Objective automate and have a spyder crawling all the time with no need of computer, (server based) uploads with no 50,000 limitation several 100 millions (mysql based). Then make 3 crosses (phone,company name, cif,fax,website) and add website 1,2 or email 1,2,3,4,5.

History

We have an application (.exe) done in visual basic were we upload txt of maximum 64000 fields that generates a xls with. Video with old process.

Mysql or php spyder script best.

Archive 1: results spyder, phone number, cif, postal code,

Fields examples.

Archivo de 50,000.

Urls Ya escrapeadas. Botikap para comparar.

Spyder Variables

Variables can be modified. (eg: phone Spanish phone number)

Input archive 2. Loadable in a php.

Alerts

Alert1 does not have urls

Alert2 weekly new results.

Archive 1 empresite sin datos.

Email 1,2,3,4,5

Archive 2

Name of field.

Different fields (phone number,name, cif) and gives the archive as a result.

Queries of archive 1 : crossings in mysql

Crawl all webpages

System that if a file is too big (maximum time)

Record of how many websites has crawled.

Poder crawlers los unsuscribe sus dominios y buscar otras opciones.

Timings to scrap urls every 10,000 urls how much time.

Regex searches.

Code of visual basic

Contenidos explicativos

Video sistema actual

Archivo resultados

Archivo con urls objetivo

Archivo con queries de cruzado (empresite)

Código web scrapper

Archivo .exe actual