
Resumen

En los últimos años se ha especulado acerca del tamaño y contenido de la parte de internet conocida como “darknet”. Este trabajo trata de comprobar la viabilidad de indexar este contenido, medir el alcance obtenible, y obtener un sistema de monitorización de contenidos.

Mediante técnicas clásicas de arañado web distribuido y el uso de diversos componentes de software se ha desarrollado un sistema de arañado e indexado de contenidos web. Con una ejecución de tres meses y limitando su alcance a tres de las darknets existentes más conocidas, se ha logrado obtener resultados satisfactorios.

Abstract

During the last years, there have been speculations about the size and content of the part of the internet known as darknet. This work aims to determine the viability of indexing that content, measuring the obtainable reach, and develop a content monitoring system.

Using classic web crawling techniques and several software components, a crawling and indexing web content system has been developed. Having it running for three months and limiting its scope to the three most used darknet systems, it has been possible to get satisfactory results.

Keywords

Darknets, crawler, indexer, monitoring