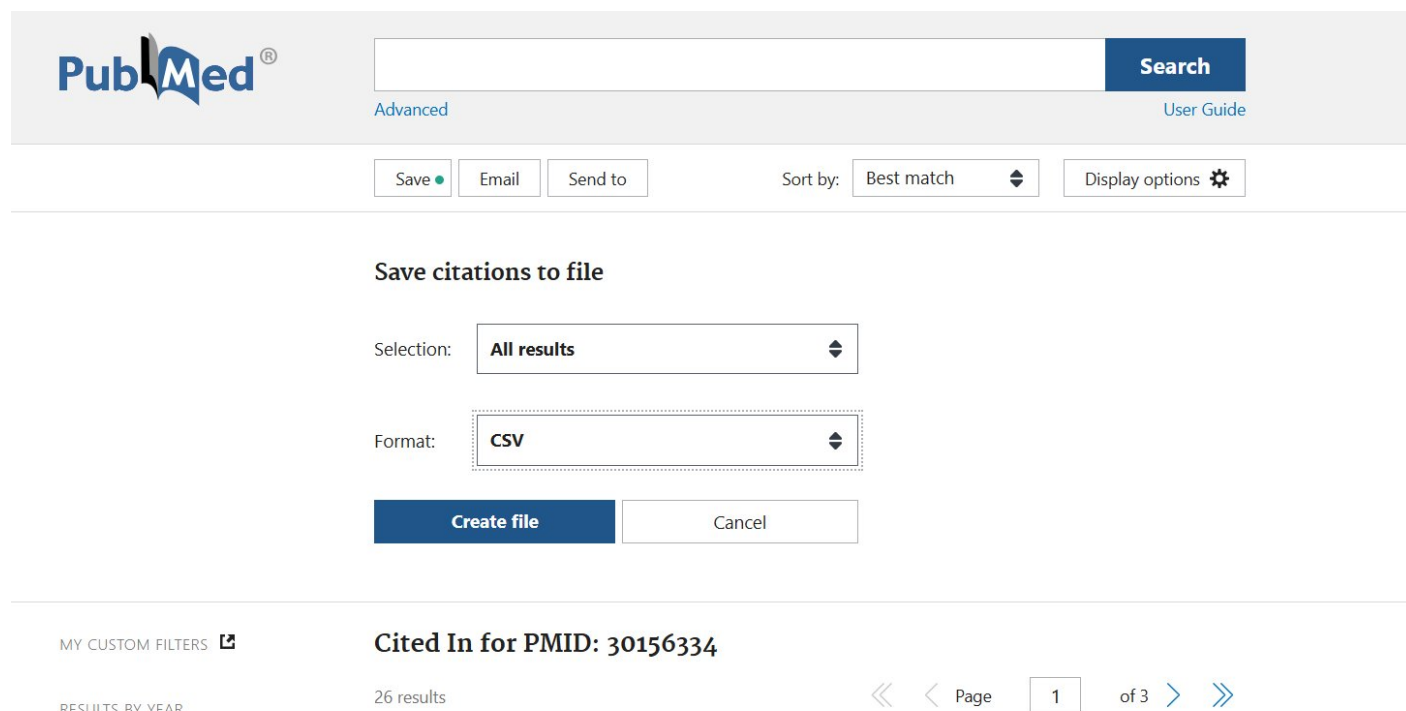


csvcut

Extracting relevant columns from csv and tab delimited files.

Exemplary workflow for quick look of papers from pubmed “similar or cited by”

Download the citation/similar articles in csv format from pubmed



The screenshot shows the PubMed search interface. At the top, there is a search bar with the PubMed logo on the left and a 'Search' button on the right. Below the search bar, there are links for 'Advanced' and 'User Guide'. A row of buttons includes 'Save', 'Email', 'Send to', 'Sort by: Best match', and 'Display options'. Below this, a section titled 'Save citations to file' contains a 'Selection:' dropdown menu set to 'All results' and a 'Format:' dropdown menu set to 'CSV'. At the bottom of this section are 'Create file' and 'Cancel' buttons. Below the 'Save citations to file' section, there is a link for 'MY CUSTOM FILTERS' and a section titled 'Cited In for PMID: 30156334'. Below this, there is a link for 'RESULTS BY YEAR', the text '26 results', and a pagination control showing 'Page 1 of 3' with navigation arrows.

