





Сайт государственных закупок (ЕИС), главная страница поиска тендеров:
<https://zakupki.gov.ru/epz/order/extendedsearch/results.html> , с постраничной разбивкой по 10 тендеров.

Нужно реализовать программу на Python, которая обходит первые две страницы по 44ФЗ:

- <https://zakupki.gov.ru/epz/order/extendedsearch/results.html?fz44=on&pageNumber=1>
- <https://zakupki.gov.ru/epz/order/extendedsearch/results.html?fz44=on&pageNumber=2>
-

При обходе, у каждого элемента списка (тендера), нужно собирать ссылку на его печатную форму:

<div>44-ФЗ Электронный аукцион</div> <div>№ 0338100004625000003 Подача заявок</div> <div><div>Объект закупки</div><div>Поставка топлива моторного</div></div> <div><div>Заказчик</div><div>ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ "КАМЧАТСКАЯ ДИРЕКЦИЯ ПО ТЕХНИЧЕСКОМУ ОБЕСПЕЧЕНИЮ НАДЗОРА НА МОРЕ"</div></div>	<div> </div> <div>Начальная цена 261 778,70 ₽</div> <div><div>Размещено</div><div>03.03.2025</div><div>Обновлено</div><div>03.03.2025</div></div> <div><div>Окончание подачи заявок</div><div>12.03.2025</div></div>
<div>44-ФЗ Запрос котировок в электронной форме</div> <div>№ 0161300006325000014 Подача заявок</div> <div><div>Объект закупки</div><div>Содержание мест массового отдыха, туристских маршрутов, подъездных путей к местам туристского показа и осмотра, местам массового отдыха</div></div> <div><div>Заказчик</div><div>АДМИНИСТРАЦИЯ МУНИЦИПАЛЬНОГО ОБРАЗОВАНИЯ ДОЛИНСКИЙ МУНИЦИПАЛЬНЫЙ ОКРУГ САХАЛИНСКОЙ ОБЛАСТИ</div></div>	<div> </div> <div>Начальная цена 2 950 000,00 ₽</div> <div><div>Размещено</div><div>03.03.2025</div><div>Обновлено</div><div>03.03.2025</div></div> <div><div>Окончание подачи заявок</div><div>10.03.2025</div></div>

Эти ссылки имеют вид

<https://zakupki.gov.ru/epz/order/notice/printForm/view.html?regNumber=0338100004625000003>.

Заменяв `view.html` на `viewXml.html`, получим ссылку на печатную XML-форму (<https://zakupki.gov.ru/epz/order/notice/printForm/viewXml.html?regNumber=0338100004625000003>). Распарсив этот XML, для каждого тендера нужно получить значение XML-поля `publishDTInEIS`, или `None` в случае его отсутствия.

```

</script>
<id>38659608</id>
<versionNumber>2</versionNumber>
▼<commonInfo>
  <purchaseNumber>0338100004625000003</purchaseNumber>
  <docNumber>ИИ1</docNumber>
  <directDT>2025-03-03T16:54:10.369+12:00</directDT>
  <plannedPublishDate>2025-03-03+12:00</plannedPublishDate>
  <publishDTInEIS>2025-03-03T19:52:24.784+12:00</publishDTInEIS>
  <href>https://zakupki.gov.ru/epz/order/notice/eaz0/view/common-
  <notPublishedOnEIS>true</notPublishedOnEIS>

```

Для простоты, результат вывести прямо в консоль (например, обычным `print()`), в виде пары “ссылка на печатную форму”-“дата публикации”.

Рекомендуемый стек:

- Python3
- пакет `Requests` - <https://docs.python-requests.org/en/latest/> для сетевых запросов
- пакет `BeautifulSoup` - <https://beautiful-soup-4.readthedocs.io/en/latest/> для парсинга HTML и получения ссылок на печатные формы
- пакет `XmlToDict` - <https://github.com/martinblech/xmltodict> для парсинга XML

Пакеты из списка выше - лишь рекомендации, если удобнее, то можно, например, парсить XML через методы стандартной библиотеки [xml.etree.ElementTree](#), или парсить HTML через [lxml/XPath](#).

Основной трудный момент этого тестового задания - нужно распараллелить задачи обхода списка и парсинга XML-форм, разбить написанный выше скрипт на отдельные таски. Для этого нужно использовать пакет `Celery` - <https://docs.celeryproject.org/en/latest/>. Рекомендуемая разбивка на (асинхронные) подзадачи:

- сбор ссылок с каждой страницы - отдельный таск (для каждой страницы)
- парсинг печатных XML-форм - отдельный таск (для каждой формы)

Для тестирования рекомендуется использовать т.н. eager-режим (синхронное выполнение) - <https://docs.celeryproject.org/en/stable/userguide/testing.html>

Плюсом при оценке программы будет использование классов для тасков Celery:

```

import celery

class Task(celery.Task):
    def run(self):
        # ваш код здесь

```

...

Результат выложить на Гитхаб или прислать на почту ir@lmx.su в виде архива.

По всем вопросам можно писать мне на почту ir@lmx.su (Иван) я буду оперативно отвечать.