Python Application-Layer Network Programming

CME451 Tutorial 2

Hao Zhang
(Graduate Teaching Fellow)

Department of Electrical & Computer Engineering
University of Saskatchewan

Jan 13, 2017

Review of Lab 1

- Basic concepts of Python programming:
  - data types, math operations, string...
- Extract a substring from an arbitrary string
  - using `str.find()`
- Sorting a list with specific rules
  - using `list.append()`
  - using `list.sort(key=rule,reverse=True)`

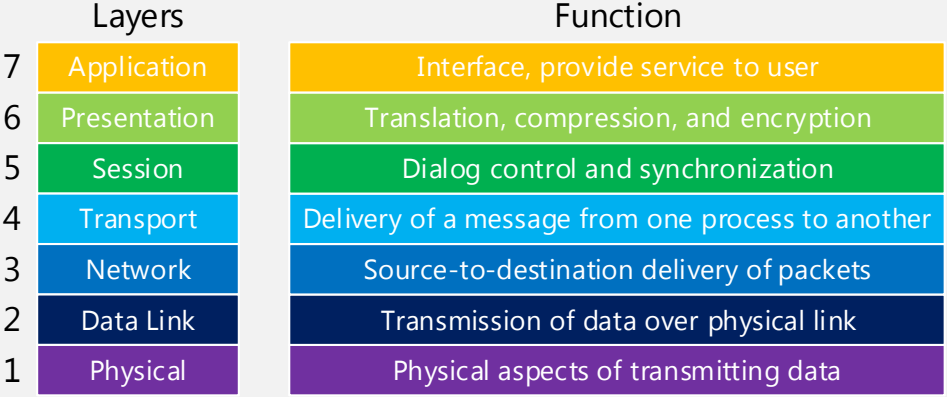
Lab 2 Objectives

Part 1: Application-Layer Programming

- Write simple application-layer program;
- Perform website data analysis;
- Retrieve information from a website.

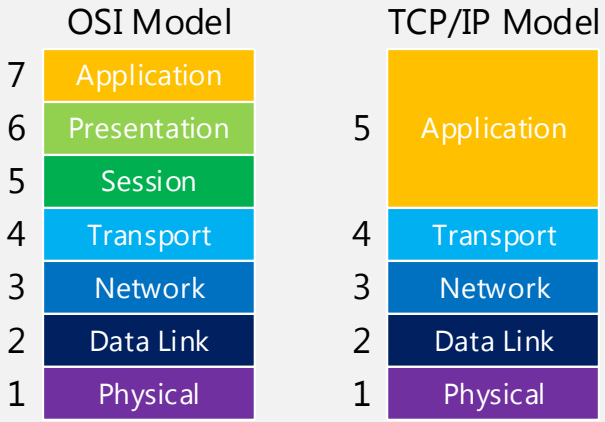
Network Layers

OSI model



Network Layers

TCP/IP model



Application-Layer Programming

Use urllib.request

- `urllib.request` module defines functions and classes that can be used in opening URLs (mostly HTTP)
- `urllib.request.urlopen(url)`
  - Open the specified URL and return a request object.
  - Acquire the content by the `read()` method, read content into a string.
  - Use decoding method `decode()` to make the content readable.
  - Analyze the website by processing the string.

```
>>> import urllib.request
>>> data = urllib.request.urlopen('http://www.usask.ca/')
>>> html_content = data.read().decode('utf-8')
>>> print(html_content)
```

- Analyze the HTML file content
- Processing HTML content as a 'large string'

```
>>> jpgIndex = html_content.find('.jpg')
>>> href = html_content.find('href=')
```

- `urllib.request.urlretrieve(url, filename)`
  - Copy a network object denoted by a URL to a local file.

```
>>> urllib.request.urlretrieve('http://www.usask.ca/img/'\
+filename, filename)
```

Lab 2 Task

- Extracting `.jpg` files from department website
- Retrieving documents from department website
- Part 2: Socket Programming