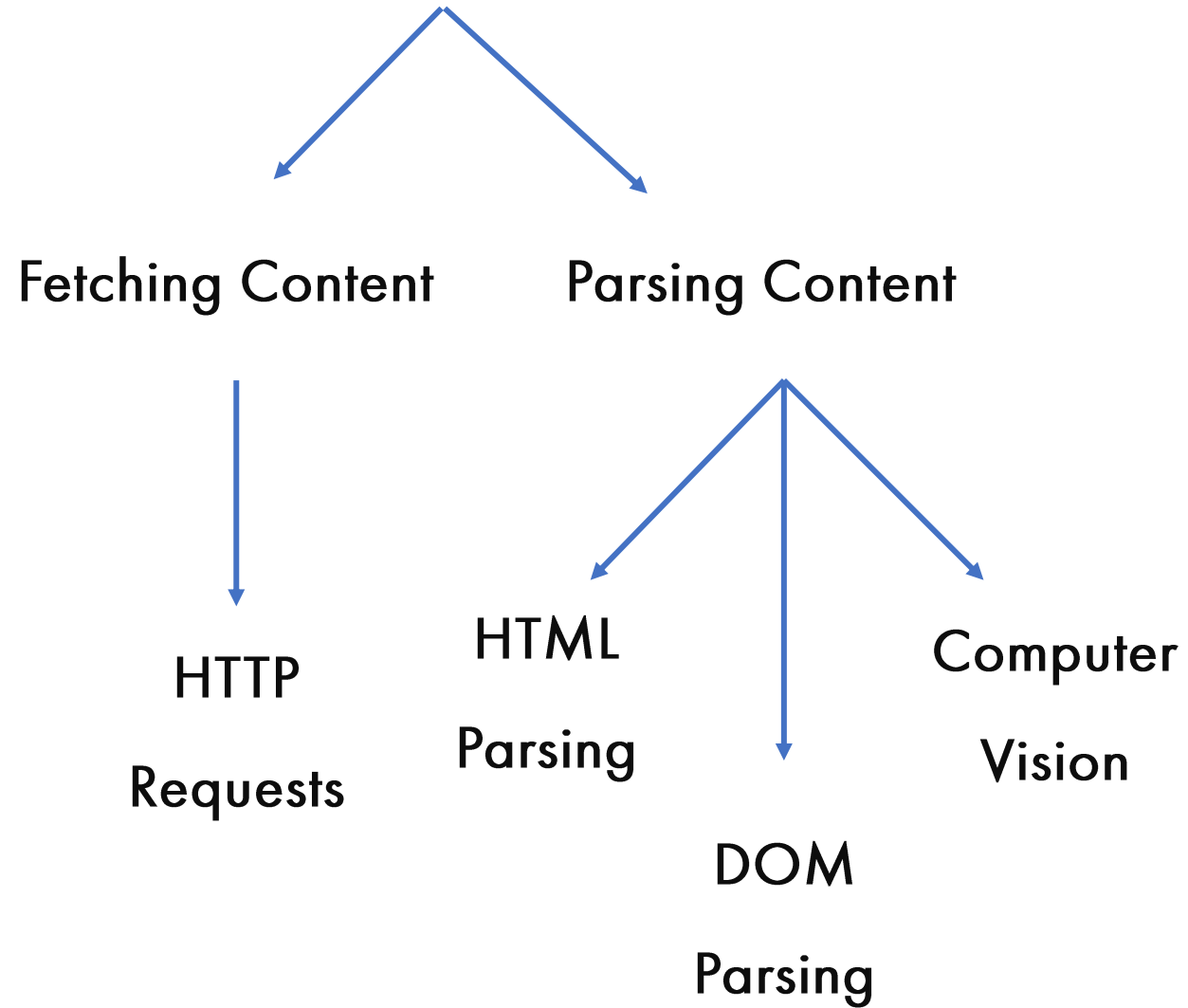


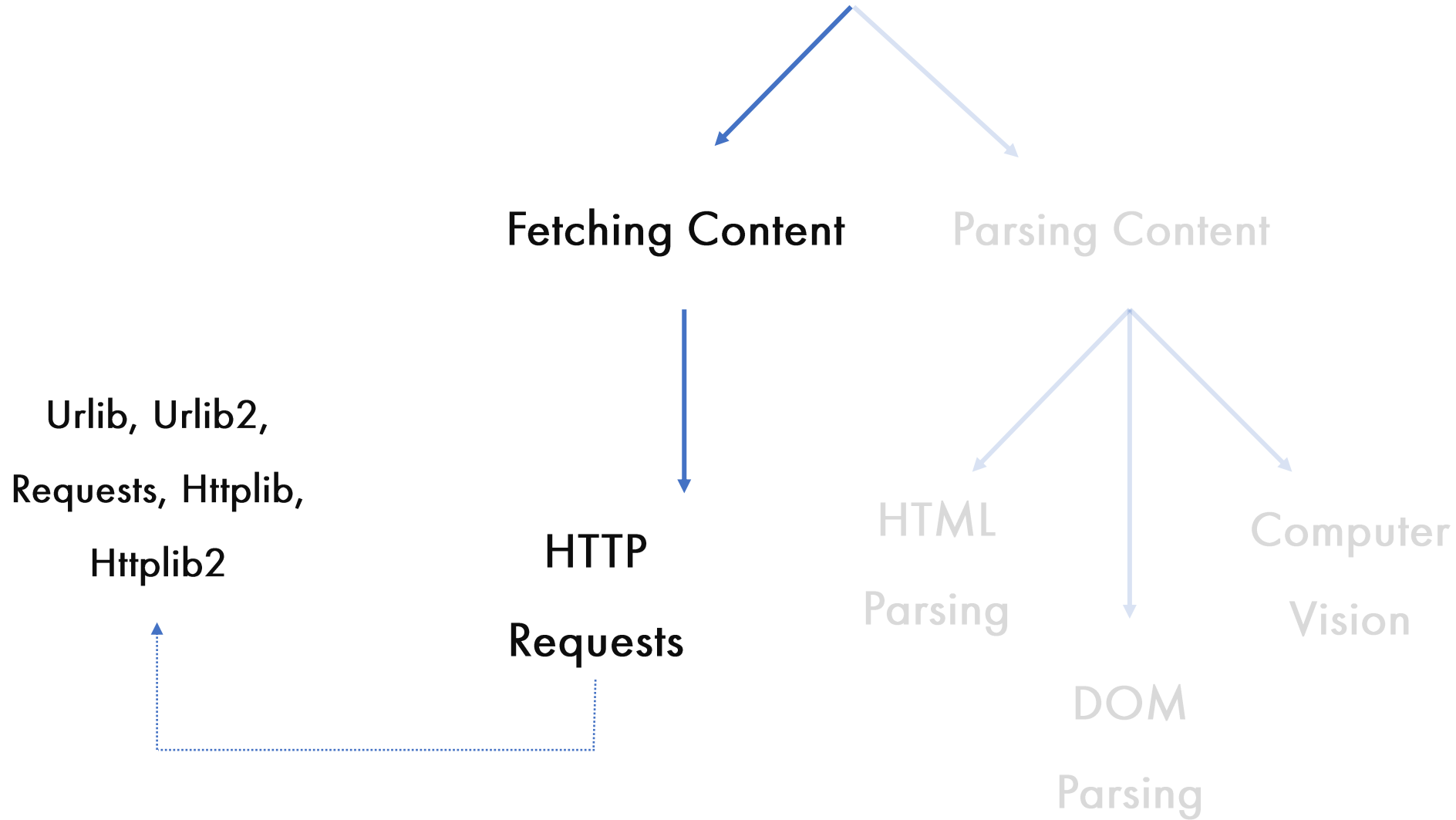
# Fetching and Parsing Content

---

# Web Scraping



# Web Scrapping



# Python HTTP Libraries

**Requests**

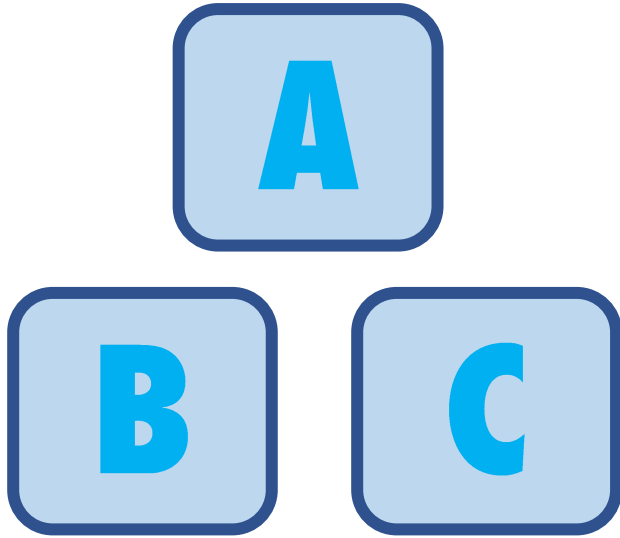
**Httplib2**

**Httpplib**

**Urllib**

**Urllib2**

# HTTP Libraries in Python



Requests

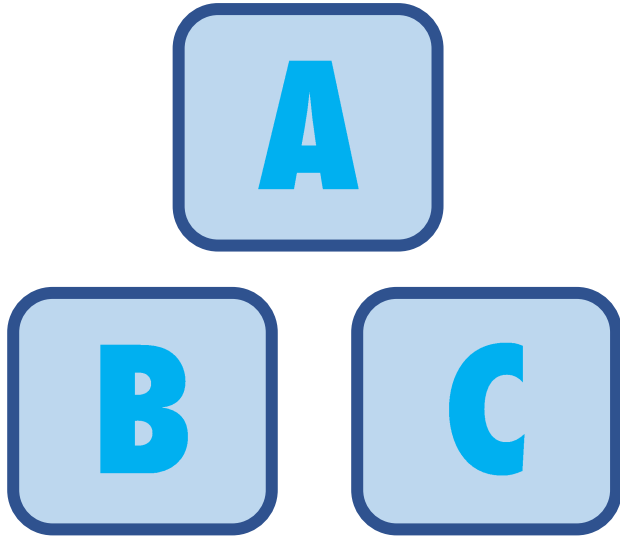
Httpplib2

- Httpplib

Urllib

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

