

Lightgrep™ - Fast Search for Forensics

Lightgrep for EnCase[[1]](#footnote-2) provides you with new options for search in a familiar environment: Keywords that are compatible with PCRE; filters for search terms to help you find the real information you're looking for faster; text extraction for Office 2007 and similar document formats.

Table of Contents

[Minimum Requirements: 4](#_Toc315731878)

[Known Issues: 4](#_Toc315731879)

[Limitations and General Information 4](#_Toc315731880)

[Search Tab 5](#_Toc315731881)

[Options 5](#_Toc315731882)

[Search for Keywords 5](#_Toc315731883)

[Search for Entities 5](#_Toc315731884)

[Search only selected items 6](#_Toc315731885)

[Ignore all items in loaded hash library (search slack) 6](#_Toc315731886)

[Ignore items with hash category ‘KNOWN’ (search slack) 6](#_Toc315731887)

[Create Bookmarks 6](#_Toc315731888)

[Create Excel Report 6](#_Toc315731889)

[Lightgrep Output Folder 6](#_Toc315731890)

[Search Status 6](#_Toc315731891)

[Keywords 7](#_Toc315731892)

[Options 7](#_Toc315731893)

[Import From EnCase 7](#_Toc315731894)

[Import From File 8](#_Toc315731895)

[Keyword Tree 9](#_Toc315731896)

[Supported Pattern Syntax 9](#_Toc315731897)

[Individual Bytes 9](#_Toc315731898)

[Metacharacters 10](#_Toc315731899)

[Character Classes 10](#_Toc315731900)

[Character Class Metacharacters 10](#_Toc315731901)

[Named Character Classes 10](#_Toc315731902)

[Repetition 10](#_Toc315731903)

[Conjunction, Alternation, Grouping 11](#_Toc315731904)

[Entity Search 12](#_Toc315731905)

[Options 12](#_Toc315731906)

[Entities 12](#_Toc315731907)

[Entity Filter 12](#_Toc315731908)

[Short Codes 13](#_Toc315731909)

[Import Entities 13](#_Toc315731910)

[Export Entities 13](#_Toc315731911)

[Reset Defaults 13](#_Toc315731912)

[Compile Filters 13](#_Toc315731913)

[Advanced Tab 14](#_Toc315731914)

[Options 14](#_Toc315731915)

[File Signature List 15](#_Toc315731916)

[Import From EnCase 15](#_Toc315731917)

[Reset Defaults 15](#_Toc315731918)

[Lightgrep executable path 15](#_Toc315731919)

[Number of Threads 16](#_Toc315731920)

# Minimum Requirements

* EnCase Forensic or Enterprise version 6.16.2 (32-bit or 64-bit)
* Windows XP SP3, Windows 7 (32-bit or 64-bit) – the application has not been tested with Windows Vista.
* 4 GB of RAM preferred minimum
* Microsoft .Net Framework 3.5 SP1 is required for Excel output

