Check for updates

SOFTWARE TOOL ARTICLE

Abstract Sifter: a comprehensive front-end system to PubMed [version 1; referees: 2 approved]

Nancy Baker [iD] [1], Thomas Knudsen[2], Antony Williams [iD] [2]

[1]Leidos, Research Triangle Park, NC, USA
[2]National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

v1   **First published:** 21 Dec 2017, **6**(Chem Inf Sci):2164 (doi: 10.12688/f1000research.12865.1)

Latest published: 21 Dec 2017, **6**(Chem Inf Sci):2164 (doi: 10.12688/f1000research.12865.1)

Abstract

The Abstract Sifter is a Microsoft Excel based application that enhances existing search capabilities of PubMed. The Abstract Sifter assists researchers to search effectively, triage results, and keep track of articles of interest. The tool implements an innovative "sifter" functionality for relevance ranking, giving the researcher a way to find articles of interest quickly. The tool also gives researchers a view of the literature landscape for a set of entities such as chemicals or genes. The Abstract Sifter is available as a Microsoft Excel macro-enabled workbook application.
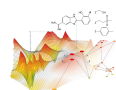
This article is included in the Chemical Information Science gateway.

Open Peer Review

Referee Status: ✔ ✔

	Invited Referees	
	1	**2**
version 1 published 21 Dec 2017	✔ report	✔ report

1 **Pauliina Damdimopoulou** [iD] , Karolinska Institutet, Sweden

Astrud Tuck, Karolinska Institutet, Sweden

2 **Qingliang (Leon) Li** [iD] , National Institutes of Health, USA

Discuss this article

Comments (2)

Introduction
Scientists in the biological and medical domains can spend considerable time searching for relevant articles in PubMed, and tools that make the searching more effective will save time and resources (Khare *et al.*, 2014). Here, we present a tool, the Abstract Sifter, built to improve efficiency in searching PubMed. Specifically, this tool was designed with the following objectives: 1) To make it quicker and easier to find relevant articles in PubMed; 2) To visualize the "literature landscape", which can help focus on key relevant articles; 3) To make it easier to evaluate and take notes on abstracts; and 4) To facilitate collaboration on literature tasks.

Methods
The Abstract Sifter application is a Microsoft Excel macro-enabled workbook that has been tested in Excel 2013 and 2016 on the Windows platform. Visual Basic for Applications (VBA) was used to develop the features that go beyond native Excel functionality. For the retrieval of PubMed query results, Entrez Programming Utilities (E-utilities) (Sayers, 2016) are called from

VBA. These utilities were developed by the National Center for Biotechnology Information (NCBI) to allow software developers to query PubMed and other NCBI databases and retrieve the results for incorporation into local applications (2017). Through implementation as an Excel workbook, the Abstract Sifter can easily be shared with collaborators.

