

PROYECTO SISTEMAS BASADOS EN CONICIMIENTO

Proyecto del primer bimestre

Nombre: Carlos Xavier Hidalgo Paredes Fecha: 26/05/2021

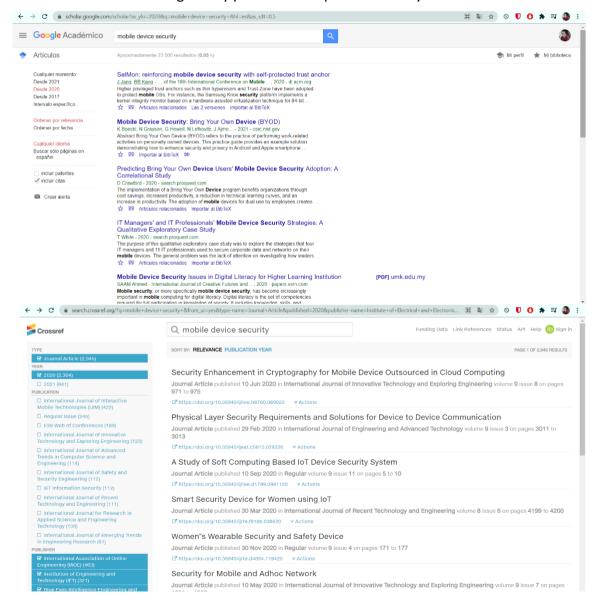
Dominio de trabajo:

Artículos científicos sobre Mobile Device Security

Fuentes de datos de dominio con las que se trabajara

Artículos científicos sobre Seguridad en los dispositivos móviles

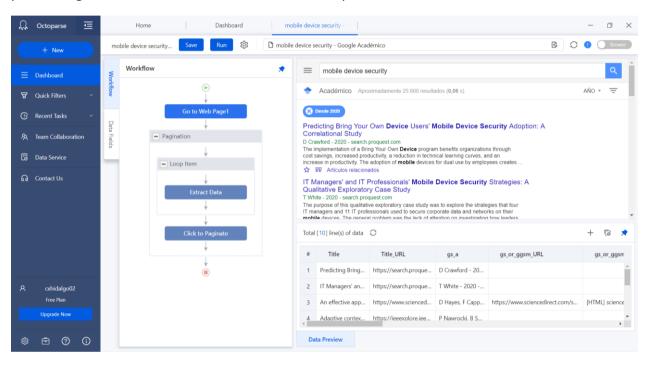
En cuanto a los recursos de donde se va a extraer la data será de la página de *Google Academic, Crossef*, obteniendo datos sobre los artículos científicos en el área de las ciencias de Mobile Device Security, obteniendo título, autores, número de páginas, resumen, autores, fecha de publicación, tipo de documento, entre otros; la búsqueda se hará de todas las regiones y países con respecto al 2020 y 2021.

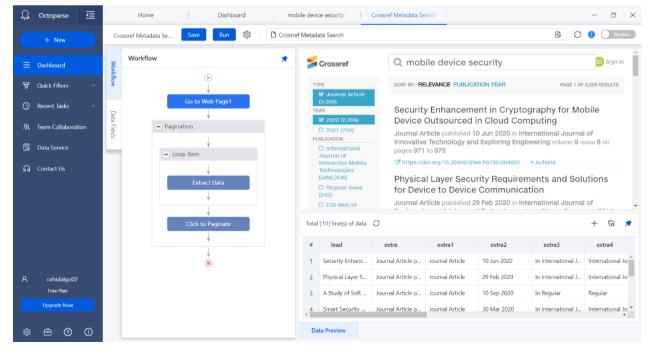




Método o Herramienta para extracción de datos

El método de extracción se hará scrapy a las páginas de *Google Academic, Crossef y Scimago*, con *Octoparse (Octopus Data)*; es una herramienta que permite realizar scrapy de forma masiva solo se ingresa el Url y se hace los filtros, una vez echo eso permite extraer la información del sitio web. Es una herramienta que permite extraer datos de una web, direcciones Ip, direcciones Ip de correo electrónico, precios, número de teléfonos móviles, extracto de imágenes, resumen y publicación de datos sobre casi cualquier sitio web y te permitirá guardarlos como datos estructurados limpios en tu formato de elección.







