|  |  |
| --- | --- |
| shell/command line/terminal | What happens |
| scrapy startproject bw2 | In your current directory generates folder bw2 with subfolders needed for the project |
| cd bw2/ | **To be able to run spiders**, you move inside your project folder bw2 |
| scrapy genspider spider34 example.com | generates spider34 inside bw2/bw2/spiders/  you can open it and modify as you want. |
|  | You can manually transfer ready-to-use spiders insider folder bw2/bw2/spiders/ |
| scrapy crawl spider34 -O file\_name.json | runs spider34, saves output to file with extension json (you can also put .csv extension), files are located in project folder (bw2) |

Explanation of spider34 code (it works also for es.38):

Go to the webpage https://www.lafeltrinelli.it/libri-inglese/science-computer-technology and press **Shift + Command + C**. This opens the developer mode. In this mode, you can inspect the page, hover over different elements of the page, and see their names

Spider34 extracts data from the webpage (see **start\_urls**). In **libri**, we extract data from the whole **div** element with the class **'cc-product-list-item'**. From each **libri** element, we extract the book title as **text** from the element '**a.cc-title**', and the image link as the **'src'** attribute from the **'img'** element.

Immagine che contiene testo, software, diagramma, Pagina Web

Descrizione generata automaticamente