Web Scraping

Extracción de datos de Internet

Fontes de datos

Administracións públicas

Portais de emprego

Redes sociais

Tendas en liña

Xornais dixitais

Casos de uso

Aprendizaxe automática

Estudos de mercado

Vixilancia

Requisitos mínimos

Cliente de HTTP

Analizador de HTML

Problemas

Dificultade

Rendemento

Seguridade

Mantemento

Heteroxeneidade

Detección

Scrapy

Facilidade

Comportamento intuitivo

CSS e XPath

Exportación

Caché

Intérprete

Rendemento

Asincronía

Control da simultaneidade

Seguridade

Límites de dominios

Límites de tamaño

Límites de tempo

Protección contra vulnerabilidades

Mantemento

Elementos

Parametrización

Arañas base

Contratos

Xestión de ficheiros

Spidermon

Scrapyd

Flexibilidade

Extensións

Python

Flexibilidade

dateparser

extruct

js2xml

parsel

price-parser

w3lib

Camuflaxe

Comportamento

Proxy

scrapy-splash

scrapy-selenium

Titorial

i.gal/scrapy