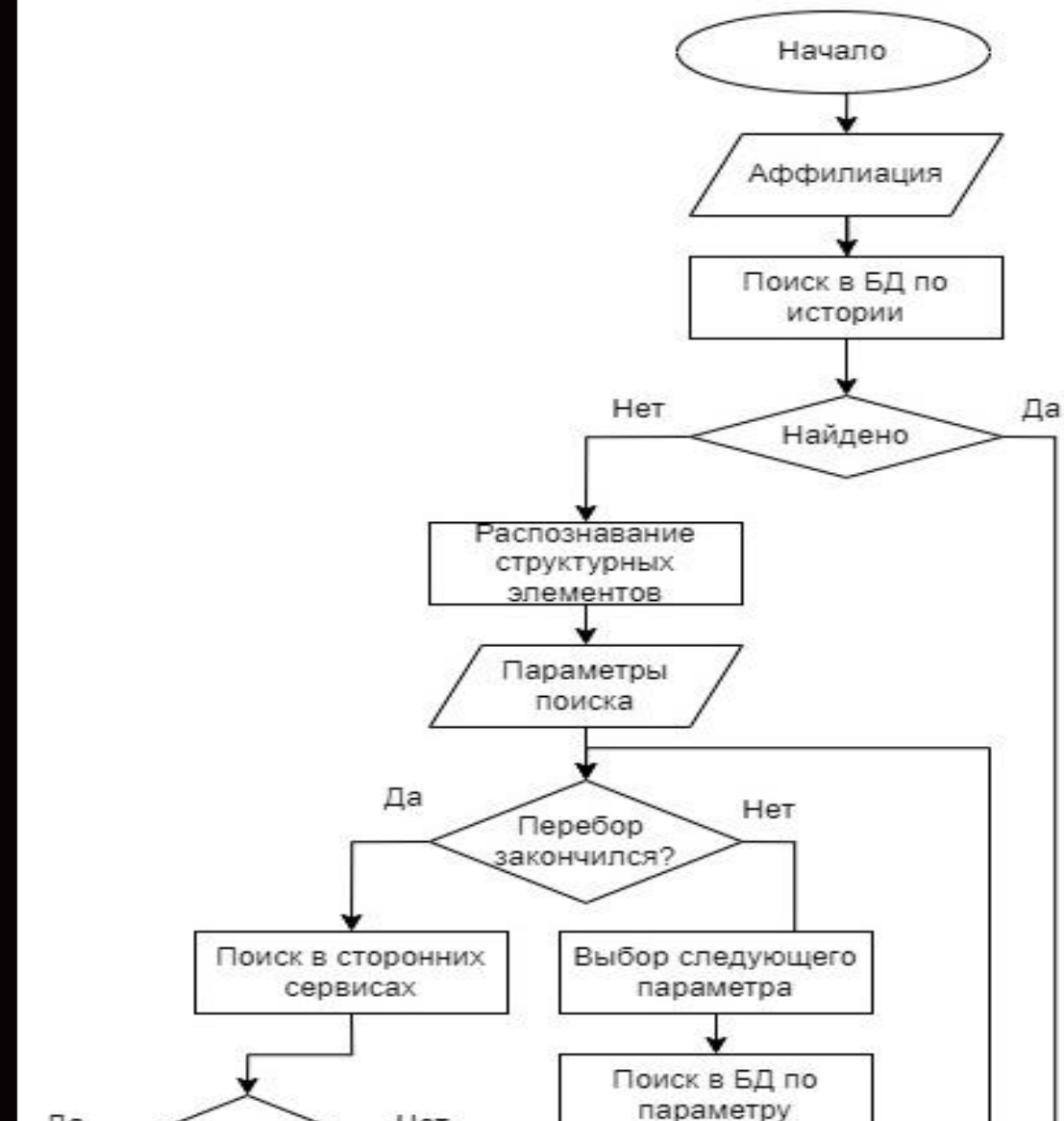


Алгоритм

Meshcheryakov Laboratory of
Information Technologies, Joint Institute
for Nuclear Research, 6 Joliot-Curie,
Dubna, Moscow region, 141980, Russia