Use case
The Abstract Sifter application workbook contains seven sheets: ReadMe, Main, Abstract, Notes, Log, and Landscape, and SampleQueries. The Main sheet is where the basic functions operate, including the novel functionality called "sifting". To start, the end-user clicks on the *Query PubMed* button at the top of the screen and enters any PubMed query of interest. For the example in Figure 1a, the end-user has entered the simple query "chlorpyrifos". However, these queries can be more complex. In fact, any query run in PubMed can be executed in the Sifter. When the query entry is finished, the user then clicks on *Submit* and the query is sent to the NCBI PubMed E-utility. The first response returned by the E-utility is the number of articles
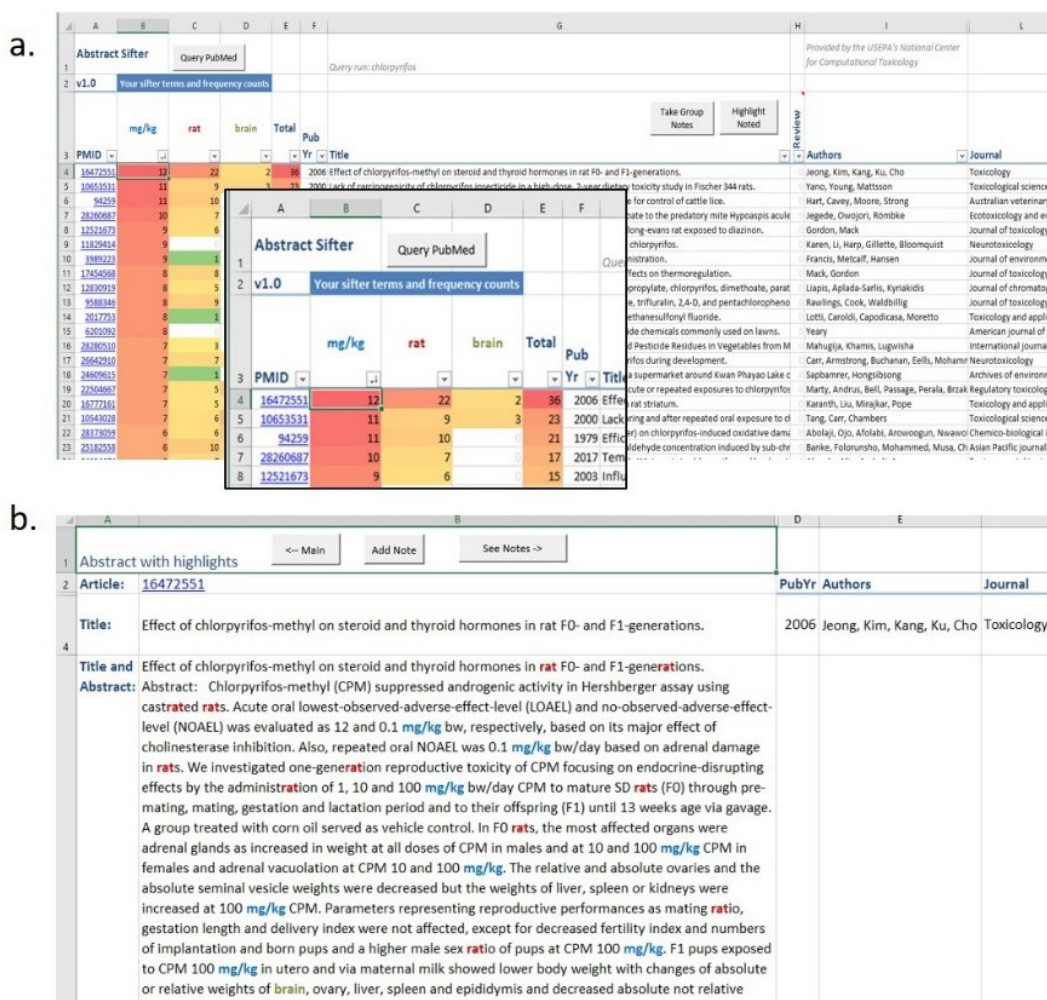

Figure 1. 1a. Abstract Sifter Main sheet and **1b**. Abstract sheet view.

found that satisfy the query. The citations are downloaded from PubMed by the Abstract Sifter, parsed by pattern matching algorithms coded in VBA. All citations are thus parsed for title, abstract, authors, publication year, journal, and PubMed identifier, and the data is inserted into rows in the Main sheet. Every new search will clear results from the previous query. For performance purposes, if the number of articles exceeds 5,000, the query will not be run and the user is encouraged to re-word the query to return fewer records.

The results of the query stored in the Main sheet can be browsed like any other data in a spreadsheet. The sifter feature provides a novel and effective way to narrow search results with large number of citations to find articles of interest. For example, the query for "chlorpyrifos" returned over 4,000 PubMed citations. If a researcher is looking for neurological effects in studies where rats were dosed with chlorpyrifos, the researcher could type the term "mg/kg" in the spreadsheet cell B3, "rat" in C3, and "brain" in D3 (Figure 1a inset). The Abstract Sifter returns the number of occurrences of each term found in the title and abstract combined. The Main sheet's citations can be sorted by these counts. Sifting by entering terms and sorting can be repeated. Similarly, new PubMed queries can be run, altered, and rerun. Double-clicking on any cell in the row (except the cell containing the PMID) takes the end-user to the Abstract sheet where the title and abstract of that citation are shown (Figure 1b). The sifter terms

are highlighted by giving each the color of the term on the Main sheet. Together, these query and sifting capabilities provide a powerful search tool.

The Abstract Sifter also incorporates functionality to allow the end-user to take notes on citations. On the Abstract sheet, for instance, the user can click on the button *Add Note*. A form appears that provides the opportunity to add short notes (tags) or long notes or to specify one of three categories (yes, no, or maybe) (Figure 2a). How these values are used is a decision of the end-user. When the user clicks on *OK*, a row is inserted into the Notes sheet with the citation information along with the notes. Alternatively, the end-user can take notes on more than one article at a time from the Main sheet. To do this, the end-user selects multiple rows of interest and then clicks on *Take Group Notes*. Each of the selected citations will be inserted into the Notes sheet with the entered notes and tags (Figure 2b).