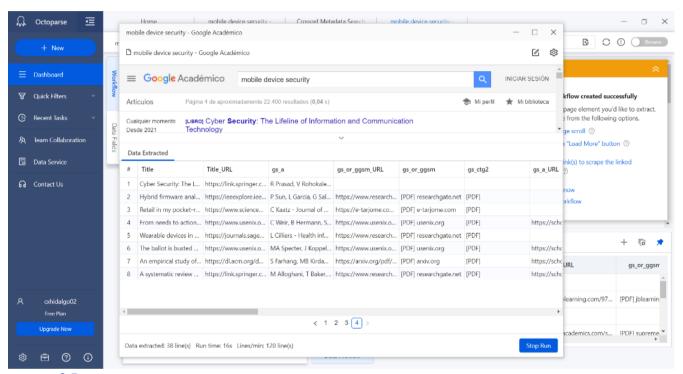
DATOS DESCARGADOS

Google Academic / 2020 - 2021

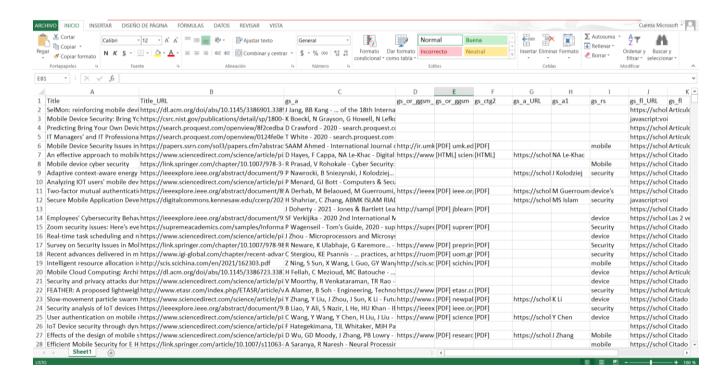
En Google academic se realizó la búsqueda de Artículos Científicos sobre Mobile Device security, ordenados, entre el 2020 y el 2021, dándonos como resultados aproximadamente 23.500 resultados (0,05 s)

Url:

https://scholar.google.com/scholar?as_vlo=2020&q=mobile+device+security+&hl=es&as_sd



t = 0,5



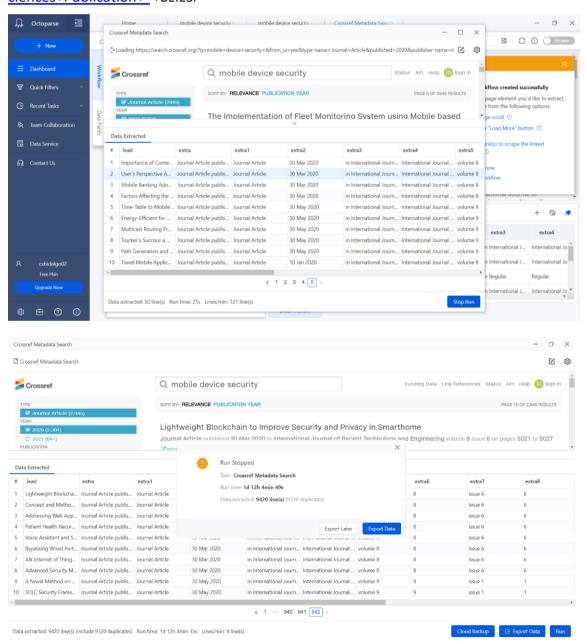


Crossref / 2020 - 2021

En Crossref se realizó la búsqueda de Artículos Científicos sobre Mobile Device security, ordenados, entre el 2020 y el 2021, publicadas el Institute of Electrical and Electronics Engineers (IEEE) (5,779), International Association of Online Engineering (IAOE) (463), Institution of Engineering and Technology (IET) (321) y Blue Eyes Intelligence Engineering and Sciences Engineering and Sciences Publication - BEIESP (299), esto nos da como resultado aproximadamente 58.102 resultados.

Url: https://search.crossref.org/?q=mobile+device+security+&from_ui=yes&type-name=Journal+Article&published=2020&publisher-

<u>name=Institute+of+Electrical+and+Electronics+Engineers+%28IEEE%29%3BInternational+Association+of+Online+Engineering+%28IAOE%29%3BInstitution+of+Engineering+and+Technology+%28IET%29%3BBlue+Eyes+Intelligence+Engineering+and+Sciences+Engineering+and+Sciences+Publication+-+BEIESP</u>

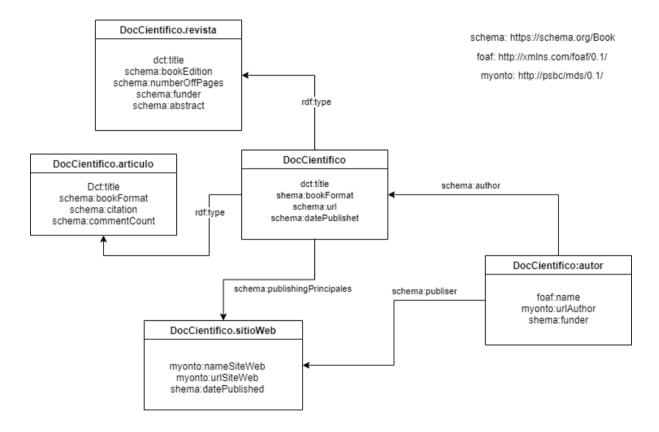




La herramienta a trabajar con los datos descargados es **OpenRefine:** Es una herramienta que permita trabajar con datos desordenados, permite limpiarlos, transformar datos, OpenRefine es una herramienta que mantiene la privacidad de los datos en la propia computadora hasta que uno desee compartirlos o brindar colaboración de los mismos. *Link:* https://openrefine.org/

