**Instituto Tecnológico de Costa Rica Sede San Carlos**

**Escuela de Computación**

**Tema: Web Scrapping**

**Jeison Saborío Rodríguez**

**Sede San Carlos**

**Fecha: 03 / 04/ 2017**

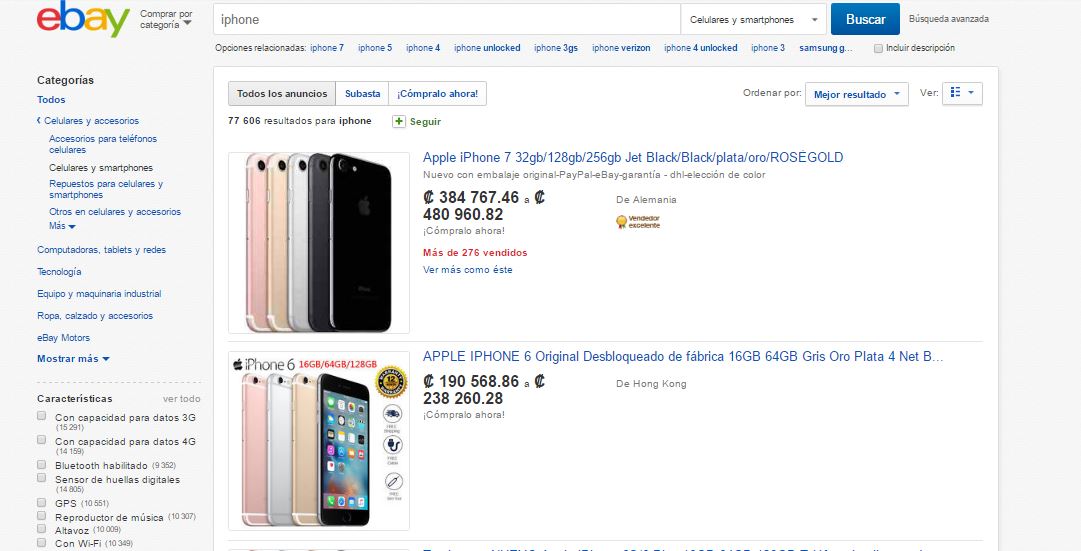
**Introducción**

Los datos están creciendo en disponibilidad para los usuarios en la web, por lo que se está ideando mediante métodos procesar automáticamente la información disponible. Este campo de Web Scrapping es una técnica de extracción para el descubrimiento y obtención de información disponible en la Web, la cual comprende varios campos como procesamiento del lenguaje natural, técnicas relacionadas con la investigación. Por otro lado se pueden aplicar diferentes técnicas para la aplicación del web Scrapping, la cual serán expuestos más adelante y sin dejar al lado las herramientas que son punto importante para la extracción de la información.

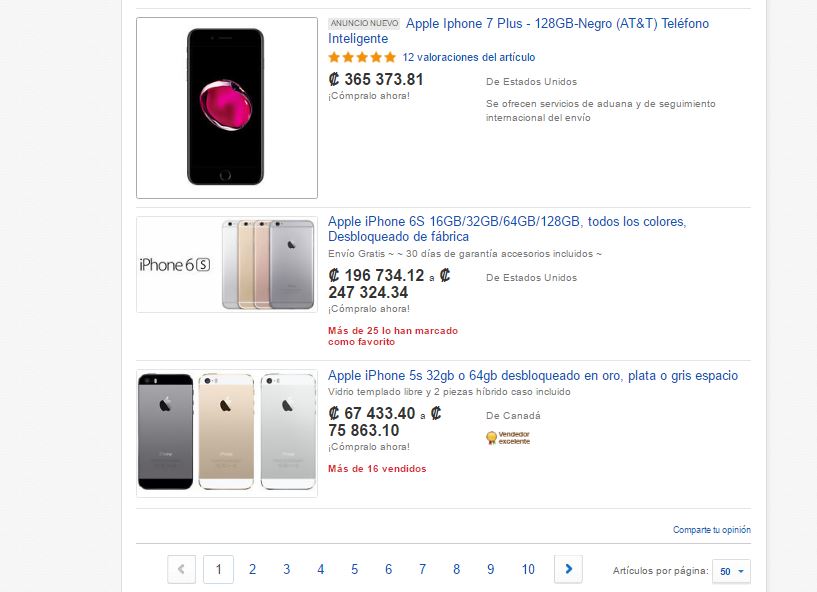
**Desarrollo**

Web Scrapping es el método de extraer información de una manera automática de la Web. La cual en esta investigación se desarrolló una solución utilizando como lenguaje nodejs, aplicada en la página <http://www.ebay.com/>.

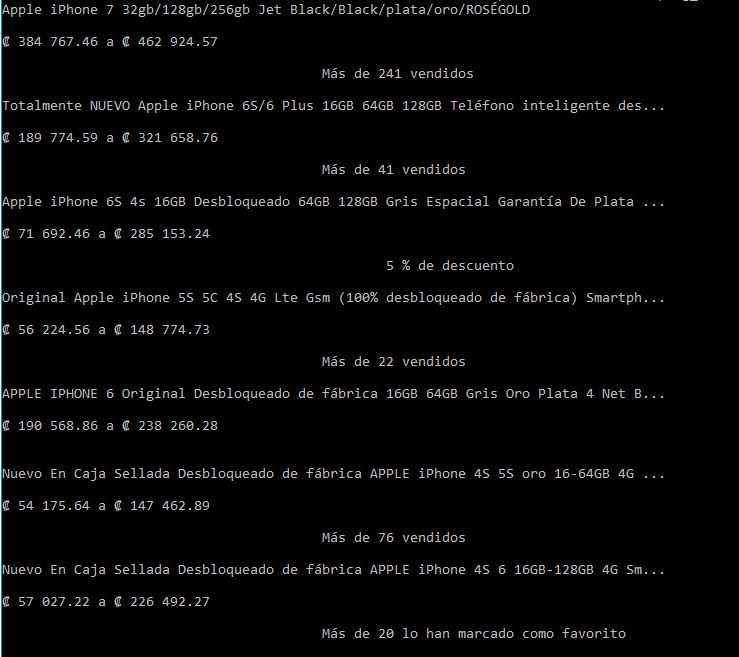
Para explicar más a fondo se utilizó express la cual es un framework de nodejs que nos permite levantar un servidor en tiempo real, de igual manera cheerio esto es una implementación simple de jquery, permitiendo realizar consultas a la página de <http://www.ebay.com/>. Intentamos extraer datos de importancia como el título de una compra, precio, vendedor, otras características de relevancia en cada producto.



A continuación se muestra la paginación de la página ebay, la cual nos permite extraer información relevante de las diferentes páginas enumeradas dentro de la página web.



Seguidamente se ejemplifica la información extraída de la página de ebay, mediante el método de Scraping que sin ninguna duda nos permite la obtención de la información a gran escala.



**Conclusión**

Sin ninguna duda el web scrapping es un proceso que tiene mucho futuro en una sociedad donde se busca es recopilar datos de la web y mostrarlo de una manera significativa, buscando representarlo una manera más sencilla. Como ya se mostró nodejs nos permite, mediante el uso de algoritmos programados o interfaz gráfica obtener datos de relevancia, de esta forma tener los datos clasificados que nos permite estudiarlos para tener un panorama de lo que sucede en el ámbito aplicado.

**Bibliografía**

<https://github.com/cheeriojs/cheerio>

<https://cheerio.js.org/>

<https://www.digitalocean.com/community/tutorials/how-to-use-node-js-request-and-cheerio-to-set-up-simple-web-scraping>