# Known Issues

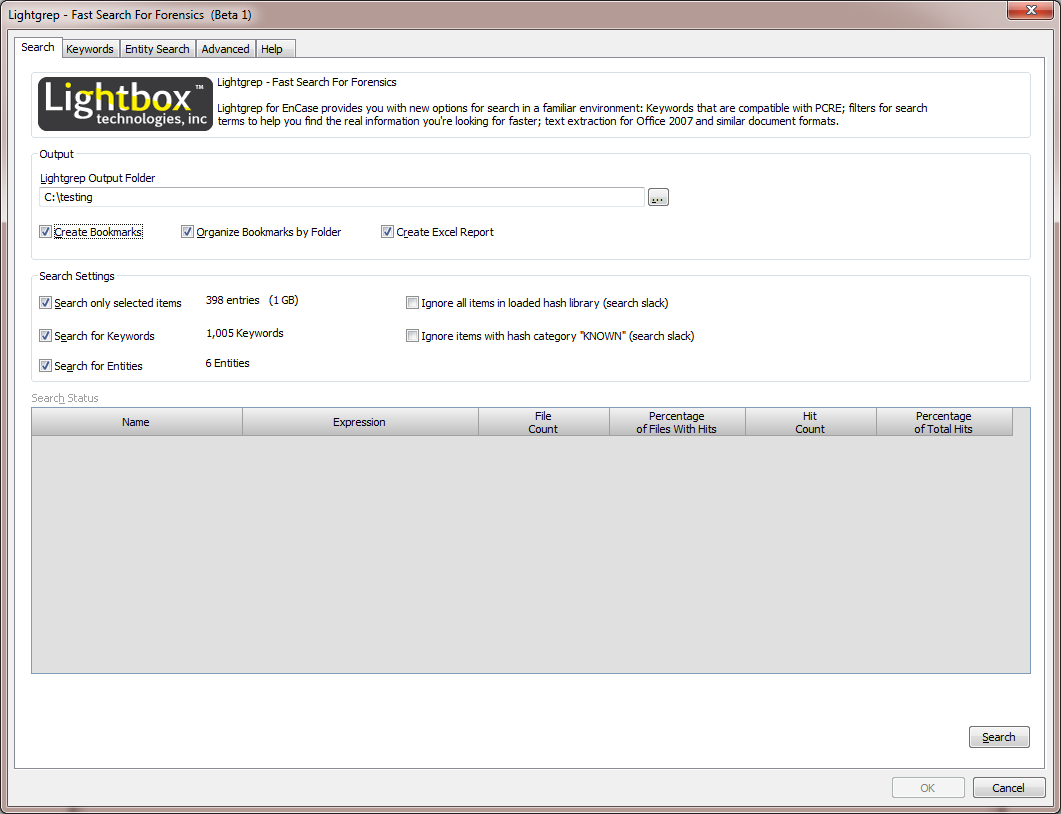
* Not compatible with EnCase version 7

# Limitations and General Information

* Sweeping Bookmarks in EnCase version 6 have a limit of 16,384 bytes. Although Lightgrep does not have a technical limit on the length of search hits, the limit within EnCase presents a practical limitation when the desired output format is an EnCase Bookmark.
* Entries without a starting extent cannot be bookmarked. These usually are files containing Invalid Cluster in the Description field, although some files marked as Internal, such as the streams $UsnJrnl·$J and $BadClus·$Bad, also fall into this category.
* The “Lightgrep Output Folder” location cannot be set to a protected location in Windows Vista or Windows 7. Protected locations include, but may not be limited to, those locations represented by the following Windows environment variables: SystemRoot, ProgramFiles(x86), ProgramW6432 (e.g. C:\Windows , C:\Program Files, C:\Program Files (x86)). Lightgrep for EnCase will attempt to automatically identify if you are trying to save output to one of these locations and ask you to correct it.
* Lightgrep stores settings in two files: *Lightgrep Search.ini* and *Lightgrep Search.xml*. The .ini file stores simple options that have been selected in the GUI; the .xml file stores Entities and Keywords. The .xml file uses approximately 70 MB of space per 100,000 keywords.
* Lightgrep uses a constant and relatively small amount of memory while searching, but can require more memory up front while parsing the keyword list depending on the number of keywords. 4 GB of RAM with no other applications running should allow you to easily parse and search several hundred thousand fixed string keywords. Approximately 8 GB of RAM is required to parse and search 1 million fixed string keywords. When searching 1 million plus keywords, it can take several minutes before the search begins while the search engine parses the list of keywords.

# Search Tab

The Lightgrep Search Tab as shown in presents the main program options that should be reviewed before performing a search, and also provides an area to display ongoing output about search statistics.



Figure

## 

## Options

### Lightgrep Output Folder

The folder specified here will be used both as a temporary folder location and as the output location for Lightgrep search results. Generally speaking, less than 10 megabytes of space is used by temporary files from the application, depending on the number and size of keywords. The size of results files can vary greatly depending on the number of hits found.

### Create Bookmarks

This option will create sweeping bookmarks of all search hits in the current EnCase Case. The output is very similar to EnCase search hits.

### Organize Bookmarks by Folder

Checking this option will organize search hits into the same folder structure as present in the Keywords tree (see Figure 2). If you have a folder named “Financial” containing 5 keywords, you will also have a bookmarks folder named “Financial” which will contain one folder for each of the 5 keywords that have hits. If you do not select this option, the results will be generated in a flat structure, with all keywords having top level folders to contain their hits.

### Create Excel Report

This option will create an Excel 2007 format overview report, containing statistics about the results of the search. The Microsoft .NET Framework 3.5 SP1 is required for output. It is not required to have Microsoft Excel installed in order to produce the report.

