Contents

About this Guide ..................................................................................................... 5
Keyword Search Quick Reference .............................................................................. 5
Clearwell Detailed Search Reference .......................................................................... 6
  Clearwell User Interface ....................................................................................... 6
  General Notes ................................................................................................ 6
Understanding Search Result Statistics .................................................................... 8
Stemmed Searches .................................................................................................. 9
Boolean Searches ................................................................................................. 10
  Grouping ......................................................................................................... 11
Wildcard Searches ............................................................................................... 12
  Single Character Wildcard ............................................................................. 12
  Multiple Character Wildcard ........................................................................... 12
Phrase Searches .................................................................................................... 13
  Additional Notes ........................................................................................... 13
Proximity Searches ................................................................................................ 14
  Nested Proximity Searches ............................................................................. 15
Transparent Searches ............................................................................................. 16
  Using the Search Preview Feature .................................................................. 17
  Using Multiple Query Analytics ..................................................................... 18
Running Transparent Searches ............................................................................... 19
  Running Transparent Searches – Search Jobs ............................................... 20
  Using Keyword Query Filters ......................................................................... 22
  Using the Search Report ............................................................................... 25
Participant Searches .............................................................................................. 28
Concept Search .................................................................................................... 31
  Concept Search Workflow ............................................................................... 32
Freeform Searches ................................................................................................. 35
  About the Freeform Search Page ..................................................................... 35
  Basic Freeform Queries ............................................................................... 35
Terms ........................................................................................................... 35

Proprietary & Confidential
Rev. 050611
# About this Guide

This guide provides an overview of the keyword search capabilities of the Clearwell E-Discovery Platform. The guide is intended for end users who want to run advanced keyword searches using Clearwell’s search query syntax.

For more information on other Clearwell capabilities including searching for date ranges, file types, and other document metadata, refer to the User Guide.

## Keyword Search Quick Reference

<table>
<thead>
<tr>
<th>Query Type</th>
<th>Syntax</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stemmed vs. Literal</td>
<td>Basic Search field: Searches are always stemmed</td>
<td>Enclosing text in quotes does not affect stemming behavior. Words in exact phrase and proximity searches will be stemmed when run as a Basic Search or an Advanced search with the stemming on.</td>
</tr>
<tr>
<td></td>
<td>Advanced Search screen: Select stemmed or literal search</td>
<td></td>
</tr>
<tr>
<td></td>
<td>using the Search all variations of the keyword terms (stemmed search) checkbox.</td>
<td></td>
</tr>
<tr>
<td>Boolean Operators &amp;</td>
<td>Logic Operators: OR, AND, NOT</td>
<td>The text operators, OR, AND, and NOT must be capitalized.</td>
</tr>
<tr>
<td>Groupings</td>
<td>Groupings: ( )</td>
<td></td>
</tr>
<tr>
<td>Wildcard</td>
<td>* for multi-character wildcard searches. Matches zero or more characters.</td>
<td>Wildcard characters can be used in the beginning, middle and end of terms.</td>
</tr>
<tr>
<td></td>
<td>? for single-character wildcard searches</td>
<td></td>
</tr>
<tr>
<td>Phrase</td>
<td>&quot;word1 word2&quot;</td>
<td></td>
</tr>
<tr>
<td>Proximity</td>
<td>term1 w/n term2</td>
<td>w/n specifies the number of words that can separate the terms. In other words, term1 is within n words of term2. The w/n operator is not case sensitive.</td>
</tr>
<tr>
<td></td>
<td>or</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&quot;term1 term2&quot; ~n</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** Because w/n is now an operator, searches containing the string, w/n, are interpreted as proximity searches. Verify that the saved searches of upgraded cases are not impacted. Upgraded cases
<table>
<thead>
<tr>
<th>Query Type</th>
<th>Syntax</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nested Proximity</td>
<td>term1 w/n (term2 w/n term3)</td>
<td>Nested proximity searches combine two query types, proximity and grouping.</td>
</tr>
</tbody>
</table>

**Clearwell Detailed Search Reference**

**Clearwell User Interface**

Keyword search can be performed using the Basic Search field or the Advanced Search page.

