ScientoPy v 2.0.1, User Manual

Juan Pablo Ruiz Rosero
jpabloruiz@unicauca.edu.co

Contents

1 Installation

1. Download the latest stable release from: `https://github.com/jpruiz84/ScientoPy/releases`

2. Unzip the downloaded file and run ScientoPyGui.exe

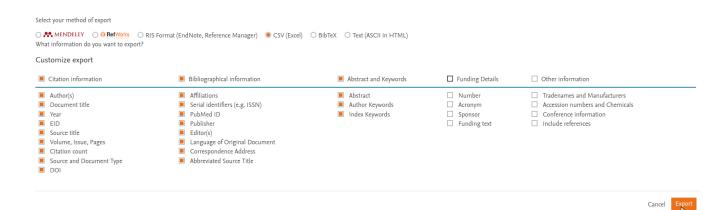# 2 Download the bibliometric dataset

This section describes how to download the proper dataset from Scopus and WoS. Define a search criteria that will be used for Scopus and WoS. For this guide and for the example dataset we are using: "Internet of thing" AND "Gateway"

2.1 Download the dataset from Scopus

1. Make your search with the defined search criteria for Article title, Abstract, Keywords.

2. Select all the results and click on Export:

3. Select as method of export **CSV (Excel)**, and select the Customize export **Citation information, Bibliographical information, Abstract and Keywords**, then click on Export:
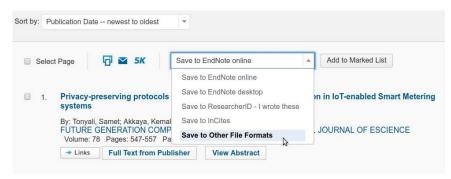
4. Save the file on the folder `/ScientoPy/dataIn`
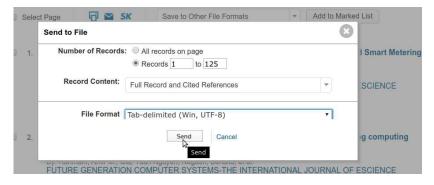
2.2 Download the dataset from WoS

1. Make your search with the defined search criteria for Topic.

2. Select **Save in Other File Formats**

3. Select the number of records to download, on Record Contented select **Full Record and Cited References**, on File Format select **Tab-delimited (Win, UTF-8)**, and click on Send.

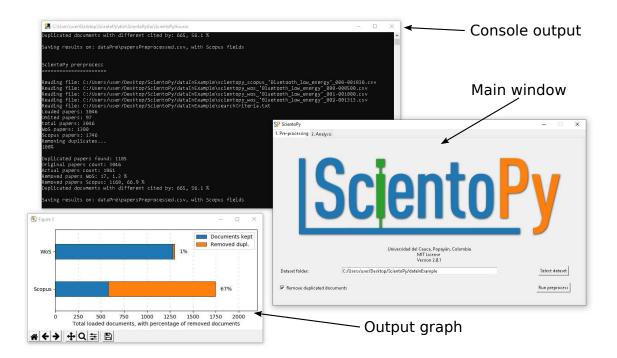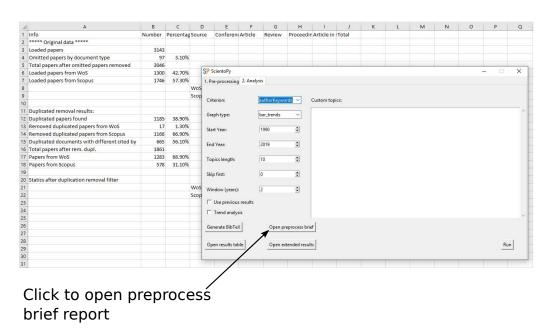4. Save the file on the folder /ScientoPy/dataIn

3 Preprocessing

First we need to preprocess the downloaded dataset. This preprocess merge all the downloaded files from the dataset folder to a single file. Also, this process remove the duplicated files. To preprocess the example dataset ("Bluetooth low energy" downloaded in March 21, 2019, located in dataInExample) open ScientoPyGui.exe and follow the next steps:

1. Click in `Select dataset` button, and select the folder that contains the downloaded dataset. For the case of run the example dataset, select the folder *dataInExample* inside ScientoPy unziped folder.