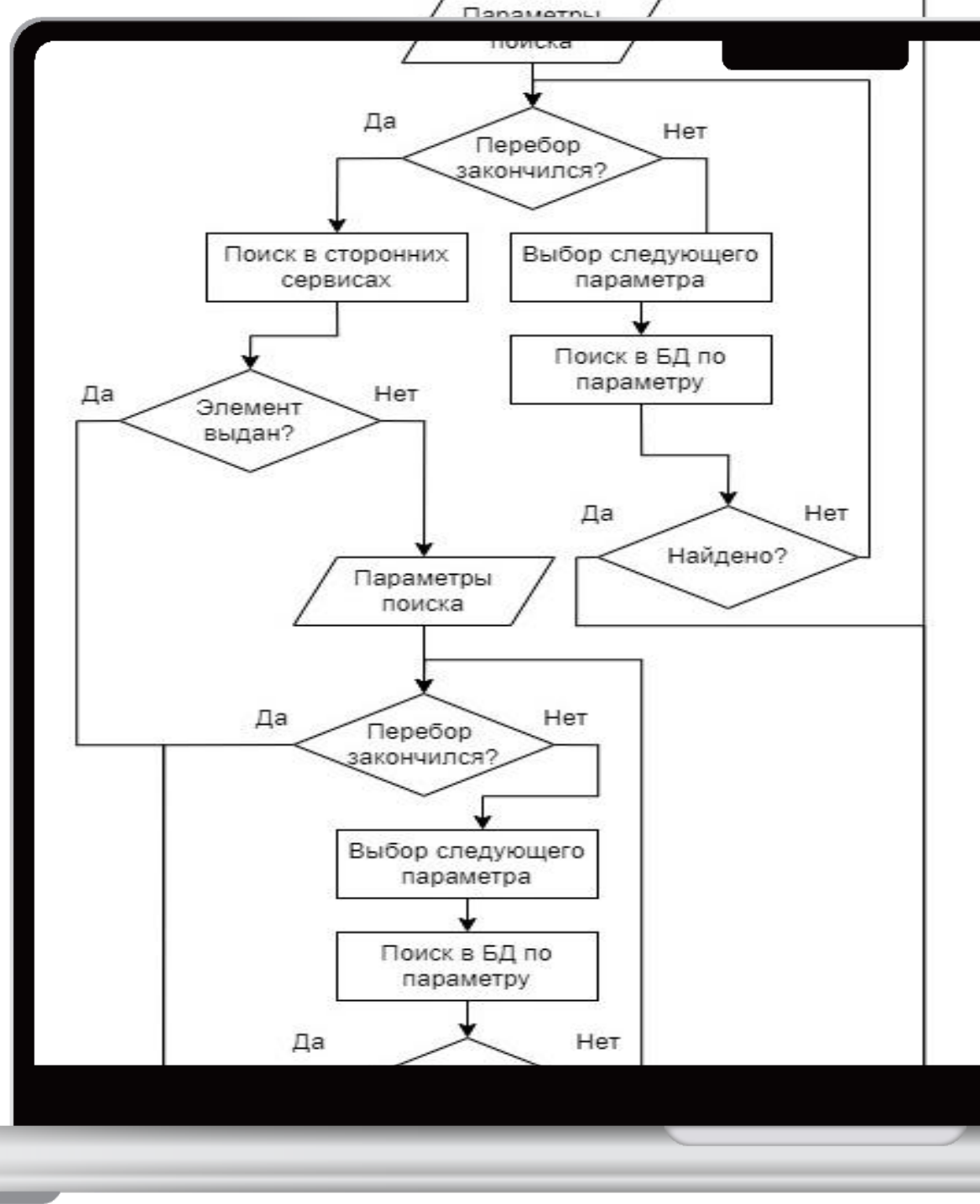
- **Laboratory:** Meshcheryakov Laboratory of Information Technologies
- **Institution:** Joint Institute for Nuclear Research
- **Addrline:** 6 Joliot-Curie
- **PostCode:** 141980
- **Settlement:** Dubna Moscow
- **Country:** Russia



