

## Curating data with OpenRefine to visualize them in VOSviewer

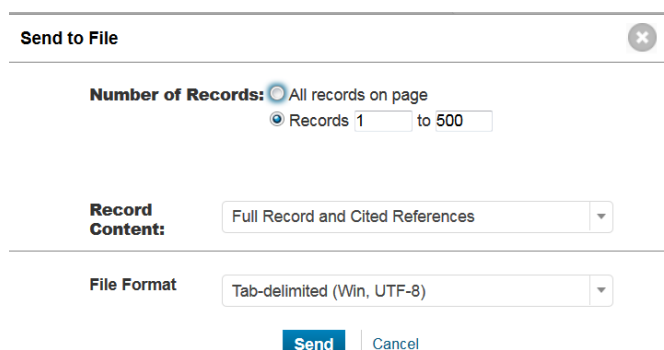
If you choose to create a map based on bibliographic data to visualize keywords co-occurrences in VOSviewer you will visualize keywords as they are indexed in the Web of Science or Scopus. If written in a different manner, occurrences of the same term will be considered as different items: e.g. "climate-change" and "climate change", or "city" and "cities", while they are the same concept.

A small number of this kind of errors should not affect too much your visualization in VOSviewer but too many of them could blur the picture and diminish the relevance of the map. Most of these issues can be solved in few steps thanks to OpenRefine. This tutorial explains how.

### Curating data from Web of Science

#### Exporting data from the Web of Science

Export data from the WoS as you usually do when working with VOSviewer. As file format, select the **Tab-delimited format (Win, UTF-8)**.



Send to File

Number of Records:  All records on page  
 Records 1 to 500

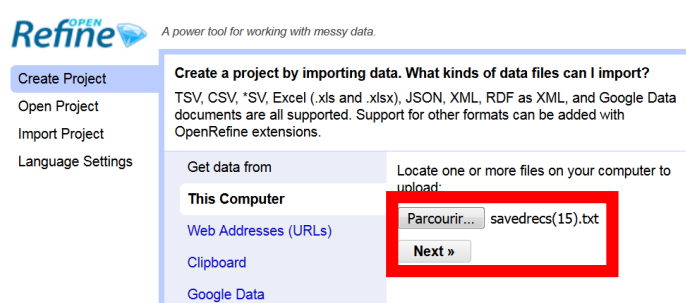
Record Content: Full Record and Cited References

File Format: Tab-delimited (Win, UTF-8)

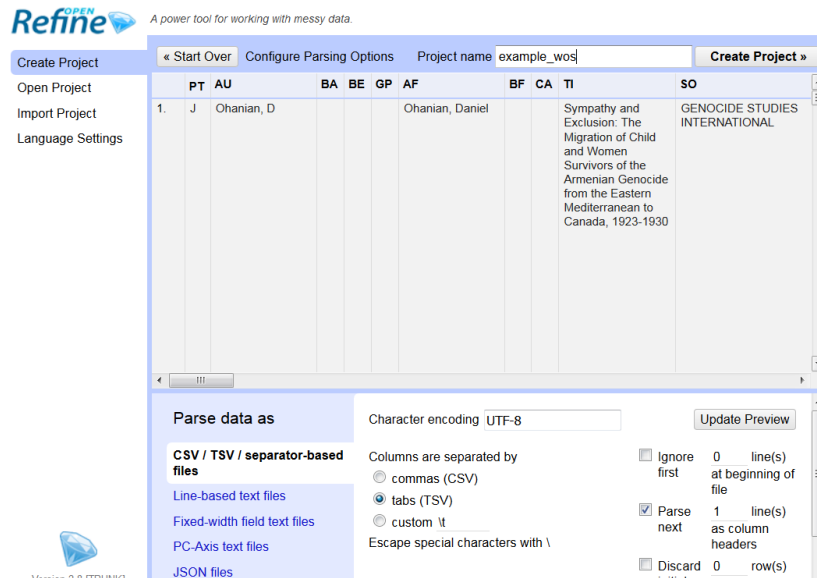
Send Cancel

#### Creating a project in OpenRefine

Download OpenRefine on your computer and launch the software. Then upload the file you want to clean in OpenRefine and click on **Next**.