Often, after entering a number of notes, the user will forget which citations have been read and evaluated and for which notes have been taken. By clicking on the *Highlight Noted PMIDs* button either on the Main or Notes sheet, the PubMed identifier (PMID) on the Main sheet will be set to the color specified in the Note form (Figure 2c). Using the built-in Excel filtering feature, the color can be selected or sorted on to view Noted citations.
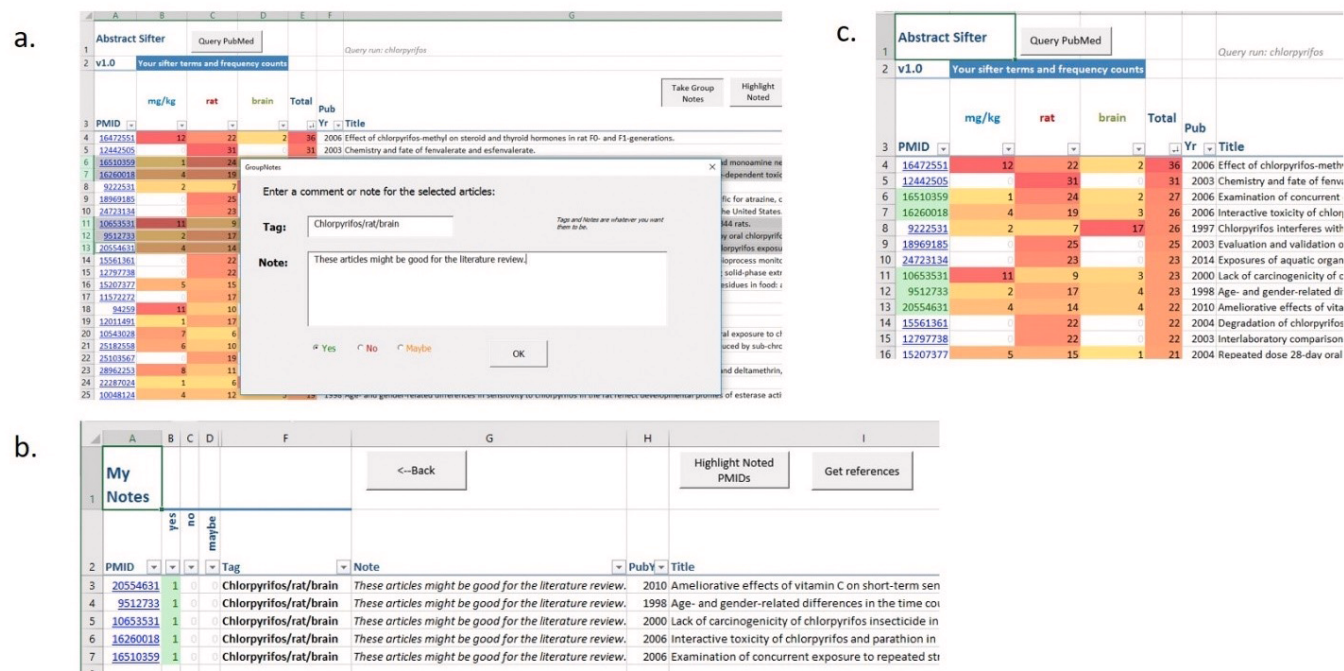

Figure 2. 2a. Taking Group notes. **2b.** Viewing Notes sheet. **2c.** Highlighting noted citations.

The Notes sheet itself can be viewed and edited and rows can be deleted. Entries on the Notes sheet can be exported to PubMed where they can then be downloaded in a number of different formats, including a format for direct import into citation management software. The button to export to a citation manager via PubMed is labelled *Get references* and appears on the Notes sheet.

Another unique feature of the Abstract Sifter is the Landscape sheet functionality. The Landscape sheet is an alternative to the Main sheet as an entry point, and provides the end-user a visualization of literature for a set of chemicals. To use this functionality, the end-user enters chemical name queries in Column C of the Landscape sheet after Row 4. An example in which the end-user has entered seven chemical queries is shown in Figure 3. The chemical queries can be extended with CAS registry numbers or synonyms. Next, the end-user enters subject matter query text in Row 3, Columns E and higher. In the example depicted in Figure 3, the end-user has entered several subject matter queries, starting with "neoplasms OR cancer".

