

BUSCADOR SEMÁNTICO

TECNOLOGÍAS EMERGENTES

Descripción Universidad Técnica Particular de Loja

Proyecto de la materia Sistemas de Basados en el conocimiento

Integrantes: Katherine Morocho María Fernanda Ordóñez

INTRODUCCIÓN

Nuestro proyecto va a ser realizado dentro del dominio de Publicaciones científicas, específicamente a los metadatos de las **tecnologías emergentes(nuevas tecnologías)**, ya que en la actualidad es un tema de suma importancia y con mucha fluidez de la información que podríamos utilizar para el aprendizaje inteligente.

FUENTE DE DATOS

Después de una lectura y análisis de las fuentes de datos para nuestro dominio, hemos decidido usar la extracción de datos de:

- Semantic Scholar
- Google Scholar

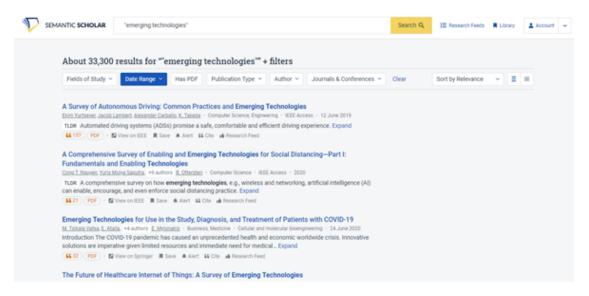
Utilizando una herramienta de web scraping muy poderosa que optimiza la extracción de datos y su eficiencia llamada Octoparse.

Para la extracción de las publicaciones científicas se utilizará los atributos de la herramienta para sacar información como: Id, abstract, título, autor, año de publicación, DOI, volúmen.

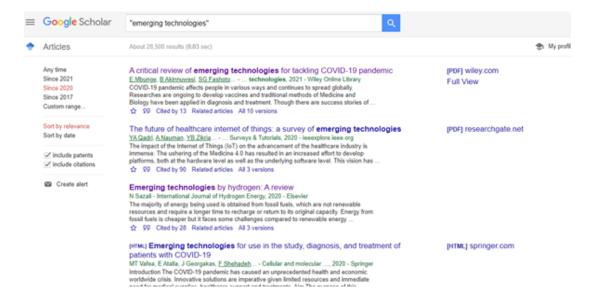
Se realizará búsquedas del último año ya que el tema de Tecnologías Emergentes es muy amplio.

El resultado de la cadena de búsqueda que se estableció para:

• Semantic Scholar nos devolvió 33300 investigaciones relacionadas a "emerging technologies" con características relacionadas a lo buscado.



 Google Scholar nos devolvió 28500 investigaciones relacionadas a "emerging technologies" con múltiples características relacionadas al artículo obtenido.



CADENA DE BÚSQUEDA

Fuente: Semantic Scholar

Cadena de busqueda: TITLE-ABS-KEY (emerging AND technologies) Fecha de obtención de

data: 01/06/2021

Total: 33300

Fuente: Google Scholar

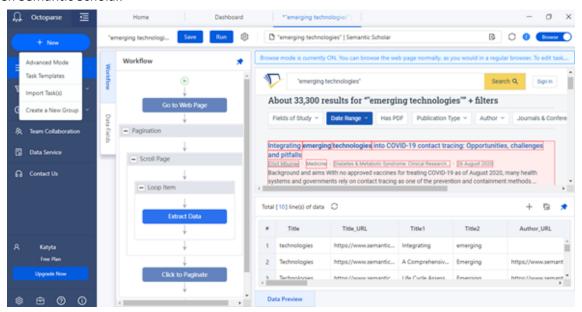
Cadena de busqueda: TITLE-ABS-KEY (emerging AND technologies) Fecha de obtención de