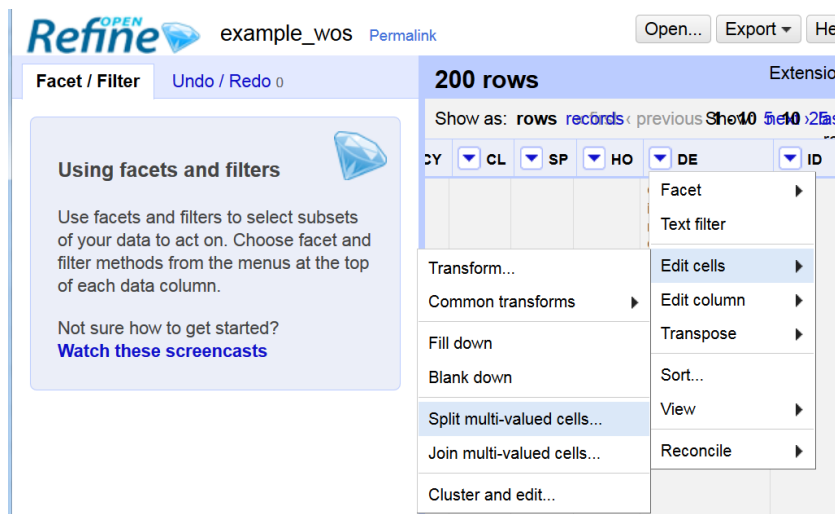


Then you must make sure that the different fields of the file are well recognized. Select the following options: CSV/TSV/separator-based files; **separated by tabs** (TSV); then choose **UTF-8** as the Encoding character and Parse next **1 line** as column header and create your project.

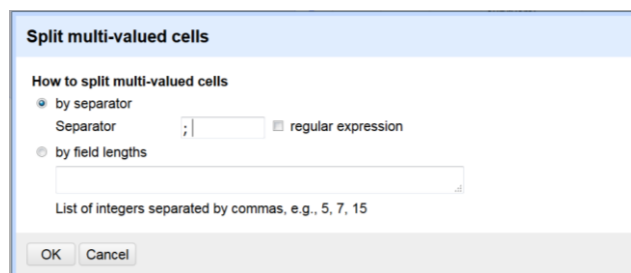


Once your project is created, each row contains information about a publication and each column is a metadata. The first line of the WoS file is the column header.

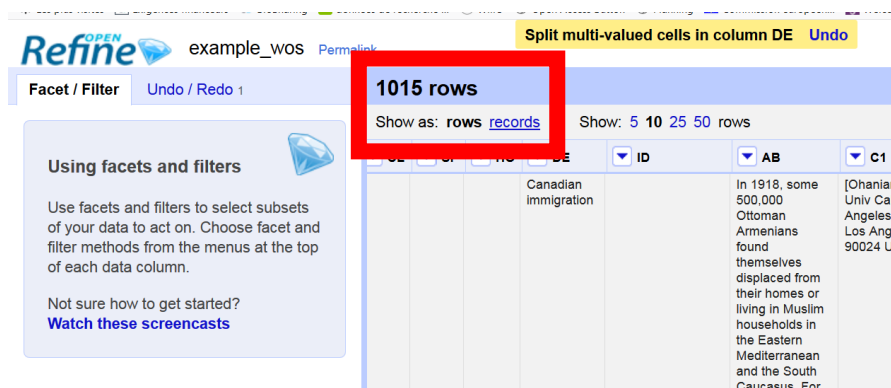
In the column "DE", go to "Edit cells", "Split multi-valued cells".



