

ScientoPy: Herramienta para análisis cientométrico

PhD.(c) Juan Pablo Ruiz Rosero
PhD. Gustavo Adolfo Ramírez González
MSc. Jesus Alberto Viveros Delgado

Universidad del Cauca, Popayán, Colombia.
jpabloriguiz@unicauca.edu.co

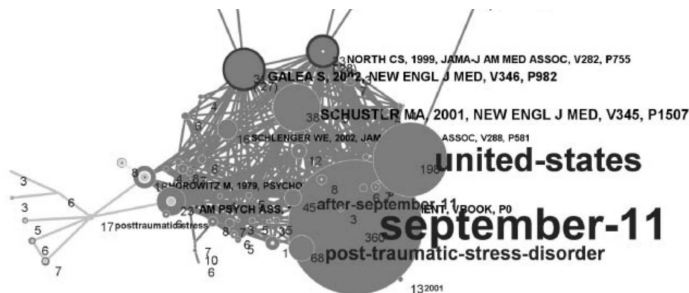


April 26, 2019



Análisis cientométrico (scientometrics)

El análisis cientométrico proporciona información sobre las **tendencias** de investigación en los últimos años al investigar la afiliación del país del autor o autores, los autores con mayores publicaciones, el **impacto** y la **correlación** diversos temas de investigación [Ruiz-Rosero et al., 2017]



CiteSpace II, Psychological-psychiatric impacts of September 11, 2001, terrorist attacks, [Chen, 2006]

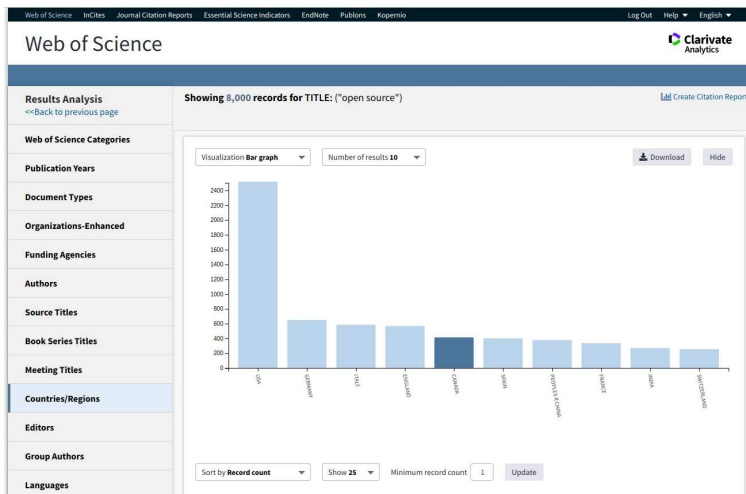
Bases de datos de artículos científicos

- ▶ Son el insumo para las herramientas de análisis cientométrico
- ▶ Scopus y Web of Science (WoS) incluyen solo journals SJR y JCR (revistas indexadas de alto impacto)



Análisis cientométrico online

Scopus y WoS ofrecen análisis de los resultados de sus búsquedas online



Sin embargo, no cuentan con análisis de tendencias y variadas opciones para representación de datos

ScientoPy es una herramienta open source (licencia MIT) de análisis cientométrico basada en Python. Tiene las siguientes características principales:

- ▶ Importe los datos de Clarivate Web of Science (WoS) y Scopus
- ▶ Encuentra y elimina documentos duplicados
- ▶ Extracción del índice H para los temas analizados.
- ▶ Extracción del país e institución de las afiliaciones de autor
- ▶ Temas principales y análisis de temas específicos basado en authorKeywords
- ▶ Búsqueda de temas con comodines
- ▶ Más de 7 gráficos de visualización diferentes: línea de tiempo, barra, paramétrica, nube de palabras, entre otros.
- ▶ Interfaz gráfica de usuario basada en ElectronJS

ScientoPy en Github

jpruiz84 / ScientoPy Unwatch 1 Star 0 Fork 1

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

ScientoPy is an open-source Python based scientometric analysis tool [Edit](#)

Manage topics

130 commits 1 branch 1 release 0 contributors MIT

Branch: master New pull request Create new file Upload files Find File Clone or download

