

# Getting started

Web scraping is where a programmer will write an application to download web pages and parse out specific information from them. Usually when you are scraping data you will need to make your application navigate the website programmatically. In this chapter, we will learn how to download files from the internet and parse them if need be. We will also learn how to create a simple spider that we can use to crawl a website.

## Tips for Scraping

There are a few tips that we need to go over before we start scraping.

- Always check the website's terms and conditions **before** you scrape them. They usually have terms that limit how often you can scrape or what you can scrape.
- Because your script will run much faster than a human can browse, make sure you don't hammer their website with lots of requests. This may even be covered in the terms and conditions of the website.
- You can get into legal trouble if you overload a website with your requests or you attempt to use it in a way that violates the terms and conditions you agreed to.
- Websites change all the time, so your scraper will break some day. Know this: You will have to maintain your scraper if you want it to keep working.
- Unfortunately the data you get from websites can be a mess. As with any data parsing activity, you will need to clean it up to make it useful to you.

With that out of the way, let's start scraping!

