Inheritance

Extending Students

In last week's tutorial, we wrote a Student class which aimed to store, retrieve, and perform computations on, information about a university student. The file students.py contains a much simpler model of a student; it includes a name, student number, and a list of courses. Each course has an associated tuition fee.

We would like to extend this model to allow for students who live in college. In addition to paying tuition fees, college students must also pay an extra fee to live on campus. Write a class CollegeStudent which inherits from Student, and implements the following functionality:

- The constructor should take an extra two arguments, the name of the student's college (as a string), and their college fee (as an integer).

- Add a method get_college which returns the name of the student's college.

- The calculate_fees method of a CollegeStudent should return the sum of all their tuition fees, plus their college fee.

The following shows an example of using this class:

```
>>> fred = CollegeStudent("fred", 43214321, "St. Leo's", 18000)
>>> fred.get_college()
"St. Leo's"
>>> fred.calculate_fees()
18000
>>> fred.enrol('CSSE1001', 1000)
>>> fred.enrol('LAWS1000', 1300)
>>> fred.calculate_fees()
20300
```

Webcrawling

If you are building a search engine, you need to start by having a list of all web pages on the Internet. The simple way to find webpages is to start with a known web page, and follow all the links to different pages; repeating the procedure on those web pages. (This method is called web crawling, since it is akin to the way a spider crawls along its web.)

For this task, we will write a program which takes a URL for a Web page, and finds all links on that page. To do this, we'll need to learn a bit about HTML, and the Python HTMLParser.

HTML Basics

In HTML, text is annotated with *tags*, which affect the way the text is displayed. For example, to make text appear in **bold face**, we put it inside `<strong>` tags. To see more tags, view the source of any web page in your browser.

Tags can have attributes; attributes have a name and a value. For example, the tag `<img src="img/spam.jpg" alt="Spam">` is an img tag with the two attributes `src="img/spam.jpg"` and `alt="Spam"`.

Of particular interest to us is the `a` tag, which describes a link. The URL to be linked is put in the attribute called `href`. For example, `<a href="http://example.com/">link</a>` produces: link.

HTMLParser

Python contains a library called HTMLParser which can help interpret HTML text. The HTMLParser class has a method feed, which takes a string. For each HTML tag within the string, the class calls its own handle_starttag method. By default, this method does nothing, but subclasses can overwrite that behaviour.

In all the examples of inheritance we have seen, we are defining two (or more) classes in the same file, and have one inherit from the other. In this task, we will instead be inheriting from a class in another library; this works in exactly the same way — we write the name of the superclass (HTMLParser) inside () parentheses next to the class definition.

Since we cannot see the code we are inheriting from, we have to be careful when overriding methods to call the superclass method, including for the __init__ method.

Start with the sample code below (available as webcrawler.py). The function find_links takes a URL (for example, the URL of this webpage), and calls the feed method of a LinkParser class (which has not yet been defined).

```python
from html.parser import HTMLParser
import urllib
import urllib.request

# Write the LinkParser class here

def find_links(url) :
    """Return a list of links from the given Web page.

    Return:
        (list[str]): List of all links found at the given URL.
    """
    # Open the Web page and read the HTML text
    fd = urllib.request.urlopen(url)
    text = fd.read()
    fd.close()

    # Create a parser instance and feed it the text
    parser = LinkParser()
    parser.feed(str(text))   # Need to convert text to a str as read gives a
                             # bytes type which feed cannot process

    # Write a return statement here
```

Make the following modifications to this code:

1. Write a class LinkParser which inherits from HTMLParser, and write a method handle_starttag(self, tag, attrs) which prints out the two parameters tag and attrs. Run the code and determine what the code is doing, and what values are being given to handle_starttag when it is called. Discuss with others at your table, or with a tutor.

2. Modify the handle_starttag method to print out the href attribute of every link (a tag) within the HTML text.

3. Modify the class so that it stores a list of URLs, and modify the handle_starttag method so that it adds the linked URLs into the list. Add a method get_urls which returns this list, and use that method in the return value of find_links.

   (Hint: think carefully about where you should create this list, so that you will be able to access it again.)