File	Description	Last commit
Manual	Put note to example paper	5 months ago
dataInExample	Not git ignore updated dataInExample	a month ago
.gitignore	Not git ignore updated dataInExample	a month ago
LICENSE.txt	Added licence and readme root files	5 months ago
README.md	Fixed readme	5 months ago
exampleGenerateGraphs.sh	Updated example Generate Graphs	5 months ago
generateBibtex.py	Fixed bug in not correctly coded papers titles	25 days ago
globalVar.py	Updated markers to avoid color and shape repetition	3 months ago
graphUtils.py	New preprocess graph	3 days ago
paperSave.py	Added porcentaje of documents in last years (PDLY)	4 days ago
paperUtils.py	New preprocess graph	3 days ago
preProcess.py	New preprocess graph	3 days ago
scientoPy.py	Added porcentaje of documents in last years (PDLY)	4 days ago

README.md

ScientoPy

ScientoPy is an open-source Python based scientometric analysis tool. It has the following main characteristics:

- Import Clarivate Web of Science (WoS) and Scopus data set

<https://github.com/jpruiz84/ScientoPy>

ScientoPy en manual y paper de ejemplo

ScientoPy v 1.3.5, Installation and User Manual



Juan Pablo Ruiz Rosero
jpabloruiz@unicauca.edu.co

Contents

1	Installation	2
2	Download the bibliometric dataset	2

Table 3. Internet of Things top 50 countries of first author's corresponding address. Country number position (N.), total number of publications (Total), average percentage growth from the last 3 years (2014 to 2016), and h-indices (h-Ind.) from 2006 to 2016.

N.	Country	Total	Average Growth	h-Ind.
1	China	4822	16%	47
2	United States	1561	116%	42
3	India	1089	169%	15
4	South Korea	894	206%	16
5	Italy	874	61%	32
6	Germany	811	64%	24
7	United King.	711	71%	25
8	France	543	126%	21
9	Spain	463	42%	23
10	Japan	449	166%	11
11	Taiwan	438	68%	16
12	Brazil	272	90%	9
13	Finland	266	50%	20
14	Canada	259	104%	15
15	Australia	249	59%	22
16	Sweden	216	68%	17
17	Switzerland	193	31%	19
18	Portugal	191	45%	13
19	Greece	180	46%	14
20	Romania	169	72%	9
21	Belgium	164	87%	11
22	Austria	146	113%	12
23	Malaysia	137	71%	9
24	Russian Fed.	134	271%	8
25	Ireland	126	116%	9
26	Netherlands	109	122%	12
27	Singapore	109	112%	8
28	Poland	104	77%	6
29	Czech Rep.	101	153%	5
30	Turkey	92	319%	5
31	Pakistan	82	210%	7
32	Saudi Arabia	80	122%	7
33	Norway	72	119%	11
34	UAE	71	180%	6
35	South Africa	60	162%	9
36	Denmark	59	39%	11
37	Tunisia	55	163%	6
38	Serbia	53	-1%	6
39	Croatia	51	159%	6
40	Hungary	51	24%	6
41	Indonesia	51	410%	3
42	Egypt	49	159%	4
43	Morocco	47	163%	4
44	Iran	42	94%	5
45	Colombia	39	146%	3
46	Algeria	38	113%	5
47	Jordan	38	108%	5
48	New Zealand	38	86%	6
49	Mexico	36	172%	5
50	Thailand	32	107%	4

To get the previous table, run the following script, and find the results in results/Country.csv:
`python scientoPy.py country --startYear 2002 -l 50 --noPlot`

”open source” cadena de busqueda para titulo de artículos

The screenshot shows the Scopus search results page for the query "open source". The top navigation bar includes "Search", "Sources", "Alerts", "Lists", "Help", "SciVal", and "Register". A blue banner at the top indicates "11,603 document results" with links to "View secondary documents" and "View 53173 patent results". Below the banner, the search criteria "TITLE ('open source')" and options like "Edit", "Save", "Set alert", and "Set feed" are visible. A search box on the left allows for refining results. The main content area, titled "Analyze search results", features a table of results. The first result is highlighted with a yellow background.

Document title	Authors	Year	Source
1 An Open-Source Software Metric Tool for Defect Prediction, Its Case Study and Lessons We Learned	Gabdrakhmanov, B., Tolkachev, A., Succo, G., Yi, J.	2020	Advances in Intelligent Syst Computing 925, pp. 74-85

The screenshot displays the Web of Science search results page. The search criteria are "TITLE: ('open source') ...More". The results section shows 8,000 documents from the Web of Science Core Collection. The first result is highlighted with a yellow background. The page includes navigation options like "Sort by: Date", "Times Cited", "Usage Count", "Relevance", and "More". There are also buttons for "Export...", "Add to Marked List", "Analyze Results", and "Create Citation Report".