![Basic Search field](image)

**Advanced Search screen**

**General Notes**

- Searches involving Boolean, phrase, wildcard, or proximity queries can be entered into the Basic Search field or Any of these words field on the Advanced Search screen. These types of searches are generally not supported in other fields within Advanced search.
  - Note that the size of the input fields on the Advanced Search page will grow as you add text.
- If you enter words in more than one field on the Advanced Search page, the search results include only documents that match all of the fields. Each term is AND’ed with every other term in the search.
<table>
<thead>
<tr>
<th>Example</th>
<th>Search Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any of these words field: energy</td>
<td>Include items that include the word “energy” and also include the phrase “nuclear power”</td>
</tr>
<tr>
<td>The exact phrase field: nuclear power</td>
<td></td>
</tr>
</tbody>
</table>

- All searches from the Basic Search field and Advanced Search screen are case insensitive. Operators (e.g. AND, OR, NOT) must be uppercase. In email and file content, Clearwell will index certain punctuation characters and treat others as spaces in order to make as many words searchable as possible. Treatment of punctuation characters has changed since version 4.5. Please refer to the Appendices for additional information.

- As of version 4.5 and beyond, all words are indexed. In prior versions, stop words (such as “and” and “the”) were ignored unless they are included in exact phrase searches with one or more additional search terms. All cases started in those versions will continue to ignore stop words. Reference Appendix C for more information on stop words in prior versions.

- Search queries without any advanced operations are limited to approximately 8,000 terms. This limit is lowered when searches include wildcard or proximity queries.
Understanding Search Result Statistics

- **Search Guide**
- **PAGE: 8**

**Total number of Emails and Loose Files searched**

**Total number and volume of Emails and Loose Files found matching the search criteria**

**Number of Discussions that contain at least one email in the Found documents**

**Number of Topics that contain at least one email in the Found documents**

**Unique number of files contained in the Found documents. A file that is attached to one or more emails in the Found documents and is a loose file counts as a single unique file. Files having identical content with or without the same filename are also counted as one unique file.**

**Number of participants or the number of unique email addresses that either sent or received emails within the set of found documents.**
Stemmed Searches

Stemmed searches find variations of words such as plurals or alternative verb forms. For example, if you search for "test", stemming will also find instances of "tests" and "testing". The Basic Search field always uses Stemming. In the Advanced Search screen, you can choose whether to run a stemmed search or a literal search by using the Search all variations of the keyword terms (stemmed search) checkbox.

Additional Notes

- Terms contained in the To, From, CC, bCC, and attachment/file name fields in an email and the filename of loose files are not stemmed during processing in order to reduce false positives. See the FAQ on Stemming vs. Wildcard searches for more information.

- Clearwell can support stemmed searches in English, Dutch, French, German, Italian, Japanese, Korean, Portuguese, Russian and Spanish. By default, only English words are stemmed. Stemming for additional languages is controlled by your administrator. When stemming is configured for more than one language, Clearwell will perform stemming for all languages on each submitted term. For example, if you enter restaurant and both English and French stemming is configured, then Clearwell will search for both English and French variants of this term. Note that Clearwell does not perform any language translation.

- Clearwell supports two methods for supporting stemmed searches in English: linguistic stemming and suffix-based stemming. Linguistic stemming uses part of speech analysis to determine stemming rules. For example, this option considers "went" as a variant of "go." Suffix-based stemming uses the Porter algorithm to strip out common word suffixes (such as "s" or "ing"). This algorithm is useful for finding nouns in their plural and singular forms. Both methods are configured by default.
### Boolean Searches

**Logic Operators**

Individual query terms can be combined together into more complex search requests by using logic operators. The following table describes the available logic operators. The text operators, **OR**, **AND**, and **NOT** must be entered in uppercase.

