ERC Project: The Early Islamic Empire at Work – The View from the Regions Toward the Center

Working Paper 02

José Haro Peralta and Peter Verkinderen

Jedli: A Textual Analysis Toolbox for Digitized Arabic Texts

جدلي

European Research Council
Established by the European Commission
Supporting top researchers from anywhere in the world

The Early Islamic Empire at Work
The View from the Regions Toward the Center

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Early Islamic Empire Working Paper Series
ISSN 2366-214X
Hamburg, February 2016

ERC Project: The Early Islamic Empire at Work –
The View from the Regions Toward the Center
Universität Hamburg
Edmund-Siemers-Allee 1
20146 Hamburg, Germany
General Editor: Stefan Heidemann
Volume Editor: Hannah Hagemann
W: https://www.islamic-empire.uni-hamburg.de/

Authors' Contact Details:

José Haro Peralta, M.A.
Research Associate, Universität Hamburg
ERC project 'The Early Islamic Empire at Work –
The View from the Regions Toward the Center'
E-Mail: jose.haro.peralta@uni-hamburg.de or joseharoperalta@gmail.com

Dr Peter Verkinderen
Research Associate, Universität Hamburg
ERC project 'The Early Islamic Empire at Work –
The View from the Regions Toward the Center'
E-Mail: peter.verkinderen@uni-hamburg.de

Content

Introduction

This paper is a manual to the textual analysis toolbox Jedli, which was developed by José Haro Peralta and Peter Verkinderen within the framework of the project *The Early Islamic Empire at Work: The View From the Regions Toward the Center*. The ERC-funded project is based at the University of Hamburg and directed by Stefan Heidemann.

Contrary to the traditional approach focused on the imperial center, Iraq, this project looks at the political, economic, and social structures of the early Islamic Empire (7^{th} – 10^{th} Centuries CE) from the perspective of five regions across the empire. Although the study of material culture (coins, seals, archaeological remains) is an integral part of the project, the analysis of (mostly Arabic) textual sources forms its main component. The researchers in the project aim at collecting as wide a source corpus as possible, containing texts from all genres, and spanning a period between the 7^{th} and 16^{th} centuries CE.

The sheer magnitude of this corpus combined with the large scope of the project's research questions forced us to develop an alternative strategy to retrieve information from the texts faster and in a more targeted way than is usually possible with the traditional means of textual research. This strategy makes use of the opportunities offered by collections of digitized classical Arabic texts, which have become available in the last two decades.

The two largest and most advanced collections of texts, *al-Jāmiʿ al-Kabīr* and *al-Maktaba al-Shamela*, offer, in addition to the texts, search interfaces that enable more or less complex searches in (subsets of) the corpus. These search interfaces allow the user to select the texts s/he wants to search, and one or more key words; the program will then provide a list of all the loci in the selected texts where all (or if desired: only one) of these key words are present on the same page. The latest version of al-Maktaba al-Shamela's search interfaces also has a number of useful search options for disregarding different combinations of *alif-hamza*, and dotted and undotted final *yāʾs* and *tāʾ marbūṭas*.

However useful these search interfaces may be, we found them too limited for our purposes and decided to develop a number of search tools that would be more flexible and better adapted to the searches we want to conduct, and that are not limited to the format of a specific collection of texts.

To develop our tools, we used the programming language Python, which is easy to learn, offers an interactive development experience of trial and error that we found particularly useful, and contains a number of modules that are very

suitable for textual analysis.[1] Since not every researcher will be interested in learning a programming language to interact with the texts, but many may be interested in using the tools, we made a user interface that makes the tools available for any user on a Windows system. We collected the tools in a toolbox which we named Jedli (Arabic for "Find for me"); additional tools will be added to the toolbox in the future.

Two of the three tools currently in the Jedli toolbox, the *Indexer* and the *Context Search*, are basically expansions of the existing search interfaces of the large collections mentioned above. They offer the possibility to search for numerous search terms (each can be a word, or a phrase), or even checklists of search terms, at the same time. Each search term, or set of search terms, can be given specific search options (e.g., which prefixes and suffixes should be allowed to be attached to the search terms, matching any *alif-hamza* combination, matching dotted and undotted final *yāʾs* and *tāʾ marbūṭas*); the user can select a number of presets (e.g., nominal prefixes, verbal suffixes, prepositions), or write his/her own search options using regular expressions. Regular expressions are, in computer programming, a way to define highly flexible and specific search patterns.

After selecting the search terms and search options, the user can choose which texts to search. Jedli uses the simple text format (.txt) for its texts, and we converted all texts from the al-Maktaba al-Shamela database to this format; the user can add any text to his/her Jedli library if it has been converted to .txt format. The source selection interface allows the user to filter the available texts by author, title, genre or date; s/he can also make his/her own groups of texts and save these selections for later re-use.

For every search term, the Indexer creates a list of locations in the selected texts in which that term is mentioned, and displays the immediate context around that search term; the user can define how many words of context s/he would like to get.

The Context Search goes one step further and allows you to filter the results of the index search based on the vocabulary Jedli finds in the context of the search term: the user can define one or more list(s) of terms that are likely to appear in contexts s/he is interested in. Search results that do not contain these

[1] For Python, see the official Python documentation (https://docs.python.org/3/library/index.html). There are a number of very good handbooks for Python, including BEAZLEY, David / JONES, Brian K. (2013), *Python Cookbook*, 3rd ed., Sebastopol, CA: O'Reilly Media; LUTZ, Mark (2013), *Learning Python*, 5th edition, Sebastopol, CA: O'Reilly Media; LUTZ, Mark (2013), *Programming Python*, 4th edition, Sebastopol, CA: O'Reilly Media.

terms within a range of words before or after the main search term are then disregarded. The user can also draw up a list of terms that denote contexts in which s/he is not interested; if such a word appears in the range of words before or after the main search term, the search result will be deleted from the list of results. The results are saved as a html file, which has the advantage of being readable across platforms while offering good formatting capabilities.

The third tool in the Jedli toolbox, the *Highlighter*, allows you to highlight search terms in a specific source. The user can select which color each (set of) search term(s) should be highlighted in, which allows the creation of a specific color scheme. The output is an html document containing the full text of the source, with the search terms marked in the designated colors. Each color has also a hidden symbol in the html file, which allows you to jump to the next result of a specific color. This tool is very useful as a reading help, to help you find relevant information faster while skimming or scanning a text. For example, when looking for information about Fārs in al-Ṭabarī's *Taʾrīkh*, you could make a list that contains the names of the main cities of Fārs, plus the term Fārs itself, to be marked in yellow; another list that contains false positives (e.g., *mi'at fāris*), to be marked in red; and a list of terms related to the topics you are interested in, to be marked in green. You could then jump from yellow section to yellow section using the hidden sign connected to it (e.g., *$a*). You could also have the phrase "*thumma dakhalat sana*" highlighted, which introduces every year in the chronicle, so that when an event catches my attention, you can immediately jump to its date by using its hidden symbol (e.g., *$d*).

The Jedli toolbox for Windows®, as well as a large selection of digital texts, can be downloaded from https://www.islamic-empire.uni-hamburg.de/en/publications-tools/digital-tools/downloads/jedli-toolbox.html. The source code of the tools has been made available on an open-source base (https://github.com/jedlitools), and we invite other researchers to contribute to the development of these and other tools, to adapt the tools to their own needs, and to use the code to develop tools of their own.

We hope users will find the Jedli toolbox useful for extracting information from large (corpora of) texts, and would welcome any feedback and requests for additional functionalities. For more information on the development and application of Jedli, see the forthcoming article by José Haro Peralta and Peter Verkinderen, “Find for me!”: Building a context-based search tool using Python”, in Elias Muhanna (ed.), *The Digital Humanities and Islamic & Middle East Studies*, Berlin: De Gruyter.

Quick Start Guide

Jedli is a data mining toolbox, designed to get more out of digitized Arabic texts.

Getting started with Jedli:

1. Download the zipped Jedli archive from the website: https://www.islamic-empire.uni-hamburg.de/en/publications-tools/digital-tools/downloads/jedli-toolbox.html
2. Unzip the Jedli archive and save the Jedli folder on your computer
3. In the Jedli folder, double-click the jedli_main.exe file to launch the program. When you launch it the first time, Jedli might take around a minute to install itself.

Jedli currently contains three main tools:

- The Indexer: creates customized index files for any number of words in any number of sources.
- The Highlighter: highlights the words the user provided in a text, in order to facilitate retrieving relevant information from lengthy texts; different colors can be used for different (sets of) search terms, allowing the user to create personal color schemes.
- The Context Search: allows you to limit your search results to relevant contexts, based on user-defined checklists.

This is the first version of Jedli, more tools will be added later to the toolbox.

The advantages of Jedli over other tools:

- checklists of search terms: you can create lists with an unlimited number of search terms, which you can use repeatedly for search operations.
- allows you to use separate, highly customizable search options for each search term.
- allows you to create your own source selections (based on genres, time periods, …).
- allows you to save all your search results.

Jedli comes with a small number of sample texts; more texts can be added by the user (see the manual).

A Guided Tour Through the Jedli Toolbox

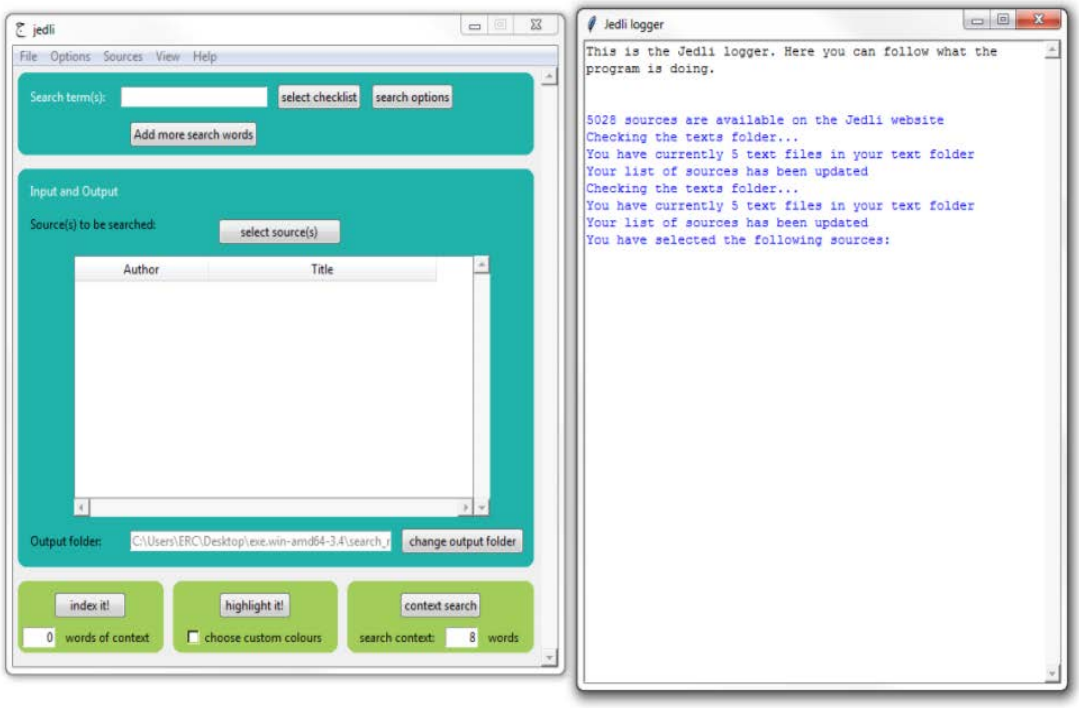
The main window of the Jedli toolbox contains four key elements:

1. the menu bar
2. the “search term” frame
3. the “input and output” frame
4. the frames for the three main Jedli tools: the Indexer, the Highlighter, and the Context Search

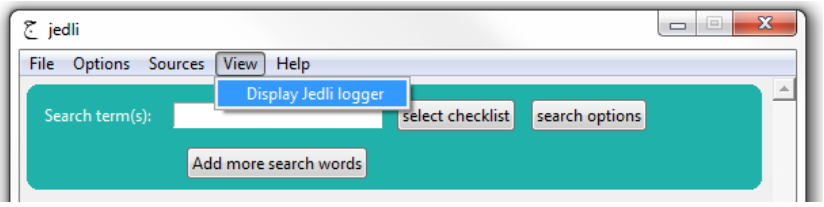
For any search, you should select

1. one or more search terms,
2. one or more sources you want to search,
3. a folder where the results of your search will be saved,
4. and the type of search you want to carry out.

Upon launching the program, a second window appears next to the main window: the Jedli logger. In this window, you can follow what the program is doing.



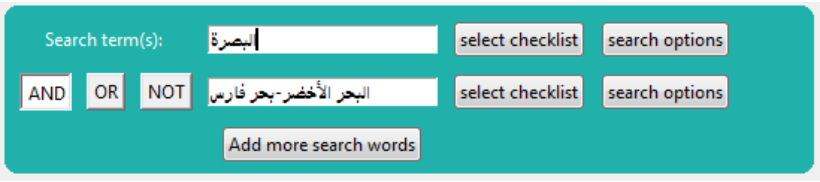
You can close the logger if you do not need it and re-open it at any time by clicking “Display Jedli Logger” in the View menu:



Selecting the Search Terms

There are a couple of ways to select search terms in Jedli:

a. You can simply type a search term (in Arabic) in the entry fields. Search terms can consist of one or more words. Clicking the "Add more search words" button will give you more entry fields to add search terms:



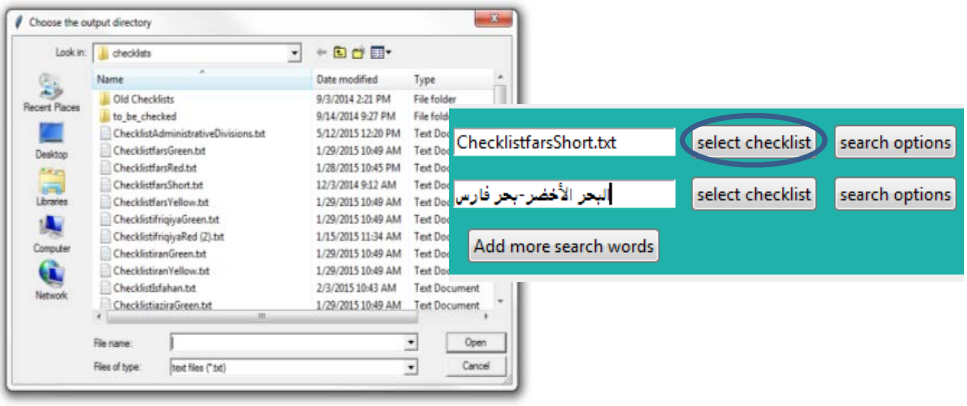
The Boolean operators AND, OR and NOT are only used for the Context Search tool. For more about the use of these operators, see below.

b. In each entry field, you can type more than one search term, divided by a hyphen ("-"). In the example above, Jedli will search for البصرة in combination with either بحر فارس or البحر الأخضر .

NB: take care not to include additional spaces between the search terms and the hyphens.

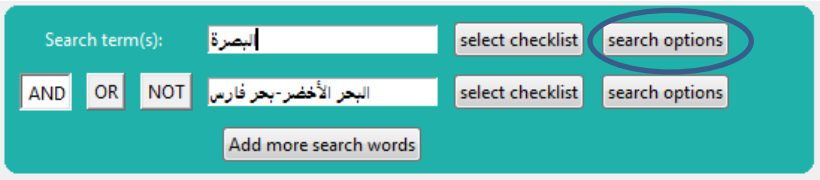
c. You can also employ user-defined checklists of search terms. These checklists are simple text documents (.txt files) that contain one search term on every new line (for more on checklists, see below):

You can open a checklist by clicking the “select checklist” button. This will open a standard “open file” dialog from which you can select your checklist:

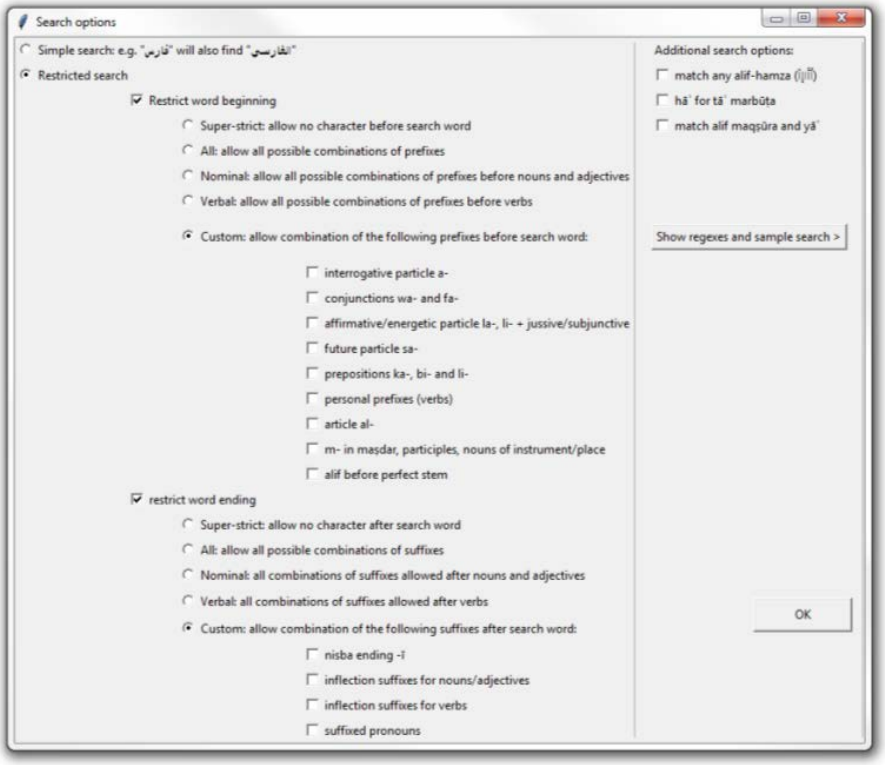


Selecting Search Options

For every (list of) search term(s), you can define specific search options by clicking the “search options” button.



This will open the following pop-up window (NB: the “custom” fields are unfolded only when “custom” is selected):



Arabic words can have prefixes and suffixes, and a number of enclitic particles and conjunctions, which have to be taken into account when conducting a search.

In the left-hand part of the “search options” window, you can define how Jedli deals with these.

You can opt for a simple search, which will simply look for the chain of characters of your search term, no matter what precedes or follows these characters. This might be useful (e.g., if you are searching for فارس, you might also be interested in someone who is called الفارسي), but it may lead to undesired results (e.g., if you are searching for كورة, you will also get results for مذكورة). In the other extreme ("super strict" search), you can tell Jedli to disregard all results where your search words are preceded or followed by any other character.

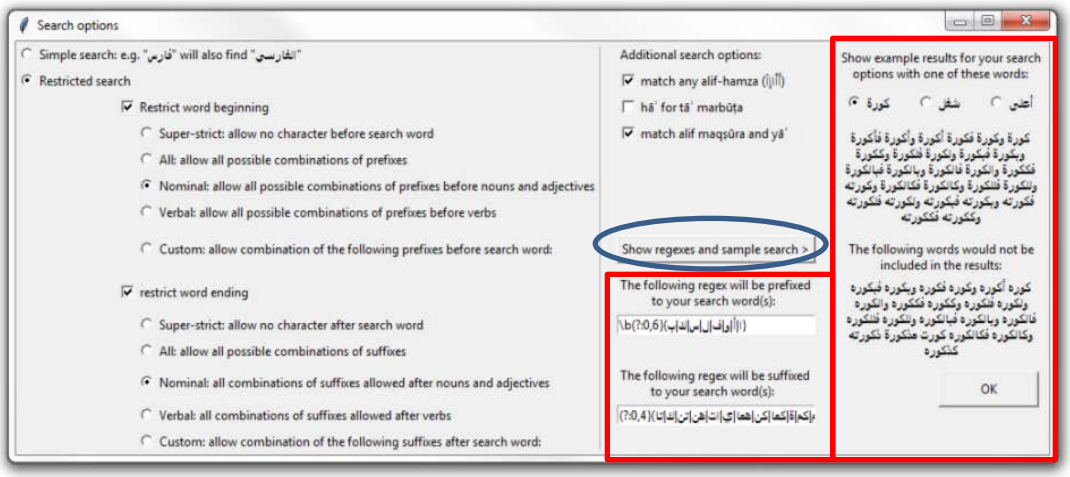
You can also define which characters should be allowed to precede and follow your search terms. Jedli groups these options in logical combinations (all, nominal, verbal), but you can also pick the prefixes and suffixes you want to allow ("custom").

Another annoyance for searches in Arabic are alternative and defective spellings. On the right-hand side of the screen, there are check boxes for additional search options that allow you to deal with three of the most common issues:

- *alif*s: since *hamzas* and *maddas* are not consistently written, it is often wise to disregard them.
- *tāʾ marbūṭa* and final *hāʾ*: in some (rare) cases, the dots on the *tāʾ marbūṭa* are omitted.
- final *yāʾ* and *alif maqṣūra*: the final *yāʾ* is sometimes written without its dots, and *alif maqṣūra* with dots.

Ticking the check boxes will tell the program to match each of the alternatives. E.g., if you search for الطبري, both الطبرى and الطبري will show up in your results.

When clicking the "show regexes and sample search" button, two new frames appear:



- Underneath the button, the regular expressions that will be used for the search are indicated. You can also write your own regular expressions here.[2]

- In the new column to the right of the button, you can see the effect of your search options on three sample search words.

Changing your search options will change the regular expressions and the search results.

[2] For more on regular expressions for searching Arabic texts, see HARO PERALTA / VERKINDEREN forthcoming.

Regular Expressions (Regex)

For the search options, Jedli uses regular expressions. A regular expression is a sequence of characters that defines a pattern. This pattern can be used to search, select, and replace sequences of characters in a text.

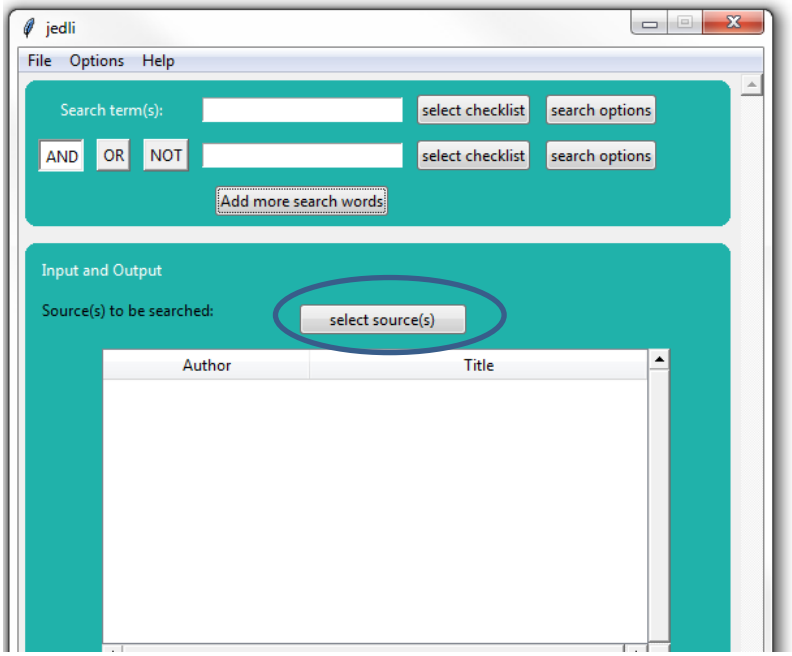
For example, if we want to find the word “color” in a text, but we do not know whether it is written in American or British spelling, we can use the following regular expression to match both forms:

colou?r

The question mark indicates that the preceding token (i.e., the “u”) is optional. Therefore, both “colour” and “color” will match this pattern and turn up in the results.

Selecting Sources

To select sources in which to search, click the “select source(s)” button in the “input and output” frame. This will open the source selection pop-up window:



By default, the table in "select your sources" contains all the sources in Jedli's "sources" folder (for more on adding sources to your sources folder, see below).

You can filter the sources displayed here by filling in one or more of the filter fields in the top part of the window ("Author", "Title", "Genre", and "Dates") and clicking the "apply filters" button. The "clear all" button deletes the input in all entry boxes and returns all sources to the "select your sources" table.

You can select a source by clicking on it; multiple sources can be selected at the same time by keeping the Ctrl or Shift key pressed while clicking. To move your selection of sources to the "selected sources" list, click the ">" button. You can move the entire list of sources in the left-hand table to the selected sources list on the right by clicking the ">>" button.

To remove sources from the selected sources list, click them and hit the "<" button; or use the "<<" button to remove all sources from the list.

When you are done selecting sources, click the "confirm selection" button to return to the main Jedli window and conduct a search in the selected sources, which will be displayed in the main Jedli window.

If you are going to use this same selection of sources more often, you can save your selection by clicking the "save selection for later" button. You will be prompted to provide a name for your selection.

The next time you use Jedli, your selection will be in the "my selections" drop-down menu in the filter field.

Jedli and al-Maktaba al-Shamela

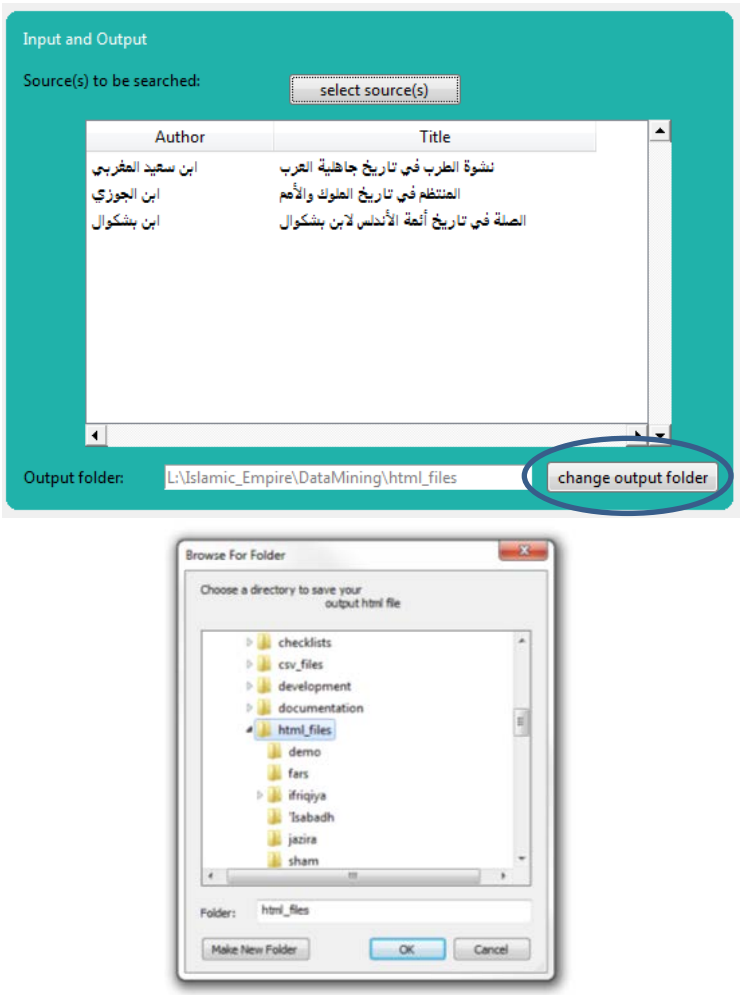
Al-Maktaba al-Shamela is currently the most extensive digital library of Arabic texts available.

Jedli is designed with the texts of al-Maktaba al-Shamela in mind. This means that for texts originating from the al-Maktaba al-Shamela collection, Jedli has the information on the author, title, genre, and date of the text in its database, and will be able to link the search results to the text on the Shamela website. The "genre" categories in al-Maktaba al-Shamela are not always very useful, but you can use the "save selection for later" button to create your own genres in Jedli.

You can also use Jedli for texts in .txt format that come from sources other than al-Maktaba al-Shamela by copying them into Jedli's "sources" folder. For these texts, you can enter the source information manually in the Jedli database (see below). If you have not entered any source information for a text, the "select sources" window will display the file name in the "Author" and "Title" fields.

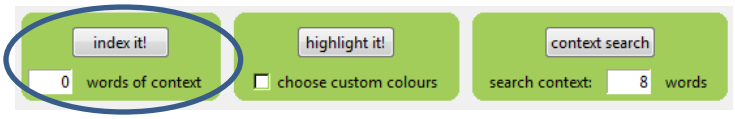
Selecting the Output Folder for the Search Results

The search results from the Jedli tools are saved to your computer as html files, which you can open in any web browser. By default, Jedli saves the html files to its "search results" folder. If you want to save the search results in another folder, click the "change output folder" button in the "input and output" frame. This will open a dialog window that lets you select the folder of your choice:

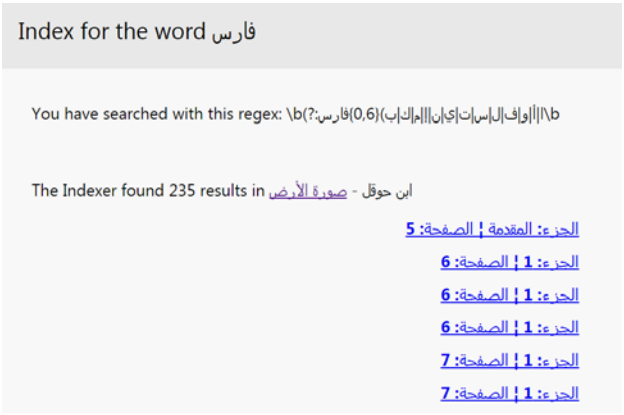


For more on how to change the default folder for your search results, see below.

The Indexer



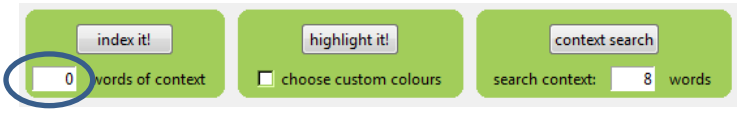
The Indexer creates an index (volume number and page number) for any of your search terms in every source you have selected. Simply click the “index it!” button to run the tool.



The page numbers are links; if you click on them, you are taken to the relevant page of the book on the Shamela website. Clicking on the book title will bring you to the book's info page.

The Indexer can also provide some context around the words to make it easier to see which results are relevant:

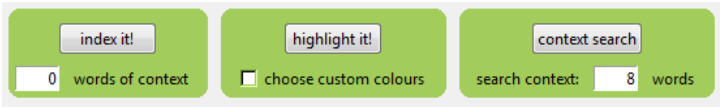
You can choose how many words of context will be included around the search word by changing the value in the "words of context" entry box:



The Highlighter

The highlighter creates a full-text version of every source text you select, highlighting the search terms in different colors to facilitate identifying relevant passages.

To run the Highlighter tool, click the "highlight it!" button:



The highlighted document opens in your browser window. Any browser works, but Google Chrome has an additional feature: if you use the browser's search function (ctrl+F) to search for the dollar sign ("$"), the browser will display colored lines in the scroll bar for every highlighted word in the document, allowing you to identify clusters of occurences of your search term(s). In the following example, we have highlighted the word "Fārs" in Ibn Ḥawqal's *Kitāb Ṣūrat al-Arḍ*, and the highlighting on the scroll bar clearly shows the sections of the work with a high concentration of hits:

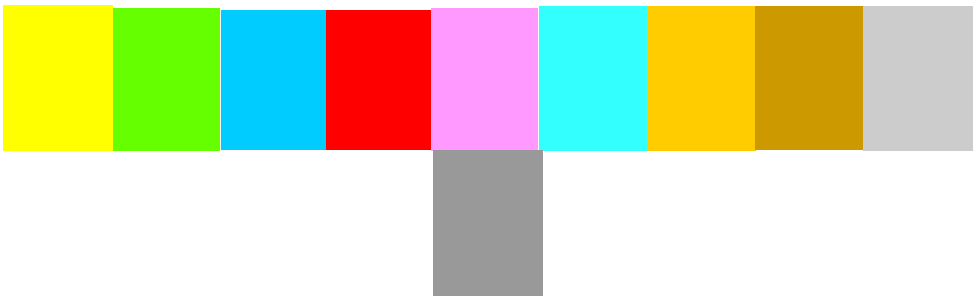

[بحر فارس]
(1) والذي يجب أن يذكر بعد ديار العرب بحر فارس لأنه يشتمل على أكثر حدودها وتتصل ديار العرب به وبكثير من بلدان الإسلام وتعتوره ثم أذكر جوامع مما يشتمل عليه هذا البحر، وأبتدئ بالقلزم وساحله مما يلي المشرق فإنه ينتهي الى ايلة ثم يطوف بحدود ديار العرب التي ذكرتها وأثبتها قبل هذا من هنا الى عبادان ثم يقطع عرض الدجلة وينتهي على الساحل الى مهروبان ثم الى جنابه ثم يمر على سيف فارس الى سيراف ثم يمتد الى سواحل هرموز من وراء كرمان الى الديبل وسواحل الملتان وهو [13 ب] ساحل السند وقد انتهى حد بلد الإسلام ثم ينتهي الى سواحل الهند ماضيا الى سواحل التبت فيقطعها الى أرض الصين، وإذا أخذت من أرض القلزم من جانب البحر الغربي على ساحله سرت في مفاوز من حدود مصر حتى تنتهي الى جزائر تعرف ببني حدان وكان بها مراكب لمن أثر الحج تخطف بالحجاج الى الجار وجدة ثم تمتد في مفاوز للبجة كان بها معدن الزمرد وشيء من معادن الذهب الى مدينة على شط البحر يقال لها عيذاب «15» وهي «16» محاذية للجار ثم يتصل السيف الى سواكن وهي ثلث جزائر يسكنها تجار الفرس وقوم من ربيعة ويدعى فيها لصاحب المغرب وهي محاذية لجدة وبين سواكن وعيذاب «18» سنجلة جزيرة بين رأس جبل دواى وجبل ابن جرشم وهي لطيفة وبها مغاص للؤلؤ «19» ويقصد في كل حين بالزاد والرجال وبينها وبين جدة

الجزء: 1 ¦ الصفحة: 42

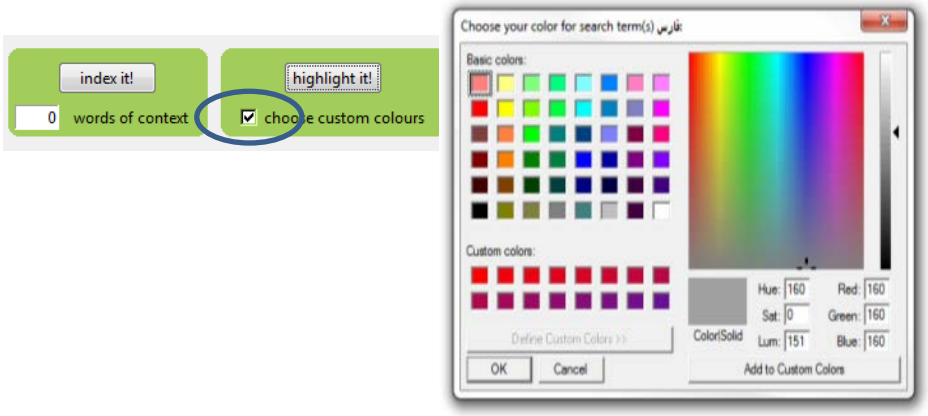
يوم واحد وليلة والمتسحل منها يصل الى جزيرة باضع وبينهما مجراوان، ثم يخطف المستحل عنها الى دهلك أربعة مجار ومن دهلك الى زيلع ستة مجار، وباضع جزيرة ذات خير ومير وماشية وهي محاذية لحلى، وجزيرة دهلك محاذية لعثر وجزيرة «4» زيلع فكأنها بين

For each "Highlighted" text, Jedli only saves the most recent search in the "search_results" output folder. If you want to run the Highlighter several times (for different search terms) using the same text (e.g., al-Ṭabarī's *Taʾrīkh*), you can save the results as an html file from the browser (right-click: "save as").

Jedli uses its default color scheme for highlighting: yellow for the first (list of) word(s), green for the next, and so on:

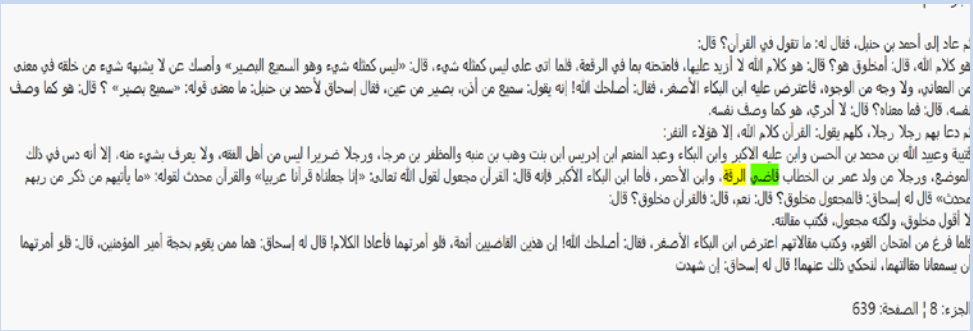


If you want to choose yourself which color is to be used for which search terms, tick the Highlighter's "choose custom colors" checkbox. This will bring up a color choice pop-up for every (list of) search term(s):



Examples for the use of the Highlighter

The Highlighter tool can be used in a variety of ways that alter and adapt to how you read your primary sources. One possible use is the search for particular terms or details you are interested in, such as certain names, dates, or places. You can look for as many search terms and/or checklists as you like; in the html file containing your search results, you can quickly scroll through the text and, for instance, watch out for highlighted combinations of the words you were looking for. An example of this can be found in the following screenshot, which captures one result from al-Ṭabarī's *Taʾrīkh* combining the search terms "al-Raqqa" and "qāḍī":



Another way in which the Highlighter can enhance our reading of a text is by looking for structural elements, such as terms introducing a *khabar* or *ḥadīth* (e.g., *akhbaranā, ḥaddathanī, samiʿtu*), or a chapter in a chronicle (e.g., *thumma dakhalat sana*, *bāb*). This is particularly useful to quickly find where a section containing a relevant search result starts or ends.

The Context Search

The Context Search tool is a more advanced version of the Indexer. It allows you to restrict your search results to specific semantic contexts, which are defined by checklists of words.

You can define one or more primary search terms ("OR" words), which are then indexed by the Indexer. In a next step, the Context Search tool checks for each of these hits if it is in a desired context, i.e., if any of the secondary search words ("AND" words) that are likely to signal a relevant context for the primary search term(s) is located within a user-defined range of words from the primary search word. If the hit fails this test, it is excluded from the search results. In order to have even more control over the final results of your search, you can also define a number of words that should not appear in the context of the primary search words ("NOT" words).

An example will elucidate the working of the Context Search tool:

Using the Context Search

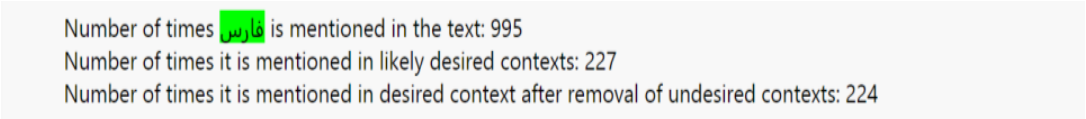
This tool was developed in the first instance to find information about provincial governors in the early Islamic empire. How we proceeded in our search for governors will illustrate how this tool can be used.

In a first step, we used the Indexer to look up all contexts in which the name of a province was mentioned in one source. We then manually selected those search results that were related to governors. In a next step, we analyzed the vocabulary composition of these search results and identified the "trigger words" in these contexts, on the basis of which we (consciously or subconsciously) had decided that the text fragment talks about a governor. The most effective trigger word was found to be *ʿalā* in combination with the name of the province (e.g., *ʿalā Ifrīqiya*, "in charge of Ifrīqiya"). Other trigger words included titles used for governors (e.g., *wālī, ʿāmil*) and terms related to appointment, dismissal, and functions of a provincial governor (e.g., *wallā, wilāya, waliya, aqarra, ʿazala*). We also analyzed how close to the name of the province these trigger words were located in order to define a word range that would limit irrelevant contexts while not excluding relevant contexts.

In the case of the province of Fārs, we had a problem that many irrelevant results turned up because of its homonym *fāris* ("horseman"). We made a list of contexts where فارس most likely refers to a horseman rather than the province (e.g., ألف فارس "1.000 horsemen"), and excluded these from our searches. Any time a new word turned up in our searches that indicated that فارس refered to a horseman, we added it to the checklist.

To use the Context Search tool, select your search words, using the buttons to the left of the input fields to indicate how you want Jedli to deal with these search words. In the example below, the Context Search will search for every mention of the word Fārs in the text. It will then select a context of 8 words (see the input field under the "context search" button at the bottom of the screen) around every occurrence of the word Fārs and check if it contains any of the words in the "key_words_governors – Fars" checklist (which we selected by clicking the "select checklist" button). If this is not the case, this hit will not be displayed in the list of result. Similarly, in case the only mention of Fārs in this context was in the word combination "*alf fāris*" or "*miʾat fāris*" (1000/100 horsemen), the result will not be considered relevant.

This is an excerpt of the html page with the results of this search in al-Ṭabarī's *Taʾrīkh* that opens in your browser once Jedli finished the search:




الجزء: 6 ¦ الصفحة: 49
وبعث عبد الله بن شريك النهدي في مائتي فارس إلى أحمر بن شميط، وثبت مكانه، فانتهوا

الجزء: 6 ¦ الصفحة: 59
واذكر فتى ماجدا حلوا شمائله ما مثله فارس في آل همدانا قال موسى بن عامر: وبعث

الجزء: 6 ¦ الصفحة: 94
فكتب المصعب إلى المهلب- وهو عامله على فارس: أن أقبل إلينا لتشهد أمرنا، فإنا نريد

الجزء: 6 ¦ الصفحة: 119
مصعبا وجه عمر بن عبيد الله بن معمر على فارس أميرا، وكانت الأزارقة لحقت بفارس

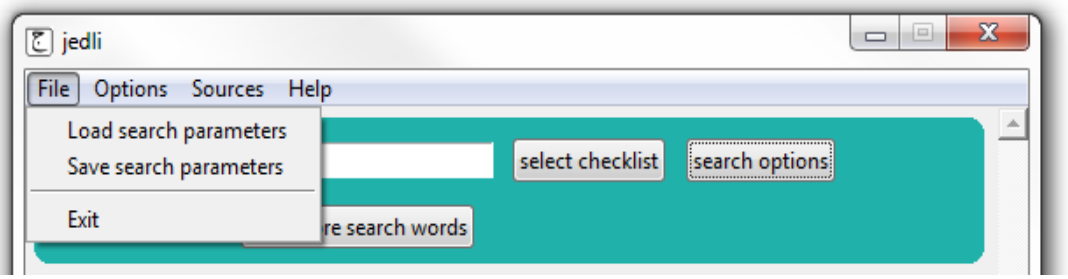
The main search word, Fārs, is highlighted in green, the search words from the governors key word list in yellow. Jedli filtered 224 hits out of the 995 mentions of Fārs in the text; more false results could be filtered out by expanding the list of NOT words (e.g., *miʾatay fārs*, "200 horsemen", on p. 49 of vol. 6).

The Menu Bar

The menu bar contains a number of additional useful functions.

Load and Save Search Parameters

If you plan to use certain combinations of search terms and/or sources more often, it might be useful to save these by clicking "save search parameters" in the "File" drop-down list in the menu bar.



This will open a dialog box that allows you to choose a file name for your search parameters. After clicking OK, the following parameters will be saved in the jedli\modules\saved_searches folder:

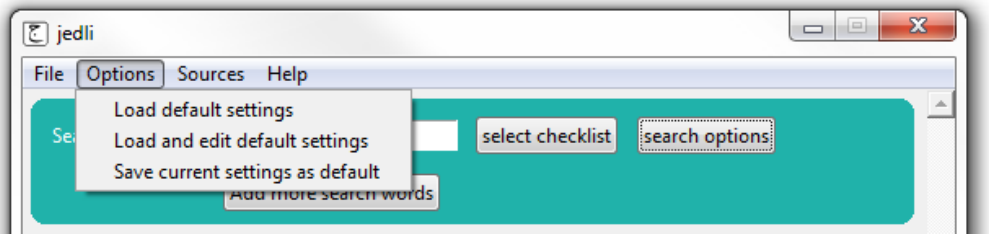
- your search terms/checklists, each with their respective search options
- your sources
- your output folder

You can later load these same parameters again by clicking “load search parameters” in the “File” drop-down list in the menu bar.

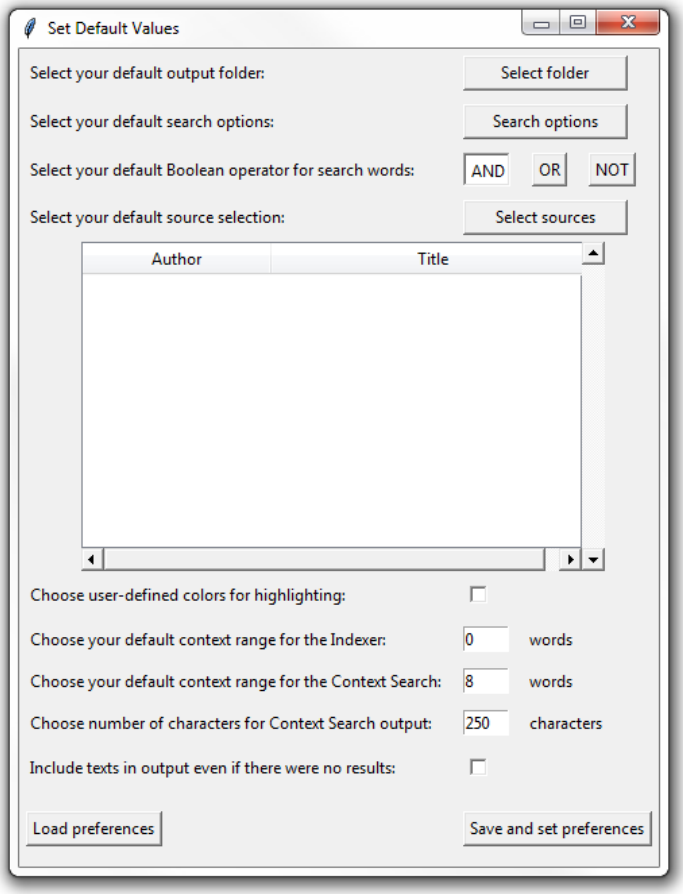
Default Settings

You can change Jedli’s default settings in order to speed up and personalize your searches.

To make new default settings, click “load and edit default settings” in the “Options” drop-down list in the menu bar.



This will open the following pop-up window:



After selecting your favorite settings, click “save and set preferences” on the bottom right of the screen, and you will be prompted to give a name to this set of preferences.

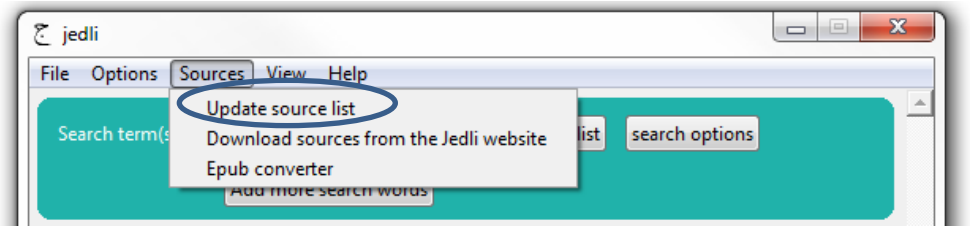
You can later load these preferences by clicking the “load preferences” button in the “Set Default Values” window. You can edit them and save them again under another name (or the same).

You can also simply load saved settings directly from the “Options” drop-down list in the menu bar by clicking “load default settings”.