Results: 8,000
(from Web of Science Core Collection)

You searched for: TITLE: ("open source") ...More

Create Alert

Refine Results

Sort by: Date **IF** Times Cited Usage Count Relevance More ▾

1 of 800

Select Page Export... Add to Marked List

1. **Open-source dataset for control-oriented modelling in diesel engines**

By: Zhao, Jinghua; Zhou, Sitong; Hu, Yunfeng; et al.
SCIENCE CHINA-INFORMATION SCIENCES Volume: 62 Issue: 7 Article Number: 077201 Published: JUL 2019

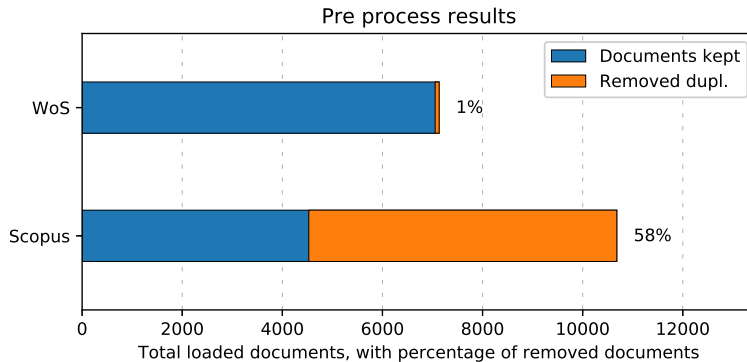
Links Free Full Text from Publisher

Analyze Results
Create Citation Report

Times Cited: 0
(from Web of Science Core Collection)

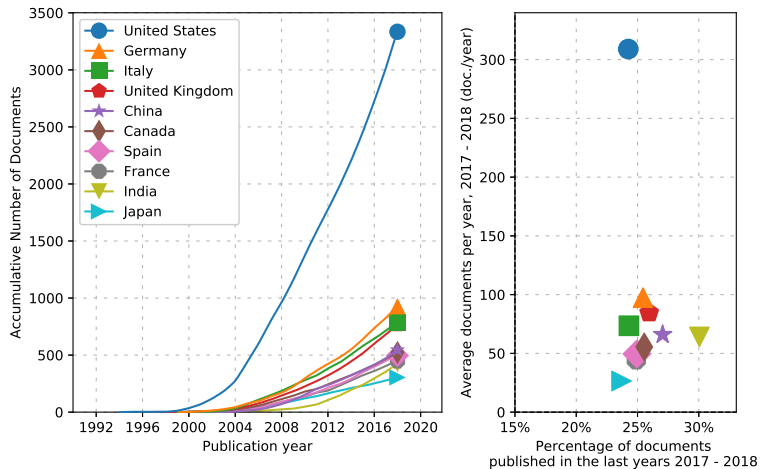
Usage Count ▾

Pre procesamiento



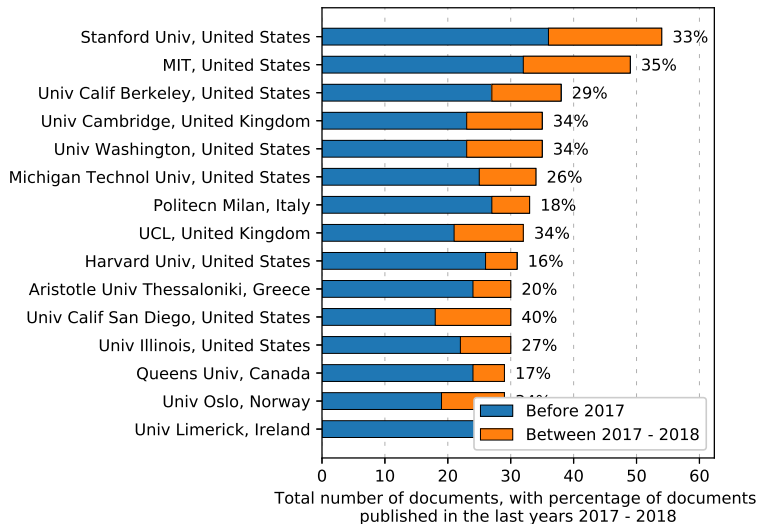
```
python3 preProcess.py dataInOpenSource --graphTitle "Pre process results"
```

