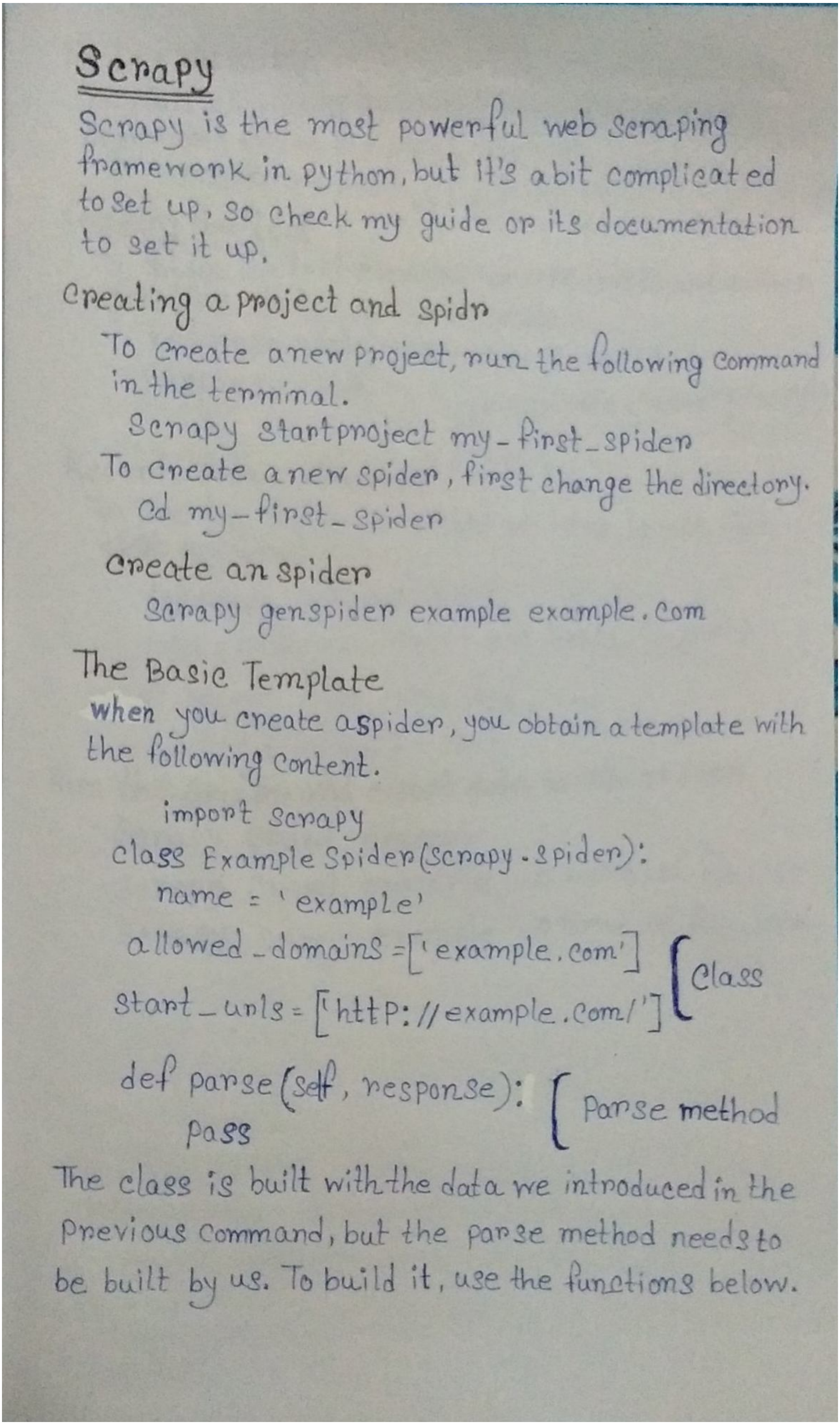

Scrapy

Scrapy is the most powerful web scraping framework in python, but it's abit complicated to set up, so check my guide or its documentation to set it up.

Creating a project and spidr

To create anew project, run the following command in the terminal.

```
Scrapy startproject my-first_spider
```

To create anew spider, first change the directory.

```
cd my-first_spider
```

Create an spider

```
Scrapy genspider example example.com
```

The Basic Template

when you create aspider, you obtain a template with the following content.

```
    import scrapy
class ExampleSpider(scrapy.spider):
    name = 'example'
  allowed_domains =['example.com']
 start_urls = ['http://example.com/']
```

Class

```
  def parse(self, response):
      pass
```

Parse method

The class is built with the data we introduced in the previous command, but the parse method needs to be built by us. To build it, use the functions below.

Finding elements

To find elements in scrapy, use the response argument from the parse method

```
response.xPath('//tag[@AttributeName="value"]')
```

Getting the text

To obtain the text element we usetext() and either .get() or .getall(). For example:

```
response.xpath('//h1/text()').get()
response.xPath('//tag[@Attribute="value"]/text()')
                    .getall()
```

Return data extracted

To see the data extracted we have to use the yield keyword

```
def parse(self, response):
title = response.xpath('//h1/text()').get()

# Return data extracted
 yield {'titles': title}
```

Run the spider and export data to csv or JSON

```
Scrapy crawl example
scrapy Crowl example -o name_of_file.csv
scrapy crowl example -o name_of_file.json
```