2. Click in `Run preprocess` button. This will start the preprocess. You can check the preprocess status and progress in the console output as shown in the following picture

3. After the preprocess is finished, the preprocess brief table will show up as the *Output graph*.

Console output

Main window

Output graph

Then, select the `2. Analysis` tab, and click in `Open preprocess brief` button. This will open the preprocess CSV file in Excel, as show in the following figure.
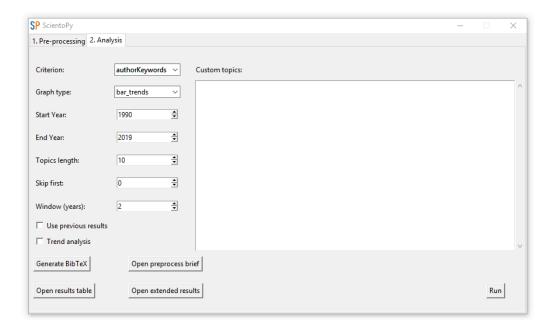

Click to open preprocess
brief report

As result, inside the folder `ScientoPy/dataPre` you will find the following files:

- **papersPreprocessed.tsv:** this file contains the information of all papers after the pre-process. This file will be used by the others scripts as the input data.
- **PreprocessedBrief.tsv:** this file briefs the pre-process statics results, such as duplicated papers removed, types of documents, and others.

4   Analysis page

The analysis page let us to perform the scientometric analysis operations. To access to this page, click in the `2. Analysis` tab. Then, you will see the following information described below:

- **Criterion:** let the user to select the search criterion, such as author keywords, or author names.
- **Graph type:** define the type of the output graph.
- **Start year:** limit the start year for the analysis and graphs.
- **End year:** limit the end year for the analysis and graphs.
- **Topics length:** number of top topics to find in the selected criterion.
- **Skip first:** number of first top topics to skip during the analysis.
- **Window (years):** window width in years for growth indicators.
- **Use previous results:** run the analysis based on the previous results.
- **Trend analysis:** get the top trending topics, with the highest average growth rate.
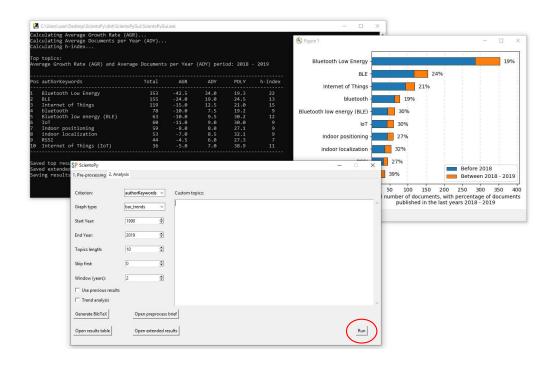- **Custom topics:** to enter custom topics for the analysis instead of the top topics.

The ScientoPy criterion are described on Table 1:

Table 1: ScientoPy criterion description

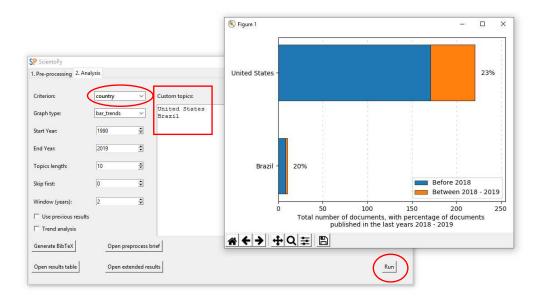| Criteron	Description
author	Authors last name and first name initial
sourceTitle	Publication or journal name
subject	Research areas, only from WoS documents
authorKeywords	Author keywords
indexKeywords	Keywords generated by the index, from WoS {Keyword Plus}, and from Scopus {Indexed keywords}
bothKeywords	AuthorKeywords and indexKeywords are used for this search
abstract	Document abstract, for use with pre-defined topics and asterisk wildcard
documentType	Type of document
dataBase	Database where the document was extracted (WoS or Scopus)
country	Country extracted from authors affiliations
institution	Institution extracted from authors affiliations
institutionWithCountry	Institution with country extracted from authors affiliations