Top países



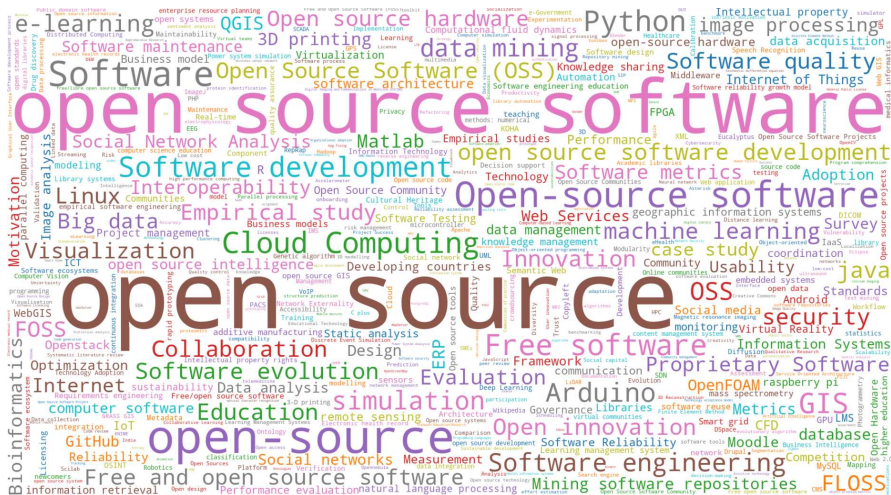
```
python3 scientoPy.py country --startYear 1990 --endYear 2018 --parametric3
```

Top instituciones con países



```
python3 scientoPy.py institutionWithCountry -l 15 --bar2
```

Nube de palabras con authorKeywords



```
python3 scientoPy.py authorKeywords -l 500 --wordCloud
```

