KAI NITSCHKE

MANUAL FOR DOI SCOUT V1.0

Contents

Abstract and Background Information

Abstract.

The DoiScout – an automatic tool for gathering information about registered clinical trials and resulting publications.

The *DoiScout* is a tool that facilitates large-scale literature reviews and citation analyses. It was developed as part of the IIT project, "Impact of therapeutic investigator initiated trials on medical practice" (DFG grant number BL 1395/2-1), in order to make the process of extensive literature searches more time effective.

The *DoiScout* offers two primary features. The first main feature is the automatic identification of publications that reference a particular study number (e.g. NCT02179424) as their source. Information about relevant publications is extracted and presented in a list that is formatted so that information can be passed on to other software programs for further processing. Note that the *DoiScout* does not automatically assess the relevance or validity of the search results.

The second primary feature of the *DoiScout* is the identification of citations. Databases behind such platforms as PubMed and Web of Science can be used to identify articles, systematic reviews and clinical guidelines in which a given article has been cited. The *DoiScout* extracts information about these texts and provides it to the user in a workable format. In addition, the *DoiScout* can be used to identify publications that go on to cite texts citing the original source, thus providing a comprehensive insight into the extent of a project's academic impact.

In addition to these two primary features, the *DoiScout* includes several secondary features, which focus on facilitating workflow.

Background and objectives. The *DoiScout* software was developed as part of the IMPACT/IIT project at the Institute of Medical Evidence at the University of Freiburg funded by the German Research Foundation (DFG, grant number BL 1395/2-1). This project aimed to investigate the impact of registered clinical trials on the academic literature as well as on medical practice. To this end, investigator initiated trials (including all DFG and BMBF funded trials), as well as a set of industry sponsored trials, all of which were started and completed within the years 2005 to 2012, were evaluated with respect to whether the results were (1) published, (2) cited by others, especially in systematic reviews and meta-analyses, and (3) found their way into clinical guidelines and subsequently into medical practice. The project required extensive and time-consuming manual searches of literature databases and study registers. Thus, the development of a semi-automatic tool to facilitate and standardize future literature reviews of this sort was a key deliverable included within the funding application.

The *DoiScout* includes functionalities for i) automatic extraction of bibliographic information, ii) automated downloading of full texts (PDFs) using the DOI (Digital Object Identifier) of a publication, iii) finding publications using the study registration number (ClinicalTrials.gov, EudraCT, ISRCT), iv)

identification of citations of publications, and v) identification of guidelines citing specific publications. The functions are modular and cross-linked so that i.e. the results of a study register search can be enriched with the bibliographic information of these publications and subsequently the full texts can be automatically downloaded. All results are presented in a spreadsheet format for further processing by the user. An export function for Endnote is also available.

The *DoiScout* was implemented using freely available, open-source tools in order to make it both accessible and modifiable for users. In order to ensure accessibility, the *DoiScout* was created using the programming language *Python* and its associated libraries. *Python* can be downloaded and used free of charge, and it works independently of any specific operating system. In addition a graphical user interface (GUI) as well as this manual are provided so that no programming knowledge or advanced computer skills are necessary to use the *DoiScout*. With respect to modifiability, the *DoiScout* is free, open-source and available on github.com: `https://github.com/kainitschke/doiscout`. It has been programmed modularly to enable experienced users to adapt, alter or extend the source code by other features (e.g. further database sources) to maximize its utility.

An Example. A researcher wants to compare 100 commercially sponsored studies to 100 publically sponsored studies with respect to their scientific impact (e.g., time between registration and publication, number of publications, numbers of citations). In a first step, all primary publications associated with study register numbers are identified; bibliographic information and register results are extracted automatically by the *DoiScout*. In addition, full-texts of all relevant texts are downloaded. The researcher must now manually assess the validity of the reported results and filter out those publications that remain relevant. Once this has been done, the *DoiScout* can be used to automatically identify and extract bibliographic information from all publications that cite the remaining primary studies of interest. With relatively little effort, the researcher now has the information needed to draw comparisons between the two groups of studies with respect to the chosen variable of interest. The example demonstrates how combining the primary and secondary features of the *DoiScout* greatly reduces the amount of time researchers need to spend on the manual searching of literature databases. This means that the number of studies included in a review can be increased feasibly with only a negligible additional workload for researchers.

Requirements and Technical Information

Python and Packages

The *DoiScout* was written in *Python 3.6*. To run, it requires that the following software is installed (note that all required software is available free of charge):

- *Python 3*. Any software version newer or equal to Python 3.6 should work. Older version might work. You can get more information on as well as download *Python 3* on `https://www.python.org/`.

- *Mozilla Firefox* is the most common free and open source web browser.

- The *geckodriver* is needed so the *DoiScout* is able to automatically operate Mozilla Firefox. It can be downloaded for free here: `https://github.com/mozilla/geckodriver/releases`. Download the file *geckodriver* and store it in the same folder that you have extracted the *DoiScout* to.

- In order to use the optional feature of the notification mails, *SSMTP* must be installed on your system. This feature has only been tested on a linux system.
  If SSMTP is not available on your operating system, the *DoiScout* will still function properly despite the loss of the email feature (E-mails). Do not select the respective options if *SSMTP* is not installed on your OS.

The Python libraries that are required to run the *DoiScout* are:
```
datetime, numbers, operator, os, pandas, pdfkit, random, re,
selenium, sys, time, tkinter, urllib3, weasyprint, webbrowser,
xlrd
```

System Requirements

Recommended system
- min. dual-core CPU
- min. 8 GB RAM
- display with min. resolution of 1024x768
- 64-bit OS (min. Windows 7, Ubuntu 16.04, Max OS X)

The *DoiScout* was programmed on a PC with the following specifications:
- Ubuntu 17.10 x64
- Intel Core i5-6500 CPU @ 3.20 GHz x 4
- 7.7 GB RAM
- Intel HD Graphics 530 (Skylake GT2) and Samsung 27" 3840x2160 @ 60 Hz monitor

Getting Started

Starting the DoiScout

The *DoiScout* is a *Python3*-based software. Hence you need to have *Python3* installed (see Requirements and Technical Information) and start it via *Python3*. How to start the *DoiScout* depends on the operating system (OS) of your computer.

