dtSearch Frequently Asked Questions

Version 3

Table of Contents

dtSearch Indexing

Q. What is dtSearch?

A. dtSearch is a product that performs index-based searches. dtSearch is the backbone of Index Searching in FTK.

Q. How can I view and change my indexing options?

A. Indexing options can only be viewed and changed when a case is first created. From the “New Case Options” window (the first window that opens when you select to create a new case), click the “Detailed Options” button, then click the “Indexing Options” button.

Q. What characters are indexed?

A. By default, the indexable alphabet includes the underscore (‘_’), numbers 0-9, and letters A-Z (upper and lower cases). Other foreign characters including the following accented characters are also indexed by default (both upper and lower cases):

Accent	**Character**					
Grave	à	è	ì	ò	ù	
Acute	á	é	í	ó	ú	ý
Circumflex	â	ê	î	ô	û	
Tilde	ã	ñ	õ			
Umlaut	ä	ë	ï	ö	ü	ÿ

Q. What characters are treated as spaces?

A. The following characters are treated as spaces by default:

Single Quote	‘
Exclamation Mark	!
Double Quotes	“
Pound Sign (Hash)	#
Dollar Sign	$
Percent	%
Ampersand	&
Right Parenthesis	(
Lift Parenthesis	)
Asterisk	*
Comma	,
Forward Slash	/

Back Slash	\
Colon	:
Semicolon	;
Question Mark	?
At Symbol	@
Left Bracket	[
Right Bracket	]
Carrot	^
Tick Mark	`
Left Brace	{
Right Brace	}
Bar	|

Tilde	~
Plus	+
Equals	=
Less Than	<
Greater Than	>
Carriage Return	
Form Feed	
Line Feed	
Space	
Tab	

Q. How are hyphens treated?

A. By default, the hyphen (dash) is treated as a space, but can be set to be ignored or treated as a hyphen. If you choose to treat hyphens as hyphens rather than as spaces, the hyphen becomes part of the indexable alphabet.

Q. What terms/words are indexed?

A. By default, any group of 2-32 characters from the indexable alphabet, surrounded by space characters, will be indexed as long as that term is not considered a Noise Word.

Q. What are Noise Words?

A. Noise words are extremely common words such as “it” and “the”, and are not added to the index. A list of all noise words can be found in the “Indexing Options” window. If you know when starting a case that you will need to search for terms that may contain noise words, you should remove those terms from the list of Noise Words in the Indexing Options.

Q. How can I index individual letters?

A. Indexing single letters can be useful if you want to search for acronyms with periods or spaces (eg. “A.C.E.” or “A C E”). Some FTK installations will show an entry that looks like “a b c d e f g h” in the Noise Words list. If you see this entry, you can simply remove it from the Noise Words list to allow FTK to index individual letters. If you do not see this entry, you can do the following:

1. Open the “Indexing Options” while creating a new case
2. Click the “Add” button next to the Noise Words list, and add the string “a b c d e f g h I j k l m n o p q r s t u v w x y z” (unquoted)
3. Click “OK”
4. Click “Save as My Defaults”
5. Go into Indexing Options again
6. Highlight the string “a b c d e f g h…” that you just entered and click “Remove”
7. Click “OK”

Note: We do not recommend making this change to the Indexing Options if eDiscovery is installed on your FTK box.

Q. What does “Merge Index” do?

A. Indexing is multi-threaded and each thread creates its own index. Because multiple files will usually contain the same words, the words in each index will overlap each other (eg. one index tracked the word “berries” in 55 different files, and another index tracked the word “berries” in another 100 files). Merging indexes will merge all the indexes into one, eliminating this overlap and resulting in faster index searches. However, the process of merging indexes can take as long as the original indexing process, so you will need to determine based on the size of your case if the trade off is worth it.