4.1 Extract top topics

To find the top 10 author keywords, kept the default fileds in the analysis tab and push the Run button (as shown in the following figure). This will generate a list with the top 10 topics on the selected criterion (in this case authorKeywords), with the number of documents per topic. Also, this script show the bar graph with the percentage of documents in the last years, and saves the quantitative results on the folder ScientoPy/results.

Also, the list of the top author keywords is printed in the *Output console*. In the same way, you can push the `Open results table` to see the quantitative results in a CSV file, or click in `Open extended results` to see the CSV file with the papers related to each author keyword.

4.2   Analyze custom topics inside a criterion

If you want to make an analysis of custom topics, such as the two selected countries papers evolution, you can write these in the `Custom topics` text box, as shown in the following figure:
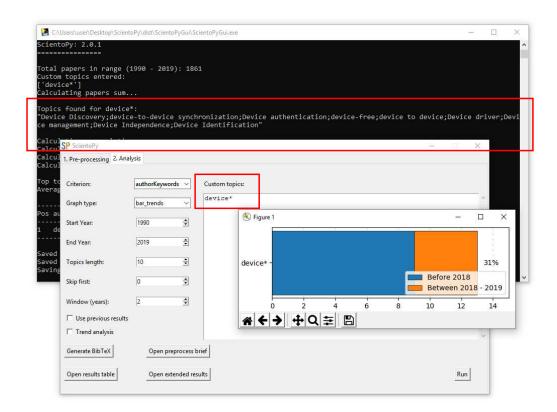

You can analyze any topic in any criterion. Also, you can integrate two or more topics in one, by dividing it with `,`. This is very useful for abbreviations and plural singulars, for example:
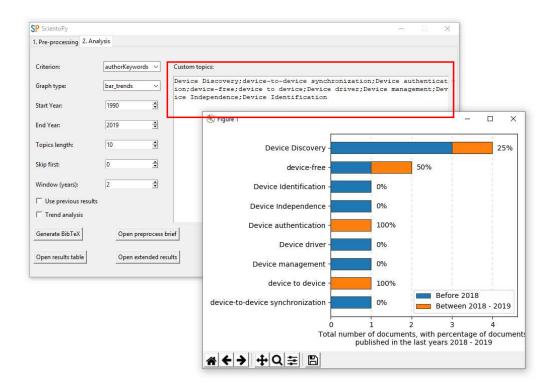
```
WSN, Wireless sensor network, Wireless sensor networks
RFID, RADIO FREQUENCY IDENTIFICATION
```

4.3   Asterisk (*) wildcard

You can use the asterisk wildcard to find phrases or words which starts or ends with the letters that you have inserted. For example, if you want to find "device", "devices", and "device integration", entering `device*` in the `Custom topics` text box:
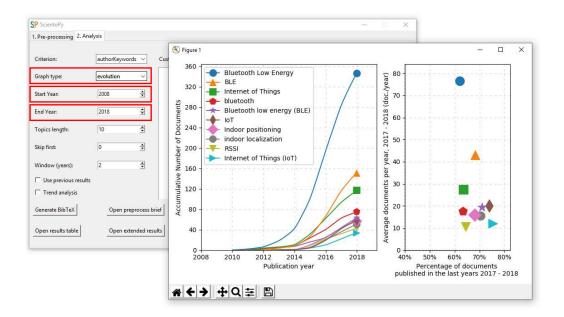
As shown in the firs red square, ScientoPy prints the topics found for the wildcard based search. You can copy this information, to analyze each specific topic found, like this (note, the ; is also used for topics separation):

4.4 Evolution plot

Also, you can see the results with an evolution plot. This option plots the accumulative documents, average documents per year (ADY), and percentage of documents in the last years, for example:

Note: select the right start and end years limit that fits better your plot lines.