<table>
<thead>
<tr>
<th>Operator</th>
<th>Description</th>
<th>Example</th>
<th>Clearwell Query Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>OR</strong></td>
<td>Includes documents that contain either of the terms connected by the OR. The OR operator is the default conjunction operator. This means that if there is no operator between two terms, the OR operator is used.</td>
<td>Search for either coffee or tea</td>
<td>coffee OR tea</td>
</tr>
<tr>
<td><strong>AND</strong></td>
<td>Includes only documents that contain both terms connected by the AND.</td>
<td>Search for espresso and cappuccino</td>
<td>espresso AND cappuccino</td>
</tr>
<tr>
<td><strong>NOT</strong></td>
<td>Excludes documents that contain the term after the NOT operator.</td>
<td>Search for &quot;french roast&quot; but not decaf</td>
<td>&quot;french roast&quot; NOT decaf</td>
</tr>
</tbody>
</table>

Note that the **NOT** operator cannot be used with just one term. For example, the following query entered with no other search criteria will return no results, even if one or more documents do not contain the term **chai**:

NOT chai

Like **AND** searches, **NOT** searches will treat messages and attachments as separate documents. In the example above, an email whose message body contained "french roast" and decaf but whose attachment contained "french roast" but did not contain decaf would still be included in the search results.
### Operator Description

Note that keywords entered in the None of these words field in the Advanced Search screen behave differently from keywords after the NOT operator. A search using None of these words will exclude messages if the email body or any of the attachments match the specified query. In the example above, an email whose message body contained "french roast" and decaf but whose attachment contained "french roast" but did not contain decaf would be excluded in the search results.

### Grouping

Use parentheses to group clauses to form sub-queries and control the Boolean logic for a query.

<table>
<thead>
<tr>
<th>Example</th>
<th>Clearwell Query Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for either coffee or tea and the word milk</td>
<td>(coffee OR tea) AND milk</td>
</tr>
</tbody>
</table>
**Wildcard Searches**

Use a `?` for single character and a `*` for multiple character wildcard searches. Wildcard characters can be used in the beginning, middle or end of a term.

**Single Character Wildcard**

The single-character wildcard matches on any single character in the wildcard position.

<table>
<thead>
<tr>
<th>Example</th>
<th>Clearwell Query Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for text or test</td>
<td>te?t</td>
</tr>
</tbody>
</table>

**Multiple Character Wildcard**

The multiple character wildcard searches matches on zero or more characters.

<table>
<thead>
<tr>
<th>Example</th>
<th>Clearwell Query Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for test, tests or tester</td>
<td>test*</td>
</tr>
</tbody>
</table>

**Additional Notes**

- The use of wildcards is not supported when used in conjunction with non-indexed characters, such as leading or trailing punctuation characters. See the Appendices on tokenization for more information on which punctuation characters are indexed and searchable.

- Wildcards can be used in the following Advanced Search fields:
  - Keywords Section
    - Any of these words, All of these words, None of these words
  - Identifiers Section
    - Source name and location
  - Email Section
    - Subject
  - Attachment/File Section
    - Any of the words

- Hit highlighting of wildcard terms via the Advanced Freeform search page is not supported.

- Searches containing non-ASCII characters and wildcards could return an error due to too many results. If this error occurs, group the non-ASCII characters and wildcards in angle brackets. This prevents the wildcard from running as a separate search.
Phrase Searches

A phrase is a group of words enclosed in double quotation marks. Phrase searches will find documents containing the terms within the quotes in the same order with no intervening other terms.

<table>
<thead>
<tr>
<th>Example</th>
<th>Clearwell Query Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for the exact phrase: grande latte</td>
<td>&quot;grande latte&quot;</td>
</tr>
</tbody>
</table>

Additional Notes

- Phrase searches can be run as stemmed or literal searches. For example, if run as a stemmed search, the phrase "energy policy" will match "energy policies" as well as "energy policy". Phrases entered in Basic search are automatically run as stemmed searches. The Basic Search field always uses Stemming. In Advanced Search, you can choose whether to run a stemmed search or a literal search.

- Searches using the Exact Phrase field on the Advanced Search page do not support the same functionality as Phrase searches using quotes entered into the Any of these words field. For example, you cannot use wildcards in the Exact Phrase field. For complex queries, it is recommended to use phrase searches in the Any of these words field instead of the Exact Phrase field.
Proximity Searches

Proximity searches find words that have a specific number of intervening words. When performing proximity searches, the word order in the phrase does not matter. Clearwell supports proximity searches containing two or more terms. You can perform a proximity search two ways:

• Separate search terms with w/n.
  
  example: budget w/10 issues

  Note: Because w/n is now an operator, searches containing the string, w/n, are interpreted as proximity searches. Verify that the saved searches of upgraded cases are not impacted. Upgraded cases containing saved searches with the string, w/n, may result in an error. Saved searches with the string, NOT w/n, are now run as a proximity search.

• Add a tilde (~) at the end of a phrase (quoted string) followed by the total number of other words that are allowed to come between the words in the phrase.
  
  example: "budget issues"~10

Both searches will find documents where there are 10 or fewer intervening words between “budget” and “issues” or where there are 10 or fewer intervening words between “issues” and “budget.”

Note: Wildcard characters (* or ?) can be used within proximity searches only in Basic search and the Advanced search Any of these words fields.

Additional Notes

• Clearwell's proximity search specifies the number of intervening words allowed between terms. Users who are running searches for others should verify with the search author as to how many intervening terms they want between the words.

• Proximity search is limited to certain fields or regions within email messages and does not span email messages and attachments. For example, proximity searches do not span the Recipient (To) and subject metadata fields or the subject and body regions of an email. Proximity searching does not span email or attachment boundaries. The Freeform Search Guide contains a list of regions within emails.

• Hit highlighting for proximity searches is not limited by the proximity number. For example, for the search budget w/10 issues, the terms "budget" and "issues" will be highlighted throughout the document not just when there are only 10 intervening terms or less.

• Proximity searches can be used to find specific number sequences, such as phone numbers or social security numbers when written according to the following example:

  <???-??-????> w/12 "social security"
• This will find an social security number in proximity to the phrase "social security".

Note: Using wildcards alone may match similar unwanted text combinations, such as the phrase "one-to-many". However, grouping the wildcards with proximity search phrasing will reduce the number of false positives in your results.

• When constructing proximity searches using the tilde format, there should be no spaces between quote marks, ~, or proximity number. For example, “budget issues” ~10 will not be recognized as a proximity search.

Nested Proximity Searches

Nested proximity searches combine two query types, proximity and grouping. Examples of nested proximity searches include:

• “apple pie” w/5 (“strawberry cheesecake” w/10 “apple tart”)
• NOT (“apple pie” w/10 “apple tart”)
• “blueberry scone” NOT (“apple pie” w/10 “apple tart”)
• 4. NOT “blueberry scone” NOT w/10 “apple tart”

The first example would find all documents that contain all three phrases “apple pie”, “strawberry cheesecake” and “apple tart” which contains at least one occurrence of “strawberry cheesecake” that is within 10 words of “apple tart” which is also within 5 words of “apple pie”.

The search in example 2 would exclude all documents that contained the phrase “apple pie” within 10 words of “apple tart”.

Similarly, example 3 would find all documents that contained the phrase “blueberry scone” but by contrast, did not also contain “apple pie” within 10 words of “apple tart”.

In example 4, this search would find all documents that contain the phrase “blueberry scone” in which “blueberry scone” does not appear within 10 words of “apple tart”.
Transparent Searches

Clearwell’s Transparent Search is designed to provide deep visibility into how searches are performed in order to improve the ability to cull irrelevant information. Transparent Search makes it easy to follow search best practices including search query testing, sampling and refining. Transparent search is comprised of four features.

- **Search Preview** - Provides visibility into matching keyword variations for wildcard and stemming searches prior to running a search. You can selectively include relevant variations or exclude false positive variations in the search query, removing irrelevant documents from search results.

- **Multiple Query Analytics** – Allows you to run multiple queries as part of a single search and get analytical data for each individual query as well as all queries combined.

- **Search Filters** - Enables filtering of search results based on individual queries or variations within a multi-query search allowing you to sample and test the results for each query in a multiple query search.

- **Search Report** - Creates a comprehensive report that documents all search criteria, including selections from search preview, and provides detailed analytics of the results for both the overall search and the individual queries within the search.
Using the Search Preview Feature

The search preview feature can be accessed by clicking on the icon to the right of the Any of These Words field on the Advanced Search page.