Linux

1. Open a console and open the folder path where you downloaded the *DoiScout*.
2. Enter the following line in the console:
   ```
   $ python3 DoiScout.py
   ```

Windows

1. Start your *Python3*-distribution (e.g. *Anaconda*) and open the folder path where you downloaded the *DoiScout*.
2. Run the *DoiScout.py* script via (e.g. by writing it into the command line):
   ```
   $ python DoiScout.py
   ```

Mac OS X

1. Open a console and open the folder path where you downloaded the *DoiScout*.
2. Enter the following line in the console:
   ```
   $ python3 DoiScout.py
   ```

Operating the DoiScout

Features. Once started, the main menu of the *DoiScout* is displayed (Fig. 1). From here, the individual subfunctions can be called up by clicking on the corresponding buttons. Detailed information on the subfunctions can be found in chapters Extracting Bibliographic Information to Convert to Endnote.

Handling.

- Select the desired function:
  1. Extracting Bibliographic Information
  2. Download PDFs from DOIs
  3. Extract Information from Study Register IDs
  4. Cited by
  5. Cited in Guidelines
  6. Convert to Endnote
  7. Manual
  8. Quit

Result. By clicking on the buttons the respective submenu is displayed.
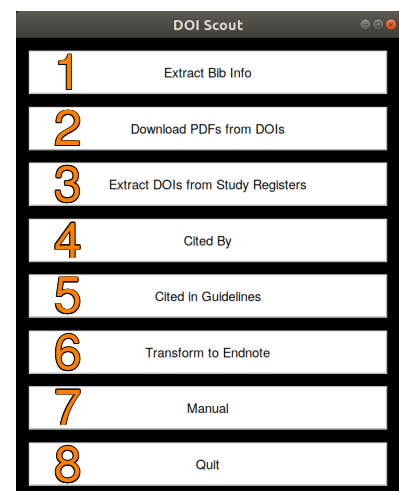

Figure 1: Main window of the *DoiScout* after starting the tool. The available functions of the *DoiScout* can be selected here.

Extracting Bibliographic Information

Description. This function searches and saves bibliographic information (DOI, title, authors, issue, etc.) for one or more DOIs or publication titles specified by the user. The bibliographic information is searched at *Pubmed* and enhanced by information from the *Web of Science*, if *Pubmed* cannot provide information for the respective field (e.g. DOI, title, author, etc.).

Handling.

1. Opens this manual.
2. Field for entering the DOIs or the publication titles you want the bibliographic information extracted for. Multiple entries must be separated by new lines.
   Note for 2-5 If the publication titles rather than DOIs are provided, titles must be accurate without any alterations or gaps (except for special characters). Otherwise no information will be identified.
3. Reads a text file. There must be nothing in this file other than the DOIs or titles. Single entries must be entered on unique lines.
   Note for 3-5 All entries imported from files are appended. Entries already in the field are not overwritten.
4. Reads a *Web-of-Science* file.
5. Reads a *Pubmed* file from NCBI.
6. Information field where current information is displayed (for more information and progress, see The Terminal Window).
7. Select if after the extraction of the bibliographic information the full texts (PDFs) should be downloaded (if access is granted through your institutional license). See Download PDFs from DOIs for further information.
8. Select the folder where the bibliographic information (and PDFs if selected in 7) should be saved.
9. Currently selected folder for the results.
10. Option for receiving mail alerts when the search is finished or failed.
11. Return to the main menu (Getting Started).
12. Run the extraction.
    Note. Only run the extraction after all previous options and settings have been selected.

Result. In the result folder (see 9) the following file(s) are created with the described content:

- *[Timestamp]_Extracted_Bib_Info.txt* (Fig. 3)
  - This file contains the result of the bibliographic extraction. Every entry that was provided by the user (Fig. 2, 2) will be represented on a new line.
  - This file comprises the following bibliographic information
    * *DOI* (digital object identifier)
    * *Pubmed ID*
    * Publication *Title*
    * *Number of authors*
    * *Authors*
    * *Country* of the first author
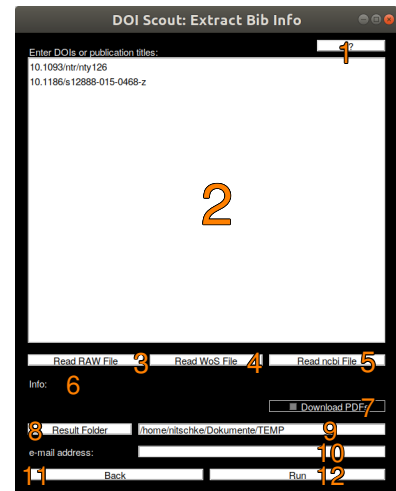   * *Year* of publication

Figure 2: The main window for the extraction of bibliographic information from *Pubmed* and the *Web of Science*.

- ⋆ *Journal* it was published in
- ⋆ *Volume*
- ⋆ *Issue*
- ⋆ *Pages*
- ⋆ *Received in Year*: Year the manuscript was submitted
- ⋆ *Received in Month*: Month the manuscript was submitted
- ⋆ *Received in Day*: Day the manuscript was submitted
- ⋆ *Accepted Year*: Year the manuscript was accepted
- ⋆ *Accepted Month*: Month the manuscript was accepted
- ⋆ *Accepted Day*: Day the manuscript was accepted
- ⋆ *epub_day*: Day the manuscript was published online
- ⋆ *epub_month*: Month the manuscript was published online
- ⋆ *epub_year*: Year the manuscript was published online
- ⋆ *print_day*: Day the manuscript was printed
- ⋆ *print_month*: Month the manuscript was printed
- ⋆ *print_year*: Year the manuscript was printed
- ⋆ *Publication_Type_Pubmed*: The publication type(s) listed in Pubmed
- ⋆ *Publication_Type_WoS*: The publication type(s) listed in Web of Science
- ⋆ *Cited_by_Pubmed*: Number of citations of the present publication listed in Pubmed
- ⋆ *Cited_by_WoS*: Number of citations of the present publication listed in Web of Science
- ⋆ *Cited_by_Cochrane*: Number of citations of the present publication listed in the Cochrane Library
- ⋆ *URL*: web address to the Pubmed- and Web-of-Science-entries
- ⋆ *abstract*: The abstract of the publication.
- • *[Timestamp]_Failed_Bib_Info.txt*
  - – This file contains every DOI or publication title by the user (Fig. 2, 2) for which no entry either in *Pubmed* or *Web of Science* was found.
  - – This file is only created if there were any DOIs or publication titles that were not found.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	DOI	PubmedID	Title	Number of Author	Author	Country	Year	Journal	Volume	Issue	Pages	Received in Year	Received
2	10.1093/ntr/nty126	29917096	Craving to Quit: A Randon	9	Garrison, KA; Pal, P; O'Mail		2018	Nicotine & tobacco rese				2018	3
3	10.1186/s12888-015-0468-z	25884648	A randomized controlled tr	6	Garrison, KA; Pal, P; Rojian	USA	2015	BMC psychiatry	15			2015	1