ScientoPy interfaz de usuario gráfica

Desarrollada por Jesus Viveros en ElectronJS



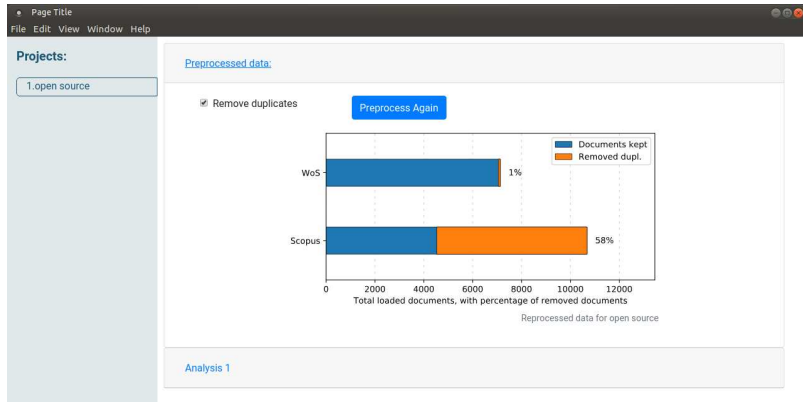
ScientoPy interfaz de usuario gráfica

Desarrollada por Jesus Viveros en ElectronJS



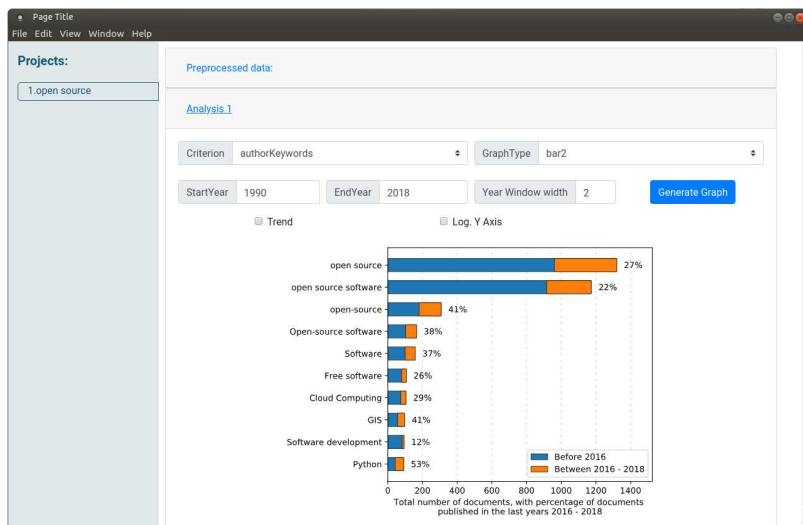
ScientoPy interfaz de usuario gráfica

Resultados de preprocesamiento



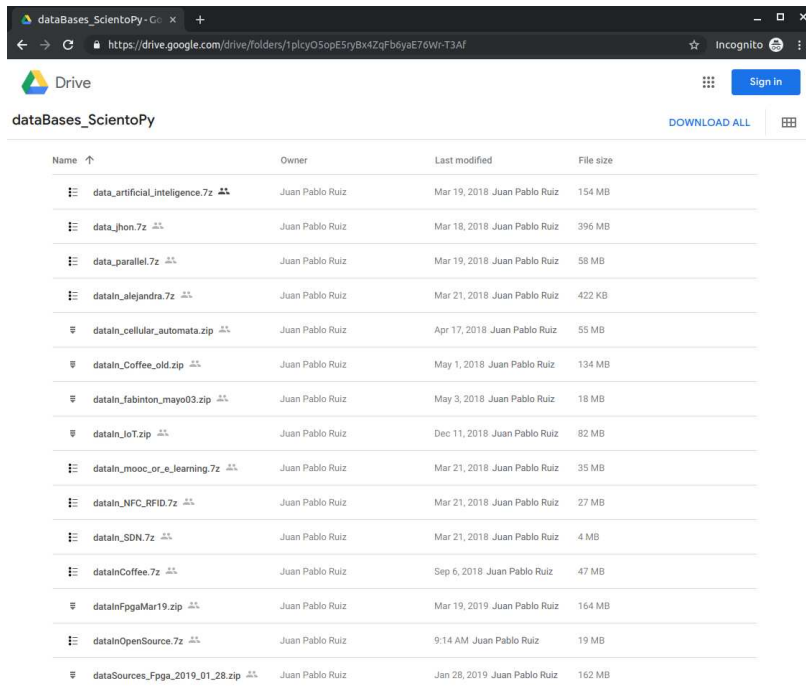
ScientoPy interfaz de usuario gráfica

Resultados de análisis cientométrico



Bases de datos previamente descargadas

<https://drive.google.com/open?id=1plcyO5opE5ryBx4ZqFb6yaE76Wr-T3Af>



The screenshot shows a Google Drive interface for a folder named 'dataBases_ScientoPy'. The folder contains 15 files, each with a name, owner, last modified date, and file size. The files are listed in a table format.

Name	Owner	Last modified	File size
data_artificial_intelligence.7z	Juan Pablo Ruiz	Mar 19, 2018	154 MB
data_jhon.7z	Juan Pablo Ruiz	Mar 18, 2018	396 MB
data_parallel.7z	Juan Pablo Ruiz	Mar 19, 2018	58 MB
dataIn_alejandra.7z	Juan Pablo Ruiz	Mar 21, 2018	422 KB
dataIn_cellular_automata.zip	Juan Pablo Ruiz	Apr 17, 2018	55 MB
dataIn_Coffee_old.zip	Juan Pablo Ruiz	May 1, 2018	134 MB
dataIn_fabinton_mayo03.zip	Juan Pablo Ruiz	May 3, 2018	18 MB
dataIn_IoT.zip	Juan Pablo Ruiz	Dec 11, 2018	82 MB
dataIn_mooc_or_e_learning.7z	Juan Pablo Ruiz	Mar 21, 2018	35 MB
dataIn_NFC_RFID.7z	Juan Pablo Ruiz	Mar 21, 2018	27 MB
dataIn_SDN.7z	Juan Pablo Ruiz	Mar 21, 2018	4 MB
dataInCoffee.7z	Juan Pablo Ruiz	Sep 6, 2018	47 MB
dataInFpgaMar19.zip	Juan Pablo Ruiz	Mar 19, 2019	164 MB
dataInOpenSource.7z	Juan Pablo Ruiz	9:14 AM	19 MB
dataSources_Fpga_2019_01_28.zip	Juan Pablo Ruiz	Jan 28, 2019	162 MB

References I

-  Chen, C. (2006).
Citespace ii: Detecting and visualizing emerging trends and transient patterns in scientific literature.
Journal of the American Society for information Science and Technology, 57(3):359–377.
-  Ruiz-Rosero, J., Ramirez-Gonzalez, G., Williams, J. M., Liu, H., Khanna, R., and Pisharody, G. (2017).
Internet of things: A scientometric review.
Symmetry, 9(12).

Preguntas