The search preview window shows all the variations for each wildcard or stemmed keyword within your search query. For example, if the query contains the keyword hir*, the window will show all terms within your data set whose first three characters are hir. If you have selected the Search All Variations of the Keyword Terms (Stemmed Search) option then the search preview window will display all stemmed variations of that term. Search preview allows you to select or de-select each shown variation including the relevant ones and excluding the non-relevant, false positive variations.

Only selected variations will be included in the search. If you do not open the search preview window and run a search with wildcard or stemmed keyword variations, then the search will run as if you had selected all variations.

Additional Notes

- The search preview feature is not available for literal searches without wildcards.
- Because terms within the To, From, CC, bCC, and attachment/file name fields are not stemmed, selected stemmed variations will not be searched within those fields. Only the unstemmed keywords entered into the Any of These Words field will be searched for within those fields.
- The counts in the search preview window are not affected by the Fields to Search setting or by visibility filters.
Using Multiple Query Analytics

Clearwell’s Transparent Search supports the ability to simultaneously run multiple queries and provide filters and analytics on each individual query plus the combination of all submitted queries. You can create a search with multiple queries by adding multiple query rows. A query row is an additional Any of These Words field on the Advanced search page and can be created by clicking on the + icon.

You can also create multiple query rows by (1) copying searches from text in another application and (2) pasting that text into the Any of These Words field. (3) A query row is created for every line of copied text.

Additional Notes

- The number of query rows allowed in a search is limited to 100.
Running Transparent Searches

You can run a Transparent Search that includes only your selected variations for each query by clicking Run Search. This will produce filters and report analytics for each query contained in the submitted search. You can generate more detailed filter and report analytics for each selected variation combinations by checking the Generate Keyword Details for Filters and Report.

Filter and Count Generation options within the Advanced Search window

- Limit filter and count generation for improved search speed. If selected, Sender, Recipient and Keyword filter information will not be generated. In addition, the Participants page will not be available and the Search Report will not display keywords or counts. To see this information, you may re-run the search at any time without this option selected.
- Normal Filter and count generation. Creates a filter for each search term entered, however, it does not create a filter for the expanded, wildcard matches of the search terms.
- Generate keyword details for filters and report.
- Creates filters for the search terms and all wildcard matches of the search terms.
- It takes significantly more resources and time to run searches with the Generate Keyword Details for Filters and Report option selected. The performance of a search with this option checked will be affected by the number of keywords within an Any of These Words query row field and the number of query rows. Currently, these searches are limited to 10,000 keyword combinations, which might take approximately 20-30 minutes to run. Keyword combinations are the number searches that are generated from a search using wildcards or stemming. For example, if the term hir* expanded to hire and hired, then the search hir* AND policy would have two keyword combinations: hire AND policy and hired AND policy. Searches that exceed that number of combinations and are likely to take longer to run will produce an error similar to the following: "Term expansion combinations count of [X] exceeds the limit of 10,000. Reduce selected expansions or disable keyword details."
Running Transparent Searches – Search Jobs

If the system determines that the search is large, the system automatically creates a job for the search, which is run in the background as shown below. When a search runs as a job, the results of the search are calculated and saved with the search in order to enable quicker access to the results of large searches.

Search jobs run in the Searches area on the Documents page and are shown with a spinning magnifying glass icon and a cancel option. Completed search jobs have a grayed magnifying glass icon and edit and refresh options. The results of a completed search job can be accessed by clicking on the search name. Searches that are not run in the background as jobs are indicated by a non-colored magnifying glass with an edit option.
Additional Notes

- If additional documents are processed or additional tags have been made and the search contains tagging search criteria, then the results of the search job can become “stale” or out-of-date. You can either review your saved results or re-run the search to update the results by clicking on the search job as shown above.

- The system will save the results of up to 50 search jobs. After the 50th search is reached, the system will delete the results associated with a job but not the query. You will still be able to access the results of a search by clicking on the search in the Searches window, but you will only be able to re-run the search. You will not be able to access the saved results.