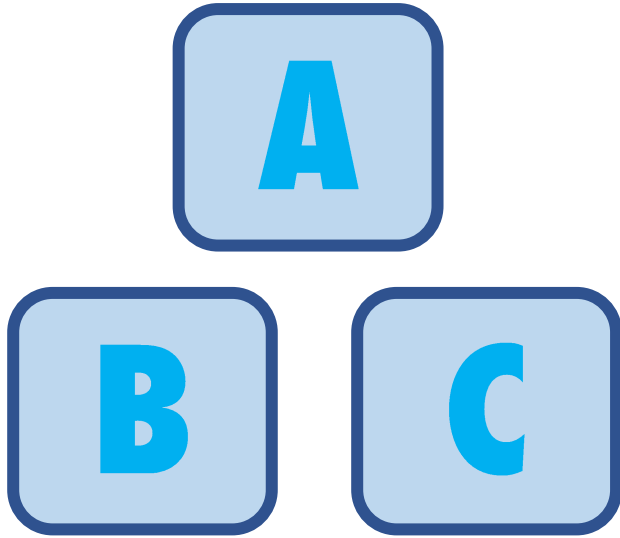
HttpLib2

- HttpLib

Urllib

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

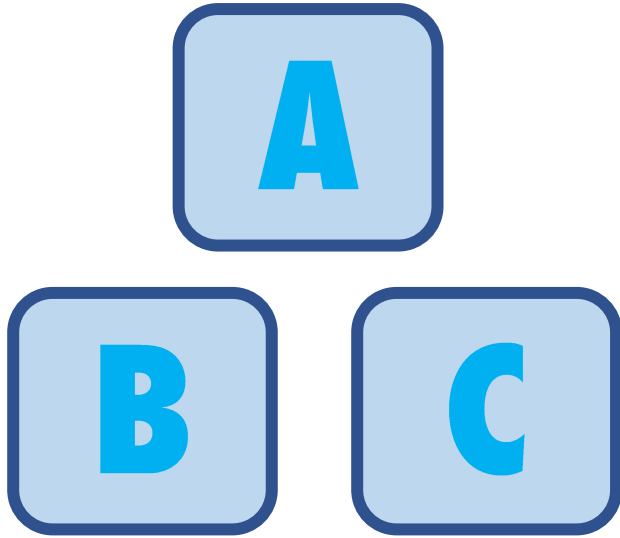
Httplib2: Google's version of Httplib

- Httplib

Urllib

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

Httplib2: Fine-grained control

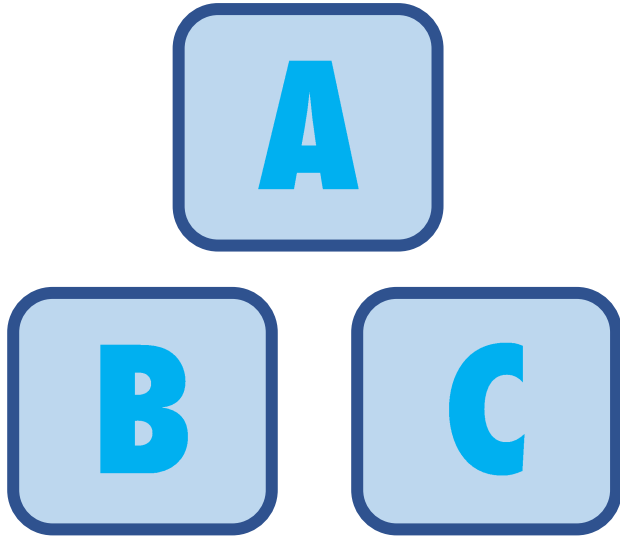
- Httplib

Urllib

- Urllib2



# HTTP Libraries in Python



Requests: Should be your default choice