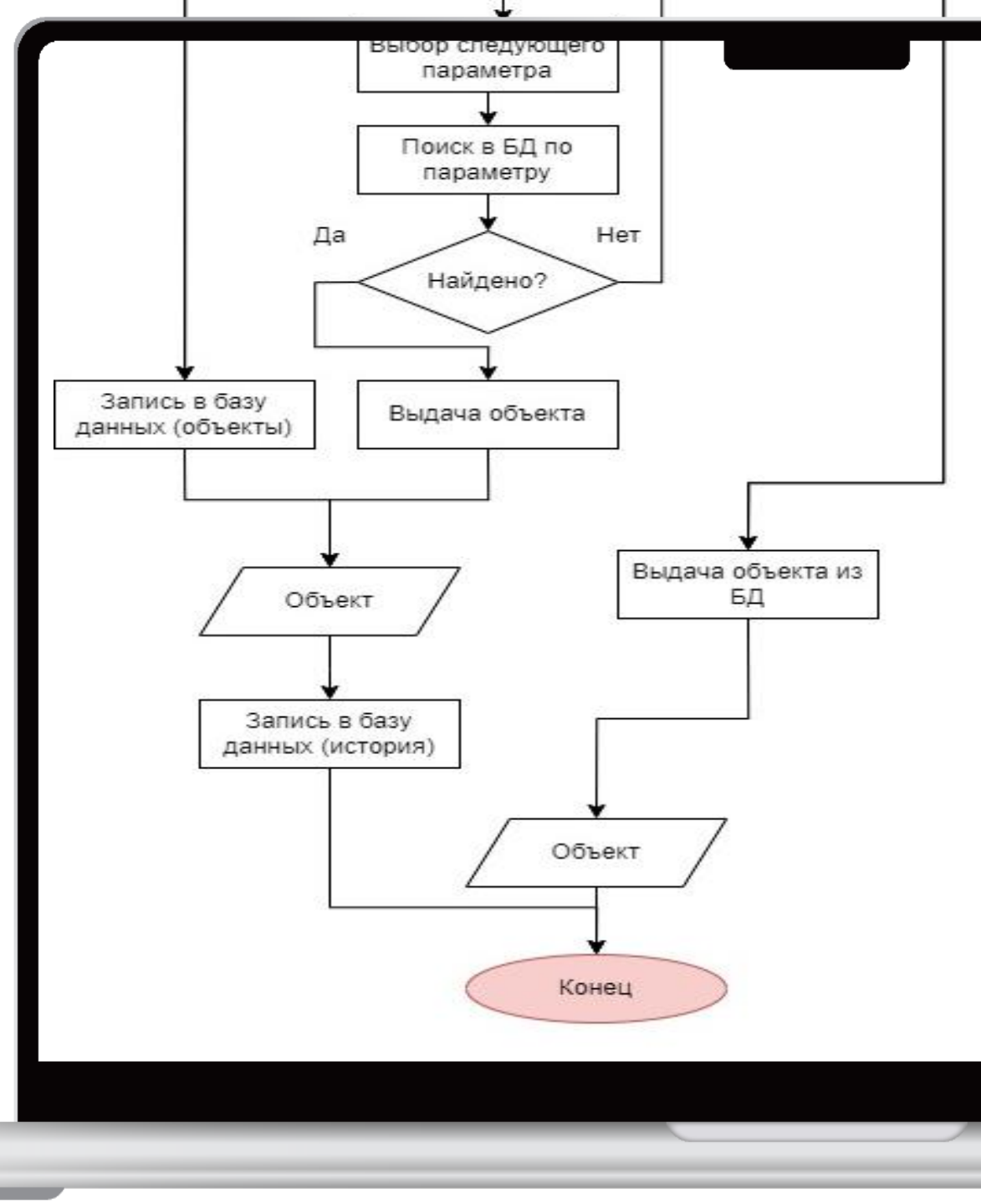
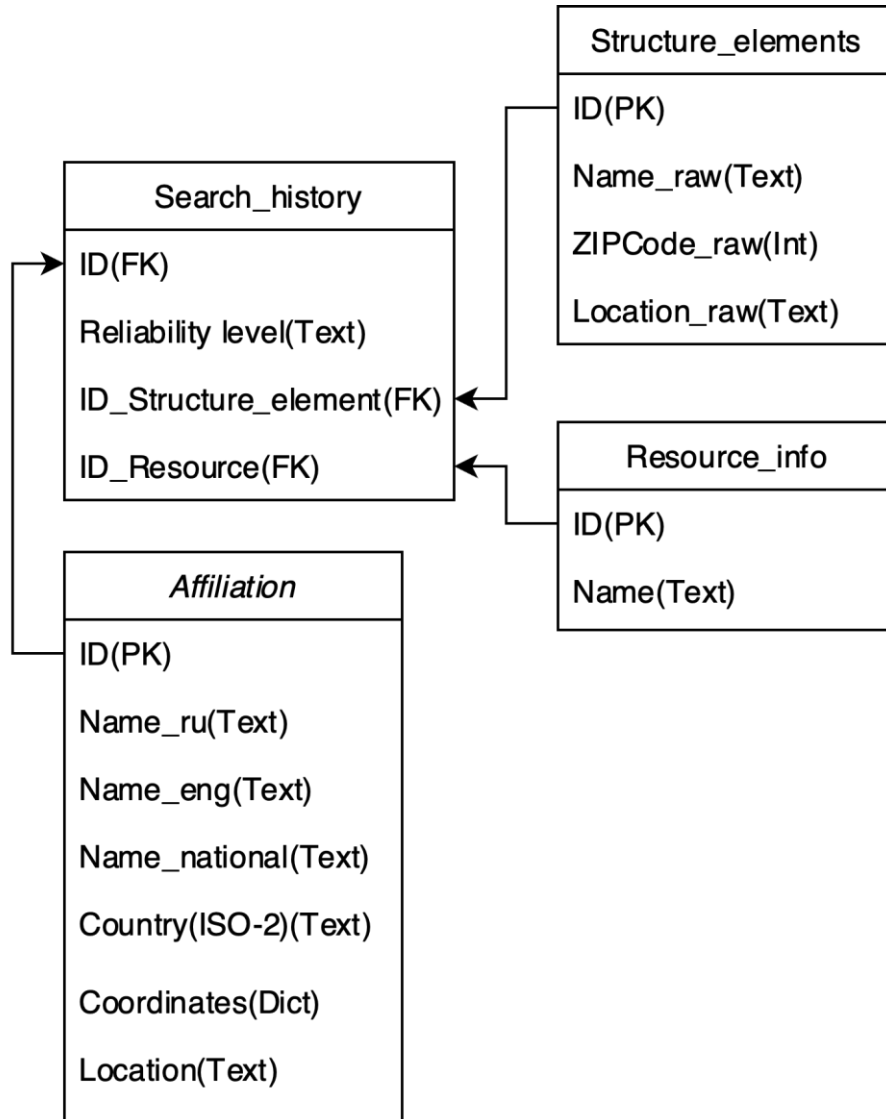
Алгоритм