Figure 3: The *[Timestamp]_Extracted_Bib_-Info.txt*-file imported into a spreadsheet software (e.g. Excel). For the sake of space last bibliographic information is truncated.

Download PDFs from DOIs

Description. This function downloads the full text PDFs from the specified DOIs or titles, if licensed by your organization. The PDFs are downloaded directly from the journal pages or from OVID.

This function is intended for downloading a great number of PDFs. If the user is only interested in single PDFs, he/she is usually much quicker downloading them manually. If the *DoiScout* was not able to download a full text, this does not necessarily mean that the user cannot obtain it from other sources (e.g. ResearchGate etc.). The *DoiScout* lists the titles that have not been found, so that they can be searched for manually.

Handling.

1. Opens this manual.
2. Field for entering the DOIs or the publication title you want the PDFs downloaded for. Multiple entries have to be separated by new lines.
   Note for 2-5 If (instead of DOIs) the publication titles are provided those titles must be accurate without any alterations or gaps (except for special characters). Otherwise no information will be identified.
3. Reads a text file. There must be nothing in this file other than the DOIs or titles. Single entries have to be listed line by line.
   Note for 3-5 All entries imported from files are appended. Entries already in the field are not overwritten.
4. Reads a *Web-of-Science* file.
5. Reads a *Pubmed* file from NCBI.
6. Information field where current information is displayed (for more information and progress, see The Terminal Window).
7. Select if bibliographic information should be extracted in addition to the downloading of PDFs (see Extracting Bibliographic Information). Extracted information is stored in a spreadsheet file.
   Bibliographic information is only obtained for PDFs that were successfully downloaded.
8. Select the folder where the bibliographic information (and PDFs if selected in 7) should be saved.
9. Currently selected folder for the results. A subfolder *fulltexts* will be created where the downloaded PDFs are stored.
10. Option for receiving mail alerts when the download is finished or failed.
11. Return to the main menu (Getting Started).
12. Start the download.
    Note. Only start the download after all previous options and settings have been selected.

Result. In the specified result folder (9) a subfolder named "fulltexts" is created if not existing. This folder contains all downloaded PDFs. The PDFs are automatically renamed after their DOIs. All / are replaced by _. In addition, the file *[Timestamp]_PDFs_not_optained.txt* is created in the result folder (9). This file lists all DOIs or titles for which no PDFs could be downloaded and is created even if all PDFs were downloaded, thus, the user can ensure no PDF was missed.
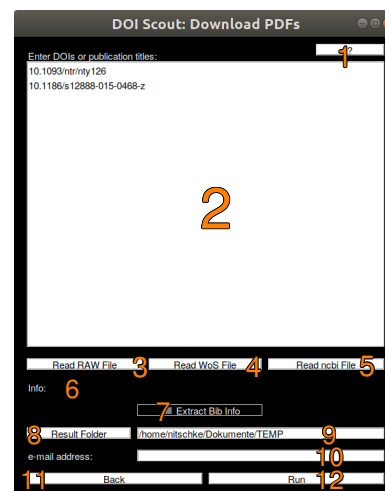

Figure 4: The main window for downloading full texts (PDFs) from the respective journal website or OVID.

Extract Information from Study Register IDs

Description. This function is used to find publications based on study registry numbers (study IDs) reported in publications. For this purpose, the databases listed below are searched according to your selection. This function always additionally searches for bibliographic information on the results (see Extracting Bibliographic Information). If no bibliographical information is found or resources that cannot be automatically processed are found, those sources are marked separately for further processing by the user. Study registry numbers that no resources are found for are listed separately. Note that failure to return any information may result from incorrect user entries.

Databases that can be searched (Fig. 5):

- study registers
  - ClinicalTrials.gov (`https://clinicaltrials.gov`)
    * provides publication citations and results
  - EudraCT (`https://www.clinicaltrialsregister.eu`)
    * does not provide publications citations, therefore, an EudraCT search will always show up in the *NoResults* -File
    * provides results
  - ISRCT (`https://www.isrctn.com`)
    * provides publication citations
    * does not provide results itself, but has a field «Participant level data». Only if this is specified as «Stored in repository» the resp. ID will be evaluated positively for *ResultsProvided*.
- Pubmed (NCBI)
- Google Scholar
- Livivo
- Trip Database
- Base-search.net
- Web of Science (Web of Knowledge)

Handling.

