In [ ]:
```
# import requests
# from bs4 import BeautifulSoup
```

In [ ]:
```
# import urllib

# page_url = 'http://tabrizu.ac.ir'
# page_content = urllib.request.urlopen(page_url).read()
# print(page_content)
```

In [ ]:
```
# import urllib.request
# with urllib.request.urlopen('http://python.org/') as response:
#    html = response.read()
# print(html)
```

In [1]:
```
import urllib.request

page_content = urllib.request.urlopen('http://tabrizu.ac.ir').read().decode('utf-
print(page_content)
```


```
tle":"موسسه علوم اسلامی - انسانی"},{"link":"http://mmtaz.tabrizu.ac.ir/fa","titl
e":"موسسه مطالعات آذربایجان شناسی"},{"link":"http://phtr.tabrizu.ac.ir/fa","title":"ن
خستین همایش ترجمه در حوزه فلسفه، کلام، ادیان وعرفان"},{"link":"http://leader.tabrizu.ac.i
r/fa","title":"نهاد رهبری"},{"link":"http://slaa9.tabrizu.ac.ir/fa","title":"نهم
ین سمینار جبرخطی و کاربردهای آن"},{"link":"http://gozinesh.tabrizu.ac.ir/fa","titl
e":"هسته گزینش دانشگاه"},{"link":"http://70th.tabrizu.ac.ir/fa","title":"هفتادمین سالگرد
تاسیس دانشگاه تبریز "},{"link":"http://cir.tabrizu.ac.ir/fa","title":"همایش انقلاب اسلام
ی، کارآمدی،فرصت ها و چالش ها"},{"link":"http://cbmc.tabrizu.ac.ir/fa","title":"همایش
مجازی مدیریت کسب و کار "},{"link":"http://precisionagriculture2017.tabrizu.ac.ir/
fa","title":"همایش کشاورزی دقیق"},{"link":"http://studentfinance.tabrizu.ac.ir/f
a","title":"واحد مالی دوره شبانه"},{"link":"http://arasp.tabrizu.ac.ir/fa","titl
e":"پردیس های دانشگاه تبریز"},{"link":"http://riapa.tabrizu.ac.ir/fa","title":"پژوهش
کده فیزیک کاربردی و ستاره شناسی"},{"link":"http://icce2018.tabrizu.ac.ir/fa","titl
e":"چهارمین کنفرانس مهندسی مخابرات ایران"},{"link":"http://lib.tabrizu.ac.ir/fa","titl
e":"کتابخانه مرکزی و مرکز اسناد"},{"link":"http://hse.tabrizu.ac.ir/fa","title":"کمی
ته ایمنی، بهداشت و محیط زیست"},{"link":"http://4thcongress.tabrizu.ac.ir/fa","titl
e":"کمیسیون اقتصاد اسلامی چهارمین کنگره علوم انسانی و اسلامی"},{"link":"http://ngtu.tabrizu.
ac.ir/fa","title":"کنفرانس ملی فن آوری ها و کاربردهای نوین ژئوماتیک"},{"link":"http://pb
s2019.tabrizu.ac.ir/fa","title":"کنگره تازه های روان شناسی و علوم رفتاری"},{"link":"h
ttp://acer.tabrizu.ac.ir/fa","title":"گروه تحقیقاتی سرامیکهای مهندسی پیشرفته"},{"lin
```


Get the First link

In [2]:
```
start_link = page_content.find('<a href=')
start_quote = page_content.find('"',start_link)
end_quote = page_content.find('"',start_quote + 1)
url = page_content[start_quote + 1:end_quote]
print(url)
```

```
http://tabrizu.ac.ir/en (http://tabrizu.ac.ir/en)
```

Get the 2nd link

In [3]:
```
page_content = page_content[end_quote:]
print(page_content)
```

```
"> English</a></li>
                        </ul>
</div>


                </div>
                <div class="lang-box col-md-7 col-sm-8 col-xs-6">
                                    <a href="http://tabrizu.ac.ir/e
n">English</a>

                                </div>
                <div class="col-md-5 col-sm-4 col-xs-6">
                    <div class="row">
                    <div class="col-sm-10 col-xs-6 hidden-xs ">
                        <div class="DateBar">
                            ۰۷ بهمن ۱۳۹۷                            </div>
                    </div>
                    <div class="col-sm-2 col-xs-12">
                        <div class="search-bar-wrapper">
```

In [4]:
```
start_link = page_content.find('<a href=')
start_quote = page_content.find('"',start_link)
end_quote = page_content.find('"',start_quote + 1)
url = page_content[start_quote + 1:end_quote]
print(url)
```

```
http://tabrizu.ac.ir/en (http://tabrizu.ac.ir/en)
```

Get the 3th link

In [5]:
```
page_content = page_content[end_quote:]

start_link = page_content.find('<a href=')
start_quote = page_content.find('"',start_link)
end_quote = page_content.find('"',start_quote + 1)
url = page_content[start_quote + 1:end_quote]
print(url)
```

```
http://tabrizu.ac.ir/fa/page/6/%D9%85%D8%B9%D8%B1%D9%81%DB%8C-%D9%88-%D8%AA%D8%
A7%D8%B1%DB%8C%D8%AE%DA%86%D9%87 (http://tabrizu.ac.ir/fa/page/6/%D9%85%D8%B9%D
8%B1%D9%81%DB%8C-%D9%88-%D8%AA%D8%A7%D8%B1%DB%8C%D8%AE%DA%86%D9%87)
```

Get the 4th link

In [6]:
```
page_content = page_content[end_quote:]

start_link = page_content.find('<a href=')
start_quote = page_content.find('"',start_link)
end_quote = page_content.find('"',start_quote + 1)
url = page_content[start_quote + 1:end_quote]
print(url)
```

```
http://tabrizu.ac.ir/fa/page/8/%D8%B1%D9%88%D8%B3%D8%A7%DB%8C-%D8%B3%D8%A7%D8%A8%D9%82 (http://tabrizu.ac.ir/fa/page/8/%D8%B1%D9%88%D8%B3%D8%A7%DB%8C-%D8%B3%D8%A7%D8%A8%D9%82)
```

what if there are 100 or more links?

by using functions

In [8]:
```
def get_next_target(page_content):
    '''
    find the next url in the page
    '''
    start_link = page_content.find('<a href=')
    start_quote = page_content.find('"',start_link)
    end_quote = page_content.find('"',start_quote + 1)
    url = page_content[start_quote + 1:end_quote]

    return url, end_quote
```