Параметры

- Название
- Адрес
- Координаты
- Страна
- ZIP-код



Хранилище



Этап 1. Поиск в Базе данных

```
{  
'Name_raw': None,  
'Location_raw': None,  
'Name_ru': None,  
'Name_eng': None,  
'Name_national': None,  
'ZIP_code': None,  
'Country(ISO-2)': None,  
'Location': None,  
'Coordinates': None,  
}
```

ВХОД

""Meshcheryakov Laboratory of Information
Technologies, Joint Institute for Nuclear Research, 6
Joliot-Curie, Dubna, Moscow region, 141980,
Russia""



ВЫХОД

None

Этап 2. Распознавание сущностей

```
{  
'Name_raw': None,  
'Location_raw': None,  
'Name_ru': None,  
'Name_eng': None,  
'Name_national': None,  
'ZIP_code': None,  
'Country(ISO-2)': None,  
'Location': None,  
'Coordinates': None,  
}
```

ВХОД

""Meshcheryakov Laboratory of Information Technologies, Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna, Moscow region, 141980, Russia""



ВЫХОД

```
<affiliation><orgName  
type="laboratory">Meshcheryakov Laboratory of  
Information Technologies</orgName><orgName  
type="institution">Joint Institute for Nuclear  
Research</orgName><address><addrLine>6 Joliot-  
Curie</addrLine><postCode>141980</postCode><se  
ttlement>Dubna, Moscow  
region</settlement><country  
key="RU">Russia</country></address></affiliation>
```