The end-user can then select cells in the intersection region (E5 through J11 in Figure 3) and click on the button *Update*

Article Counts. This action causes the tool to iterate through the cells, build a query using the chemical terms in column C appended to the effect terms in Row 3, and then send the query to PubMed for execution. The counts of the articles satisfying each query are returned from PubMed and are inserted into the corresponding cells. To see the PubMed records, the user double-clicks on a cell in the intersection region. This action starts the PubMed query process and sends the results to the Main page for sifting. More chemicals, entries and/or and additional queries can be added by the user on this Landscape sheet. Buttons are available on the Landscape sheet to help with formatting results, such as applying heat-map coloring to the article counts. The SampleQueries sheet has some text that can be used as a starting point for Landscape queries. To use, the end-user selects rows and clicks on the *Send Queries to Landscape* button to have the queries appended to Row 3 on the Landscape page.

The Log sheet contains a row for each query run. The query text is inserted into the row along with date and time information and the number of records returned. Queries can be easily rerun by double-clicking on the query text in column C.

	Preferred Name	Chemical / entity query	Cancer	Dermatitis	Algae	Dust exposure	Food exposure	Water exposure
5	Tetrahydrocurcumin	Tetrahydrocurcumin OR 36062-04-1	43	5	0	0	0	0
6	(6)-Gingerol	(6)-Gingerol OR 23513-14-6	116	13	0	0	0	0
7	Curcumin	Curcumin OR 458-37-7	4139	424	4	0	16	7
8	Doxorubicin	Doxorubicin OR 23214-92-8	51667	9376	7	1	7	12
9	Hydroquinone	Hydroquinone OR 123-31-9	519	515	12	1	8	23
10	Coumarin	Coumarin OR 91-64-5	1070	307	5	1	6	9
11	PFOA	Perfluorooctanoic acid OR PFOA OR 335-67-1	107	16	7	52	242	525

Figure 3. Landscape sheet.

Discussion

The Abstract Sifter can facilitate many PubMed literature tasks by enabling rapid identification, triage, and tracking of relevant articles. The literature landscape viewing and navigating capabilities give researchers distinctive insight into characteristics of a literature corpus.

Software availability

The Abstract Sifter Excel workbook and user documentation are available at: https://github.com/USEPA/CompTox-Chemistry-Dashboard-Abstract-Sifter/

Archived source code as at the time of publication: http://dx.doi.org/10.5281/zenodo.1040961 (nancycolebaker, 2017)

License: CC0 1.0 Public Domain Dedication.

Competing interests

No competing interests were disclosed.

Grant information

References

Khare R, Leaman R, Lu Z: **Accessing biomedical literature in the current information landscape.** *Methods Mol Biol.* 2014; **1159**: 11–31.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

nancycolebaker: **nancycolebaker/LitInfo: PubMed Abstract Sifter public release v.1. (Version v1).** *Zenodo.* 2017.
Data Source

NCBI Resource Coordinators: **Database Resources of the National Center for Biotechnology Information.** *Nucleic Acids Res.* 2017; **45**(D1): D12–D17.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Sayers E: **A General Introduction to the E-utilities.** 2016; Retrieved 2016.
Reference Source

Open Peer Review

Current Referee Status: ✔ ✔

Version 1

Referee Report 06 February 2018

✔   **Qingliang (Leon) Li** 🆔

National Institutes of Health, Bethesda, MD, USA

The authors reported an Excel front-end for PubMed search by using NCBI E-Utilities. The tool supports searching literature in PubMed, importing search results into excel, adding notes into individual literature etc. Here are some minor issues that the authors may consider to improve the usabilities of the tool:

1. Currently, the tool has a limit of 5000 results for a search. This makes sense when considering excel performance, however, it will improve the usability if the tool can support pagination and/or have the user to select which literature to import.

2. The term search/count feature in the main sheet can be improved by restricting the "searched term" to be a "real term". For example, "gene**rat**ion" that has nothing related to "rat" should not be highlighted as a 'rat term'.

3. In the landscape view, the authors may consider linking chemical name to chemical databases, such as PubChem, to view chemical structures.

