

ШИНЖЛЭХ УХААН ТЕХНОЛОГИЙН ИХ СУРГУУЛЬ

Мэдээлэл, холбооны технологийн сургууль



Тайлан

Хэл ба мэдээлэл

2021-2022 хичээлийн жилийн намрын улирал

Шалгасан багш:

Батзолбоо

Гүйцэтгэсэн:

B200900802 Анужин Баатарцогт

Улаанбаатар хот
2022 он

▼ Daalgavar 1

1. requests.get(url) - вебсайтнаас хүсэлт авна.
2. BeautifulSoup(content, 'html.parser') - парс хийнэ.
3. if 'http' in link.get('href') or 'https' in link.get('href'): - линкийг нь ялгаж авна
4. urlList - ялгаж авсан линкуудээ лист-д хадгална.
5. allText += para.get_text(): - хадгалсан линкээрээ ороод сайтныхаа текстг нь ялгаж аваад allText-д хадгална. Дараа нь allText-ээ файлд хадгална.

```
import requests
from bs4 import BeautifulSoup
```

```
import urllib.request
def parseLinks(url):
    page = requests.get(url)
    content = page.content
    soup = BeautifulSoup(content, 'html.parser')
    counter = 1;
    firstSiteUrl = ''
    text = ""
    allText = ""
    urlList = []
    for link in soup.find_all('a'):
        if 'http' in link.get('href') or 'https' in link.get('href'):
            # print("{0}} {1}".format(counter, link.get('href')))
            urlList.append(link.get('href'))
            text = text + str(counter) + ") " + link.get('href')
            text += "\n"
            if counter == 1:
                firstSiteUrl = link.get('href')
            counter = counter + 1
    print()
    page = requests.get(str(firstSiteUrl))
    content = page.content
    soup = BeautifulSoup(content, 'html.parser')
    counter = 1
    # print("{0} links".format(firstSiteUrl))
    text += "\n"
    text = text + firstSiteUrl + " links:\n"
    for link in soup.find_all('a'):
        row = link.get('href')
        if 'http' in row or 'https' in row:
            # print("{0}} {1}".format(counter, link.get('href')))
            urlList.append(link.get('href'))
```

```

        text = text + str(counter) + ") " + link.get('href') + "\n"
        counter = counter + 1
for url in urlList:
    # opening the url for reading
    html = urllib.request.urlopen(url)

    # parsing the html file
    htmlParse = BeautifulSoup(html, 'html.parser')

    # getting all the paragraphs
    for para in htmlParse.find_all("p"):
        allText += para.get_text()
    file = open("sites.txt", "a", encoding = "utf-8")
    file.write(allText)
return text

```

```

url = "http://gogo.mn"
print(parseLinks(url))

```

- 1) <https://mongolia.gogo.mn>
- 2) <https://www.yolo.mn/>
- 3) <http://www.maamuu.mn/>
- 4) <https://mongolia.gogo.mn>
- 5) <https://www.maamuu.mn>
- 6) <https://www.yolo.mn>
- 7) <https://mongolia.gogo.mn>
- 8) <https://www.maamuu.mn>
- 9) <https://www.yolo.mn>

<https://mongolia.gogo.mn> links:

- 1) <https://gogo.mn>
- 2) <https://www.yolo.mn/>
- 3) <http://www.maamuu.mn/>
- 4) <https://gogo.mn>
- 5) <https://www.maamuu.mn>
- 6) <https://www.yolo.mn>
- 7) <https://gogo.mn>
- 8) <https://www.maamuu.mn>
- 9) <https://www.yolo.mn>

```

file = open("sites.txt", "r", encoding = "utf-8")
file.read()

