# Web Scraping - Quality Assurance – Utils

Table des matières

[Web Scraping - Quality Assurance - Utils 1](#_Toc157515771)

[Tree Structure 2](#_Toc157515772)

# Tree Structure

* **Src**
* **scrapyd**
* **dbs**
* **electrical\_data.db** => database for raw scraping data
* **items**
* **electrical\_data**
* **ieso\_general\_xml**
* **ieso\_hourly\_demand** => JSON extraction of the scraping process (task extraction)
* **task\_1\_2024-01-22T11\_11\_00.jl** => JSON item file extraction
* **…**
* **initial\_test**
* **logs**
* **electrical\_data**
* **ieso\_general\_xml**
* **ieso\_hourly\_demand**
* **task\_1\_2024-01-22T12\_33\_00.json** => JSON information of the task extraction
* **…**
* **initial\_test**
* **stats.json**
* **scrapyd.conf**
* **utils**
* **\_\_init\_\_.py**
* **Item.py** => Base functions for generals items
* **web\_interface**
* **scrapydweb\_settings\_v10.py**
* **web\_scraping**
* **electrical\_data**
* **electrical\_data**
  + **items**
    - **ieso**
* **\_\_init\_\_.py**
* **demand.py**
  + - **utils**
* **\_\_init\_\_.py**
* **item.py**
* **\_\_init\_\_.py**
* **spiders**
* **\_\_init\_\_.py**
* **items.py**
* **ieso\_general\_xml.py**
* **ieso\_hourly\_demand.py**

**initial\_test.py**

* **\_\_init\_\_.py**
* **items.py**
* **middlewares.py**
* **pipelines.py**
* **settings.py**
* **run\_spider.py** => Script to RUN spider
* **scrapy.cfg** => Data scraper and setting location

# Data sources

## Reliability

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |