Document Repository



Users Guide

March 19, 2021



NOAHworks



Library	
Overview	
How to Ask Your Question	
Searching	6
Stopwords	8
Entry & Uploading	
Search and Modify	
Add New	15
Dictionaries	16
Administration	17



LIBRARY

OVERVIEW

The Library feature creates a repository for documents of all types combined with an ability to search a text description or content of those files for words or combinations of words of interest.

Find for me all documents with "San Diego"

The result of the search is a list of documents containing the search values ordered by relevance. A simple click on any item causes it to appear in another browser window where it can be viewed and/or downloaded.

A simple 1-2-3 process:

- 1. Enter what you are looking for.
- 2. Select a document from the list of hits.
- 3. Read or download the selected document.

For the administrator, the Library plugin provides a page for uploading new documents and their associated text descriptions and saving the information in the database in a way that permits full text searching of the descriptions. For Adobe .pdf documents, the entire Adobe file is copied into the text description. The system has a large list of "stopwords" that are skipped when indexing the description so that searches are not bogged down with "noise" words.

In addition to the document and its description, the Library plugin collects the title of the document, its category as contained in a user-defined dictionary, the date it was authored, and the date it was uploaded.

HOW TO ASK YOUR QUESTION

There is one page on the website where you ask your question. All that is required is a list of words (and operator(s)) that tell the computer what you would like it to find.





The first field is the search argument that you author according to the rules presented in the next section.

Documents are organized by category...in this example, the categories are Newsletters, the Ship, the Crew, Tributes and About Us. You define these categories in a dictionary available to your system administrator. The list is presented automatically, and you may use it to narrow the search by checking the boxes for the areas to be searched.

When finished, click on the SEARCH icon and a results list will appear of the documents that match your criteria and they are listed in relevance order with the documents that most satisfy the search on top.





In this sample only one document was found, but if there are more the list will expand to fill the page. To see any of the documents listed, just click on the

green arrow in the first column of the document you want to see.

In this example, a five-page document was found in the "The Ship" category with the word "Fletcher" appearing one or more times.

That document is retrieved and presented in a new window. You may read it in place, use the find command (Ctrl / f) to find specific references, and/or download the document to your computer.





SEARCHING

The Library plugin uses a simplified version of searching that uses words you provide to formulate the question to ask. Words like:

Fletcher or snipes or "San Diego"

Even though upper- and lower-case letters were used above, the search is case insensitive and will return san diego as easily as San Diego.

The following symbols are used as you author the question.

Symbol	Usage
+	A leading or trailing plus sign indicates that this word <i>must</i> be present in each document that is returned
-	A leading or trailing minus sign indicates that this word must <i>not</i> be present in any of the rows that are returned. Note the - operator acts only to exclude rows that are otherwise matched by other search terms. Thus, a search that contains only terms preceded by - returns an empty result. It does not return "all rows except those containing any of the excluded terms."
No sign	By default (when neither + nor - is specified), the word is optional, but the documents that contain it are rated higher and thus appear at the top of the result list
< or >	Documents are rated based on their relevance to the question being asked. You can increase or decrease the relevance of any word in your question using these operators. In the question One <two>Three, the One is a normal weight, the Two is less important and the Three is more important.</two>
()	Parentheses group words into subexpressions. Parenthesized groups can be nested
~	A leading tilde acts as a negation operator, causing the word's contribution to the row's relevance to be negative. This is useful for marking "noise" words. A row containing such a word is rated lower than others, but is not excluded



Symbol	Usage
	altogether, as it would be with the - operator
*	The asterisk serves as the wildcard operator. Unlike the other operators, it is <i>appended</i> to the word to be affected. Words match if they begin with the word preceding the * operator.
	If a word is specified with the wildcard operator, it is not ignored when it is too short or a stopword. The wildcarded word is considered as a prefix that must be present at beginning of one or more words. If the minimum word length is 4, a search for '+word +the*' could return fewer rows than a search for '+word +the', because the second query ignores the too-short search term the.
11 11	A phrase that is enclosed within double quote (") characters matches only rows that contain the phrase literally, as it was typed. The full-text engine splits the phrase into words and performs a search in the FULLTEXT index for the words. Nonword characters need not be matched exactly: Phrase searching requires only that matches contain exactly the same words as the phrase and in the same order. For example, "test phrase" matches "test, phrase".
	If the phrase contains no words that are in the index, the result is empty. The words might not be in the index because of a combination of factors: if they do not exist in the text, are stopwords, or are shorter than the minimum length of indexed words

The following examples demonstrate some search values:

- 'apple banana'
 Find rows that contain at least one of the two words.
- '+apple +juice'
 Find rows that contain both words.
- '+apple macintosh'
 Find rows that contain the word "apple", but rank rows higher if they also contain "macintosh".



'+apple -macintosh'

Find rows that contain the word "apple" but not "macintosh".

'+apple ~macintosh'

Find rows that contain the word "apple", but if the row also contains the word "macintosh", rate it lower than if row does not. This is "softer" than a search for '+apple -macintosh', for which the presence of "macintosh" causes the row not to be returned at all.

'+apple +(>turnover <strudel)'

Find rows that contain the words "apple" and "turnover", or "apple" and "strudel" (in any order), but rank "apple turnover" higher than "apple strudel".

'apple*'

Find rows that contain words such as "apple", "apples", "applesauce", or "applet".

"'some words"

Find rows that contain the exact phrase "some words" (for example, rows that contain "some words of wisdom" but not "some noise words"). Note that the "characters that enclose the phrase are operator characters that delimit the phrase. They are not the quotation marks that enclose the search string itself.

STOPWORDS

The words below are automatically skipped by the Library plugin when you upload and index a new document.

a's	able	about	above	according
accordingly	across	actually	after	afterwards
again	against	ain't	all	allow
allows	almost	alone	along	already
also	although	always	am	among
amongst	an	and	another	any
anybody	anyhow	anyone	anything	anyway
anyways	anywhere	apart	appear	appreciate
appropriate	are	aren't	around	as
aside	ask	asking	associated	at
available	away	awfully	be	became
because	become	becomes	becoming	been
before	beforehand	behind	being	believe
below	beside	besides	best	better
between	beyond	both	brief	but

8



by	c'mon	c's	came	can
can't	cannot	cant	cause	causes
certain	certainly	changes	clearly	СО
com	come	comes	concerning	consequently
consider	considering	contain	containing	contains
corresponding	could	couldn't	course	currently
definitely	described	despite	did	didn't
different	do	does	doesn't	doing
don't	done	down	downwards	during
each	edu	eg	eight	either
else	elsewhere	enough	entirely	especially
et	etc	even	ever	every
everybody	everyone	everything	everywhere	ex
exactly	example	except	far	few
fifth	first	five	followed	following
follows	for	former	formerly	forth
four	from	further	furthermore	get
gets	getting	given	gives	go
goes	going	gone	got	gotten
greetings	had	hadn't	happens	hardly
has	hasn't	have	haven't	having
he	he's	hello	help	hence
her	here	here's	hereafter	hereby
herein	hereupon	hers	herself	hi
him	himself	his	hither	hopefully
how	howbeit	however	i'd	i'll
i'm	i've	ie	if	ignored
immediate	in	inasmuch	inc	indeed
indicate	indicated	indicates	inner	insofar
instead	into	inward	is	isn't
it	it'd	it'll	it's	its
itself	just	keep	keeps	kept
know	known	knows	last	lately
later	latter	latterly	least	less
lest	let	let's	like	liked
likely	little	look	looking	looks
ltd	mainly	many	may	maybe
me	mean	meanwhile	merely	might
more	moreover	most	mostly	much
must	my	myself	name	namely
nd	near	nearly	necessary	need
needs	neither	never	nevertheless	new
next	nine	no	nobody	non
none	noone	nor	normally	not
nothing	novel	now	nowhere	obviously
of	off	often	oh	ok
okay	old	on	once	one
ones	only	onto	or	other
others	otherwise	ought	our	ours
ourselves	out	outside	over	overall
own	particular	particularly	per	perhaps
placed	please	plus	possible	presumably
probably	provides	que	quite	qv
rather	rd	re	really	reasonably
regarding	regardless	regards	relatively	respectively
right	said	same	saw	say
saying	says	second	secondly	see
	•		•	

9



seeing seen serious she	seem self seriously should	seemed selves seven shouldn't somebody	seeming sensible several since somehow	seems sent shall six
SO	some	sometimes	somewhat	someone
something	sometime	specified	specify	somewhere specifying
soon still	sorry sub	such		sure
t's	take	taken	sup tell	tends
th	than	thank	thanks	thanx
that	that's	thats	the	their
theirs	them	themselves	then	thence
there	there's	thereafter	thereby	therefore
therein	theres	thereupon	these	they
they'd	they'll	they're	they've	think
third	this	thorough	thoroughly	those
though	three	through	throughout	thru
thus	to	together	too	took
toward	towards	tried	tries	truly
try	trying	twice	two	un
under	unfortunately		unlikely	until
unto	up	upon	unitkely	use
used	useful	uses	using	usually
value	various	very	via	viz
VS	want	wants	was	wasn't
way	we	we'd	we'll	we're
we've	welcome	well	went	were
weren't	what	what's	whatever	when
whence	whenever	where	where's	whereafter
whereas	whereby	wherein	whereupon	wherever
whether	which	while	whither	who
who's	whoever	whole	whom	whose
why	will	willing	wish	with
within	without	won't	wonder	would
wouldn't	yes	yet	you	you'd
you'll	you're	you've	your	yours
yourself	yourselves	zero	,	,
,	,			

ENTRY & UPLOADING

You may retrieve and modify any library entry already on file or upload a new document and add it to the library.

The standard fields available are:

Field	Usage
Title	A short title for the document, used in displays



Field	Usage
Source	A category code defined by the using organization that separates the documents into areas of interest
Date Published	The date the document was first published
Date Last Changed	The date the library entry was last modified and saved
Link to the Document	The file name of the document. The path is assigned by the Library plugin and is not needed by the user
Document Text	The clear text version of the document. For Adobe .pdf documents this is the entire document stripped of internal Adobe commands. For images and other documents, it is text you enter to be used by the Library search engine.
User Defined	There are additional user-defined fields that may be activated by your system administrator.

SEARCH AND MODIFY

In this example we have searched for the "The Ship" document with the word "Fletcher" in its text. It appears on the results list and we click on the green arrow to select it for modification.

At this point you may modify any of the fields except the PDF Link to the Document.

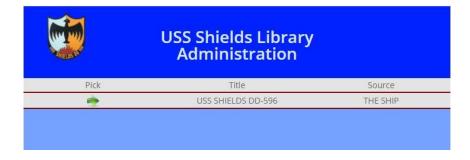
If you wish to replace the document, you may upload a new version and in doing so will replace the current document and, it the document is an Adobe .pdf, the Library plugin will update the Document Text.

11





The Search results in one document as before.



We click on the green arrow to select the document and it opens for our review and modification. The PDF Link field is locked but all other fields are open to your modification.





Below the form are five buttons. Click the VIEW button if you wish to view the source document. It will appear in a new window as you saw for Library users.

If you wish to refresh the Search Text, you can position your cursor in the pdf document, press the key combination of Ctrl / a to highlight all of the text, then press Ctrl / c to copy the text, move to the Search Text field as shown above and press Ctrl / v to paste the copied text into the Search Text field.



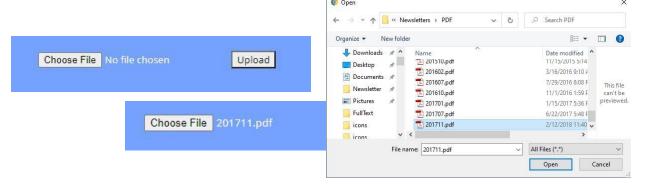
When you have made all the changes you want, click on the SAVE button and your changes will be written to the database and a message "SAVED" will appear to the left of the buttons.



If you make some changes and decide you made a mistake (and have not yet saved the record), you can click on the CANCEL button to restore the record without your changes.

If you decide you no longer want this entry in the Library, click on the DELETE button. The system will change the Source code to "PURGE", save the record, and a message "PURGED" will appear to the left of the buttons. Only the system administrator can completely delete entries from the Library, but they are gone from your sight and searches.

The last button is used to upload a new version of the document while retaining the initial Library record. When clicked two new buttons will appear in the open area. The first opens your computer and allows you to select the new version of the document.



The selected file name then appears to the right of the button. Click on the Upload button and the file will be transferred from your computer to the Library and you will see a message confirming the upload's success (or failure). The Library record will appear as before, but the record will now be using the new version. You can click on the VIEW button to confirm this.

You will also need to manually update the Search Text as before by using the Ctrl / a >> c >> v sequence to copy and paste the text into the Search Text field. When finished, click on SAVE.

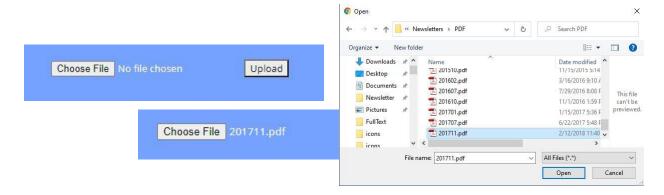
At the bottom of the page are two green buttons. The one on the left returns you to the starting search page and the one on the right returns you to the hit list to select another document in your search.





ADD NEW

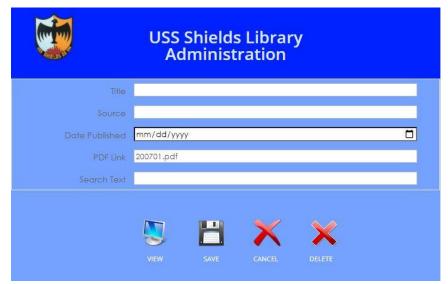
Adding a new entry is like modifying an existing record. You start by clicking on the ADD NEW button on the starting page. When you click on the green Plus sign button the file choose and upload buttons appear.



Click the Choose File button to open a window into your computer where you select the file to upload. That file then appears to the right of the button. Click on the Upload button to copy the file from your computer to the system. The system will confirm the success (or failure) of the upload.

If successful, you will then see the Library entry form with only the name of the document in the PDF Link field. You should fill in all the other fields and use the Ctrl / a >> c >> v sequence to copy and paste the text into the Search Text field. When finished, click on SAVE. Your record will be written to the database and the Search Text will build the full text index used for searching.







DICTIONARIES

NOAHworks



ADMINISTRATION