- Saved results in search jobs are not affected by visibility filters. If this is a concern, save these searches as Private Saved Searches.
Using Keyword Query Filters

Clearwell generates keyword query filters for each search. These filters enable you to restrict your overall results to the documents that match a single query row within your Advanced search. To quickly filter search results, simply select the filter and clicking Apply Filters. In the following example, selecting hir* AND policy restricts the filtered results to the 56 documents that only match the query.
You can also build complex filters using multiple criteria. In this example, one Sender Domain filter value has been checked and the Keyword Query filter for the search `hir* AND policy` has been unchecked. This will filter results to find emails sent from the selected domain and will not include emails that only match the `hir* AND policy` query.

Highlighted filters applied to the search.
Checking the *Generate Keyword Details For Filters and Report* option when you perform your search will generate additional keyword query filters. For example, without this option, you have the option to filter on all of the documents that match the query `hir* AND policy`. With this option checked, you also have the ability to match all of the query expansions of this query, such as `hired AND policy` or `hire AND policies`.

![Keyword Query Example](image)
Using the Search Report

The Search Report provides information on the specified search criteria and results of a search.

The Search Report has three sections: Search Report, Results, and Keywords.

**Note:** If you have run a concept search, the Search Report will include a Concepts section displaying the total concept terms applied in the search. See Concept Search on page 31.

- **Search Report** – The first section lists information related to the case and search query including all of the specified search criteria. The keywords used in a search are shown by default. All other search criteria are hidden by default. Click on Show Search Detail to show all of the specified search criteria.

- **Results** – The results section provides the following counts:

<table>
<thead>
<tr>
<th>Document Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>Total number of emails and loose files</td>
</tr>
<tr>
<td>Emails</td>
<td>Emails and their attachments (note that an email with 2 attachments counts as a single email)</td>
</tr>
<tr>
<td>Loose files</td>
<td>Files that are not attached to emails</td>
</tr>
<tr>
<td>Matching Emails</td>
<td>Emails whose content matches the search criteria</td>
</tr>
<tr>
<td>Non-matching Emails</td>
<td>Emails whose content does not match the search criteria but which has an attachment whose content does match</td>
</tr>
<tr>
<td><strong>Attachments</strong></td>
<td>Total number of files attached to emails</td>
</tr>
<tr>
<td>----------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td><strong>Matching Unique Files</strong></td>
<td>The number of unique files, which can be attachments, loose files or both, whose content matches the search criteria</td>
</tr>
<tr>
<td><strong>Non-matching Unique Files</strong></td>
<td>The number of unique files whose content does not match the search criteria but which is attached to an email whose content does match.</td>
</tr>
<tr>
<td><strong>Unique Files</strong></td>
<td>Total number of unique files in the search results. A unique file can be an attachment that is attached to multiple emails and/or a loose file. These attachments or loose files have the exact same content but may have different file names or modified dates.</td>
</tr>
<tr>
<td><strong>Discussions</strong></td>
<td>Total number of email discussion threads</td>
</tr>
<tr>
<td><strong>Topics</strong></td>
<td>Total number of groupings of conceptually similar emails</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>Total number of unique email addresses which have sent and/or received emails</td>
</tr>
<tr>
<td><strong>Reviewable items</strong></td>
<td>Total number of emails, attachments and loose files</td>
</tr>
</tbody>
</table>
• **Keywords** – The final section shows the number of documents that each keyword query would match if run individually. To see additional details on the keyword query, click Show Keyword Detail. The keyword details section documents the stemmed or wildcard word variations that were searched or not searched based on the selections or de-selections made using the Search Preview feature. If you checked Generate Keyword Details For Filters and Report for a search, then keyword details will include a new Results section that lists all the keyword query expansions and the number of documents that match the query.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>interview</td>
<td>1,185</td>
</tr>
<tr>
<td><em>hr</em> AND policy</td>
<td>56</td>
</tr>
</tbody>
</table>

Results:
- hire AND policies 32
- hiring AND policies 28
- hire AND policy 27
- hired AND policy 21
- hired AND policies 20
- hires AND policy 6
- hires AND policies 6
- fired AND policy 3
- fired AND policies 3
- hire@emt@company.com AND policy 2
- hires@emt@company.com AND policy 2
- hires @ company.com AND policies 2
- hire AND policy 1
- hire AND policies 1

Variations searched for:
- *hr* (stemmed): hire, hired
- hire AND policies

Not searched for:
- *hr* (stemmed):

```
recruit
54
```

```
"employee" contract
52
```

*Keyword detail in a Search Report. The Results section (highlighted) displays when you select Generate Keyword Details for Filters and Report.*

**Additional Notes**

- The information listed in the Search Report is not affected by any applied or saved filters
Participant Searches

As of version 6.1, there are two ways to perform a participant search on the Advanced Search page. The Participants search area is an expandable alternative to using the static keyword search fields on the left side (such as "Any of these words"), allowing users to perform a more robust participant search.