4. The authors may consider connecting this tool with "My NCBI" (https://www.ncbi.nlm.nih.gov/account/?back_url=https%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fmyr)

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Partly

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 15 January 2018

Pauliina Damdimopoulou 🆔 , **Astrud Tuck**
Unit of Toxicological Sciences, Swetox, Karolinska Institutet, Södertälje, Sweden

This article describes a Microsoft Excel workbook based tool for rapid PubMed literature searches where the results can be easily sifted and annotated. With the ever-growing body of literature, tools for finding and sorting literature become increasingly more important. I am not a software developer myself, and cannot therefore evaluate the technical aspects of the tool (codes, macros, algorithms…). However, I am a keen end-user of tools such as Abstract Sifter, because my work is highly interdisciplinary and I often need to rapidly find information about unfamiliar topics, and can therefore estimate the user friendliness and usefulness of this new tool.

A clear benefit of Abstract Sifter is its Excel format. Most researchers are familiar with Excel and therefore starting to work with Abstract Sifter does not require any extra effort to understand the interphase. It is convenient that different searches can be saved and shared with colleagues like any file. I can also vision sharing a Sifter file through a cloud service and simultaneously annotating and going through literature for example for a review or grant application.

I have carefully read through the article (last updated 02 JAN 2018) and the User Guide (version 1.0) as well as tried the different functions of Abstract Sifter (v1), and find only minor issues to comment on. I think the work is of acceptable standard to be approved.

Minor comments
1. Building queries in the Sifter. It would be helpful to have an "advanced query builder" option available. When I was trying different queries, I ended up building the more complicated ones using the PubMed advanced search builder and copy pasting the query line to Sifter.

2. Does the Sifter function take article keywords and authors into consideration in the counts?

3. Page 3: "Highlight Noted PMIDs" should be "Highlight Noted".

4. The "For HAWC" button on the Notes sheet is not explained in the article (please also see comment 7).

5. The Helpful Tips in the user guide are very useful! One to add could be how to retrieve selected references from PubMed with the "Get references" button. For example, if I have gone through and annotated the 4,000 chlorpyrifos citations and only want to retrieve the 150 I've marked "yes"

(green) in the Notes sheet, how should I do that?

6. The Landscape function is an exciting way to quickly get an idea of how well certain topics have been covered in the literature. What is Column D "Link" in the Landscape sheet? What is the relationship between Rows 3 and 4? Is row 4 just user defined short names for row 3 queries, and not taken into account in the PubMed based search for citations? This could be clarified in the article text. The Landscape heatmap colours could also be explained too together with their cut-offs (I assume it's based on % instead of absolute counts?). Should 0 labelled with white or blue?

7. On the ReadMe sheet in the Shifter file, "Chemical sheet" is mentioned, although the Sifter does not have one with that name. Do you maybe mean Landscape sheet? ReadMe page also gives a link to a HAWC video, although HAWC is not explained in the article (see also comment 4).

8. User Manual page 10: "…short version of the chemical name is in Column A" should be "…Column B". Column A is called "DSSTOX link to Dashboard" in the file.

9. I wonder, if I were to use this tool with a colleague to select citations for a review, for example, is there a way to add "comments" to the notes? For example, if I have classified a citation as "yes" but my colleague does not agree, how could that be easily marked into the Notes page?

10. Does the Shifter work both in Mac and PC environment? What about Open Office?

Is the rationale for developing the new software tool clearly explained?
Yes

Is the description of the software tool technically sound?
Yes

Are sufficient details of the code, methods and analysis (if applicable) provided to allow replication of the software development and its use by others?
Yes

Is sufficient information provided to allow interpretation of the expected output datasets and any results generated using the tool?
Yes

Are the conclusions about the tool and its performance adequately supported by the findings presented in the article?
Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Molecular biology, environmental medicine, endocrine disrupters, hormone signalling, female fertility

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Discuss this Article

Version 1

Author Response 02 Jan 2018

Nancy Baker, Leidos, USA

Thanks to feedback from helpful users, the user guide has been updated and corrected. Be sure to download the latest version.

Competing Interests: I am an author on this work.

Reader Comment 24 Dec 2017

Emma Schymanski, LCSB, University of Luxembourg, Luxembourg

I was made aware of the Abstract Sifter by the authors prior to this publication - while this article is quite short, the user documentation contains more information and steps users through the features. I encourage people to try it and see if it could be helpful for them, I found it surprisingly helpful, very intuitive and it was interesting to adjust the examples to some own examples and see which papers were found and which not - this will be very helpful to refine searches. The highlighting, while simple conceptually, supports interpretation a lot. While this is likely not the only way to enhance PubMed searching, it could help many and I could see some possible uses beyond the likely original scope of the authors. Being excel based means it is not completely platform independent, but this is unavoidable and will make these features at least very accessible to excel users.

Competing Interests: No competing interests were disclosed.