### Search only selected items

This option determines whether all Entries or only selected (blue checked) Entries will be searched. Note that Records are not searched directly, but their content can be searched by selecting the appropriate Entries. Information about the number and size of selected Entries is displayed for reference.

### Search for Keywords

When this is selected, Lightgrep will search for the keywords selected in the Keywords tab. See Figure 2.

### Search for Entities

When this option is selected, Lightgrep will search for entities using the options on the Entity Search tab. Keyword hit results will be restricted based on each entity’s filter code. See the Entity Search tab description for further information.

### Ignore all items in loaded hash library (search slack)

This option will prevent searching of any item that is present in the currently loaded hash library. To learn more about loading hash sets into the active hash library, please consult the EnCase Help file sections named “Hash Sets” and “Rebuilding a Hash Library.” Slack space will always be searched regardless of this setting. Selecting this item disables the following option.

### Ignore items with hash category ‘KNOWN’ (search slack)

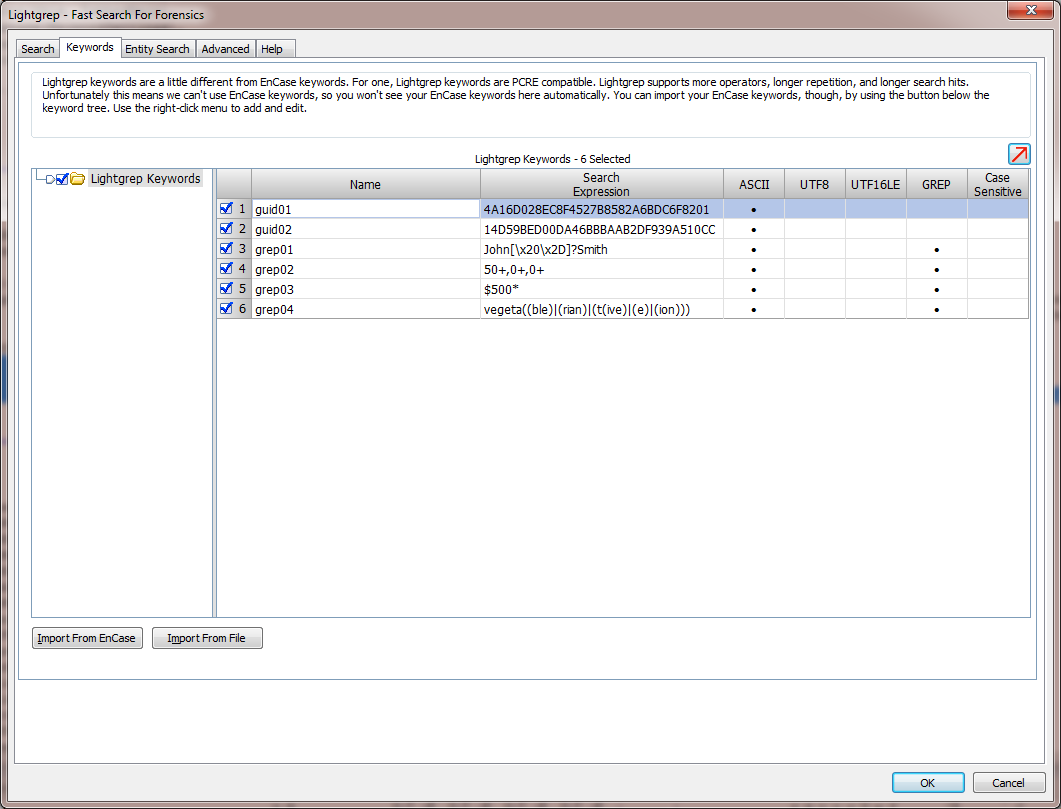
This option will prevent searching of any item with a hash category of ‘KNOWN’ (case insensitive) that is present in the currently loaded hash library. To learn more about loading hash sets into the active hash library, please consult the EnCase Help file sections named “Hash Sets” and “Rebuilding a Hash Library.” Slack space will always be searched regardless of this setting. This item will be disabled if the preceding option is selected.

## Search Status

The Search Status window displays statistics about the currently running search. You will see the keyword name, search expression, a count of how many files have hits for each keyword, the percentage of files with hits, the total count of hits produced by that keyword, and the percentage of total hits related to that keyword.

