CredB

One Happy Crawler

Abstract

A crawler is a program that visits Web sites and reads their pages and other information in order to create entries for a search engine index.Crawlers are typically programmed to visit sites that have been submitted by their owners as new or updated. Entire sites or specific pages can be selectively visited and indexed. Crawlers apparently gained the name because they crawl through a site a page at a time, following the links to other pages on the site until all pages have been read.

In our computer labs, during the time of practical examinations, one of the lab members has to manually go through each and every workstation and check for any file that can be maliciously used during the times of practical. This process is very time consuming. To overcome this scenario and to reduce human efforts, we have developed a directory crawler, which will crawl all the directories of every workstation, have all the file details which are present in the workstation compiled in a single file. End user will be provided with GUI using which all the files as well any single file or piece of information can be retrieved on a single click.

This offline search engine, the name we like to call it, has it's roots in python. This crawler employs PHP and MySQL on different instances.