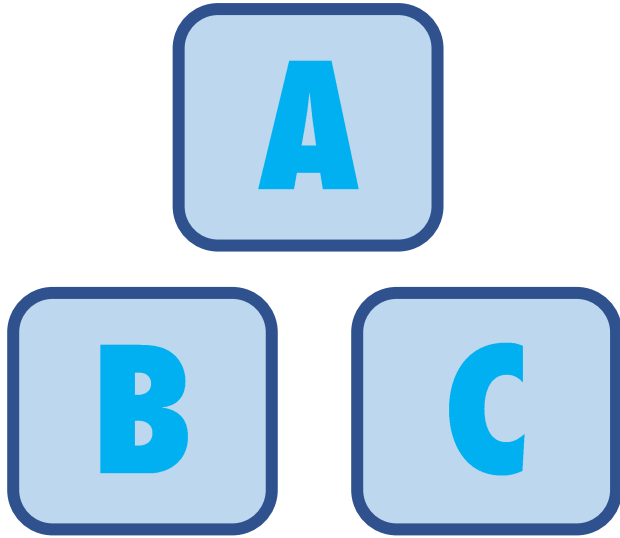
Httplib2: Fine-grained control

- ~~Httplib: Little reason to use directly~~

Urllib

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

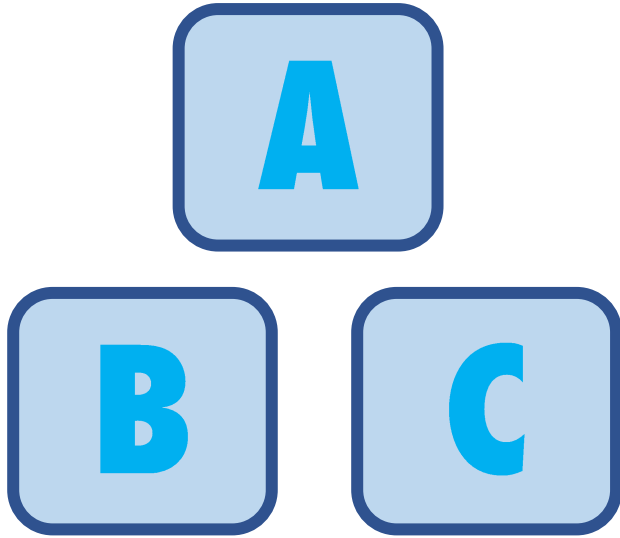
HttpLib2: Fine-grained control

- ~~HttpLib: Little reason to use directly~~

Urllib: Part of Python standard library

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

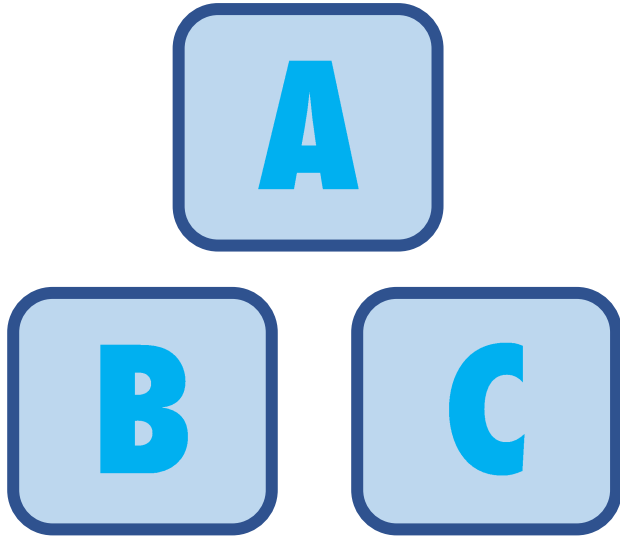
Httplib2: Fine-grained control