# Keywords

Lightgrep keywords are a little different from EnCase keywords. For one, Lightgrep keywords are PCRE compatible. Lightgrep supports more operators, longer repetition, and longer search hits. Unfortunately this means we can't use EnCase keywords, so you won't see your EnCase keywords here automatically. Supported keyword syntax is discussed in the section Supported Pattern Syntax.



Figure

## 

## Options

### Import From EnCase

Clicking this button will open the *Import Keywords from EnCase* dialog as shown in Figure 3. The leftmost tree pane shows folders in both Global Keywords and Case Keywords from EnCase. Use this pane and the pane on the right to select (blue-check) the keywords you would like to import.

The right tree pane shows Lightgrep Keywords folders. The folder selected (highlighted) here will receive the imported keywords.



Figure

At the bottom of the dialog, a checkbox labeled “Retain folder structure when importing” allows you to keep the imported keywords organized as they were in EnCase. The folders named “\*\*GLOBAL KEYWORDS\*\*” and “\*\*CASE KEYWORDS\*\*” are not retained, and are displayed only for informational purposes. In the example from Figure 3, the root of Lightgrep Keywords is selected on the right; this will produce folders as shown below in Figure 4.



Figure

When done, click the OK button and the EnCase keywords will be imported and converted into Lightgrep Keywords. A summary of the import will be presented as in Figure 5.

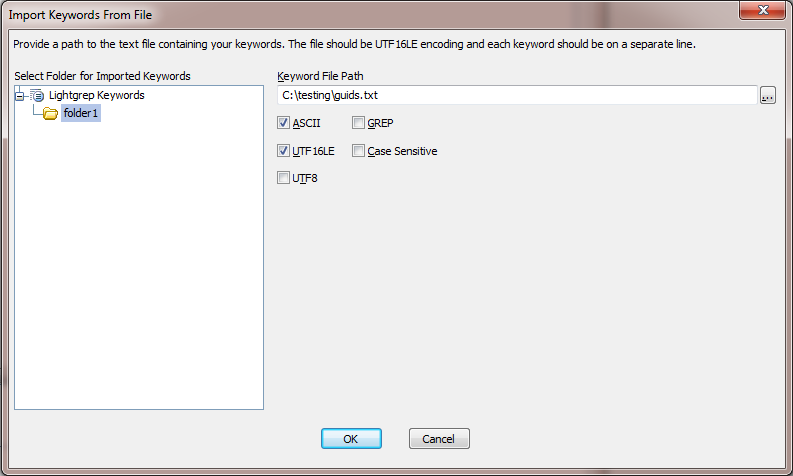


Figure

### 

### Import From File

Clicking this button will open the *Import Keywords from File* dialog as shown in Figure 5. The leftmost tree pane shows Lightgrep Keywords folders. The folder selected (highlighted) here will receive the imported keywords. Select a file in the Keyword File Path box, select the appropriate options for encodings and whether the keywords should be treated as GREP (vs fixed strings) or Case Sensitive.



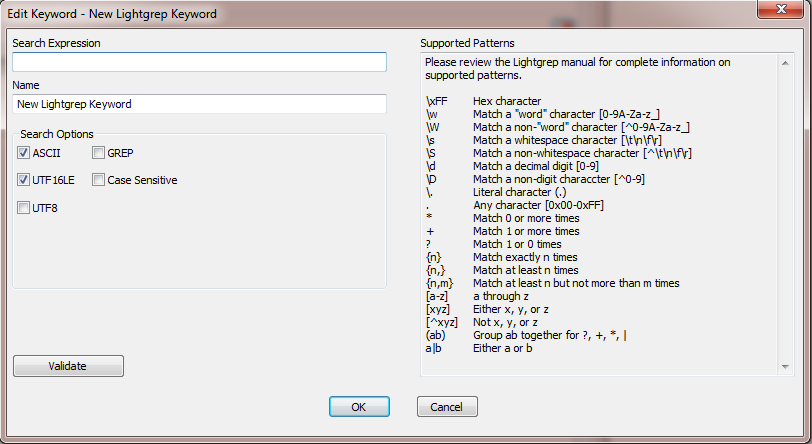
Figure

### 

### Keyword Tree

The Keyword Tree as shown in gives a familiar interface to view, organize, edit, and delete keywords. Keywords can be organized into folders on the left (tree) side of the interface just as any other dialog within EnCase. The heading / title directly above the keyword tree shows how many keywords are currently selected. **Only keywords which are blue checked will be searched by Lightgrep.**

You can create a new keyword by right-clicking anywhere in the tree and selecting “New” from the context menu or edit an existing keyword either by right-clicking on the keyword and selecting “Edit” or by double clicking on the keyword. Upon creating a new keyword or opening one to edit, you will be presented with the *Edit Keyword* dialog as shown below in .



