**User Agents, Headers, and Proxies**

**User Agents**

* A user agent identifies the scraper as a browser or tool. Websites use it to determine how to serve content. Its purpose is to make our scraper appear as a legitimate browser.

Example:

USER\_AGENT = 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/93.0 Safari/537.36'

**Request Headers**

* Headers provide extra context (e.g., cookies, referers) to mimic real browsing behavior. It helps avoid detection by including additional request details.

**Example**:

headers = {'User-Agent': '...', 'Referer': 'https://example.com'}

yield scrapy.Request(url, headers=headers)

**Proxies**

* Proxies act as intermediaries, masking our IP address to avoid bans. These distribute requests across multiple IPs, thus bypassing restrictions placed on such addresses.

**Example**:

request.meta['proxy'] = 'http://proxy-server:port'

Let me try to clarify how user agents, headers, and proxies differ using one analogy:

Imagine you are a spy and your assigned task is to gather data on a suspect, so you have to investigate their house at night.  
Think of User Agents as your **disguise**. It tells the website (cctv camera) who you are (e.g., a Chrome or Mozilla browser). If you always use Scrapy’s default disguise, websites can recognise and block you.

Now, imagine you're carrying **extra documents** like a fake ID to prove you're someone else. These details make you blend in even better. This is the task of the Request Headers.

Picture entering the house through different doors every time you enter so the owner doesn’t realise it’s the same person repeatedly. This is the job of the Proxies. They keep your IP address fresh, avoiding suspicion.

When combined, these tools help you investigate the suspects house in a stealthier way i.e. scrape websites stealthily thus helping you mask your identity better.