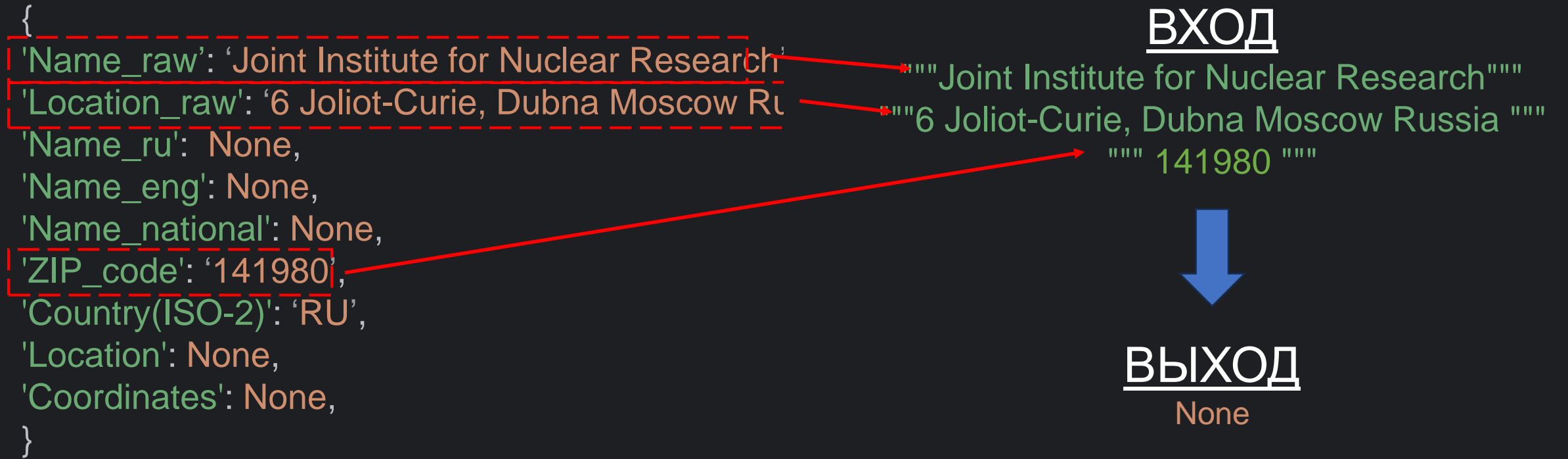
Этап 2. Распознавание сущностей

```
{  
'Name_raw': 'Joint Institute for Nuclear Research',  
'Location_raw': '6 Joliot-Curie, Dubna Moscow Rus  
'Name_ru': None,  
'Name_eng': None,  
'Name_national': None,  
'ZIP_code': '141980',  
'Country(ISO-2)': 'RU',  
'Location': None,  
'Coordinates': None,  
}
```

```
1 import requests  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20
```

```
1 usage  
def NER(affiliation):  
    headers = {  
        'Content-Type': 'application/x-www-form-urlencoded',  
    }  
    data = f'affiliations={affiliation}'  
    url = 'http://<host>:<port>'  
    r = requests.post(url, headers=headers, data=data)  
    if r.status_code == 200:  
        print(r.text)  
  
NER(  
    """Meshcheryakov Laboratory of Information Technologies,  
    Joint Institute for Nuclear Research, 6 Joliot-Curie, Dubna,  
    Moscow region, 141980, Russia"""  
)
```

Этап 3. Поиск в базе данных



Этап 4. Поиск в сервисах геокодирования

```
{  
'Name_raw': 'Joint Institute for Nuclear Research'  
'Location_raw': '6 Joliot-Curie, Dubna Moscow Ru  
'Name_ru': None,  
'Name_eng': None,  
'Name_national': None,  
'ZIP_code': '141980',  
'Country(ISO-2)': 'RU',  
'Location': None,  
'Coordinates': None,  
}
```

ВХОД
"Joint Institute for Nuclear Research"



ВЫХОД
Location(Объединённый институт ядерных исследований, аллея имени Н. С. Амаглобели, Институтская часть, Дубна, городской округ Дубна, Московская область, Центральный федеральный округ, 141980, Россия, (56.735897800000004, 37.18739802956689, 0.0))