Adding New Sources to the Source List

Jedli comes with a small selection of sample texts that originate from al-Maktaba al-Shamela. These texts have been converted from the epub versions that can be downloaded from their website.

If you want to add more sources to your source list, you have a couple of options. In all cases, you will have to update the sources folder before you can use these new texts in Jedli, by clicking "Update source list" in the Sources menu:



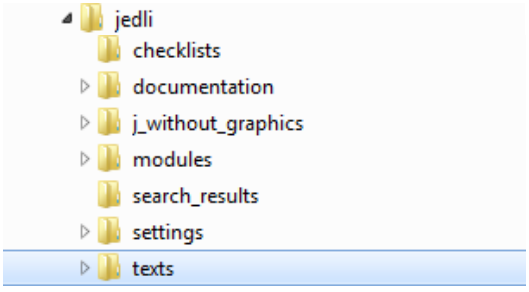
Batch Download Texts from the Jedli Website

All available texts from al-Maktaba al-Shamela have been converted by the Jedli team to .txt format and can be downloaded via the following link:

https://www.islamic-empire.uni-hamburg.de/en/publications-tools/digital-tools/downloads/jedli-toolbox.html

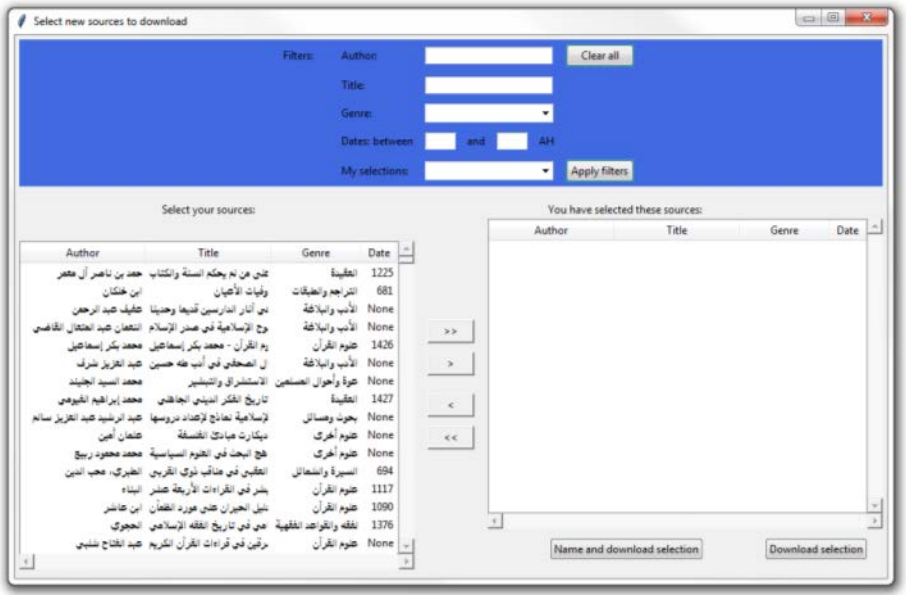
You can either download all the texts or specific batches.

The texts are zipped; unzip the zip archive and put the .txt files into Jedli's "text" folder:



After updating the source list (see above), you can now use the newly added texts in Jedli.

Select Specific Texts for Download from Jedli's "Download" Window



The "Download" window in Jedli lists all the texts that are available on the Jedli website and are not yet in your sources folder. You can filter and select the sources in the same way as you do in the "Source selection" window.

You can access the Download interface in two ways:

- From the Sources menu: choose "Download sources from the Jedli website"

- With the “download additonal sources” button in the Source selection window:

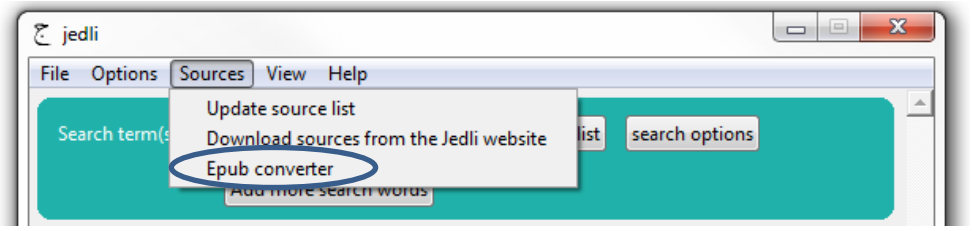
Download Texts from al-Maktaba al-Shamela and Convert Them

If you want to download a text from al-Maktaba al-Shamela and convert it yourself, this is how you can go about it:

1. Go to the website http://shamela.ws/ and search for your text in the search box (البحث) in the right-hand menu.

2. Once you have selected the book you want, you can download the epub file by clicking this icon:  (right-click; “save file as”). Save the file to the Jedli epub folder (jedli\texts\epub_files).

3. Once you have downloaded all the epubs you need, you can click “epub converter” in the “Sources” drop-down list in the menu bar.



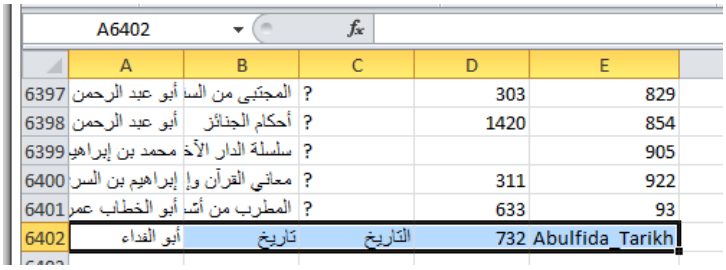
4. Select the epubs you want to convert and click “open”.
5. The epub files have now been converted to text files (.txt) and saved in Jedli’s text folder (jedli\texts).

6. If you want the new texts to show up in the list of sources in Jedli, click “update source list” in the “Sources” drop-down list in the menu bar. The information about author, title, date, and genre of the recently converted epub files are now added to the list.

Add Texts from Other Sources

You can also manually add text files in .txt format to the Jedli texts folder and update the source list in the same way by clicking “update source list” in the “Sources” menu.

If your new text files do not originate from al-Maktaba al-Shamela, or do not have the original Shamela text ID number as their file name, Jedli will not be able to retrieve the author, title, date, and genre information for these files. You can add this information manually in Jedli’s source information spreadsheet (jedli\texts\source_info\source_info.xlsx):



Column A contains the name of the author, B the title of the book, C the genre (as defined by al-Maktaba al-Shamela), D the date the source was written, and E the filename (without the .txt extension).

Do not forget to click “update source list” in the “Sources” menu after adding the information for the changes to take effect.

Volume and Page Numbers

Jedli uses a standardized volume and page numbering system; if you want to be able to use Jedli's Indexer and Context Search for texts that are not derived from al-Maktaba al-Shamela, you will have to adapt the formatting of the volume and page numbers to this format (at the bottom of the page):

الجزء: ١١ ¦ الصفحة: ٢٣٢

A future version of Jedli will have a function to automatically format the page numbers.