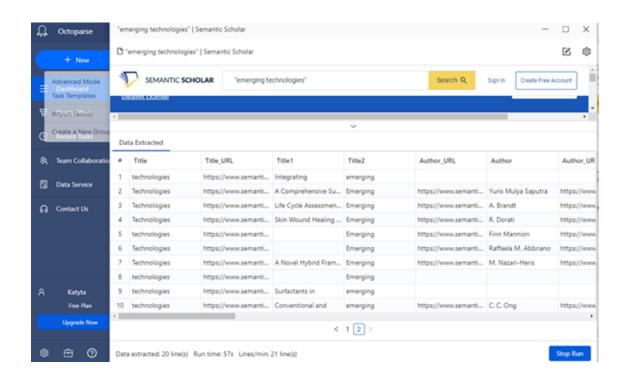
data: 01/06/2021

Total: 28500

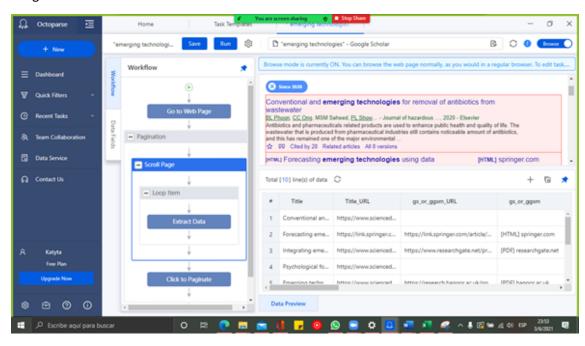
EXTRACCIÓN Y LIMPIEZA DE DATOS

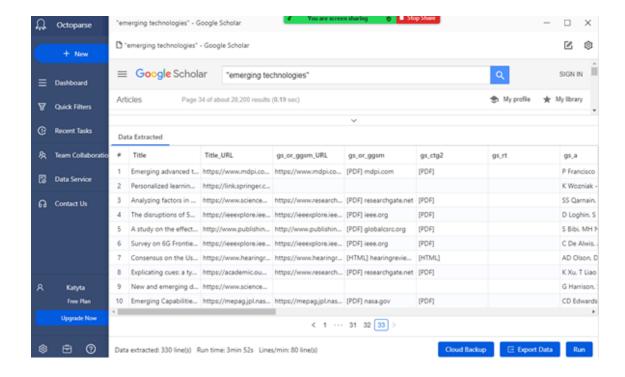
Utilizando la herramienta Octoparse con el link de búsqueda realizamos la extracción de datos en Semantic Scholar.





Utilizando la herramienta Octoparse con el link de búsqueda realizamos la extracción de datos en Google Scholar.





MODELO CONCEPTUAL

Nuestro modelo conceptual, se basa en algunos vocabularios (ontologías) existentes y uno creado para el proyecto para las cuales voy a citar a cada una de ellos:

Schema (schema.org) es un vocabulario estructurado de datos que define entidades, acciones y relaciones en Internet (webs, emails, etcétera). Este vocabulario hace posible que los motores de búsqueda entiendan el significado que hay detrás de diferentes temas (entidades) en la web y, como consecuencia, sirve para proporcionar una mejor experiencia para los usuarios de Internet.

FOAF (*Friend Of A Friend*, literalmente "Amigo de un Amigo") esta es una ontología que describe a personas , objetos y sus relaciones de manera entendible para el computador, utilizando un marco de descripción RDF y el lenguaje OWL.

Dublin Core al igual que Schema este vocabulario abarca propiedades, clases para relacionarlas entre sí, además de ser compatible con otros vocabularios. Los términos que se expresan el el DC son mediante RDF para la vinculación con otros datos.

Prims, Es una fuente de metadatos. define un estándar para la descripción, el intercambio y la reutilización de contenido interoperable en contextos de publicación tradicionales y electrónicos. PRISM recomienda el uso de ciertos estándares existentes, como XML, RDF, Dublin Core y varias especificaciones ISO para ubicaciones.

Fabio es un vocabulario descriptivo expresado utilizando el marco de descripción de recursos (RDF) y el lenguaje de ontología web (OWL).

Cada perfil tiene un identificador único (como las direcciones de correo electrónico de la persona, el número de teléfono internacional, el nombre de la cuenta de Facebook, un Jabber ID o un URI de la página de inicio o el weblog de la persona), que se utiliza al definir estas relaciones.

