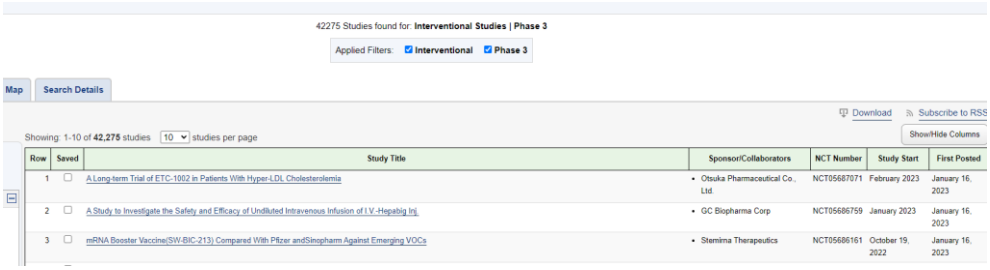
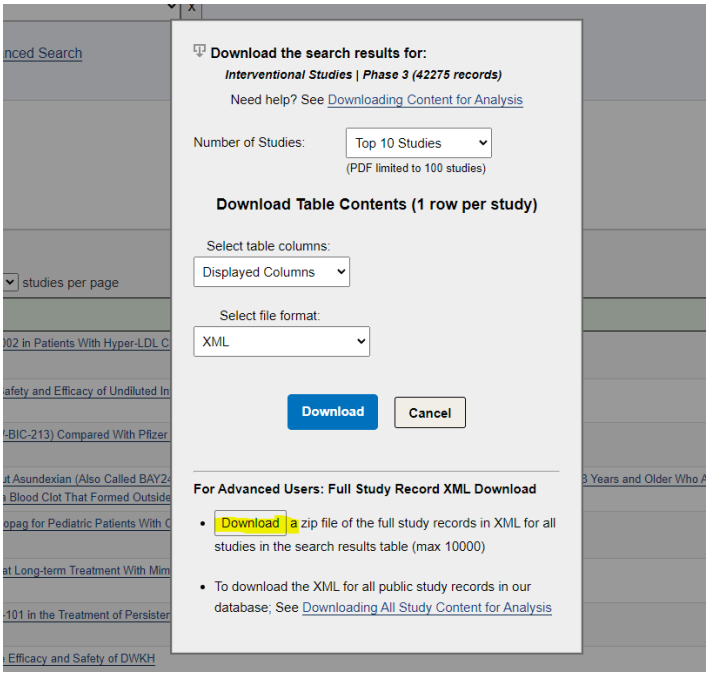
This toolkit was developed since clinicaltrial.gov doesn't provide any access for searching in the main text of trial documents. The main idea of this tool is to download the trial documents via the XML files provided by clinical.gov which contain the area name of trial documents.

Therefore, the working process includes three main parts:

1. Search the clinicaltrial.gov with some sketchy conditions and download XML files of all potential qualified cases via clinicaltrial.gov manually.

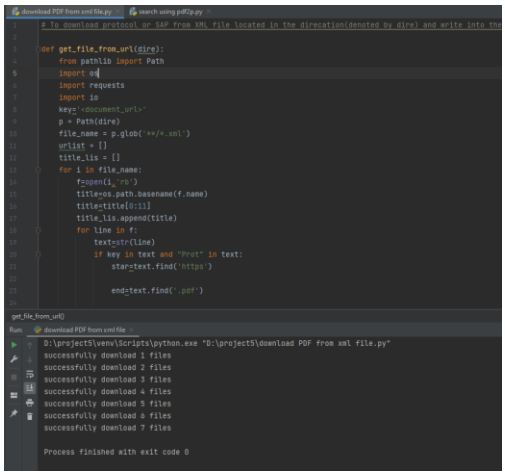


(all interventional phase 3 study)



(Download a zip for all XML files, then unzip these files to another folder.)

2. Read and parse the XML files with the python script titled "*download PDF from XML file.py*", the targeted PDF files will be downloaded through which.



venv	2022/12/16 18:56	File folder	
.idea	2023/1/18 10:55	File folder	
searching key words in PDF file.py	2023/1/18 10:18	PY File	6 KB
search using pdf2p.py	2022/5/29 4:07	PY File	6 KB
download PDF from xml file.py	2022/12/16 17:19	PY File	3 KB
User guide for clinicaltrial.gov searching ...	2023/1/18 10:19	Microsoft Word ...	0 KB
sss	2022/12/17 15:37	File	1 KB
i+key	2022/3/25 3:23	File	1,106 KB
dsds	2022/12/17 14:39	File	0 KB
NCT03886025.pdf	2023/1/18 10:55	Adobe Acrobat 文...	309 KB
NCT03877627.pdf	2023/1/18 10:55	Adobe Acrobat 文...	290 KB
NCT03617445.pdf	2023/1/18 10:55	Adobe Acrobat 文...	8,053 KB
NCT03486444.pdf	2023/1/18 10:55	Adobe Acrobat 文...	2,923 KB
NCT03473483.pdf	2023/1/18 10:55	Adobe Acrobat 文...	793 KB
NCT03430973.pdf	2023/1/18 10:55	Adobe Acrobat 文...	357 KB
NCT01917604.pdf	2023/1/18 10:55	Adobe Acrobat 文...	1,891 KB

(Execute the python script, if it works, a similar log will appear, and you will find the downloaded PDF file in the same pathway as the script.)