1. After selecting *Extract DOIs from Study Registers* from the main menu the user selects (Fig. 5) which sources should be searched afterwards.
2. Opens this manual.
3. Input field for the study register numbers (IDs). Multiple entries are possible and have to be separated by linebreaks.
4. Reads a textfile and enters its content in the input field (3).
   Note. All entries imported from files are appended. Entries already in the field are not overwritten.
   Note. The study registration numbers should be as precise and consistent as possible. Spaces should be avoided. Failure to do so may result in a considerably longer *DoiScout* runtime and hundreds of pointless results, which must then be processed manually.
5. Information field where current information is displayed (for more information and progress, see The Terminal Window).
6. Option to download PDFs of results stored in study registers.
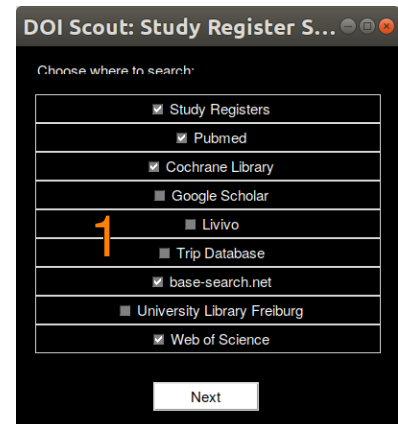7. Option to subsequently download full texts (PDFs) of the DOIs found in this

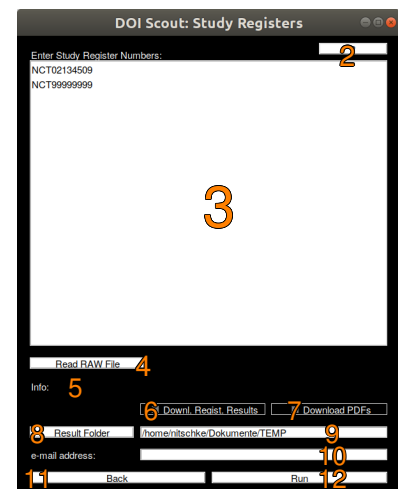Figure 5: Choice of the resources that will be searched during the subsequent step.

Figure 6: The main window for the extraction of information from study registers.

search (see Download PDFs from DOIs).

8. Select the folder where the results (and PDFs if selected 6) should be saved.

9. Currently selected folder for the results.

10. Option for receiving mail alerts when the search is finished or failed.

11. Return to the main menu.

    Note. If you want to change the selection of the sources to be searched (1), you can return to the main menu (Getting Started) here and call (Fig. 1, 3) again.

12. Run the search.

    Note. Only run the search after all previous options and settings have been selected.

Result. In the result folder (see 9) the following files are created with the described content:

- *[Timestamp]_Results_Register_Search.txt* (Fig. 7)
  - Contains the results of the register search.
  - The file contains the following information, listed by column number:
    * Column 1 (13): *SearchTerm* (in this case study registry number; e.g. NCT02134509) that was used to identify the publications .
    * Columns 2 - 31 (14): Bibliographic information (see Extracting Bibliographic Information for more information).
    * Columns 32 - ? (15): (the number of respective columns changes in accordance with number of selected sources):
      FV_[source name] (e.g. *FV_Pubmed*). Indicates whether the result in the respective line (e.g. the study with the DOI 10.1186/s12888-015-0468-z) was found in the named ressource: *1* indicates the study was found (e.g. for Pubmed); *0* indicates the study was not found (e.g. in the Cochrane Library). A publication identified by a single study register number, but found in multiple sources is still only listed once but with several *1*s in the *FV_[source name]* columns (e.g. the study with the DOI 10.1186/s12888-015-0468-z was found via the study register https://clinicaltrials.gov/ as well as in the pubmed search).
    * Penultimate column: contains source_urls of the original references to simplify further manual processing by the user.
    * Last column: *attention_required* indicates whether the resource found requires further processing by the user (*1* if yes; otherwise *0*). This occurs if no DOI was found, or if the publication was not identified in either Pubmed or Web-of-Science.
- *[Timestamp]_NoResults_Register_Search.txt*
  - List of all study registry numbers for which no results were found in any of the specified sources (Fig. 5, 1). This might be caused by a spelling error or any other error in the input (Fig. 6, 3).
- *[Timestamp]_Errors_Register_Search.txt*
  - If an error during a search occurs and the automatic search for a given study registration number cannot be continued in the current source (Fig. 5, 1), the study registration number is added to the list within this file.

    Note. A manual search is required for these study registry numbers, otherwise the systematic search is incomplete and, therefore, invalid.

Figure 7: The *[Timestamp]_Results_-Register_Search.txt*-file imported into a spreadsheet software (e.g. Excel). For the sake of space most bibliographic information is hidden in this figure.

- **–** Structure

    1st column: Source in which the error occurred.

    2nd column: Study registration number for which the error in the source (1st column) occurred

- *[Timestamp]_Results_Provided_in_Study_Registers_.txt*
    - **–** A list indicating whether study results were reported in the corresponding study registers for each study registry number.

        1 means that results were reported.

        0 means that no results were reported.

        -1 means that the corresponding study register is not implemented in the *DoiScout* or that the corresponding number was not found in the respective study register.

Cited by

Description. This function searches the *Pubmed* and *Web of Science* for the specified DOIs and titles (Original-DOIs, also parents) and extracts other publications that have cited them (children). All citing sources and their bibliographic information are stored in a spreadsheet readable file.

The search depth option (7) allows the *cited-by*-function to be run recursively over several levels. The standard selection is *1* where the citation tree is followed for one level, i.e. all information from all publications citing the Original-DOIs is extracted. If a search depth of *0* is selected, only the bibliographic information of the Original-DOIs (the DOIs specified by the user) is extracted. If a search depth of *2* is selected, information from the Original-DOIs (also level-0-parents) is extracted, as well as information from from all publications citing the Original-DOIs (children of the Original-DOIs, but also level-1-parents of the following DOIs) as well as all information from publications citing the publications that cited the Original-DOIs (the grandchildren, i.e. children's children, of the Original-DOIs, but also the children of the level-1-parent-DOIs). Therefore, all children are also parents of the grandchildren.

Theoretically, for the search depth option every number larger or equal to *0* can be chosen. However, with increasing search depth citation-trees grow exponentially and so does the runtime of the *cited-by*-search. In most cases the default setting of *1* will suffice.