Using Index Search

Q. What search operators can be used in index searches?

A. Search operators along with examples are listed below:

Name	Operator	Example	Result
And	AND	apple AND pear	Matches files containing both “apple” and “pear”
Or	OR	apple OR pear	Matches files containing either “apple” or “pear”
Within	W/	apple w/4 pear	Matches “apple” within 4 words of “pear”
Not	NOT	apple AND NOT pear	Matches files containing “apple” but not “pear”
		apple NOT w/4 pear	Matches files containing “apple” but not within 4 words of “pear”
Contains	CONTAINS	name CONTAINS adam	Matches files whose metadata “name” field contains “adam”
Wildcard	*	apple *	Matches “apple” followed by any word
		appl*	Matches “appl” followed by any number or characters (“apples”, “applied”, etc.)
Single Character Wildcard	?	appl?	Matches “appl” followed by any character (“apple”, “apply”, etc.)
Stemming	~	apply~	Matches different stems/conjugations of “apply” (“applies”, “applied”, etc.)
Fuzzy	%	ba%nana	Matches different spellings of “banana” (“banana”, “bannana", etc.)
Phonic	#	#pear	Matches words that sound like “pear” (“pare”, “pair”, etc.)
Synonym	&	fast&	Matches synonyms for “fast” (“quick”, “speedy”, etc.)
Numeric Range	~~	12~~20	Matches any number between 12 and 20

Q. How can I use Regular Expressions in index searches?

A. As of FTK 4, we have added the ability to use Regular Expressions in index searches. We use the TR1 implementation of Regular Expressions. The full documentation of TR1 with its capabilities can be accessed at http://msdn.microsoft.com/en-us/library/bb982727.aspx.

Rules:

- A Regular Expression in a search request must be quoted and must begin with ##.
  Example: "##appl." will match "apple
- A Regular Expression search term will only match a single word.
  Example: "##app.*ie" will not match "apple pie"
- Multiple Regular Expression search terms can be used together by using other search operators.
  Example: "##appl." AND "##[a-z]ie" will match "apple pie"

Some special characters in Regular Expressions:

Regular Expression	Effect	Example	Result
.	A period matches any single character	"##appl."	Matches "apple" or "apply"
[abc]	Brackets indicate a set of characters, one of with must be present	"##car[se]"	Matches "cars" or "care", but not "carb"
[a-z]	A dash inside brackets indicates a range of characters, one of with must be present	"##car[a-f]"	Matches "carb" or "care", but not "cars"
[^a-z]	A carrot indicates characters that must *not* be present	"##appl[y]"	Matches "apple" but not "apply"
*	An asterisk must follow another regular expression and means "0 or more" of something.	"##app.*"	Matches "app", "apps", "apple", or "applied" ("app" followed by any string or characters, *or by nothing*)
+	A plus must follow another regular expression and means "1 or more" of something	"##app.+"	Matches "apps", "apple", or "applied", but not "app" ("app" followed by any string or characters)

Q. How do I search for a phrase?

A. The most reliable way to search for a phrase is to enter it without quotes. For example, if you want to search for the phrase “angry brown fox” you would enter “angry brown fox” (unquoted) into the Terms box.

Q. How are search hits calculated?

A. Search hits are calculated over indexed terms, not search terms. Search terms consist of one or more indexed terms, so any match to a search term can yield 1 or more hits depending on how many indexed terms are in the search term. For example, every match to the search term “fox” will result in 1 hit, while every match for “brown fox” will result in two hits (because “brown” and “fox” are two different indexed terms).

Q. What does “Accumulate Results” do?

A. “Accumulate Results” shows you how many hits your search will yield. This can give you a rough idea of how long it will take to perform your search (the more hits a search finds, the longer it may take to perform the search and populate the results).

Q. Can I import a list of search terms?

A. Yes. By clicking the “Import” button to the right of the Search Criteria field you can import a text file with a list of search terms. Make sure that each search term is on a new line in the text file.

Q. How do I search for accented and Unicode characters?

A. Currently, you can't type accented and Unicode characters directly into the Search Terms field. However, you can use these characters by importing a list of search terms or copy-and-pasting search terms into the Search Terms field.

Q. Can I have a noise word in my search term?

A. No. Any noise words in a search term will be treated as a wildcards when searching. For example, searching for "statue of liberty" will match any three-word phrase beginning with "statue" and ending with "liberty" (because "of" is a noise word by default). If you know when starting a case that you will need to search for terms that may contain noise words, you should remove those terms from the list of Noise Words in the Indexing Options. In addition, the terms "and", "not", "or," and "contains" will always be treated as operators, so you cannot search for these terms even if you remove them from the list of Noise Words.

Q. Can I use punctuation and symbols in search terms?

A. No. Any punctuation or symbols will be treated as a string of wildcards when searching. For example, searching for "microsoft.com" will match any length phrase beginning with the word "microsoft" and ending with "com". Likewise, searching for "user@microsoft.com" will match any length phrase starting with the word "user" and ending in "com" that also contains the word "microsoft". If you would like to search for a phrase that contains punctuation or symbols, use spaces instead of symbols in the search term (eg. searching for "user microsoft com" will match "user@microsoft.com").