Using the Participant search feature provides greater control and flexibility in the types of searches you can perform. A complex participant search can be expanded using multiple rows, each of which contains three drop-down boxes and a text field.

- **General Rules** – Multiple names, email addresses, or domains must be separated by semicolons (;).

- **Email Addresses** – To search for an alias, enter alias(<aliasname>) (or click to select any email address (primary or secondary) for any known participant.

- **Participants** – The name order is not critical. To find all documents sent or owned by a participant, enter the first, last or last, first:

<table>
<thead>
<tr>
<th>Example</th>
<th>Participant Option with Text Field Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for the participant: john smith</td>
<td>john smith or smith john</td>
</tr>
</tbody>
</table>

- **Domains** – If entering a partial domain, specify the right-most portion of the name, and include the full text between period delimiters.

<table>
<thead>
<tr>
<th>Examples</th>
<th>Domain Option with Text Field Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Broad]: Search all documents from: yahoo</td>
<td>yahoo.com  [includes all yahoo domains, including yahoo.com and segments such as images.yahoo.com (but not images.yahoo.co.uk)]</td>
</tr>
<tr>
<td>[Narrower]: Search all documents from: images.yahoo.com.hk</td>
<td>images.yahoo.com.hk  [includes only the images.yahoo.com.hk domain]</td>
</tr>
</tbody>
</table>
### Field/Option Description

**Any/and any/or any/not any**

Finds documents that have the specified participant names, email addresses, or domains according to the following operators:

- **any** (in the first row) specifies that for the text entered, only one of the criteria must match in a document for the entire row to be considered a match. and any specifies that the criteria in that row are required in the search. or any (in subsequent rows) is optional, indicating that the same documents can contain the text entered in that row. (However, one row must be required if all others are optional.) not any indicates that the documents must not contain the (prohibited) criteria that follow in that row. (If documents contain any participants in a prohibited row, those documents will not appear in your results.)

**(All)/ (Recipients)/ From/ To/ Cc/ Bcc**

Finds documents that have the specified names, email addresses, or domains according to the following rules:

- **(All)** searches all fields: From, To, Cc, or Bcc (fields are blank on loose files).
- **(Recipients)** finds documents in the To, Cc, or Bcc fields.
- To search any single sender or recipient field, select From, To, Cc, or Bcc. These fields represent the specified individual and search all documents from all of that individual’s email addresses.
- If the “Search in contained senders and/or recipients” option is selected, the equivalent contained fields are also searched.

**Note:** See Participants/E-mail address/Domain field options for usage.

**Participant/E-mail address/Domain name**

Specifies the search type:

- **Participant** — searches for all documents from an individual by primary email address. Results will contain the primary email address of the selected participant. (A participant search on a secondary email address will not return any results.)

  **Note:** The “primary” email address is determined by the first address found for a given participant when data is indexed by Clearwell.

- **E-mail address** — searches for documents with an exact match of the original email address (finds all messages from or to a single email address). The email selector can be used to identify documents from the participant with the exact email address.

- **Domain** — searches for documents with part or all of a domain from the original email address. (Sender and recipient domains are generated using the original email address domains so that the domain will appear in the appropriate field of the filtered documents in your results.)