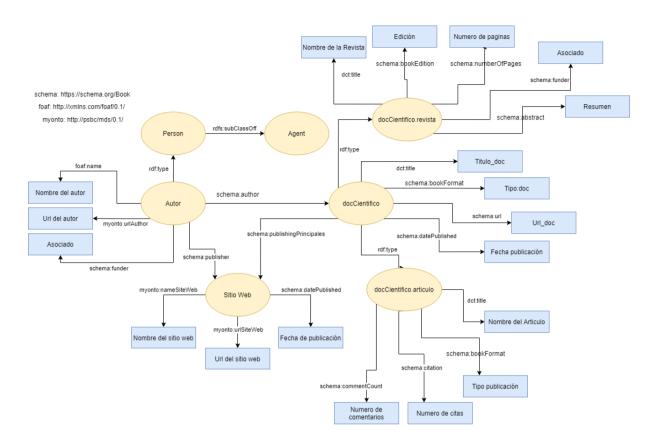
Código Fuente Java: utilizaremos una liberia en Java que se llama **POI**, esta permite trabajar con documentos Microsoft en creación, escritura y lectura. Esa librería se la utilizara para la lectura de los archivos generados por la herramienta Octoparse para realizar las tripletas correspondientes y ser almacenadas en una base de datos *MySql*.

DIAGRAMA DE INTEGRACIÓN





DESCRIPCION SEMANTICA / ONTOLOGIA



VOCABULARIO

PREFIJOS

psbc: ProyectoMDS

owl: https://www.w3.org/2002/07/owl#Class

foaf: http://xmlns.com/foaf/0.1/

myonto: http://psbc/mds/0.1/

Subjet	Predicate	Objet
psbc:autor	rdf:type	rdfs:Class
psbc:autor	rdf:type	foaf:Person
psbc:autor	rdfs:subClassOf	foaf:Agent
psbc:autor	rdfs:label	"writer"@en
psbc:autor	rdfs:label	"escritor"@es
psbc:autor	foaf:name	Nombre del escritor
psbc:autor	myonto:urlAuthor	Url del escritor
psbc:autor	schema:funder	Asociado / fundador
psbc:docCientifico	rdf:type	rdfs:Class
psbc:docCientifico	rdfs:label	"document"@en
psbc:docCientifico	rdfs:label	"documento"@es
psbc:docCientifico	schema:author	psbc:autor



psbc:docCientifico	rdfs:subClassOf	myonto:docCientifico.revista
psbc:docCientifico	rdfs:subClassOf	myonto:docCientifico.articulo
psbc:docCientifico	rdf:type	psbc:docCientifico.revista
psbc:docCientifico	rdfs:label	"Scientific Magazine"@en
psbc:docCientifico	rdfs:label	"Revista Cientifica"@es
psbc:docCientifico	rdf:type	psbc:docCientifico.articulo
psbc:docCientifico	rdfs:label	"Scientific Article"@en
psbc:docCientifico	rdfs:label	"Articulo Cientifico"@es
psbc:docCientifico	dct:title	Nombre del documento
psbc:docCientifico	schema:bookFormat	Tipo de documento
psbc:docCientifico	schema:url	Url del documento
psbc:docCientifico	schema:datePublished	Fecha de publicacion
psbc:docCientifico.revista	dct:title	Nombre de la revista
psbc:docCientifico.revista	schema:numberOffPages	Numero de paginas
psbc:docCientifico.revista	schema:bookEdition	Numero de edicion
psbc:docCientifico.revista	schema:funder	Asociado / fundador
psbc:docCientifico.revista	schema:abstract	Resumen
psbc:docCientifico.articulo	dct:title	Nombre del articulo
psbc:docCientifico.articulo	schema:bookFormat	Tipo de publicacion
psbc:docCientifico.articulo	schema:citation	Numero de citas
psbc:docCientifico.articulo	schema:commentCount	Numero de cometarios
psbc:autor	schema:publisher	psbc:sitioWeb
psbc:docCientifico	schema:publishingPrincipales	psbc:sitioWeb
psbc:sitioWeb	myonto:nameSiteWeb	Nombre del sitio web
psbc:sitioWeb	myonto:urlSiteWeb	Url del sitio web
psbc:sitioWeb	schema:datePublished	Fecha de publicacion