

Pipeline to literature

A Pipeline for Obtaining Relevant Literature Based on Given Keywords

It's a pipeline to help researchers accelerate literature searches and information acquisition

Let's start following the steps!

Step 1

Syntax for obtaining query syntaxes for databases such as PubMed based on keywords

1. Common approach

Take PubMed as an example.

Take the subject keywords of our current study (e.g. **Mycotoxin, enzyme, degrade**, degradation, etc.) as an example.

Website: <https://pubmed.ncbi.nlm.nih.gov/advanced/>

Search based on search keyword statements

An official website of the United States government [Here's how you know](#) ✓

NIH National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed Advanced Search Builder

PubMed[®]
User Guide

Add terms to the query box

1. Selecting the target term

Title/Abstract ▾ Enter a search term

AND ▾

Show Index

Query box

2. Query syntax


3. Clicking it

((Mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (degrad*[Title/Abstract])


Search ▾

Note: When you use a literature database to search for relevant literature resources, we recommend that you optimize your keywords. For example, if your research area of interest is a physician topic, you should perform keyword validation at the MeSH URL (<http://www.nlm.nih.gov/mesh/>). This is to ensure that the most accurate research vocabulary is used. This maximizes the chance of ensuring that the literature resources searched in the database are the most accurate and relevant.

Download all retrieved literature information

 **National Library of Medicine**
National Center for Biotechnology Information

Log in



Search

Advanced Create alert Create RSS User Guide

Save Email Send to Sort by: Best match Display options

1. Clicking it

Save citations to file

Selection:

All results

2. Selecting it

Format:

CSV

3. Selecting it

4. Clicking it

Create file

Cancel

For Web of Science:

Website: <https://www.webofscience.com/wos/woscc/advanced-search>

Clarivate English Products

Web of Science™ Search Sign In Register

Advanced Search Query Builder

DOCUMENTS RESEARCHERS

Search in: Web of Science Core Collection Editions: All

Add terms to the query preview

All Fields Example: liver disease india singh And Add to query

More options Query Preview

1. Query syntax

2. Clicking it

Booleans: AND, OR, NOT Field Tags: TS=Topic, TI=Title, AB=Abstract, AU=Author, AI=Author Identifier, Keywords, C=City, P=Province/State, CU=Country/Region, ZP=Zip/Postal Code, F=Grant Number, FD=Funding Details, TI=Title, DOI=DOI, PY=Year Published, CF=Conference, AD=Address, OC=Organization, OO=Organization Number, SA=Street Address, PS=Province/State, CU=Country/Region, ZP=Zip/Postal Code, F=Grant Number, FD=Funding Details, TI=Title, DOI=DOI, PY=Year Published, WC=Web of Science Categories, SI=ISI/ISI/ISI, UT=Accession Number, PMID=PubMed ID, DOI=PubMed ID, LI=Index Date, PU=Publisher, AL=All Fields, PP=Final publication year, SD=Sustainable Development Goals

Clarivate English Products

Web of Science™ Search Sign In Register

Advanced Search Results for ((ALL=(Mycotoxin)) AND ALL=(enzyme)) AND ALL=(degrad*)

346 results from Web of Science Core Collection for:

Quick add keywords: ENZYMATIC DETOXIFICATION

Publications You may also like...

Refine results Search within results...

Quick Filters Review Article 51, Early Access 9, Open Access 167, Enriched Cited References 65, Publication Years 2024 2

Export

1. Clicking it

2. Clicking it

Effect of the combined compound probiotics with mycotoxin degradation enzyme on detoxifying aflatoxin B₁ and zearalenone

Huang, W; Chang, J.; Gao, T.Z

2018 | JOURNAL OF TOXICOLOGICAL SCIENCES 43 (4-6), pp.377-385

Aflatoxin B₁ (AFB₁) and zearalenone (ZEA) are the secondary toxic metabolites of fungi which contaminate a wide range of food and feedstuffs. Limiting exposure of humans and livestock to them is very essential. Among numerous methods of mycotoxin degradation, biodegradation by microorganisms and enzymes is an effective and promising approach to eliminate their hazards. The present study aims ... Show more

Free Full Text from Publisher

36 Citations, 33 References, Related records

You can also supplement the relevant literature in other databases such as Google Scholar, Science Direct, etc.

2. Common approach

To minimize manual operations, here we have created a homemade Python script that automatically generates all possible lexical variations and PubMed and Web of Science query syntaxes and corresponding download links based on keywords provided by the user.

Python script name:

generate_query_statements_and_links_to_literature_database_searches_based_on_keywords.py

Required Modules:

nltk, inflect, argparse, itertools

If your machine does not have the corresponding module, use **pip install module** to install it successfully.

Usage:

Enter the following command in the terminal to see help on using the program:

```
python
generate_query_statements_and_links_to_literature_database_searches_based_on_key
words.py -h
```

```
usage: generate_query_statements_and_links_to_literature_database_searches_based_on_keywords.py [-h] -m selectionMode -i keywordsFile -o outputFile
Generating query statements and links to literature database searches based on keywords
optional arguments:
  -h, --help            show this help message and exit
  -m selectionMode, --inputMode selectionMode
                        For the first run, enter init, and after that, enter run
  -i keywordsFile, --inputFilePath keywordsFile
                        Enter a text file path that contains only keywords
  -o outputFile, --outputFilePath outputFile
                        Output file path
```

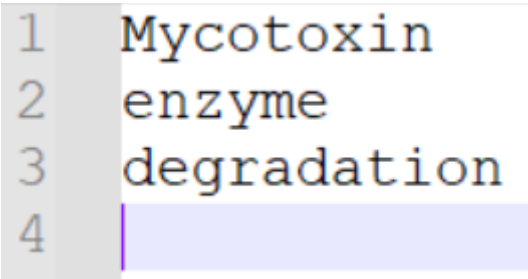
All parameters and descriptions are listed below:

Parameters	Descriptions
-m	When running the script for the first time, use -m init to download the dictionary library first. Once downloaded, use -m run for subsequent run parameters.
-i	Setting the path to a file containing only keywords.
-o	Setting the output file path.

Enter the file format:

keyword 1
keyword 2
keyword 3
...

As shown in the figure below:



Practical training:

```
python
generate_query_statements_and_links_to_literature_database_searches_based_on_key
words.py -m run -i keywords.txt -o my_result.txt
```

Outputs the contents of the file:

```
1 Mycotoxin:
2 mycotoxin | Mycotoxins | MYCOTOXIN | mycotoxins | Mycotoxin
3 enzyme:
4 Enzyme | enzyme | enzymes | ENZYME
5 degradation:
6 debasement | Debasement | abasement | abasements | degradations | DEBASEMENT | DEGRADATION | degradation | Degradation | Abasement | debasen
7 Note: Please determine the most professional keywords that you need. For example, through professional thesaurus such as MeSH, use * to redu
8
9 PubMed database search terms are as follows:
10 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (debasement[Title/Abstract])
11 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (Debasement[Title/Abstract])
12 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (ab
13 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (ab
14 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (degradations[Title/Abstract])
15 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (DEBASEMENT[Title/Abstract])
16 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (DEGRADATION[Title/Abstract])
17 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (degradation[Title/Abstract])
18 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (Degradation[Title/Abstract])
19 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (Abasement[Title/Abstract])
20 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (debasements[Title/Abstract])
21 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (Abjection[Title/Abstract])
22 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (ABJECTION[Title/Abstract])
23 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (ABASEMENT[Title/Abstract])
24 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (abjections[Title/Abstract])
25 ((mycotoxin[Title/Abstract]) AND (Enzyme[Title/Abstract])) AND (abjection[Title/Abstract])
26 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (debasement[Title/Abstract])
27 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (Debasement[Title/Abstract])
28 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (abasement[Title/Abstract])
29 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (abasements[Title/Abstract])
30 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (degradations[Title/Abstract])
31 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (DEBASEMENT[Title/Abstract])
32 ((mycotoxin[Title/Abstract]) AND (enzyme[Title/Abstract])) AND (DEGRADATION[Title/Abstract])
```


If your machine does not have the corresponding module, use **pip install module** to install it successfully.

Usage:

Enter the following command in the terminal to see help on using the program:

```
python remove_duplicates.py -h
```

```
usage: remove_duplicates.py [-h] -i targetExcelFile -o outputFile
Remove duplicate literature

optional arguments:
  -h, --help            show this help message and exit
  -i targetExcelFile, --inputFilePath targetExcelFile
                        Input the path to an excel file with a .xlsx extension
  -o outputFile, --outputFilePath outputFile
                        Output file path
```

All parameters and descriptions are listed below:

Parameters	Descriptions
-i	Setting the path to MS Excel files ending in .xlsx extension
-o	Setting the output file path.

Practical training:

```
python remove_duplicates.py -i all_database_literatures_data.xlsx -o
all_database_literatures_data_single.txt
```

Outputs the contents of the file:

Title	DOI
Recent advances in biosynthesis of mycotoxin-degrading enzymes and their applications in food and feed	10.1080/10408398.2023.2294166
Small Peptides in the Detection of Mycotoxins and Their Potential Applications in Mycotoxin Removal	10.3390/toxins14110795
Characterization and application of the enzyme peroxidase to the degradation of the mycotoxin DON	10.1080/03601234.2017.1356672

Step 3

Download literatures

Based on the entirety of the relevant literature obtained earlier, a pdf of each piece of literature was downloaded.

Note: In order to get all the above literature as fast as possible, we suggest that a one-time batch download can be realized by tools such as **EndNote**, **crawler**, **scihub2pdf**, and so on. Please note that at all times, **please respect the copyrights of the authors and publishers of the literature. That is, the acquisition of the target literature is carried out through legal channels.**

Here, we provide a crawler script that can batch download pdf format literature. Just for reference.

Python script name:

batch_download_literatures_pdf_alpha_test.py

Required Modules:

pandas, selenium, time, os, random, argparse

If your machine does not have the corresponding module, use **pip install module** to install it successfully.

Usage:

Enter the following command in the terminal to see help on using the program:

```
python batch_download_literatures_pdf_alpha_test.py -h
```

```
usage: batch_download_literatures_pdf_alpha_test.py [-h] -i targetExcelFile -c chromedriverPath
Batch download pdf from sci-hub database according to DOI number
optional arguments:
  -h, --help            show this help message and exit
  -i targetExcelFile, --inputFilePath targetExcelFile
                        Input the path to an text file
  -c chromedriverPath, --inputChromedriverPath chromedriverPath
                        Specify the path to chromedriver.exe
```

Note: This script is for test use by interested parties only, and in order to comply with the publisher's copyright, please download it from the official link of the literature publisher, or purchase the target literature you need.

Step 4

Convert pdf documents to text files

After downloading all the documents (pdf), use the Python script for batch processing to convert all the documents into text files.

Python script name:

batch_pdf_file_to_text_file.py

Required Modules:

os, argparse

If your machine does not have the corresponding module, use **pip install module** to install it successfully.

Usage:

Enter the following command in the terminal to see help on using the program:

```
python batch_pdf_file_to_text_file.py -h
```

```
usage: batch_pdf_file_to_text_file.py [-h] -m selectionMethod -i targetPdfLiterature -o outputFolder
Batch pdf to text
optional arguments:
  -h, --help            show this help message and exit
  -m selectionMethod, --inputMethodNumber selectionMethod
                        Input pdf to text method number, there are 1, 2, 3, 4 optional, can only set one of them
  -i targetPdfLiterature, --inputFolderPath targetPdfLiterature
                        Enter the path to the folder that contains only pdf literatures
  -o outputFolder, --outputFolderPath outputFolder
                        Output folder path
```

All parameters and descriptions are listed below:

Parameters	Descriptions
-m	The script provides four kinds of pdf files into text files, respectively, numbered 1, 2, 3, 4, the user can set up according to their own preferences. A run, only one of the methods can be set. The purpose of such a design is that when some of the pdf documents can not be converted into text files, you can put these documents into a separate directory, try another method of conversion.
-i	Setting the path to the folder that includes only pdf-formatted literatures.
-o	Setting the path of output folder, all the text files which are converted successfully will be stored in this directory.

Practical training:

```
python batch_pdf_file_to_text_file.py -m 4 -i literatures_pdf -o
literatures_text
```

View a text-formatted document from the literatures_text folder as follows:

```
1 Article
2 Efficient Degradation of Aflatoxin B1 and Zearalenone by
3 Laccase-like Multicopper Oxidase from Streptomyces
4 thermocarboxydus in the Presence of Mediators
5
6 Xing Qin, Yanzhe Xin, Jiahuan Zou, Xiaoyun Su, Xiaolu Wang, Yaru Wang, Jie Zhang
7 Huiying Luo * and Huoqing Huang *
8
9 , Tao Tu , Bin Yao,
10
11 State Key Laboratory of Animal Nutrition, Institute of Animal Sciences, Chinese Academy of Agricultural
12 Sciences, Beijing 100193, China; qinxing@caas.cn (X.Q.); xyanzhe@163.com (Y.X.); zjh0512@126.com (J.Z.);
13 suxiaoyun@caas.cn (X.S.); wangxiaolu@caas.cn (X.W.); wangyaru@caas.cn (Y.W.); zhangjie09@caas.cn (J.Z.);
14 tutao@caas.cn (T.T.); yaobin@caas.cn (B.Y.)
15 * Correspondence: luohuiying@caas.cn (H.L.); huanghuoqing@caas.cn (H.H.)
16
17 Abstract: Multicopper oxidases (MCOs) are a diverse group of enzymes that could catalyze the
18 oxidation of different xenobiotic compounds, with simultaneous reduction in oxygen to water. Aside
19 from laccase, one member of the MCO superfamily has shown great potential in the biodegradation
20 of mycotoxins; however, the mycotoxin degradation ability of other MCOs is uncertain. In this
21 study, a novel MCO-encoding gene, StMCO, from Streptomyces thermocarboxydus, was identified,
22 cloned, and heterologously expressed in Escherichia coli. The purified recombinant StMCO exhibited
23 the characteristic blue color and bivalent copper ion-dependent enzyme activity. It was capable of
24 oxidizing the model substrate ABTS, phenolic compound DMP, and azo dye RB5. Notably, StMCO
25 could directly degrade aflatoxin B1 (AFB1) and zearalenone (ZEN) in the absence of mediators.
26 Meanwhile, the presence of various lignin unit-derived natural mediators or ABTS could significantly
27 accelerate the degradation of AFB1 and ZEN by StMCO. Furthermore, the biological toxicities of their
28 corresponding degradation products, AFQ1 and 13-OH-ZEN-quinone, were remarkably decreased.
29 Our findings suggested that efficient degradation of mycotoxins with mediators might be a common
30 feature of the MCOs superfamily. In summary, the unique properties of MCOs make them good
31 candidates for degrading multiple major mycotoxins in contaminated feed and food.
32
33 Keywords: multicopper oxidase; mycotoxin; aflatoxin; zearalenone; degradation; mediator
34
35 Key Contribution: The laccase-like multicopper oxidase StMCO could effectively degrade aflatoxin B1
```

The document was successfully converted into a text file.

Note: The file name of the document is logged in the terminal for failed conversions. Convenient for users to follow up.

Access to large language modeling tools

After that, following the process described in our article, the research question is prepared manually and then the text file is copied and pasted into the input box of a big language model such as **ChatGPT**. The goal of capturing information from the literature by big language models instead of manually can be realized.

Finally, I sincerely hope that this pipeline can accelerate your research process and wish the best of luck in research.