Marist Digital Education Workshop
Web Scrapping Demo

Objective:

- To scan through the contents of the aforementioned website link and retrieve all the workshop information which can be mailed later to the college's faculties.

Features:

- Scrap the workshop information
- Mail the scrapped information to the faculties

Future Tweaks – On Progress:

- Increase the program ability to recognize the desired content info precisely
- Create a website so that this can be managed easily
- Send mails automatically based on the meeting status – available, cancelled or rescheduled

Project Repo Location:

github.com/vivekVells/WebScrapMailer/tree/master/MaristDigitalEducationWorkshop

Vivek Vellaiyappan Surulimuthu

File last updated: March 27th 2018

Website Page (as of March 27th 2018): https://www.marist.edu/it/digitaleducation/workshops.html

MARIST

Search...

Digital Education
CTE Workshops
Digital Resources
Innovation
Certification
Faculty Showcases
Services & Support

Contact:

Digital Education
845.575.3252
For Assistance: x4357
Room 305, 3rd Floor, Library

Workshops

All workshops are available either face-to-face or through a webinar. Register for workshops by clicking the links below. Our workshops are broken into certificate tracks. Below, you see the next available workshop date and time for each workshop in each certificate track.

Note: If you are unable to log-in to the registration site or have any questions, please email to Digital.Education@Marist.edu

Student Engagement	Blended Classroom	iLearn Level 1	iLearn Level 2	Open Labs + New Tools	E3 - Enhance, Engage, Empower *
Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register Here
	Collaboration through iLearn March 22nd, 11:00 AM - 12:00 PM	Applying the Quality Matters Rubric March 28th, 11:00 AM - 12:00 PM	Optimizing WebEx in your Class April 16th, 10:00 AM - 11:00 AM	Open Labs in March Every Tuesday from 10:00 AM - 12:00 PM & Every Thursday from 2:00 PM - 4:00 PM	All week online, interactive course Next session begins on TBD
	Optimize your Face to Face Time April 27th, 12:00 PM - 1:00 PM	Design Assessments in iLearn April 4th, 11:00 AM - 12:00 PM	Quality Matters - Dig Deeper April 24th, 2:00 PM - 3:00 PM	Open Labs in April Every Tuesday from 12:00 PM - 2:00 PM & Every Friday from 10:00 AM - 12:00 PM	
			Mobile Apps for Education April 25th, 11:00 AM - 12:00 PM		

* Note - E3 - Enhance, Engage, Empower certificates will need to be refreshed every 2 years.

The workshop information we are interested in

Student Engagement	Blended Classroom	iLearn Level 1	iLearn Level 2	Open Labs + New Tools	E3 - Enhance, Engage, Empower *
Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register Here
	Collaboration through iLearn March 22nd, 11:00 AM - 12:00 PM	Applying the Quality Matters Rubric March 28th, 11:00 AM - 12:00 PM	Optimizing WebEx in your Class April 16th, 10:00 AM - 11:00 AM	Open Labs in March Every **Tuesday** from 10:00 AM - 12:00 PM & Every **Thursday** from 2:00 PM - 4:00 PM	All week online, interactive course Next session begins on TBD
	Optimize your Face to Face Time April 27th, 12:00 PM - 1:00 PM	Design Assessments in iLearn April 4th, 11:00 AM - 12:00 PM	Quality Matters - Dig Deeper April 24th, 2:00 PM - 3:00 PM	Open Labs in April Every **Tuesday** from 12:00 PM - 2:00 PM & Every **Friday** from 10:00 AM - 12:00 PM	
			Mobile Apps for Education April 25th, 11:00 AM - 12:00 PM		

Objective:
- We are interested to retrieve the workshop information alone
- To retrieve the workshop
  - Title
  - Date & Timing information

For Example:

- **Workshop Title:** Mobile Apps for Education
- **Scheduled Info:** April 25th, 11:00 AM - 12:00 PM

Program Execution Output

Scrapper (1)

Project

- WebScrapMailer E:\kevDev\ProjectWork
  - gitHub
    - MaristDigitalEducationWorkshop
      - demo
      - utility
        - __init__.py
        - Scheduler.py
        - SendEmail.py
      - webscrap
        - __init__.py
        - Scrapper.py
    - README.md
  - External Libraries

Scrapper.py

```python
from gitHub.MaristDigitalEducationWorkshop.utility import SendEmail

def scrapDigitalEd():
    webSite = "https://www.marist.edu/it/digitaleducation/workshops.html"
    # request used here to retrieve entire content of the website arg
    # type of parser should be mentioned whenver bs4 used
    soup = bs4.BeautifulSoup(requests.get(webSite).content, "html.parser")
    # retrieving web contents that have <tr>
    rows = soup.findAll('tr')
    # print(rows)

    workshopPairs = {}
    for index in range(len(rows)):
        if index > 1:
            # tdTags contain all the contents that have '<p>' tag associat
            tdTags = bs4.BeautifulSoup.findAll(rows[index], 'p')
            # obtained in tdTags above => <p><b>Collaboration through iLea
            #for tdTag in tdTags:
                #print("tdTag: ", tdTag, end="\n\n")


            for tdTag in tdTags:
                # .text used here to extract only the text content excludi
                getInfo = str(BS(str(tdTag), "html.parser").text)
                getTitle = str(BS(str(BS(str(tdTag), "html.parser").find('
                # print("getInfo: ", getInfo)
                # print("getTitle: ", getTitle)
                getDurationInfo = getInfo.replace(getTitle, "")
                # print("getDurationInfo: ", getDurationInfo)

                '''

                getInfo:  Mobile Apps for EducationApril 25th, 11:00 AM - 
                getTitle:  Mobile Apps for Education
                getDurationInfo:  April 25th, 11:00 AM - 12:00 PM
                '''


                if (len(getTitle) > 0) & (getTitle != 'None'):
```

