Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Programa de Ingeniería Electrónica y Telecomunicaciones
ScientoPy, Installation and User Manual

By: Juan Pablo Ruiz Rosero jpabloruiz@unicauca.edu.co

1. Installation

1. Download and install the last version of Python 2.7 (for example Python 2.7.14) from: `https://www.python.org/downloads/`
2. Install the matplotlib library for Python the automatic installation tool **pip**. From Windows, enter in the command line (Windows + R, cmd, and Enter), go to the folder `C:\Python27\Scripts`, and run the installation script:

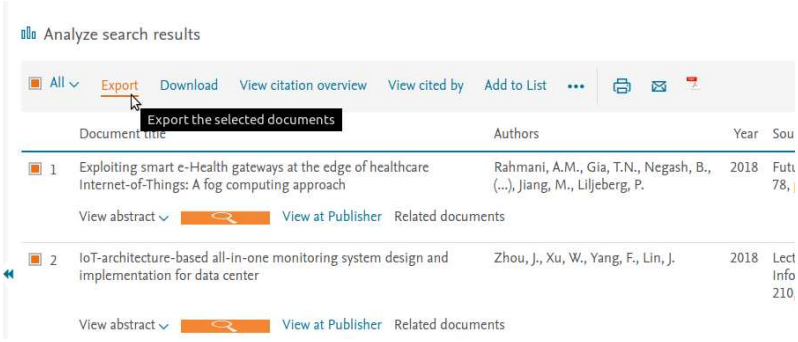
```
cd C:\Python27\Scripts
pip install matplotlib
```

2. Download the bibliometric dataset

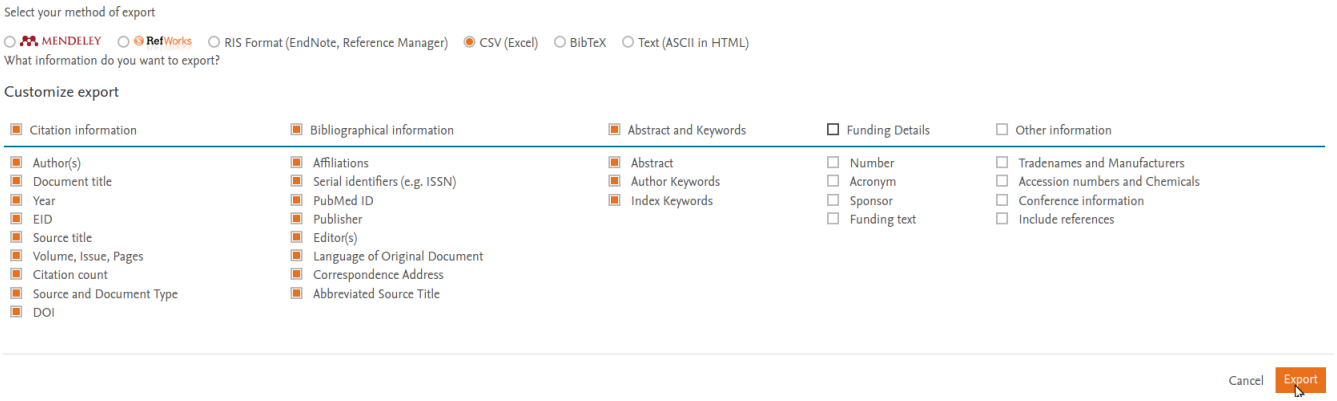
This section describes how to download the proper dataset from Scopus and WoS. Define a search criteria, it will be used for Scopus and WoS. For this guide we are using: Ïnternet of thing.ᴬND "Gateway"

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:



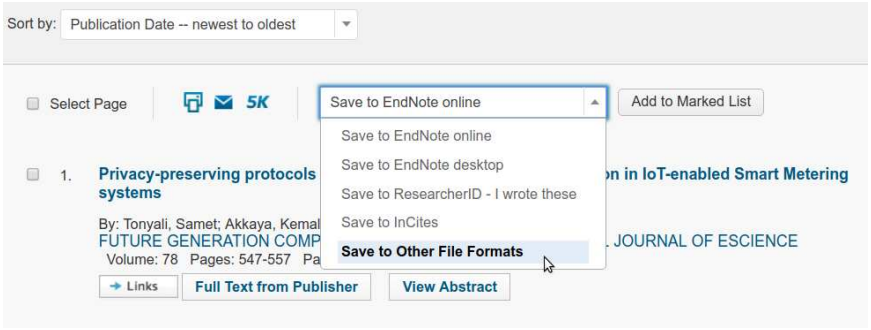
3. Select as method of export **CSV (Excel)**, and select the Customize export **Citation information, Bibliographical information, Abstract and Keywords**, then click on Export:



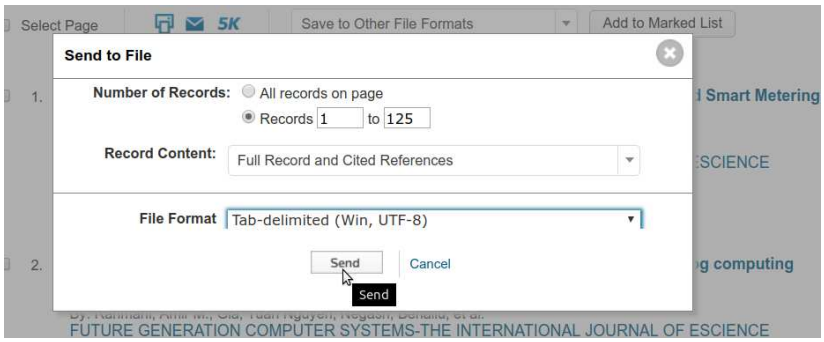
4. Save the file on the folder `/ScientoPy/dataIn`

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. Running the ScientoPy scripts

*3.1. Pre-processing*First we need to pre-process the downloaded data. This pre process joint all the downloaded files from one folder to a single file. Also, this process remove the duplicated files. To pre-process the example dataset run this command inside ScientoPy folder:

```
python preProcess.py dataInExample
```

On the folder `ScientoPy/dataPre` you will find the following files:

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

To find more options of the pre-processing script you can run:

```
python preProcess.py -h
```

*3.2. Extract the top topics*With this script you can extract the top topics of a selected criterion. The ScientoPy script criteria are:

- authors
- source
- subject
- authorKeywords
- indexKeywords
- documentType
- dataBase
- country

For example, to find the top author's keywords you can run this script:

```
python topResults.py authorKeywords
```

This will generate a list with the top 10 topics on the criterion author's keywords, with the number of documents per topic, and the h-index associated to each one. Also, this script will graph the evolution of each topic across the year, and will save the quantitative results on the folder `ScientoPy/results`.

This script have more options like, save the plot on a file, or increase the number of topic results. For more information you can run:

```
python topResults.py -h
```

*3.3. Analyze pre defined topics inside a criterion*If you want to make an analysis of pre defined topics, such as the number of papers evolution of two countries, you should use the `analizeTopic.py` script:

```
python analizeTopic.py country -t "United States; Brazil"
```

You can analyze any topic in any criterion. Put the topics on the `-t` argument. Divide the topics with the `;`. 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:

```
python analizeTopic.py authorKeywords -t \
"WSN, Wireless sensor network, Wireless sensor networks; RFID, RADIO FREQUENCY IDENTIFICATION"
```

This script have more options like, save the plot on a file, or others. For more information you can run:

```
python analizeTopic.py -h
```