```

'Filo киноны сэтгүүл саяхан олны анхаарлыг татаад буй Bong найруулагчийн 2021 оны шилдэг
 0\n
 443\n
 \nYOLLO.mn сайтын редакц нь ямар нэгэн асуудал тулгарсан уншигч хүүхдүүдээс захиа хүлээн
 22\n
 204\n
 \nМонгол Улс, Америкийн Нэгдсэн Улс хооронд дипломат харилцаа тогтоосны 35 жилийн ой 2021
 14\n
 94\n
 \nХайртай хүн чинь чамд ч бас хайртай эсэхэд хараахан бүрэн итгэхгүй, эргэлзэж байна уу?
 28\n
 3419\n
 \nБид “Хүүхэд бүрийг боловсролд тэгш хамруулах зарчмыг дэмжих нь” төслийн кампанит ажлын
 262\n
 526\n
 \nЧиний бие үнэхээр гайхалтай-бие чинь чамайг бүтээж, хийхийг хүссэн зүйлсийг чинь хийлгэж
 79\n
 520\n
 \nБидэнд мэдээлэл авах, бусадтай харилцах гэх мэтчилэн дэлгэцтэй ойр байх олон шалтгаан
 54\n
 196\n
 \nХүн бүрийн хайртай жүжигчин, бидний “нөхөрсөг хөрш” Andrew Garfield саяхан “Entertainment
 14\n
 286\n
 \nUSAID/DAI-гаас санаачлан сурталчилж буй “Миний минийх” кампанит ажлын хүрээнд хийгдсэн
 42\n
 141\n
 \nСтресс, бухимдалтай өдрүүд амьдралд чинь багагүй тохиолдож байна уу? Үүнээс улбаалаад
 262\n
 1624\n
 \nK-поп одод тайз, дэлгэц, зар сурталчилгааны үеэр үргэлж өөрсдийн стилистүүдийн сонгож
 71\n
 473\n
 \nCONJOO: Чиний дотор ямар хүн нуугдаж байгаа бол? Амралт эхэлчихсэн нойр хүрэхгүй, шөнө
 байгаа бол хийж болох сонирхолтой 5...\n\n\n
 \n\n\n 34416\n
 \nӨнөөдөр буюу хоёрдугаар сарын 16-наас зургадугаар сарын 1 хүртэл орон даяар “Илүүдэл х
 690\n
 \n\n 2022-02-16\n
 \nХотын
 Яармагт байрлалтай 59 сургуулийн хичээл эхлэх цагт өөрчлөлт оруулах...\n\n\n
 274\n
 \n\n 2022-02-18\n
 \nӨөртөө
 хүрч чадахаа сайн мэддэг. Тиймдээ ч тавьсан зорилгодоо тууштай байж...\n\n\n
 516\n
 \n\n 2022-02-11\n
 \nМэргэж
 арын 7-нд цахимаар, Ерөнхий боловсролын сургуулийн хичээл 14-нөөс 5...\n\n\n
 1288\n
 \n\n 2022-01-27\n
 \nБСШУЯ-
 раах байдлаар эцэслэн шийдвэрлэсэн.\n\n\n
 \n\n\n 299072\n
 \n\n
 \nЧи ямар мэргэжил эзэмшихээ шийдсэн үү. Хэт сэтгэлийн хөөрлөөр нэг мэргэжил сонгодог ч
 37386\n
 \n\n\n 2016-06-22\n
 \nӨнөө
 яах аргагүй хэлний мэдлэг ордог болсон цаг үе. Тиймээс ч нийгэмтэ...\n\n\n
 33637\n
 \n\n\n 2020-09-07\n
 \nIELTS-

▼ Daalgavar 2

```
!pip install pdfminer.six
!pip install PyPDF2
```

```
Collecting pdfminer.six
  Downloading pdfminer.six-20211012-py3-none-any.whl (5.6 MB)
    |████████████████████████████████████████| 5.6 MB 5.0 MB/s
Collecting cryptography
  Downloading cryptography-36.0.1-cp36-abi3-manylinux_2_24_x86_64.whl (3.6 MB)
    |████████████████████████████████████████| 3.6 MB 41.0 MB/s
Requirement already satisfied: chardet in /usr/local/lib/python3.7/dist-packages (from pdfminer.six)
Requirement already satisfied: cffi>=1.12 in /usr/local/lib/python3.7/dist-packages (from cryptography)
Requirement already satisfied: pycparser in /usr/local/lib/python3.7/dist-packages (from cryptography)
Installing collected packages: cryptography, pdfminer.six
Successfully installed cryptography-36.0.1 pdfminer.six-20211012
Collecting PyPDF2
  Downloading PyPDF2-1.26.0.tar.gz (77 kB)
    |████████████████████████████████████████| 77 kB 3.0 MB/s
Building wheels for collected packages: PyPDF2
  Building wheel for PyPDF2 (setup.py) ... done
```

```
Created wheel for PyPDF2: filename=PyPDF2-1.26.0-py3-none-any.whl size=61102 sha256=a0
Stored in directory: /root/.cache/pip/wheels/80/1a/24/648467ade3a77ed20f35cfd2badd3213
Successfully built PyPDF2
Installing collected packages: PyPDF2
Successfully installed PyPDF2-1.26.0
```