All trees and sub-brachnes are treated as fully independent trees by the *DoiScout* even if – in reality – they might intertwine as a child can have multiple parents, i.e. a publication citing multiple other publications that are also in the tree. The results are stored with all information so that the user can establish these connections later on by him/herself.
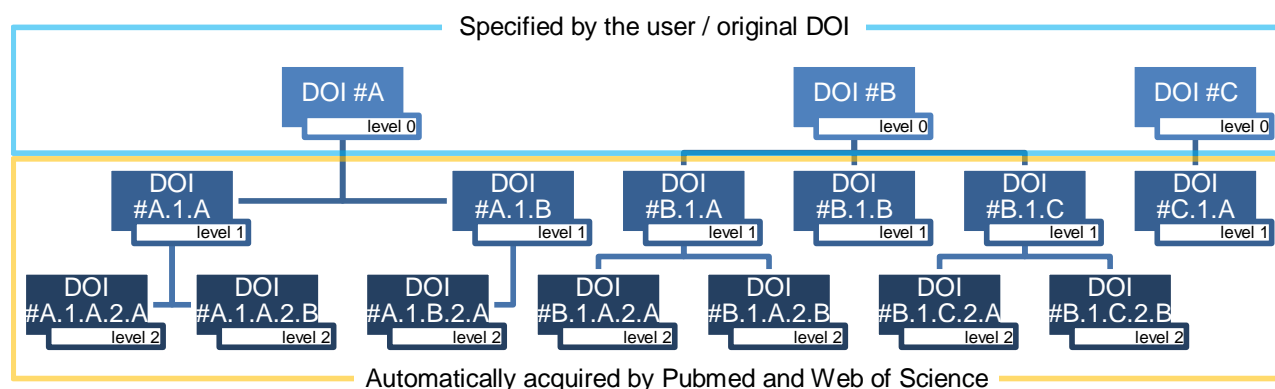

Figure 8: Example tree of a *cited-by*-search where the user specified three DOIs.

In Fig. 10 an example *cited-by*-search is illustrated of a search depth of 2.

Level 0. The user specified three parent-DOIs (*DOI #A*, *DOI #B*, and *DOI #C*). Those are also called Original-DOIs. All Original-DOIs are the root of their own, independent tree.

Level 1. *DOI #A* was cited by two children-publications (*DOI #A.1.A* and *DOI #A.1.B*). *DOI #B* was cited by three children-publications (*DOI #B.1.A*, *DOI #B.1.B*, and *DOI #B.1.C*). *DOI #C* was cited by one child-publication (*DOI #C.1.A*). A *cited-by*-search of search depth 1 would end at this stage.

Level 2. However, as the user specified a search depth of 2, the *cited-by-*search is continued for another level. In this regard, every now child-DOI from Level 1 can be seen as a new parent-DOI (however, they do not become Original-DOIs as those remain the ones specified by the user) and, therefore, its own new tree (within a tree). The *DOI #A.1.A* was cited by two DOIs (*DOI #A.1.A.2.A* and *DOI #A.1.A.2.B*). Each of these can be seen as a grandchild of the Original-DOI (DOI #A), or as a child of the new parent DOI #A1A. Had DOI #A1A been specified by the user as an Original-DOI with a search depth of 1, the resulting tree would be identical to the branch here, but would be represented at a higher level. Both, *DOI #B.1.A* and *DOI #B.1.C* were cited by two publications (*DOI #B.1.A.2.A* and *DOI #B.1.A.2.B* respectively *DOI #B.1.C.2.A* and *DOI #B.1.C.2.B*). *DOI #B.1.B* and *DOI #C.1.A* were not cited by any publication. As the search depth was set to 2, the *cited-by-*search ends at this stage. Had it been set to 3, another level with all children of the seven DOIs at this level would be added.

Handling.

1. Opens this manual.

2. Field for entering the DOIs or the publication title you want to find citations for. Multiple entries have to be separated by new lines.

    Note for 2-5 If the publication titles rather than DOIs are provided, titles must be accurate without any alterations or gaps (except for special characters). Otherwise no information will be identified.

3. Reads a text file. There must be nothing in this file other than the DOIs or titles. Single entries have to be listed on separate lines.

    Note for 3-5 All entries imported from files are appended. Entries already in the field are not overwritten.

4. Reads a *Web-of-Science* file.

5. Reads a *Pubmed* file from NCBI.

6. Information field where current information is displayed (for more information and progress, see The Terminal Window).

7. Select the search depth (see Description for further information).

8. Select if after the *cited-by-*search the full texts (PDFs) of all parents and children should be downloaded (if access is granted through your institutional license). See Download PDFs from DOIs for further information. Beware of enormous runtimes for huge trees.

9. Select the folder where the *cited-by-*search results (and PDFs if selected in 8) should be saved.

10. Currently selected folder for the results.

11. Option for receiving mail alerts when the search is finished or failed.

12. Return to the main menu (Getting Started).

13. Start the *cited-by-*search.

    Note. Only start the search after all previous options and settings have been selected.

Result. In the result folder (see 10) the following file is created with the following content:

- *[Timestamp]_Results_Cited_By.txt*
    - The file contains the following information, listed by column number:
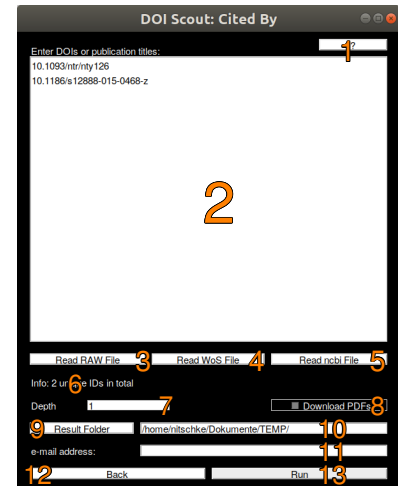   * Column 1 (14): Original-DOI respectively the root of the tree. Always one of the user-specified DOIs.

Figure 9: The main window for finding citations of specific publications. *Pubmed* and *Web of Science* are searched in.

⋆ Column 2 (15): The parent DOI of the current publication. For all levels smaller or equal to 1, the parent DOI and the Original-DOI (14) are identical.

⋆ Column 3 (16): The search depth level at which the current publication was found.

⋆ Column 4 (17): This column indicates whether the current publication was found in Pubmed.

Note for 17-18: *0* indicates it was not found by the respective source; *1* indicates it was found by the respective source. The Original-DOIs on search depth level 0 were provided by the user and were, therefore, not identified by any source, indicated by the value *-1*.

⋆ Column 5 (18): This column indicates whether the current publication was found in Web of Science.