Figure

## Supported Pattern Syntax

### Individual Bytes

***b*** where ***b*** any literal byte except **0x00** and metacharacters

**\a** **0x07**, bell

**\e** **0x1B**, escape

**\f** **0x0C**, form feed

**\n** **0x0A**, new line

**\r** **0x0D**, carriage return

**\t** **0x09**, horizontal tab

**\*o*** **0*o***, where *o* is an octal digit

**\*oo*** **0*oo***, where each *o* is an octal digit

**\*ooo*** **0*ooo***, where each *o* is an octal digit, less than or equal to 0377

**\x*hh*** **0x*hh***, where each h is a hexadecial digit

**\c** ***c***, where c is any character not listed above

### Metacharacters

Outside of character class specifications, the following characters have special meanings: \, |, (, ), ?, +, \*, ., [. To specify a literal backslash, pipe, left parenthesis, right parenthesis, question mark, plus, asterisk, period, or left square bracket, escape it with a backslash. Additionally, literal left curly braces ({) must be escaped outside of character class specifications if they could otherwise be interpreted as the start of ranged quantifiers, but need not be escaped otherwise.

### Character Classes

[*ccspec*] matches any byte in *ccspec*  
[^*ccspec*] matches any byte not in *ccspec*

Where *ccspec* is one or more of the following:

* a byte,
* a byte escape,
* a range formed by bytes or byte escapes, or
* a named character class (excerpt .).

### Character Class Metacharacters

* The right square bracket ] must be escaped inside character class specifications unless it is the first member.
* The hyphen - forms ranges inside character class specifications using the characters immediately preceding and following it, except when - is the first or last character in the class specification or if the previous character is already part of a range, in which case - stands for itself.
* A literal \ must always be escaped.
* All other characters stand for themselves.

### Named Character Classes

**.** equivalent to **[0x00-0xFF]**  
**\d** equivalent to **[0-9]**  
**\D** equivalent to **[^0-9]**  
**\s** equivalent to **[\t\n\f\r]**  
**\S** equivalent to **[^\t\n\f\r]**  
**\w** equivalent to **[0-9A-Za-z\_]**  
**\W** equivalent to **[^0-9A-Za-z\_]**

### Repetition

***S*\*** equivalent to***S*{0,}*****S*+** equivalent to***S*{1,}*****S*?** equivalent to ***S*{0,1}*****S*{*n*,}** equivalent to***S*{*n*,*M*} *S*{*n*,*m*}** matches at least ***n*** and at most ***m*** repetitions of ***S***, greedily***S*\*?** equivalent to***S*{0,}?*****S*+?** equivalent to***S*{1,}?*****S*??** equivalent to***S*{0,1}? *S*{*n*,}?** equilvalent to***S*{*n*,*M*}?*****S*{*n*,*m*}?** matches at least *n* and at most *m* repetitions of *S*, reluctantly

Where ***S*** is an atomic pattern, 0 ≤ ***n*** ≤ ***m***, and ***M*** is a largish integer which presently depends on the size of the function stack on the machine where Lightgrep is run.