- ~~Httplib: Little reason to use directly~~

**Urllib: Four distinct namespaces**

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

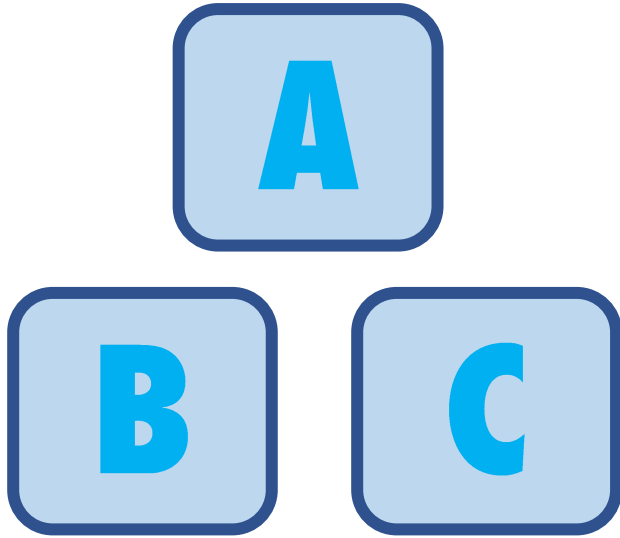
HttpLib2: Fine-grained control

- ~~HttpLib: Little reason to use directly~~

Urllib: In Python 3.x, subsumes old urllib2

- Urllib2

# HTTP Libraries in Python



Requests: Should be your default choice

Httplib2: Fine-grained control

- ~~Httplib: Little reason to use directly~~

Urllib: In Python 3.x, subsumes old urllib2

- ~~Urllib2: Python 2 only~~

# Fetching Web Content



Web servers make content available on HTTP endpoints

Browsers make HTTP request under-the-hood to get web pages

Web scraping usually involves making such requests programmatically

Many libraries and utilities available

# Fetching Web Content



## Command-line HTTP requests

- cURL

## Python libraries for programmatic access

- Requests
- Httpplib2
- Urllib