⋆ Columns 6 - 24 (19): The bibliographic information of the current publication (see Extracting Bibliographic Information).

| | 14 | | 15 | | 16 | 17 | 18 | | | 19 | | |
Original_DOI	Parent_DOI	SearchDepth	Cited_in_Pubmed	Cited_in_WoS	DOI	PubmedID	Title
10.1093/ntr/nty126	10.1093/ntr/nty126	0	-1	-1	10.1093/ntr/nty126	29917096	Craving to Quit: A Rand
10.1093/ntr/nty126	10.1093/ntr/nty126	1	0	1	10.1111/add.14652	31083767	Effectiveness of an offe
10.1093/ntr/nty126	10.1093/ntr/nty126	1	0	1	10.1007/s12671-019-01188-6		Baseline Craving Streng
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	1	10.2196/13059	31237242	Mindfulness-Based Smo
10.1093/ntr/nty126	10.1093/ntr/nty126	1	0	1	10.1007/s12671-019-01145-3		Ancient Indian Educatio
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	1	10.2196/11498	30912755	Combining Real-Time R
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	1	10.1186/s12889-019-6605-8	30849943	Mass media promotion
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	1	10.1002/14651858.CD006611.pub5	31638271	Mobile phone text mess
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	0	10.12688/f1000research.16148.2	30728950	Relative effectiveness o
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	0	10.1016/j.copsyc.2018.12.012	30682663	Mindfulness-based inter
10.1093/ntr/nty126	10.1093/ntr/nty126	1	1	0	10.2147/SAR.S145201	30532612	Mindfulness meditation
10.1093/ntr/nty126	10.2196/13059	2	1	0	10.1016/j.copsyc.2018.12.012	30682663	Mindfulness-based inter
10.1093/ntr/nty126	10.2196/11498	2	1	1	10.1016/j.cct.2019.03.006	30923022	Proactive text messagin
10.1093/ntr/nty126	10.12688/f1000research.16148.2	2	1	0	10.1177/2055207619880676	31620306	Does the addition of a s
10.1093/ntr/nty126	10.2147/SAR.S145201	2	0	1	10.1080/14659891.2019.1640305		Mindfulness-based relap
10.1093/ntr/nty126	10.2147/SAR.S145201	2	0	1	10.1111/acer.13993	30817013	Impaired Specificity of F
10.1093/ntr/nty126	10.2147/SAR.S145201	2	1	1	10.1016/j.jsat.2019.01.013	30898331	Mindfulness-based relap
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	0	-1	-1	10.1186/s12888-015-0468-z	25884648	A randomized controlled
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1007/s12671-019-01188-6		Baseline Craving Stren
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1016/j.invent.2019.100250	31110950	Smartphone-based emo
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1038/s41386-019-0403-y	31039580	Quitting starts in the bra
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1016/j.neuroimage.2019.03.066	30953836	Mediation analysis of tri
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1002/14651858.CD012659.pub2	31684699	Real-time video counse
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1002/14651858.CD006611.pub5	31638271	Mobile phone text mess
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1093/ntr/ntx171	29065202	What Do Smokers Want
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.2196/10893	30355563	Effect of a Mobile Phone
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.3389/fpsyg.2018.01418	30250438	Can Mindfulness Addres
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1016/j.addbeh.2017.11.027	29174666	Combining ecological m
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1016/j.dcn.2018.03.008	29636283	Current, future and pote
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1016/j.invent.2017.07.002	30135772	Reducing procrastinatio
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1093/tbm/ibx031	29447386	Kick.it: The developmen
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1007/s10865-017-9884-5	28918456	Testing a mobile mindfu
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1177/1460458216656472	27354394	A randomized controlled
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1371/journal.pone.0191661	29370256	Effects of a brief mindfu
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1080/14639947.2018.1443553		Feeling is Believing: The
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1037/adb0000319	29072477	Mindfulness-based inter
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1016/j.ijmedinf.2017.05.018	28750912	Development and initial
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1007/s13142-017-0492-2	28527027	A systematic review of s
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	0	1	10.1016/j.cmpb.2017.02.022	28325449	TControl: A mobile app
10.1186/s12888-015-0468-z	10.1186/s12888-015-0468-z	1	1	1	10.1016/bs.pbr.2017.08.013	29054291	Neuroscience-informed

Figure 10: *[Timestamp]_Results_Cited_By.txt* example of a *cited-by*-search with two Original-DOIs specified by the user and a search depth of 2.

Cited in Guidelines

Description. This function searches the databases AWMF, Trip (Medical) database, and NICE if selected for guidelines that cite specified publications. As these databases use inconsistent citing styles that do not necessarily include the DOI, this function requires a previous step in order to work (so the DOIs and titles are provided for the search).

Note. The user needs to run a bibliographic extraction first (Extracting Bibliographic Information) on all DOIs or titles he/she is interested in. Afterwards the resulting file can be read in this subfunction. An alteration within the result file from the bibliographic extraction function will most likely cause crashes.

Handling.

1. After selecting *Extract DOIs from Study Registers* from the main menu the user selects (Fig. 5) which databases should be searched afterwards.
2. Opens this manual.
3. Input field for the path to the file containing the bibliographic information for the search.
4. Opens a file explorer so the path to the file containing the bibliographic information can be chosen (3).
5. Information field where current information is displayed (for more information and progress, see The Terminal Window).
6. Option to search *Web of Science* and *Pubmed* for bibliographic information.
7. Select the folder where the results should be saved.
8. Currently selected folder for the results.
9. Option for receiving mail alerts when the search is finished or failed.
10. Return to the main menu.

    Note. If you want to change the selection of the sources to be searched (1), you can return to the main menu (Getting Started) here and call (Fig. 1, 5) again.
11. Runs the search.

    Note. Only run the search after all previous options and settings have been selected.