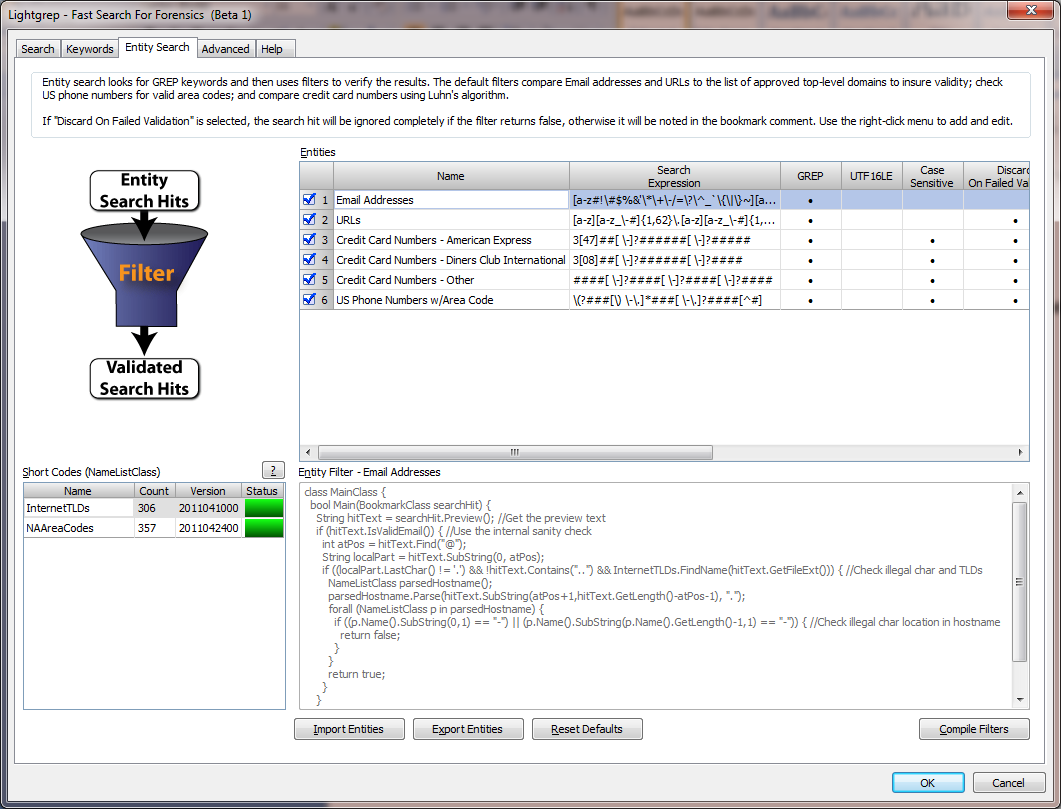
### Conjunction, Alternation, Grouping

***ST*** matches something matching ***S*** followed by something matching ***T******S*|*T*** matches ***S*** or ***T***, preferring matches for ***S***  
**(*S*)** equivalent to ***S***, but is atomic

Individual bytes, character classes, and groups are atomic.

# Entity Search

Entity search looks for GREP keywords and then uses filters to verify the results as shown in Figure 6. The default filters compare Email addresses and URLs to the list of approved top-level domains to insure validity; check US phone numbers for valid area codes; and compare credit card numbers using Luhn's algorithm. If "Discard On Failed Validation" is selected, the search hit will be ignored completely if the filter returns false, otherwise it will be noted in the bookmark comment.



Figure

## Options

### Entities

Table

### Entity Filter

Read only

### Short Codes

Short Codes are NameListClass objects that are pre-populated with useful lists of information. For example, the 'InternetTLDs' list is a NameListClass containing every IANA approved top level domain, such as com, edu, uk, etc. The list count and version (a string typically representing the date of the most recent update) are displayed, along with a visual indicator of the code's status: green if it's safe to use, red if it's not active or has an issue.

Short Codes can be used as NameListClass objects directly in the bool Main() function of entity filters without initialization:

NameListClass comDomain = InternetTLDs.FindName("com");

Current Sources:

InternetTLDs  
List of valid Internet top level domains, maintained by IANA (com, edu, uk, etc)  
http://data.iana.org/TLD/tlds-alpha-by-domain.txt

NAAreaCodes  
List of active North American area codes, maintained by the North American Numbering Plan Administration  
http://www.nationalnanpa.com/nas/public/npasInServiceByNumberReport.do?method=displayNpasInServiceByNumberReport"

### Import Entities

Stuff

### Export Entities

export

### Reset Defaults

Clicking on the Reset Defaults button will remove any custom entities that have been created and replace them with the default values as shown in Figure 6.