Run: Scrapper (1)

```
C:\Developer\Python36\python.exe
E:/kevDev/ProjectWorks/WebScrapMailer/gitHub
/MaristDigitalEducationWorkshop/webscrap/Scrapper
.py
{'All week online, interactive courseNext
 session begins on': '',
 'Applying the Quality Matters Rubric': '',
 'Collaboration through iLearn': '',
 'Design Assessments in iLearn': '',
 'Friday': 'Every  from 10:00 AM - 12:00 PM',
 'Mobile Apps for Education': '',
 'Open Labs inApril': '',
 'Open Labs inMarch': '',
 'Optimize your Face to Face Time': '',
 'Optimizing WebEx in your Class': '',
 'Quality Matters - Dig Deeper': '',
 'TBD': '',
 'Thursday': 'Every  from 2:00 PM - 4:00 PM',
 'Tuesday': 'Every  from 12:00 PM - 2:00 PM'}
Email sent!

Process finished with exit code 0
```

6: TODO   9: Version Control   Python Console   Terminal   Event Log

Mail Reception Output

askkeviv@gmail.com   6:00 PM (0 minutes ago) ☆ ↩ ▾

to ▾

\

From: askkeviv@gmail.com
To: techengineervivek@gmail.com
Subject: CTE Workshop Meeting Information - Notification

{'Collaboration through iLearn': '', 'Applying the Quality Matters Rubric': '', 'Optimizing WebEx in your Class': '', 'Open Labs inMarch': '', 'Tuesday': 'Every from 12:00 PM - 2:00 PM', 'Thursday': 'Every from 2:00 PM - 4:00 PM', 'All week online, interactive courseNext session begins on': '', 'TBD': '', 'Optimize your Face to Face Time': '', 'Design Assessments in iLearn': '', 'Quality Matters - Dig Deeper': '', 'Open Labs inApril': '', 'Friday': 'Every from 10:00 AM - 12:00 PM', 'Mobile Apps for Education': ''}

Thanks & regards,
Digitial Education Representative
Marist College
XXXX-XXX-XXX

Acutal Website Content Screenshot VS Program Execution Output

Actual Website Content Screenshot

Student Engagement	Blended Classroom	iLearn Level 1	iLearn Level 2	Open Labs + New Tools	E3 - Enhance, Engage, Empower *
Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register for Workshops Here	Register Here
	Collaboration through iLearn March 22nd, 11:00 AM - 12:00 PM	Applying the Quality Matters Rubric March 28th, 11:00 AM - 12:00 PM	Optimizing WebEx in your Class April 16th, 10:00 AM - 11:00 AM	Open Labs in March Every Tuesday from 10:00 AM - 12:00 PM & Every Thursday from 2:00 PM - 4:00 PM	All week online, interactive course Next session begins on TBD
	Optimize your Face to Face Time April 27th, 12:00 PM - 1:00 PM	Design Assessments in iLearn April 4th, 11:00 AM - 12:00 PM	Quality Matters - Dig Deeper April 24th, 2:00 PM - 3:00 PM	Open Labs in April Every Tuesday from 12:00 PM - 2:00 PM & Every Friday from 10:00 AM - 12:00 PM	
			Mobile Apps for Education April 25th, 11:00 AM - 12:00 PM		

Program Execution Output

```
Scrapper

C:\Developer\Python36\python.exe
E:/kevDev/ProjectWorks/WebScrapMailer/gitHub
/webscrap/Scrapper.py
{'All week online, interactive courseNext session
 begins on': '',
 'Applying the Quality Matters Rubric': '',
 'Collaboration through iLearn': '',
 'Design Assessments in iLearn': '',
 'Friday': 'Every  from 10:00 AM - 12:00 PM',
 'Mobile Apps for Education': '',
 'Open Labs inApril': '',
 'Open Labs inMarch': '',
 'Optimize your Face to Face Time': '',
 'Optimizing WebEx in your Class': '',
 'Quality Matters - Dig Deeper': '',
 'TBD': '',
 'Thursday': 'Every  from 2:00 PM - 4:00 PM',
 'Tuesday': 'Every  from 12:00 PM - 2:00 PM'}
Email sent!

Process finished with exit code 0
```

Inference:
- Able to retrieve the workshop title and scheduled information; but, some contents are not proper.
- The content mismatch is because **the website html tags usage are inconsistent**

Thank you so much for viewing this working demo sheet.
Hope you got an idea of how this application works.
Highly appreciate the feedbacks if any.

Thanks & Regards,
Vivek Vellaiyappan Surulimuthu
https://github.com/vivekVells
vivekvellaiyappans@gmail.com