Result. In the result folder (see 8) the following file is created with the described content:

- *[Timestamp]_Guideline_Search.xlsx* (Fig. 13)
  - The file contains the following information, listed by column number:
    * Column 1 (12): The DOIs and titles that were listed in the imported file (Fig. 12, 3/4).
    * Column 2 (13): Some databases calculate a accuracy quote of the search term and the identified guideline. This might be useful to filter very low match score. The value is between 0 and 1.
    * Column 3 (14): The internal id within the respective database.
    * Column 4 (15): The evidence level or the kind of entry within the database.
    * Column 5 (16): The society that published the guideline.
    * Column 6 (17): The URL that leads to the guideline.
    * Column 7 (18): Was the entry identified via the AWMF database.
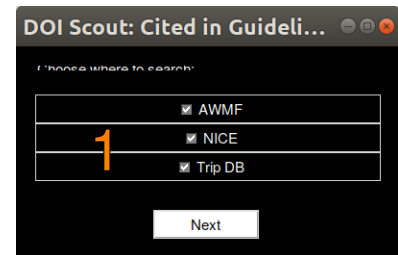

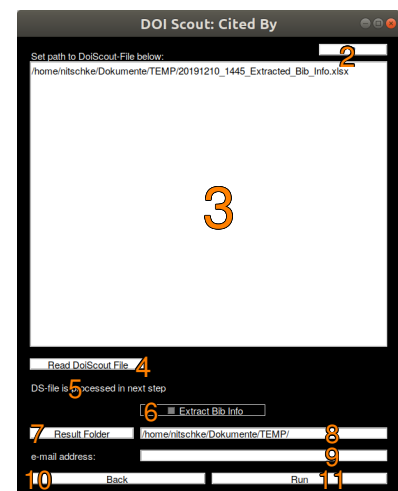Figure 11: Choice of the databases that will be searched during the subsequent step.

Figure 12: The main window for the search of citations in guidelines.

Note for 18-20: *0*: it was not identified via this database. *1*: it was identified via this database.

* Column 8 (19): Was the entry identified via the NICE database.
* Column 9 (20): Was the entry identified via the Trip database.
* Column 10-28 (21): The bibliographic information of the current guideline or publication (see Extracting Bibliographic Information).
  Note: These columns will be empty if Fig. 12 – 6 was not selected.

12	13	14	15	16	17	18	19	20	21		
Parent-DOI	hit quot	id	evidence-level	society	guideline-url	in_awmf	in_nice	in_tripdb	PubmedID	DOI	Title
10.2105/ajph.86.2.214			Guidance	National Institute for Health and Clinical Excellence - Clinical Guidelines	https://www.nice.org.uk/c	0	1	0			Smoking prevention in schools : guid
10.2105/ajph.86.2.214	0.023	9007181	Uk Guidelines		http://www.nice.org.uk/gu	0	0	1			Smoking prevention in schools
10.1183/09031936.00060806											
10.1001/archpsyc.55.8.683	1.000	128-001OL	S3	Deutsche Gesellschaft für Palliativmedizin (DGP)	https://www.awmf.org/up	1	0	0			Palliativmedizin für Patienten mit ein
10.1001/archpsyc.55.8.683	0.420	020-005	S3	Deutsche Gesellschaft für Pneumologie und Beatmungsmedizin e.V. (DGP)	https://www.awmf.org/up	1	0	0			Tabakentwöhnung bei COPD
10.1001/archpsyc.55.8.683	0.260	076-006	S3	Deutsche Gesellschaft für Suchtforschung und Suchttherapie e.V. (DG-Sucht)	https://www.awmf.org/up	1	0	0			Tabakkonsum (Rauchen), abhängige
10.1001/archpsyc.55.8.683			Primary Research	Annals of Internal Medicine	http://annals.org/article.a	0	1	0	24042367	10.7326/0003-4819-159-6-201309170-00005	Effects of varenicline on smoking ces

Figure 13: Example result file (*[Timestamp]_-Guideline_Search.xlsx*) of a guideline search with three specified DOIs by the user.

Convert to Endnote

Description. The original result file of other sub functions (e.g. the Extracting Bibliographic Information) is a spreadsheet file. As many users prefer *Endnote* to manage their literature, this function transforms the results into a file that can imported into *Endnote*.

Handling.

1. After pressing the *Convert to Endnote* button in the main menu (Fig. 1) ...
2. ... the user will be confronted with a file selection prompt where the user selects the *DoiScout* file that he/she wishes to converts to Endnote.
3. The conversion file is been written in the same folder as the original file.
4. Open Endnote.
5. Select: *File » Import » File*. Fig. 14 appears.
6. Select the conversion file.
7. Select *Tab Delimited*.
8. Select *Import All*.
9. Select *No Translation*.
10. Press *Import*.

Result. One new file is written in the folder of the original chosen file. This file is named *OriginalFileName_EndnoteImport.txt*. If a file with that exact name already exists (e.g. from a previous conversion), it is not overwritten but a random sequence of letters is added to the filename. As Endnote has fixed format of columns (fields) some information has to be written into «Custom» fields. Moreover, Endnote is very limited regarding the number of fields. Therefore, not all information extracted by the *DoiScout* is importable into Endnote. All information is only included in the original file. The following information is written in the respective field:

- *Number of Authors » Custom 1*
- *Country » Custom 2*
- *Issue » Number*
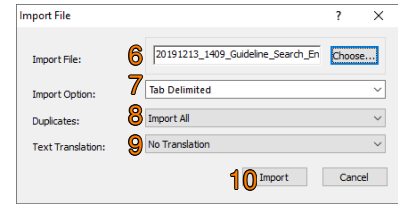- *Cited_By_Pubmed » Custom 3*
- *Cited_By_Wos » Custom 4*

Figure 14: Import results into Endnote after the conversion.

Manual

Description. Pressing the *Manual* button opens this manual in an external viewer. You can also find the manual pdf in the folder into which the *DoiScout* was downloaded.

Quit

Description. The *DoiScout* is closed by pressing the *Quit* button.

Further Information

The Terminal Window

While the *DoiScout* is performing its task in the terminal window (Fig. 15), information about the progress as well as an estimation of the remaining time needed for the present step are displayed. Upon finishing all processes, the word *Finished* is presented.

The Browser Window

For a trouble-free operation the *DoiScout* requires *Firefox* and the *Python* library *Selenium* (see Requirements and Technical Information) to be installed. During various sub functions, a Firefox window is being opened in which various forms of searches are performed.

Attention: This window can be minimized without creating problems. However, this window must remain completely unchanged by the user, otherwise false results may occur or the software will crash. Closing the window also causes these faults or crashes.