4.5 Finding trending topics

This script finds the top trending topics based on the higher average growth rate (AGR) over the others. The AGR is calculated on two years periods, using the following Equation (4.5):

$$AGR = \frac{\sum\limits_{i=Y_s}^{Y_e} P_i - P_{i-1}}{(Y_e - Y_s) + 1},$$
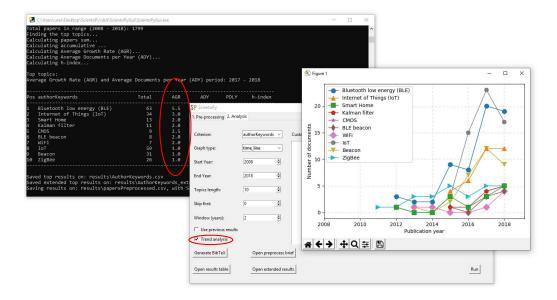
where:
AGR = Average growth rate;
Y_s = Start year;
Y_e = End year;
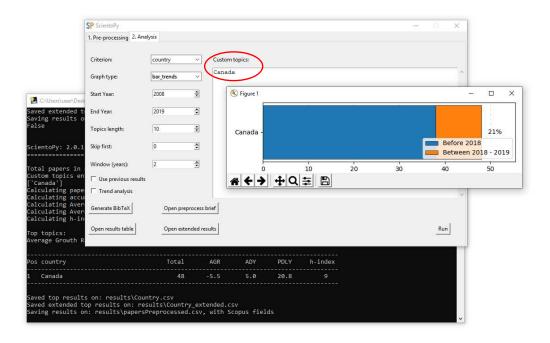P_i = Number of publications on year i.

To find the top trending topics on author keywords criterion, you can run the following configuration:
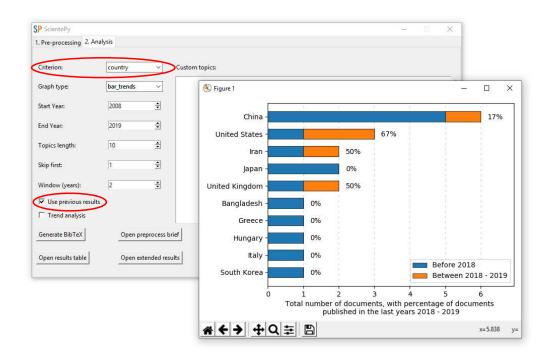
10

This script will find the top 200 topics, then it calculates the AGR for the last 2 years (*Window (years): 2*). Finally, the 200 top topics are sorted from the highest AGR in the last 2 year period to the lower.

4.6   Analysis based on the previous results

ScientoPy generates an output file with all the output documents from the last run script. For example if we run the following configuration:

ScientoPy will create a documents output file (`results/papersPreprocessed.tsv`) with all documents that have authors with affiliation in Canada. This output file can be used by ScientoPy to perform an analysis based on this, in that way if we check the option `Use previous results`, and we select in `Criterion` authorKeywords we will obtain the top author keywords from papers where the author affiliation correspond to Canada. Also, we can run the following configuration to know which are the countries that have more common documents with Canada:
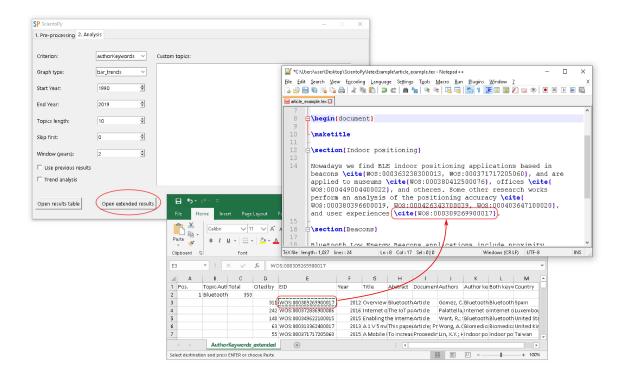
Note: the ScientoPy documents output file is only generated when the `Use previous results` is not marked. In that way, if we run many times a ScientoPy command with this option, the documents output file will not overwritten.
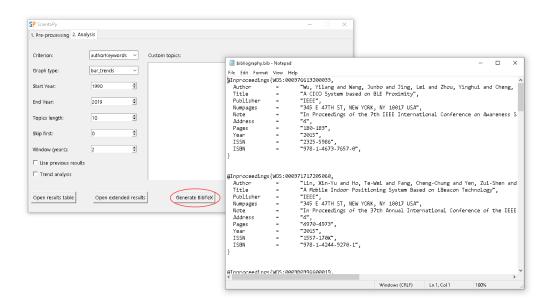
5   Generate BibTeX

With ScientoPy you can generate a LaTeXBibTeX file for your reference. For this follow the next steps:

1. Open the extended results from your analysis, by clicking the `Open extended results` button.

2. Copy the EID value of the paper or papers that you want to cite to the latex document inside the `\cite{}` preamble, as shown in the following figure:

3. Save your LaTeX document

4. Push the button `Generate BibTeX` and select the LaTeX document where you have added the EID

5. ScientoPy will look for all EID in the document and will generate the BibTeX information from the dataset preprocessed. Finally it will open the BibTeX file in a text editor in order, as shown in the following figure:
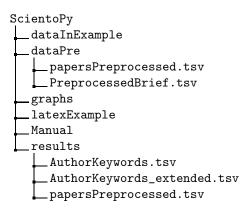

Note: for your reference we have included an article LaTeX file example in the location `latexExample/article_example.tex`. This have cites from the example dataset. To test the Generate BibTeX capability, you can follow the next steps:

- Preprocess the example dataset located in `dataInExample`,

- Click in `Generate BibTeX`

- Select the file `latexExample/article_example.tex`

- Save the generated `bibliography.bib` file in the `latexExample` folder

- Compile the `latexExample/article_example.tex` file, with the command `latexmk -pdf article_example.tex`

6 Output files and directories

After run ScientoPy analysisyou will find the following folder and files structure:

```
ScientoPy
├──dataInExample
├──dataPre
│  ├──papersPreprocessed.tsv
│  └──PreprocessedBrief.tsv
├──graphs
├──latexExample
├──Manual
└──results
   ├──AuthorKeywords.tsv
   ├──AuthorKeywords_extended.tsv
   └──papersPreprocessed.tsv
```

These folders and output files are described bellow:

- **dataInExample:** contains Scopus and WoS example data set for the search criteria "Internet of things" AND "Gateway" downloaded in 27 November 2017. This is the input example for preprocess script.

- **dataPre:** output folder for the preprocess results, and input folder for scientoPy script.

- **papersPreprocessed.tsv:** preprocesed papers data with all input documents merged, filtered, and duplication removed. This is the input file that scientoPy script uses.

- **PreprocessedBrief.tsv:** preproceses brief table that shows the preprocess results related to total papers found per data base, the omitted papers, the duplicated papers count per data base, and the total number of papers per paper type (Conference paper, article, review...)

- **graphs:** graphs output folder for preprocess and scientoPy scripts

- **latexExample:** folder with the LaTeX file article example to generate the BibTeX from the preprocess dataset.

- **Manual:** folder with the pdf manual and example paper with scientoPy commands highlighted used for graph and tables generation.