Check the number of lines in the file

```
wc -l file.csv
```

See the names of columns of the file

```
csvcut -n file.csv
```

Create an output for a quick visual check with title and doi, most recent first

Option 1: for the output in terminal, comma separated

```
cat file.csv | csvsort -r -c 7 | csvcut -c 2,11 -l
```

Option 2: for the output in terminal, a bit more nice (?)

```
cat file.csv | csvsort -r -c 7 | csvcut -c 2,11 -l | csvlook
```

Option 3: create a table using pandoc in landscape format

```
cat file.csv | csvsort -r -c 7 | csvcut -c 2,11 -l | csvlook | pandoc --variable geometry:"landscape, margin=
```

Option 4: create a table using pandoc in default (portrait) format

```
cat file.csv | csvsort -r -c 7 | csvcut -c 2,11 -l | csvlook | pandoc -o table_portrait.pdf
```

Outputs:

1. Articles count

```
vld@LAPTOP-B4G00IKB MINGW64 ~/Documents/Grnt_GACR-Zdenka_2026/papers
$ wc -l file.csv
27 file.csv
```

2. Column names

```
vld@LAPTOP-B4G00IKB MINGW64 ~/Documents/Grnt_GACR-Zdenka_2026/papers
$ csvcut -n file.csv
1: PMID
2: Title
3: Authors
4: Citation
5: First Author
6: Journal/Book
7: Publication Year
8: Create Date
9: PMCID
10: NIHMS ID
11: DOI
```

3. Table in terminal (Option 1)

```
vld@LAPTOP-B4G00IKB MINGW64 ~/Documents/Grnt_GACR-Zdenka_2026/papers
$ cat file.csv | csvsort -r -c 7 | csvcut -c 2,11 -l
line_number,Title,DOI
1,Topological defects in self-assembled patterns of mesenchymal stromal cells in vitro are predictive attributes of condensation and chondrogenesis,10.1371/journal.pone.0297769
2,Confinement promotes nematic alignment of spindle-shaped cells during Drosophila embryogenesis,10.1242/dev.202577
3,Socket Array Irregularities and Wing Membrane Distortions at the Eyespot Foci of Butterfly Wings Suggest Mechanical Signals for Color Pattern Determination,10.3390/insects15070535
```

3. Table in terminal (Option 2)

```
vld@LAPTOP-B4G00IKB MINGW64 ~/Documents/Grnt_GACR-Zdenka_2026/papers
$ cat file.csv | csvsort -r -c 7 | csvcut -c 2,11 -l | csvlook
C:\Users\vld\AppData\Local\Packages\PythonSoftwareFoundation.Python.3.12_qbz5n2kfra8p0\LocalCache\local-packages\Python312\site-packages\agate\table\from_csv.py:83: RuntimeWarning: Error sniffing CSV dialect: Could not determine delimiter
  kwargs['dialect'] = csv.Sniffer().sniff(sample)
| line_number | Title | |
|---|---|---|
| 1 | Topological defects in self-assembled patterns of mesenchymal stromal cells in vitro are predictive attributes of condensation and chondrogenesis | 10.1371/journal.pone.0297769 |
| 2 | Confinement promotes nematic alignment of spindle-shaped cells during Drosophila embryogenesis | 10.1242/dev.202577 |
| 3 | Socket Array Irregularities and Wing Membrane Distortions at the Eyespot Foci of Butterfly Wings Suggest Mechanical Signals for Color Pattern Determination | 10.3390/insects15070535 |
```

3. Table in pdf landscape (Option 3)

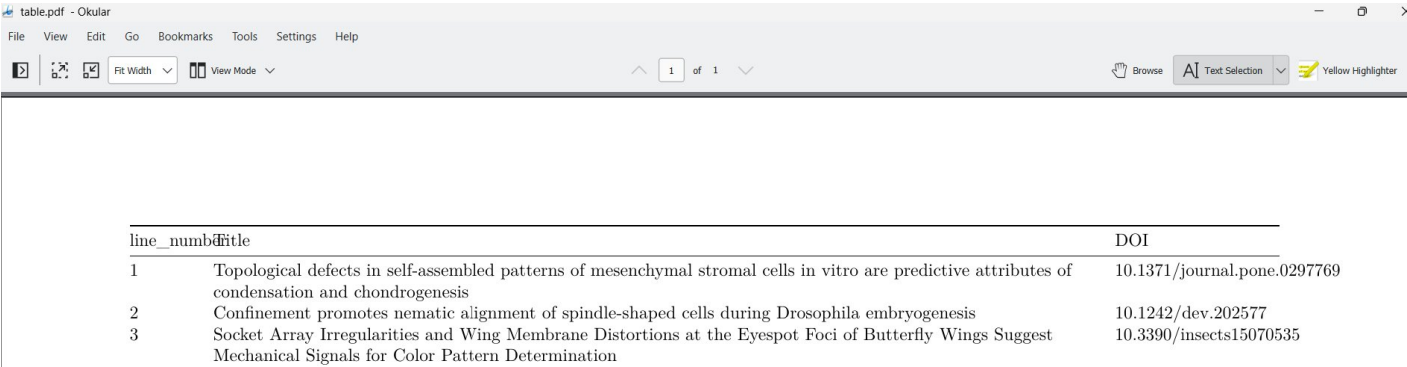


table.pdf - Okular

File View Edit Go Bookmarks Tools Settings Help

Fit Width View Mode

1 of 1

Browse Text Selection Yellow Highlighter

line_num	title	DOI
1	Topological defects in self-assembled patterns of mesenchymal stromal cells in vitro are predictive attributes of condensation and chondrogenesis	10.1371/journal.pone.0297769
2	Confinement promotes nematic alignment of spindle-shaped cells during Drosophila embryogenesis	10.1242/dev.202577
3	Socket Array Irregularities and Wing Membrane Distortions at the Eyespot Foci of Butterfly Wings Suggest Mechanical Signals for Color Pattern Determination	10.3390/insects15070535