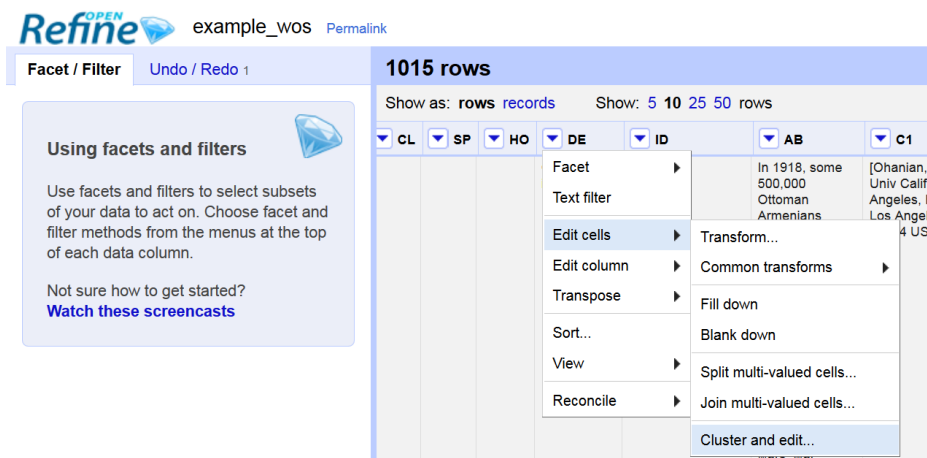
As separator, type : "; ". Beware not forgetting the space after the semi-colon.



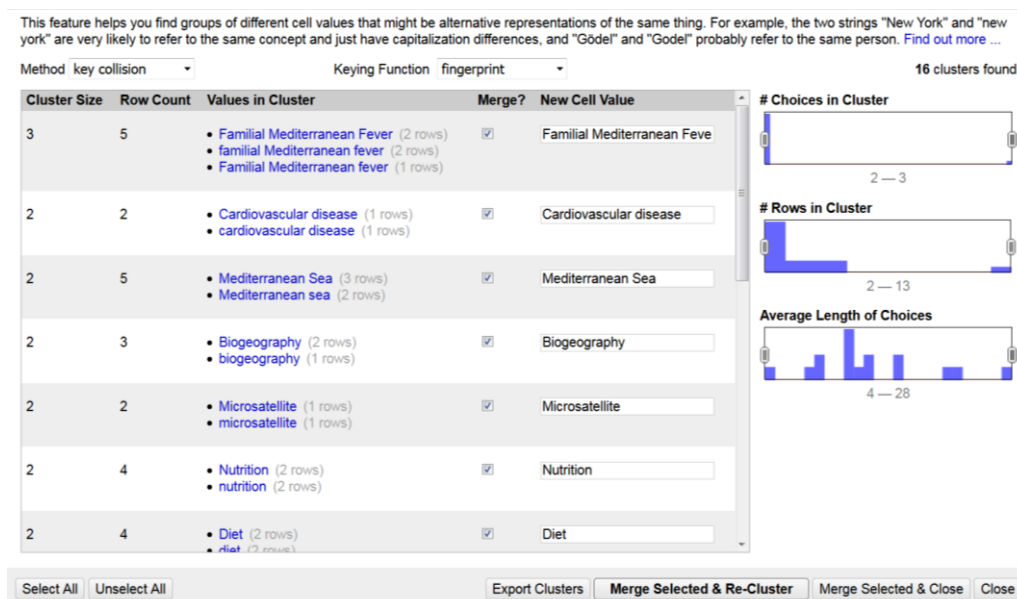
From now you have a difference between the number of rows (keywords) and the number of records (publications).



Then, still in the column "DE", go to "Edit cells" and "Cluster and edit".

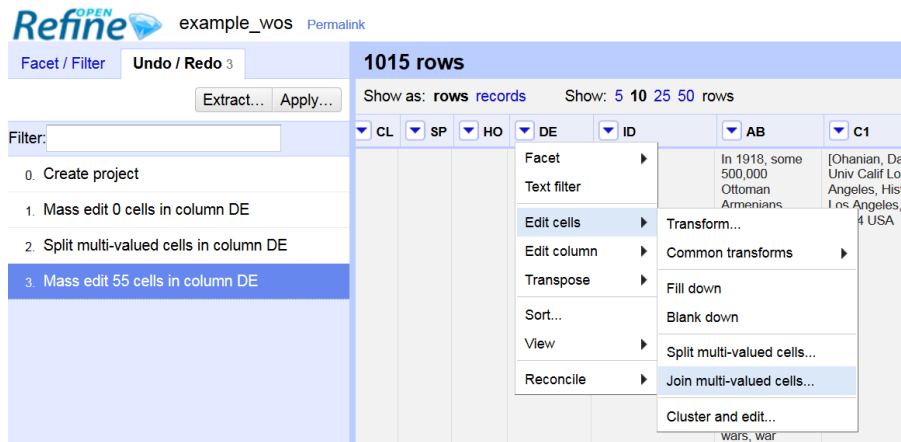


A new window opens. You can select which terms you want to merge. First click on "Select all", then uncheck any words whose merging is irrelevant. By default, the new cell value is the value that occurs the most often but you may desire to change it. Once you made your choices, click on "Merge selected & re-cluster".