1. glob ашиглаж фолдер доторх файлууд нэвтэрнэ.
2. PDFParser - файлаа парс хийнэ.
3. TextConverter - парс хийгдсэн файлаа текст хэлбэрлүү хувиргана.
4. Хуудас бүрийг нь хөрвүүлнэ. Үр дүнг result-д хадгална.
5. result-г файлд хадгална.

```
import glob

import re

from io import StringIO

from pdfminer.converter import TextConverter
from pdfminer.layout import LAMParams
from pdfminer.pdfdocument import PDFDocument
from pdfminer.pdfinterp import PDFResourceManager, PDFPageInterpreter
from pdfminer.pdfpage import PDFPage
from pdfminer.pdfparser import PDFParser
```

```
root_dir = '/content/'

for filename in glob.iglob(root_dir + '**/*.pdf', recursive=True):
    print(filename)

/content/Biy-daalt.pdf
/content/sample_data/СЕМИНАРИЙН АЖЛЫН ТАЙЛАН бичих загвар.pdf
```

```
root_dir = '/content/'

result = ''

for filename in glob.iglob(root_dir + '**/*.pdf', recursive=True):
    output_string = StringIO()

    with open(filename, 'rb') as in_file:
        parser = PDFParser(in_file)
        doc = PDFDocument(parser)
        rsrcmgr = PDFResourceManager()
        device = TextConverter(rsrcmgr, output_string, laparams=LAMParams())
```

```
interpreter = PDFPageInterpreter(rsrcmgr, device)
for page in PDFPage.create_pages(doc):
    interpreter.process_page(page)
    result += '\n' + output_string.getvalue()

with open('output.txt', 'w', encoding='utf-8') as out_file:
    out_file.write(result)
```

```
print(result)
```

ШИНЖЛЭХ УХААН ТЕХНОЛОГИЙН ИХ СУРГУУЛЬ
Мэдээлэл, холбооны технологийн сургууль

Шалгасан багш:

Гүйцэтгэсэн:

Бие даалтын ажил 1
Компьютерийн ухааны үндэс (F.CS212)
2021-2022 хичээлийн жилийн намрын улирал

маг. Х.Хулан

О.нэр: Ц.Санчир О.код: B200900046

М.Төрбаяр

B200900041

Б.Анужин

B2009000802

Б.Дарамрагчаа

B200900043

Сем.цаг:Мягмар 5

Багийн дугаар: 3

Улаанбаатар хот 2021 он

Зэрэгцээ тооцоолол (Parallel computing) ба Тархсан тооцоолол (Distributed computing)

- Танилцуулга
- Зэрэгцээ тооцоолол гэж юу вэ?
- Яагаад зэрэгцээ тооцоолох хэрэгтэй вэ?
- Зэрэгцээ тооцооллын программчлалын онцлог

- Параллель тооцоолох төрлүүд
- Тархсан тооцоолол гэж юу вэ?
- Тархсан системийн давуу тал болон сул тал
- Client/Server системүүд
- Зэрэгцээ тооцоолол болон тархсан тооцооллын ялгаа
- Жишээ:ЗЭРЭГЦЭЭ ТООЦООЛЛЫН ЗАРЧИМД ТУЛГУУРЛАСАН ГАЛТ ТЭРЭГНИЙ ХЯНАЛТ УДИРДЛАГЫН СИС
- Ашигласан материал

✓ 0s completed at 1:56 PM