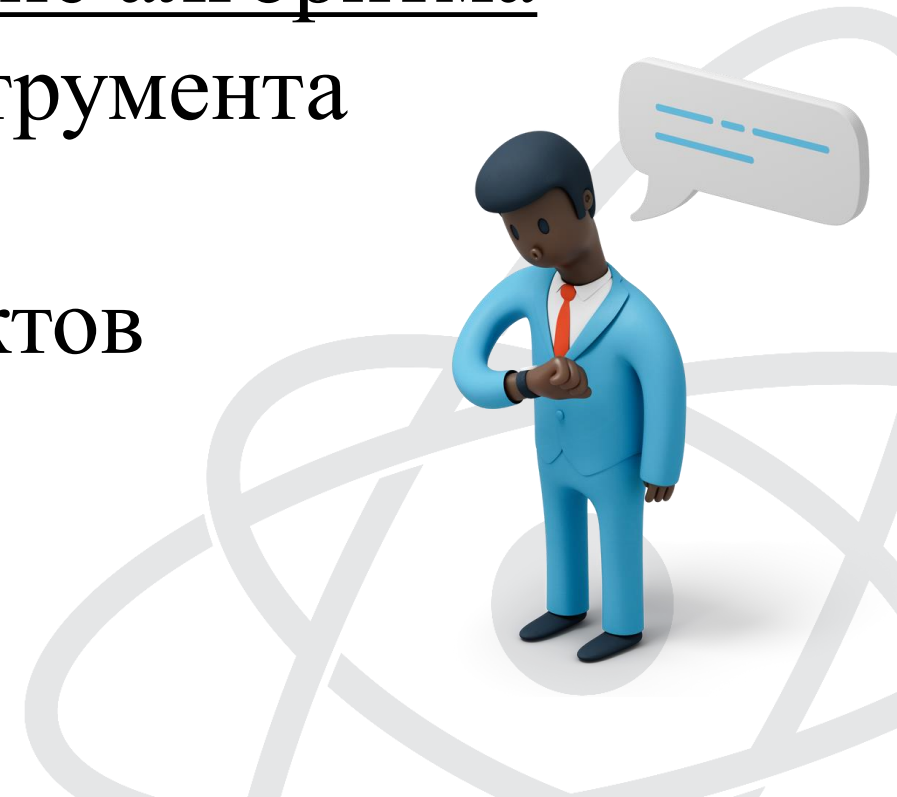
Этап 4. Поиск в сервисах геокодирования

```
{
  'Name_raw': 'Joint Institute for Nuclear Research',
  'Location_raw': '6 Joliot-Curie, Dubna Moscow Russia',
  'Name_ru': 'Объединённый институт ядерных исследований',
  'Name_eng': 'Joint Institute for Nuclear Research',
  'Name_national': 'Объединённый институт ядерных исследований',
  'ZIP_code': '141980',
  'Country(ISO-2)': 'RU',
  'Location': 'Аллея имени Н. С. Амаглобели,
  Институтская часть, Дубна, городской округ Дубна,
  Московская область, Центральный федеральный
  округ',
  'Coordinates': '56.735897800000004,
  37.18739802956689',
}
```

```
def Service(org_name: str) -> Union[dict, None]:
    org_location: object = GEOCODER.geocode(org_name, addressdetails=True, namedetails=True)
    print(org_name, org_location)
    location = {}
    country_code = None
    if org_location:
        location['lat'] = org_location.latitude
        location['lon'] = org_location.longitude
    if location:
        country = geopoint_to_country(lat=location['lat'], lon=location['lon'])
        country_eng = translate_to_eng(country)
        if country_eng:
            country_code = get_country_code([country_eng])
        name = org_location.raw.get('namedetails', {}).get('name:en', '')
        if not name:
            name = org_location.raw.get('namedetails', {}).get('name', '')
        return {
            'name': name,
            'country': country_eng,
            'country_code': country_code,
            'location': location,
            'stage': 'geopy'
        }
    return None
```

Заключение

1. Реализован базовый алгоритм
2. Разработана схема базы данных
3. Планируется усовершенствование алгоритма
 - Апробация программного инструмента
 - Работа с синонимией
 - Обработка близлежащих объектов



**СПАСИБО ЗА
ВНИМАНИЕ**

