

# Лабораторна работа №5.

## ІП-13 Ал Хадам Мурат Резгович

```
In [36]:
from selenium import webdriver
import spacy
```

```
In [37]:
driver = webdriver.ChromeEdge()

news_link = "https://pestrecy-rt.ru/news/tag/list/specoperaciia/"

driver.get(news_link)
```

```
In [38]:
nlp = spacy.load("ru_core_news_md")
```

```
In [39]:
def get_news_info(news_li):
    news_info_title = news_li.find_element("xpath", ".*//a/div[1]/h2").text
    news_info_subtitle = news_li.find_element("xpath", ".*//a/div[1]/p").text
    return news_info_title + "\n" + news_info_subtitle

def get_persons(news_text_item):
    doc = nlp(news_text_item)
    for pers in doc.ents:
        if pers.label_ == "PER":
            print(pers.text)

def scrape_page():
    news_xpath = "/html/body/main/ul"
    news_ul = driver.find_element("xpath", news_xpath)
    news = news_ul.find_elements("xpath", ".*//li")
    news_parsed = [get_news_info(news_item) for news_item in news]
    for news_item in news_parsed:
        get_persons(news_item)
```

```
In [40]:
while True:
    scrape_page()
    paging_buttons = driver.find_elements("xpath", "/html/body/main/div[2]/div/a/span")
    next_button = paging_buttons[-1]

    if next_button.text == "Далее >":
        next_button.click()
    else:
        break
```

Минмолодежи РТ  
Айгуль Сабилова  
Владислав Кузнецов  
Юрия Ивановича  
Марии Ивановны М.  
Валерий  
Алексей  
Валера  
Рафката Амировича  
Раушании Габдрахмановны  
Алмаз К.  
Рустам Сафиуллин  
Ильдара Насыбуллина  
Пестрецов  
Ивана Додосова  
Раис  
Рустам Минниханов  
Фаниль Аглиуллин  
Александра  
Лейла Фазлеева  
Гузель Удачина  
Ивана Додосова  
Пестрецов  
Елена Корчагина  
Сергей Корчагин  
Лейла Фазлеева  
Евгений Токмаков  
Альберт Ибятуллин  
Евгений Токмаков  
Лейла Фазлеева  
Лейла Фазлеева  
Гузель Удачина  
Ивана Додосова  
Пестрецов  
Елена Корчагина  
Сергей Корчагин  
Лейла Фазлеева  
Евгений Токмаков  
Альберт Ибятуллин  
Евгений Токмаков  
Лейла Фазлеева  
Артема Прокопчука  
Эдуард Вафин  
Ринат Садыков  
Сергей  
Сергей Шойгу  
Эдуард Вафин  
Кабмина РТ  
Евгений Токмаков  
Марат Нуриев  
Соцфонд  
Раиса Республики Татарстан  
Рустама Минниханова  
Путин  
Расиму Баксикову  
От  
Неустроев  
Верховному Главнокомандующему  
Евгений Токмаков  
Эдуард Шарафиев  
Сергей  
Родину  
Александром Агафоновым  
Ему  
Путин  
Татьяна Голикова  
Николая Чудотворца  
Жукова  
Валерий Максимов  
Валерием Межва  
Путин  
Рустам Сафиуллин  
Евгений Токмаков  
Александр Владимиров  
Евгений Токмаков  
Фарит Валиев  
Иван Додосов  
Иван Додосов  
Марс Бикбов  
Галина Тимофеева  
Александр  
Эдуард Вафин  
Минниханов  
Камиль Самигуллин  
Куюков  
Ильхама Кашапова  
Эдуард Вафин  
Путин  
Вячеслав Володин  
Виталий Беляев  
Тамара Лаптева  
Ринат Садыков  
Минниханов  
Ильхам Кашапов  
Путин  
Эдуард Шарафиев  
Путин  
Шихазды  
Куюков  
Виталий Беляев  
Эльмира Зарипова  
Иван Додосов  
Пестрецов  
Тинчурина  
Программа  
Дедом  
Снегурочкой  
Тимур Сулейманов  
Ильхам Кашапов