- **results:** output folder for scientoPy result output files

- **AuthorKeywords.tsv:** scientoPy output file for the selected criterion (in this case authorKeywords) that shows the top topics or the custom topics with the total number of documents, the Average Growth Rate (AGR), the Average Documents per Year (ADY), the h-index, and the documents per each year.

- **AuthorKeywords_extended.tsv:** scientoPy output file for the selected criterion (in this case author-Keywords) that show the top or custom topics with the documents related to each one.

- **papersPreprocessed.tsv:** inside the results folder, this file contains the output papers from the last scientoPy used script. This is used as an input for scientoPy script when we mark the option `Use previous results`.

7   ScientoPy graph types

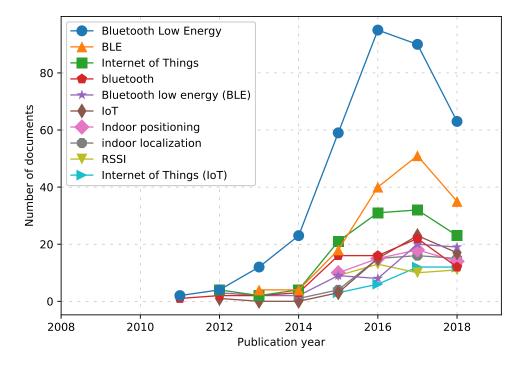ScientoPy has 5 different ways to graph the results described on Table 2.

Table 2: ScientoPy output graphs types

Graph type	Abreviation	Description
Time line	`time_line`	Graphs the number of documents of each topic vs the publication year
Horizontal bars	`bar`	Graphs the total number of documents of each topic in horizontal bars
Horizontal bars trends	`bar_trends`	Graphs the total number of documents of each topic in horizontal bars, with the percentage of document published in the last years
Evolution	`evolution`	Graphs two plots, one with the accumulative number of documents vs the publication year, and other with the average papers per year vs the percentage of documents in the last years
Word cloud	`word_cloud`	Generate a word cloud based on the topic total number of publications

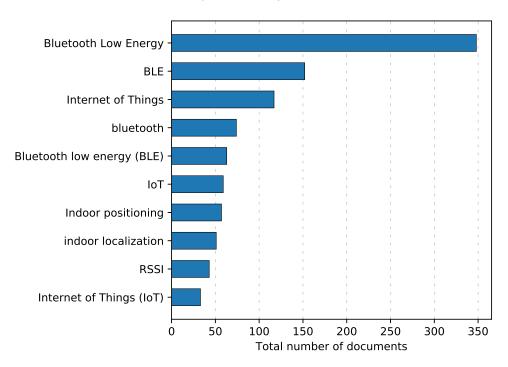Below are showed some examples of these graphs types, with the command used.

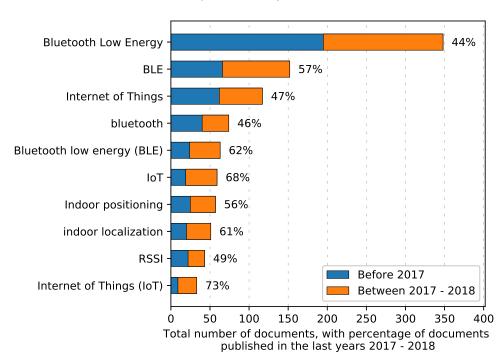## 7.1   Time line graph

Top author keywords

7.2 Horizontal bars graph
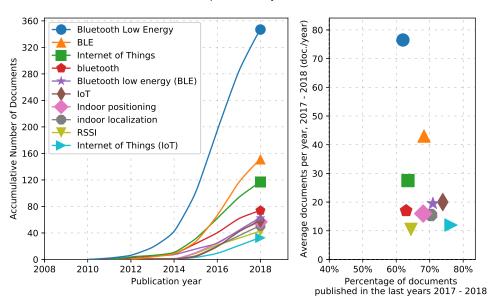
Top author keywords

7.3 Horizontal bars trends

Top author keywords

7.4   Evolution graph

Top author keywords

7.5   Word cloud graph