



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.

The screenshot shows the Semantic Scholar search interface. At the top, there's a search bar with the query "emerging technologies" and a search button. Below the search bar, it says "About 33,300 results for 'emerging technologies' + filters". There are several filter buttons: "Fields of Study", "Date Range", "Has PDF", "Publication Type", "Author", "Journals & Conferences", and "Clear". A "Sort by Relevance" button is also present. The search results are listed below, showing the title, authors, publication date, and a brief abstract. The first result is "A Survey of Autonomous Driving: Common Practices and Emerging Technologies" by Kim Yurkew, Jacob Lambert, Alexander Cardillo, K. Tanaka, published in IEEE Access on 12 June 2019. The second result is "A Comprehensive Survey of Enabling and Emerging Technologies for Social Distancing—Part I: Fundamentals and Enabling Technologies" by Coro T. Nguyen, Yurii Mihai Sautu, et al., published in IEEE Access in 2020. The third result is "Emerging Technologies for Use in the Study, Diagnosis, and Treatment of Patients with COVID-19" by M. Tsouros, Y. Vafea, et al., published in Cellular and molecular biotechnology on 24 June 2020. The fourth result is "The Future of Healthcare Internet of Things: A Survey of Emerging Technologies".

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

The screenshot shows the Google Scholar search interface. At the top, there's a search bar with the query "emerging technologies" and a search button. Below the search bar, it says "Articles" and "About 28,500 results (9.83 sec)". There are several filter buttons: "Any time", "Since 2021", "Since 2020", "Since 2017", and "Custom range...". There are also buttons for "Sort by relevance" and "Sort by date". There are checkboxes for "Include patents" and "Include citations". A "Create alert" button is also present. The search results are listed below, showing the title, authors, publication date, and a brief abstract. The first result is "A critical review of emerging technologies for tackling COVID-19 pandemic" by E. Mbunga, B. Akinyemi, S. G. Fashoto, et al., published in Wiley Online Library in 2021. The second result is "The future of healthcare internet of things: a survey of emerging technologies" by Y. A. Qadri, A. Nauman, Y. B. Zikria, et al., published in IEEE Explore in 2020. The third result is "Emerging technologies by hydrogen: A review" by N. Sazali, published in International Journal of Hydrogen Energy in 2020. The fourth result is "Emerging technologies for use in the study, diagnosis, and treatment of patients with COVID-19" by M. Tsouros, Y. Vafea, J. Georgakakos, E. Shabaneh, et al., published in Springer in 2020.

CADENA DE BÚSQUEDA

Fuente: Semantic Scholar

Cadena de búsqueda: TITLE-ABS-KEY (emerging AND technologies) **Fecha de obtención de data:** 01/06/2021

Total: 33300

Fuente: Google Scholar

Cadena de búsqueda: 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.

The screenshot displays the Octoparse web interface. On the left is a sidebar with navigation options like 'New', 'Advanced Mode', 'Task Templates', 'Import Task(s)', 'Create a New Group', 'Team Collaboration', 'Data Service', 'Contact Us', and user information for 'Katyta'. The main area is divided into a 'Workflow' panel on the left and a 'Data Preview' panel on the right. The workflow consists of the following steps: 'Go to Web Page', 'Pagination', 'Scroll Page', a 'Loop Item' containing 'Extract Data', and 'Click to Paginate'. The 'Data Preview' panel shows the results of a search for 'emerging technologies' on Semantic Scholar. It indicates 'About 33,300 results' and displays a table with 10 lines of data. The table has columns: #, Title, Title_URL, Title1, Title2, and Author_URL.

#	Title	Title_URL	Title1	Title2	Author_URL
1	technologies	https://www.semantic...	Integrating	emerging	
2	Technologies	https://www.semantic...	A Comprehensive	Emerging	https://www.semantic...
3	Technologies	https://www.semantic...	Life Cycle Assess...	Emerging	https://www.semantic...

This screenshot shows the 'Data Extracted' section of the Octoparse interface. It displays a table with 10 rows of extracted data from Semantic Scholar. The table has columns: #, Title, Title_URL, Title1, Title2, Author_URL, Author, and Author_UR. The data includes various research papers related to emerging technologies. At the bottom, it shows 'Data extracted: 20 line(s)', 'Run time: 57s', and 'Lines/min: 21 line(s)'. There is a 'Stop Run' button on the right.