Timestamp

All files and emails written as the result of a sub function called by the user start with a timestamp. The timestamp consists of the date and time when the "RUN" button of the respective function was pressed. The format reads as follows:
YYYYMMDD_HHMMSS
e.g. *20191203_141517* for the 3rd December 2019, 2:15 pm.

E-mails

If you enter your email address in the individual functions, you will be notified about events of the *DoiScout*. This feature is only available on a *Linux* OS. The events might be:

- **Errors** occurring during the process. An *Python* error is printed in the terminal window (The Terminal Window). Experienced *Python* users may identify the source of the error themselves. Other users please write to the current supporter of the *DoiScout* (see license at the start of the manual).

- **Successes.** The chosen task of the *DoiScout* is finished. The mail summarizes the most important results of the task. If the option of downloading full texts was selected, this will only start after the currently selected task has been successfully completed. A separate mail will be sent after downloads were completed.

Figure 15: Example terminal output from a register search.

```
SUCCESS: 20191203_1415 - Searching Study Registers
+++++++++++++++++++++++++++++++++++++++++++++
Result Folder: /home/nitschke/Downloads/TEMP/
+++++++++++++++++++++++++++++++++++++++++++++
Total number of IDs: 3

Total number of identified resources: 31

Number of resources that require further attention: 21

Number of errors occured during searches that require

manual search: 0

Number of searches that did not yield any result: 0
```

- **Attentions.** Here the attention and the input of the user is required (Attention and Input Prompts). Before the respective prompt is answered, the *DoiScout* pauses.
  Example:

```
ATTENTION: 20180125_0928 - Action needed (captcha)
+++++++++++++++++++++++++++++++++++++++++++++
ATTENTION: The doi scout needs your attention
There is a Captcha you need to solve
Process is stopped until Captcha is solved
Switch to the browser and solve it please
```

Attention and Input Prompts

Input prompts might occur during a register number search (Extract Information from Study Register IDs). The search pauses and waits for the user to answer the input prompt in order to continue. The user will either be notified by e-mail (if an e-mail address has been specified) or should check at regular intervals whether an entry is necessary. Currently such prompts are only required by Google Scholar. For this reason this source is searched first, so the remaining automatic searches run smoothly. However, in most cases Google Scholar runs the first few searches and asks for a user prompt later. As soon as the search is completed using Google Scholar, no further control is necessary. This can be seen in the terminal (The Terminal Window) as `Searching Google Scholar` was concluded with 100%.

   The input prompts are opened in the browser window (The Browser Window) and must be answered there. The window must then remain unchanged (it can be minimized to a task bar, but do <u>not</u> close it).

Reading a Web of Science File

To read the DOIs or (if not available) the titles of publications found in *Web of Science* follow these steps:

1. Perform your search within the *Web of Science* with any internet browser you want until you found the results you want to import into the *DoiScout*.
2. At the top of the site click on *More* (Fig. 16).
3. Click on *Other File Formats* (Fig. 16).
4. Select *Records from ...* and enter the number of records you want to read into the *DoiScout* (Fig. 17).
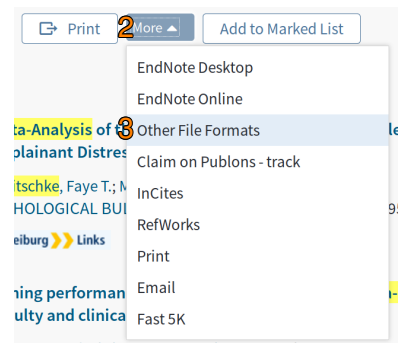

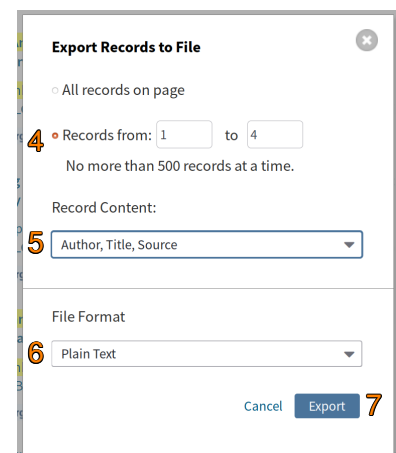Figure 16: WoS: Export the search results.

Figure 17: WoS: Print the results to a file readable by the *DoiScout*.

5. Select *Author, Title, Source* (Fig. 17).
6. Select *Plain Text* (Fig. 17).
7. Press *Export* (Fig. 17).
8. Save the file somewhere you will find it later.
9. Select the file in the *DoiScout* after pressing the *Read WoS file* button.

Reading a Pubmed File

To read the DOIs or (if not available) the titles of publications found in *Pubmed* follow these steps:

1. Perform your search within the *Pubmed* with any internet browser you want until you found the results you want to import into the *DoiScout*.
2. At the top of the site click on *Sent to* (Fig. 18).
3. Select *File* (Fig. 18).
4. Select *CSV* (Fig. 18).
5. Press *Create File* (Fig. 18).
6. Save the file somewhere you will find it later.
7. Select the file in the *DoiScout* after pressing the *Read ncbi file* button.

2 Send to ▾   **Filters:** Manag

Choose Destination

3 ○ File   ○ Clipboard
   ○ Collections   ○ E-mail
   ○ Order   ○ My Bibliography
   ○ Citation manager

Download 5 items.

Format

4 CSV
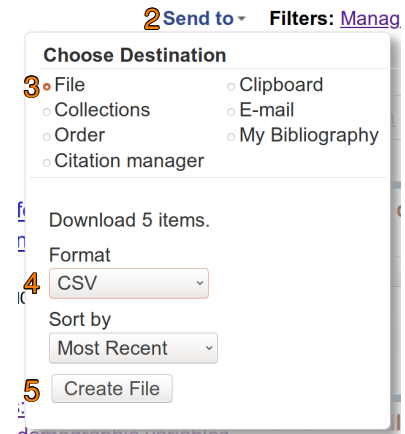
Sort by

Most Recent

5 Create File

Figure 18: Pubmed: Export the results to a file readable by the *DoiScout*.