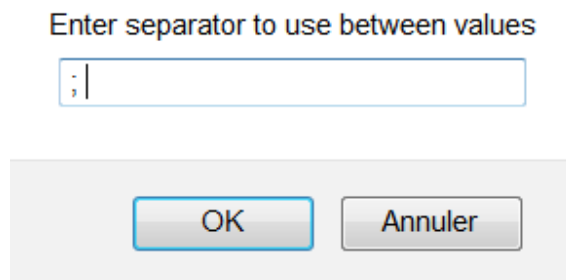


A second method of clustering is available: "nearest neighbor", you may try it. Then close this window.

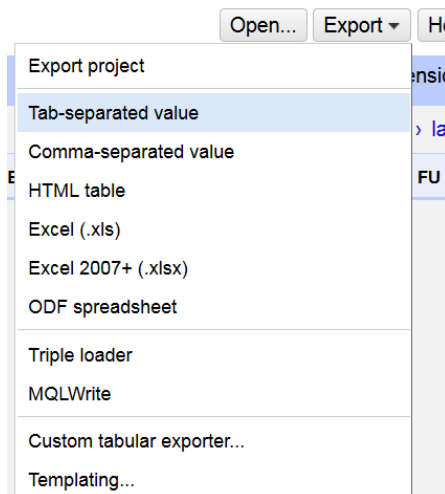
Once this is done, in the column "DE", go to "Edit cells", "Join multi-valued cells".



Enter ";" as separator.



You should now have as many rows as records. You can now save your project and export the file in TSV (Tab-separated value).



This file should be readable in VOSviewer as an original Web of Science file. The same cleaning may be done with Scopus files, in a slightly different way.

## Curating data from Scopus

### Exporting data from Scopus

Export data from Scopus as you usually do when working with VOSviewer in a comma-separated value format (CSV).

Export document settings

You have chosen to export 44530 documents

Select your method of export

MENDELEY  RefWorks  RIS Format  CSV  BibTeX  Plain Text

EndNote, Reference Manag Excel ASCII in HTML

What information do you want to export?

Citation information  Bibliographical information  Abstract & keywords  Funding details  Other information

Author(s)  Affiliations  Abstract  Number  Tradenames & manufacturers

Document title  Serial identifiers (e.g. ISSN)  Author keywords  Acronym  Accession numbers & chemicals

Year  PubMed ID  Index keywords  Sponsor  Conference information

EID  Publisher  Funding text  Include references

Source title  Editor(s)  Language of original document

volume, issue, pages  Correspondence address

Citation count  Abbreviated source title

Source & document type

DOI

Cancel **Export**

### Creating a project in OpenRefine

Upload the file you want to clean in OpenRefine and click on **Next**. Choose UTF-8 as character encoding and CSV separator-based file and create the project.

Refine A power tool for working with messy data.

Create Project « Start Over Configure Parsing Options Project name scopus 45 csv Create Project »

	Authors	Title	Year	Source title	Volume	Issue	Art. No.	Page start
1.	Eberhart Russell, Kennedy James	New optimizer using particle swarm theory	1995	Proceedings of the International Symposium on Micro Machine and Human Science				39
2.	Swain M.J., Ballard D.H.	Color indexing	1991	International Journal of	7	1		11