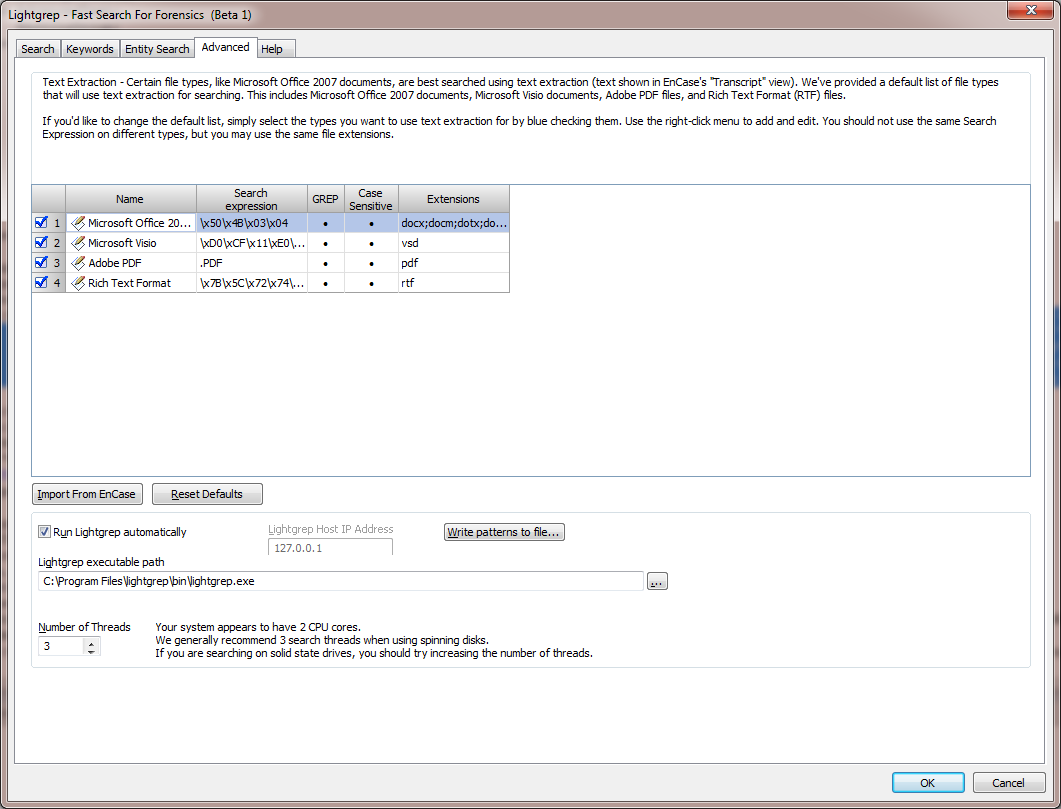
### Compile Filters

Compile stuff

# Advanced Tab

Text Extraction - Certain file types, like Microsoft Office 2007 documents, are best searched using text extraction (text shown in EnCase's "Transcript" view). We've provided a default list of file types that will use text extraction for searching. This includes Microsoft Office 2007 documents, Microsoft Visio documents, Adobe PDF files, and Rich Text Format (RTF) files.

If you'd like to change the default list, simply select the types you want to use text extraction for by blue checking them. You should not use the same Search Expression on different types, but you may use the same file extensions.



Figure

## Options

### File Signature List

The File Signature List, as shown in the top portion of Figure 7, contains a list of file types that will be used to identify files for text extraction. This list contains standard EnCase file signatures. You can add or edit file signatures just as you would in EnCase.

During a search, any file types selected here will have text extracted using EnCase’s Oracle Outside-In technology. Lightgrep will search the extracted text for keywords instead of the binary content of the file. It will also search binary file slack of any files identified for text extraction, so you don’t have to worry about missing anything.

If you deselect any types, you will receive a warning: “You have deselected one or more file types. Any file type that is not selected will be searched using standard native content instead of extracted text.”

### Import From EnCase

Clicking this button will open the *Import File Signatures from EnCase* dialog as shown in Figure 8. The dialog shows the standard EnCase File Signatures list. Use this window to select (blue-check) the file signatures you would like to import. When done, click the OK button and they will be imported into Lightgrep.



Figure

### Reset Defaults

Clicking on the Reset Defaults button will remove any custom file signatures that have been created or imported from EnCase and replace them with the default values as shown in Figure 7.

### Lightgrep executable path

The path to lightgrep.exe on the examiner system. The Lightgrep search server is run using this path.

### Number of Threads

The script will attempt to auto-detect how many cores you have on your box. Regardless of the number of cores, we generally recommend 3 search threads when using spinning / magnetic hard disks. If you are searching on solid state drives, you should try increasing the number of threads.

1. EnCase® is a registered trademark of Guidance Software, Inc. [↑](#footnote-ref-2)