An example of executing the finished code is as follows:

```
>>> find_links('http://www.itee.uq.edu.au/')
['https://www.uq.edu.au/contacts/',
 'https://future-students.uq.edu.au/',
 'https://www.uq.edu.au/maps/',
 'https://www.uq.edu.au/news/',
 'https://www.uq.edu.au/events/',
 'http://jobs.uq.edu.au/',
 'https://library.uq.edu.au/',
 'https://my.uq.edu.au/',
 'http://www.uq.edu.au/',
 '/',
 'https://www.facebook.com/UQEngTech/',
 'https://www.instagram.com/uqengineeringtech/',
 'https://twitter.com/UQ_EAIT',
 'https://www.linkedin.com/groups/4123315',
 'https://www.youtube.com/channel/UCn7-OSiaW162YraLKzs4fzA/videos',
 '/',
 '/about',
 '/occupational-health-and-safety',
 '/facilities',
 'http://www.itee.uq.edu.au/etsg',
 '/social-media',
 '/news',
 '/our-people',
 '/future-students',
 '/itee-undergraduate-programs',
 '/first-year',
 '/itee-careers',
 '/grad-cert',
 '/grad-dipl',
 '/itee-honours',
 '/itee-industry-experience',
 '/internships',
 '/postgraduate-coursework',
 '/future-postgraduate-research',
 '/publications',
 '/advanced-programs',
 '/future-international',
 '/future-scholarships',
 'http://www.eait.uq.edu.au/why-study-engineering-uq',
 '/why-uq',
 '/current-students',
 '/student-faqs-general',
 '/academic-advice',
 '/student-resources',
 '/itee-guidelines',
 '/itee-undergraduate-programs',
 '/postgraduate-coursework',
 '/rhd-students',
 '/scc',
 '/study-overseas',
 '/courses',
```

```
  '/tutoring',
  '/thesis',
  '/career-resources-and-jobs',
  '/research',
  '/itee-school-based-ethics-review',
  '/seminars',
  '/industry',
  '/student-engagement',
  '/industry-advisory-board',
  '/rd-collaboration',
  '/consultancy-services',
  '/shared-facilities',
  '/commercial-ventures',
  '/donations',
  '/community',
  '/activities-outreach',
  '/contact-us',
  '/our-people',
 'https://future-students.uq.edu.au/study/find-a-program/studyarea/engineering-and
-information-technology-1581?career=%5B%22ugrd%22%5D',
  'https://future-students.uq.edu.au/study/find-a-program/studyarea/engineering-and
-information-technology-1581?career=%5B%22pgcw%22%5D',
  'https://future-students.uq.edu.au/study/find-a-program/studyarea/engineering-and
-information-technology-1581?career=%5B%22pgrs%22%5D',
  'https://future-students.uq.edu.au/study/program/Bachelor-of-Computer-Science-242
5',
  '/node/7060/',
  'http://eait.uq.edu.au/ingenuity',
  '/current-students',
  '/current-students',
  '/current-students',
  '/research',
  '/research',
  '/research',
  '/itee-industry',
  '/itee-industry',
  '/itee-industry',
  '/community',
  '/community',
  '/community',
  '/news',
  '/news/article/scholarships-awarded-commonwealth-games-competitors',
  '/news/article/scholarships-awarded-commonwealth-games-competitors',
  '/news/article/uq-aces-world-subject-rankings',
  '/news/article/uq-aces-world-subject-rankings',
  '/news/article/big-data-centre-energy-web',
  '/news/article/big-data-centre-energy-web',
  '/news/article/researchers-help-fight-rising-power-costs',
  '/news/article/researchers-help-fight-rising-power-costs',
  '/news/article/opie-visits-arnhem-land',
  '/news/article/opie-visits-arnhem-land',
  '/news/article/portable-3d-brain-scanner-set-save-lives',
  '/news/article/portable-3d-brain-scanner-set-save-lives',
  '/news/article/older-australians-shape-how-new-technologies-are-developed',
```

```
 '/news/article/older-australians-shape-how-new-technologies-are-developed',
 '/news/article/uq-wins-two-acs-digital-disruptors-awards',
 '/news/article/uq-wins-two-acs-digital-disruptors-awards',
 '/scholarships-giving',
 'http://www.uq.edu.au/maps/index.html?menu=1&id=160',
 'mailto:enquiries@itee.uq.edu.au',
 'mailto:studentenquiries@itee.uq.edu.au',
 'mailto:rhdstudentenquiries@itee.uq.edu.au',
 'http://eait.uq.edu.au/eng-academic-advice',
 'http://eait.uq.edu.au/',
 'http://eait.uq.edu.au/current-engineering-students',
 'http://future-students.uq.edu.au',
 'https://internal.itee.uq.edu.au/',
 'tel:+6173365111',
 'http://uq.edu.au/contacts',
 'http://www.uq.edu.au/about/cricos-link',
 'tel:+61733653333',
 'http://www.uq.edu.au/terms-of-use/',
 'http://www.uq.edu.au/feedback',
 '/eait_login']
```