#	Title	Title_URL	Title1	Title2	Author_URL	Author	Author_UR
1	technologies	https://www.semantic...	Integrating	emerging			
2	Technologies	https://www.semantic...	A Comprehensive Su...	Emerging	https://www.semantic...	Yuris Mulya Saputra	https://www...
3	Technologies	https://www.semantic...	Life Cycle Assessmen...	Emerging	https://www.semantic...	A. Brandt	https://www...
4	Technologies	https://www.semantic...	Skin Wound Healing ...	Emerging	https://www.semantic...	R. Dorati	https://www...
5	technologies	https://www.semantic...		Emerging	https://www.semantic...	Finn Mannion	https://www...
6	Technologies	https://www.semantic...		Emerging	https://www.semantic...	Raffaella M. Abbriano	https://www...
7	Technologies	https://www.semantic...	A Novel Hybrid Fram...	Emerging	https://www.semantic...	M. Nazari-Heris	https://www...
8	technologies	https://www.semantic...		Emerging			
9	technologies	https://www.semantic...	Surfactants in	emerging			
10	technologies	https://www.semantic...	Conventional and	emerging	https://www.semantic...	C. C. Ong	https://www...

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

The screenshot shows the Octoparse web interface. On the left is a sidebar with navigation options like 'New', 'Dashboard', 'Quick Filters', 'Recent Tasks', 'Team Collaboration', 'Data Service', and 'Contact Us'. The main area displays a 'Workflow' editor with a sequence of steps: 'Go to Web Page', 'Pagination', 'Scroll Page', 'Loop Item', 'Extract Data', and 'Click to Paginate'. To the right, a 'Data Preview' section shows a table of extracted data from a Google Scholar search for 'emerging technologies'. The table includes columns for '#', 'Title', 'Title_URL', 'gs_or_ggsm_URL', and 'gs_or_ggsm'. The preview shows the first 10 results of the search.

#	Title	Title_URL	gs_or_ggsm_URL	gs_or_ggsm
1	Conventional an...	https://www.scienced...		
2	Forecasting eme...	https://link.springer.c...	https://link.springer.com/article/...	[HTML] springer.com
3	Integrating eme...	https://www.scienced...	https://www.researchgate.net/pr...	[PDF] researchgate.net
4	Psychological fo...	https://www.scienced...		
5	Forecasting eme...	https://www.scienced...	https://www.researchgate.net/pr...	[PDF] researchgate.net

The screenshot shows the 'Data Extracted' section of the Octoparse interface. It displays a table with 10 rows of extracted data from Google Scholar. The table includes columns for '#', 'Title', 'Title_URL', 'gs_or_ggsm_URL', 'gs_or_ggsm', 'gs_ctg2', 'gs_rt', and 'gs_a'. The bottom of the interface shows the total data extracted (330 line(s)) and the run time (3min 52s). The interface also includes a 'Data Preview' section at the top right, which shows a preview of the extracted data.

#	Title	Title_URL	gs_or_ggsm_URL	gs_or_ggsm	gs_ctg2	gs_rt	gs_a
1	Emerging advanced t...	https://www.mdpi.co...	https://www.mdpi.co...	[PDF] mdpi.com	[PDF]		P Francisco
2	Personalized learnin...	https://link.springer.c...					K Wozniak -
3	Analyzing factors in ...	https://www.science...	https://www.research...	[PDF] researchgate.net	[PDF]		SS Qarnain,
4	The disruptions of 5...	https://ieeexplore.iee...	https://ieeexplore.iee...	[PDF] ieee.org	[PDF]		D Loghin, S
5	A study on the effect...	http://www.publishin...	http://www.publishin...	[PDF] globalcsrc.org	[PDF]		S Bibi, MH F
6	Survey on 6G Frontie...	https://ieeexplore.iee...	https://ieeexplore.iee...	[PDF] ieee.org	[PDF]		C De Alwis, .
7	Consensus on the Us...	https://www.hearingr...	https://www.hearingr...	[HTML] hearingrevie...	[HTML]		AD Olson, D
8	Explicating cues: a ty...	https://academic.ou...	https://www.research...	[PDF] researchgate.net	[PDF]		K Xu, T Liao
9	New and emerging d...	https://www.science...					G Harrison, .
10	Emerging Capabilitie...	https://mepag.jplnas...	https://mepag.jplnas...	[PDF] nasa.gov	[PDF]		CD Edwards

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.