Parse data as

Character encoding UTF-8 Update Preview

CSV / TSV / separator-based files

Columns are separated by

commas (CSV)  tabs (TSV)  custom ,

Escape special characters with \

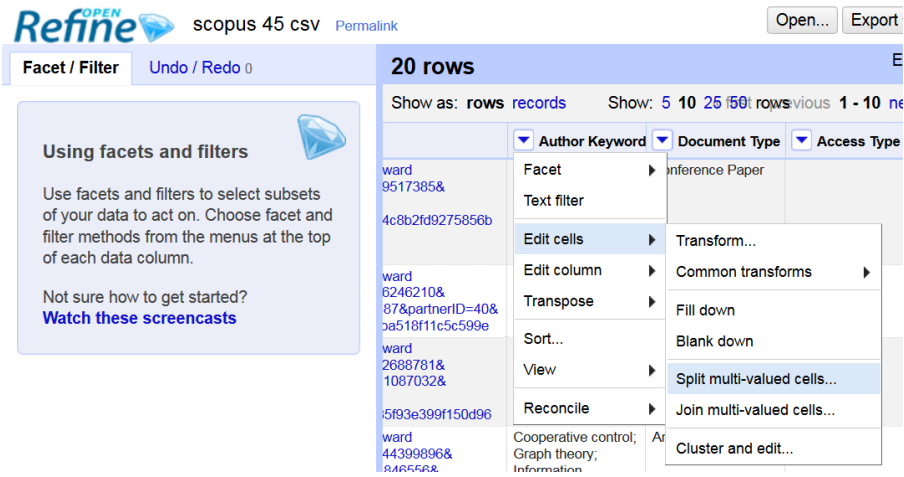
Ignore first 0 line(s) at beginning of file

Parse next 1 line(s) as column headers

Discard initial 0 row(s) of data

Load at most 0 row(s) of data

In the column « Author keywords », « Edit cells », and click on « **Split multi-valued cells** », write ";" as separator. Then merge the terms in the same way than with the Web of Science file.

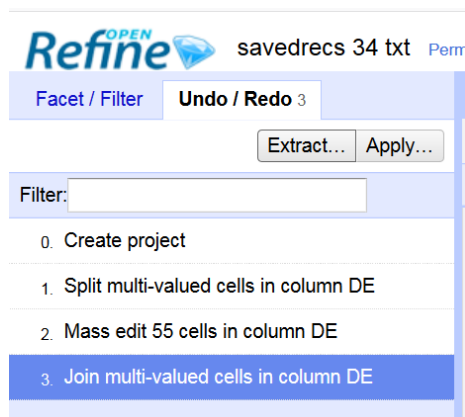


Finally join the multi-valued cells with “; ” as separator and export the file as a csv file. This file should be readable in VOSviewer.

## Clean several files with the same rules

If you want to view several files in VOSviewer, the files curation in OpenRefine could take some time, and the merging choices could not always be the same since the maximum number of occurrences for a concept may change from one file to the other.

To be sure to apply always the same rules you can go in the “Undo/Redo” tab and click on “Extract”.



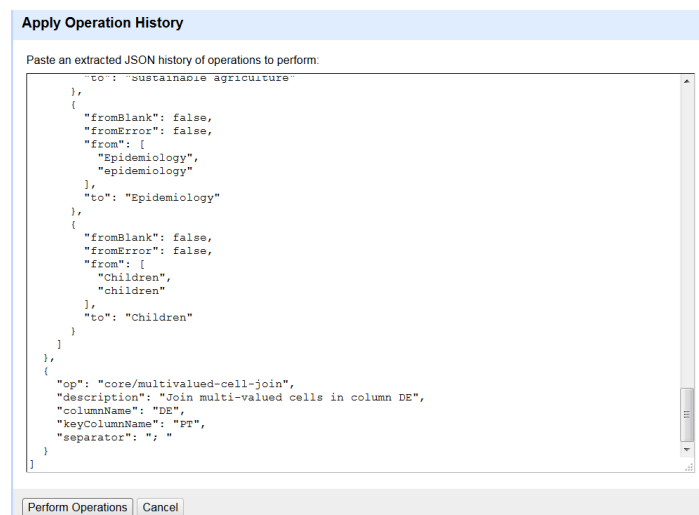
Then the different operations you have just made appear in JSON format. You may copy-paste it as text.



Then, when you have created a new project for another file, in the “Undo/Redo” tab, click on “Apply”.



Paste the text with the history of operations you have done with the previous file and click on “**Perform operations**”.



The file has been curated in the same way than the previous one. You may now export it in the format you want to read it in VOSviewer.