3. Comprehensive searching by the two scripts titled '*searching keywords in PDF file.py*' and '*search using pdf2p.py*', the corresponding searching results will be automatically sorted and stored.

Notice: the search terms are input by list variable 'lisx', users could customize the search by modifying the search list.

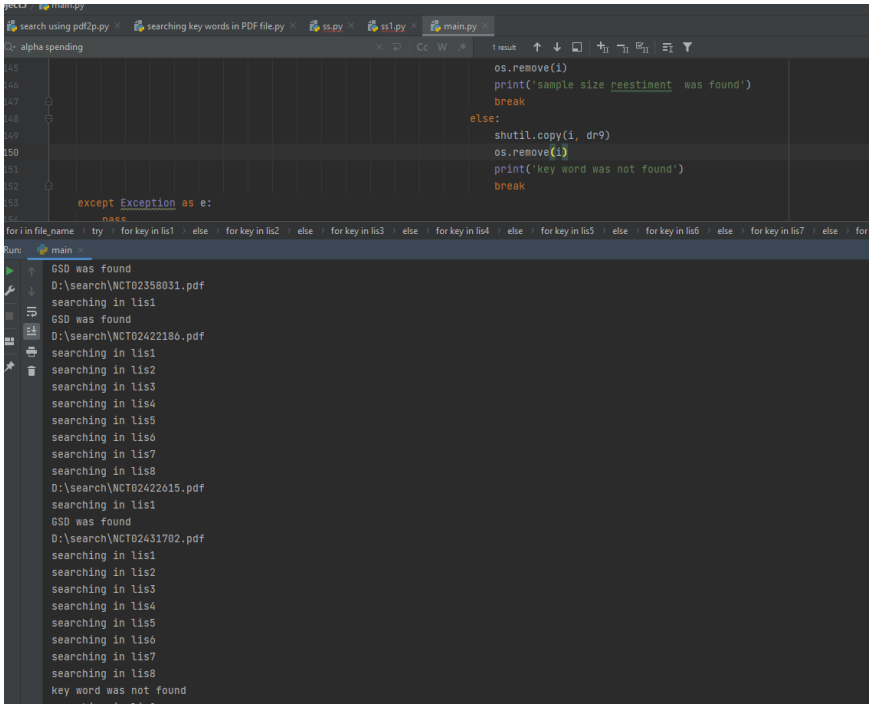
Before the execution, please modify the absolute pathway variable 'drx' to customize the output.

The function of the two scripts 'searching keywords in PDF file.py' and '*search using pdf2p.py' are similar but were realized by different packages. Since PDF is not a stable data format, the author will recommend using the two scripts twice for quality control.*

```python
import shutil

logging.propagate = False
logging.getLogger().setLevel(logging.ERROR)

lis1=['group sequential','conditional power','predictive power','conditional error','alpha spending','interim ana
lis2=['multi-arm multi-stage','multiarm multistage','multi stage']
lis3=['adaptive platform','adaptive platform']
lis4=['pick the winner','drop the loser','adaptive dose selection','seamless','adaptive design']
lis5=['adaptive enrichment','population enrichment','patient enrichment','enrichment design','biomarker adaptive'
lis6=['response adaptive','adaptive randomisation','adaptive randomization','outcome adaptive']
lis7=['adaptive hypothesis','adaptive hypotheses']
lis8=['sample size adjustment','sample size re estimation','sample size reestimation','sample size modification',
    'sample size re assessment','sample size re-calculation','sample size recalculation']

dir =r'D:\search'
dr1 = r'D:\results\GSD'
dr2 = r'D:\results\MAMS'
dr3 = r'D:\results\Platform'
dr4 = r'D:\results\Dose selection'
dr5 = r'D:\results\Enrichment'
dr6 = r'D:\results\ARD'
dr7 = r'D:\results\Adaptive hypo'
dr8 = r'D:\results\sample size reestimation'
dr9 = r'D:\results\Not adaptive'
```

```
search using pdf2p
range(1, 167)
D:\search\NCT02422615.pdf
GSD was found
range(1, 69)
D:\search\NCT02509078.pdf
sample size reestiment was found
D:\search\NCT02509078.pdf
range(1, 163)
D:\search\NCT02569398.pdf
sample size reestiment was found
D:\search\NCT02569398.pdf
range(1, 73)
D:\search\NCT02586155.pdf
```

(an example of successful execution of 'search using pdf2p.py')



(an example of successful execution of 'searching keywords in PDF file.py')