3. Table in pdf portrait (Option 4)

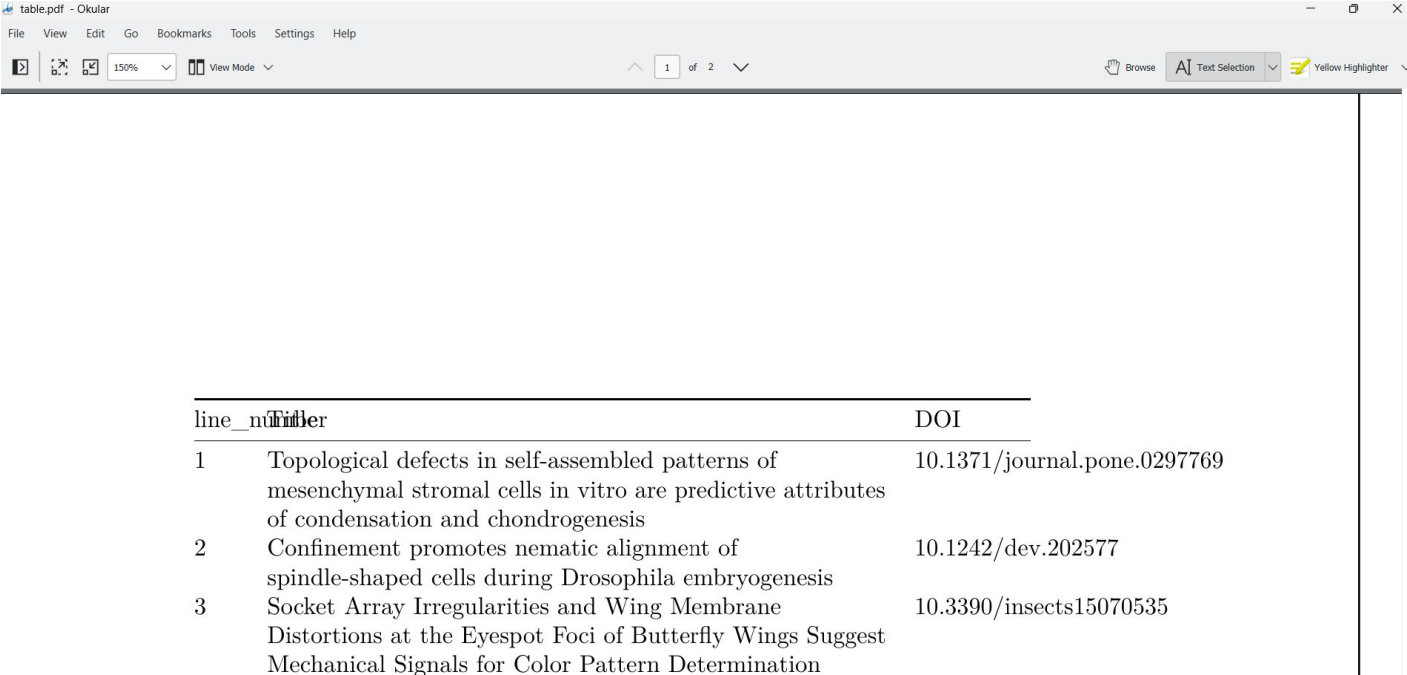


table.pdf - Okular

File View Edit Go Bookmarks Tools Settings Help

150% View Mode

1 of 2

Browse Text Selection Yellow Highlighter

line_num	title	DOI
1	Topological defects in self-assembled patterns of mesenchymal stromal cells in vitro are predictive attributes of condensation and chondrogenesis	10.1371/journal.pone.0297769
2	Confinement promotes nematic alignment of spindle-shaped cells during Drosophila embryogenesis	10.1242/dev.202577
3	Socket Array Irregularities and Wing Membrane Distortions at the Eyespot Foci of Butterfly Wings Suggest Mechanical Signals for Color Pattern Determination	10.3390/insects15070535

TROUBLESHOOTING

The tutorial is very good

1. Set the correct encoding

When parsing the csv files from pubmed, some characters are not recognized by default => proper encoding need to be set

> UnicodeEncodeError: 'charmap' codec can't encode character '0144' in position 110: character maps to <undefined>

Solution:

Set the encoding variable for csvkit in the current terminal, would be used by shell for all other commands.

export PYTHONUTF8=1