  **Note:** See Additional Notes.

### Field/Option Description

<table>
<thead>
<tr>
<th>Field/Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| Any/and any/or any/not any | Finds documents that have the specified participant names, email addresses, or domains according to the following operators:  
- **any** (in the first row) specifies that for the text entered, only one of the criteria must match in a document for the entire row to be considered a match. and any specifies that the criteria in that row are required in the search. or any (in subsequent rows) is optional, indicating that the same documents can contain the text entered in that row. (However, one row must be required if all others are optional.) not any indicates that the documents must not contain the (prohibited) criteria that follow in that row. (If documents contain any participants in a prohibited row, those documents will not appear in your results.) |
| (All)/ (Recipients)/ From/ To/ Cc/ Bcc | Finds documents that have the specified names, email addresses, or domains according to the following rules:  
- **(All)** searches all fields: From, To, Cc, or Bcc (fields are blank on loose files).  
- **(Recipients)** finds documents in the To, Cc, or Bcc fields.  
- To search any single sender or recipient field, select From, To, Cc, or Bcc. These fields represent the specified individual and search all documents from all of that individual’s email addresses.  
- If the “Search in contained senders and/or recipients” option is selected, the equivalent contained fields are also searched.  

  **Note:** See Participants/E-mail address/Domain field options for usage. |
| Participant/E-mail address/Domain name | Specifies the search type:  
- **Participant** — searches for all documents from an individual by primary email address. Results will contain the primary email address of the selected participant. (A participant search on a secondary email address will not return any results.)  

  **Note:** The “primary” email address is determined by the first address found for a given participant when data is indexed by Clearwell.  

- **E-mail address** — searches for documents with an exact match of the original email address (finds all messages from or to a single email address). The email selector can be used to identify documents from the participant with the exact email address.  

- **Domain** — searches for documents with part or all of a domain from the original email address. (Sender and recipient domains are generated using the original email address domains so that the domain will appear in the appropriate field of the filtered documents in your results.)  

  **Note:** See Additional Notes. |
Search in contained senders and/or recipients (Checkbox)

Finds messages with senders or recipients that are in contained emails (email messages that have been replied to or forwarded).

Additional Notes

- Differences between the two participant search methods

Searches executed from the left side of the Advanced Search screen are broader in scope. Participants are included if “All fields” is selected (by default), or “Senders and recipients”, but cannot be limited to, for example, only senders. This search will find original email addresses (or, in upgraded cases, both primary and original email). For example, searches for “j smith@yahoo.com” in “senders/recipients” return documents containing that email address. However, if another document was sent by “j smith@acme.com”, no results are returned, even if the “acme” address was John Smith’s primary email address.

Note: To refine your search, use the Participants search area to specify the sender and/or recipients, participant email addresses (including the participant picker to select from a list of existing individuals and email addresses), and/or domains.

- How domains in participant searches are tokenized

Tokenization is done by splitting domains on the period delimiter (to provide additional flexibility of not requiring users to enter the entire domain).

- Wildcard searches in participant search types

All three search types (participant, email address, and domain) support the use of wildcard searches using ? and *. However, use of wildcards in Participant searches will not initiate a background search and could considerably slow performance. Additionally, you cannot choose term variations, and expansions are limited to 100,000 terms.

Note: Avoid leading wildcard searches, such as *gma* (for gmail), as this can significantly slow the search process.

- Participant Search filters

The Sender Name and Recipient Name filters are generated the same as they were in previous versions, representing the individual with that name, including all messages from all of that individual’s email addresses. (There is no set of filters for original email addresses.) Prior to version 6.1, these filters were generated using the primary email address domains for each document in the search results, but Clearwell now applies original email address domain filters. Essentially, when a domain filter is applied, the domain of interest will be present in the appropriate field of each of the filtered documents.
Concept Search

As of version 6.6, Clearwell’s Concept Search adds a visually transparent and intuitive way to identify potentially relevant documents based on a concept. You can perform a basic or advanced search of a concept. In Advanced Search, selecting Concept allows you to enter multiple concept terms and custom-refine your search.

There are three main areas of (Advanced) Concept search:

- Concept Search Preview
- Concept Search Explorer
- Concept Search Report

Clicking the “edit” icon next to the box containing your concept terms opens the Concept Builder, allowing you to refine your concept by building on related terms in Search Preview and Explorer.
1. Based on your original concept, start in Search Preview to select (or de-select) terms that are relevant only to your case.

   For example, searching for the concept “pay-off” will list all terms found to contain or be related to its meaning, such as “evidence”, “profits” or “government”

2. As you select terms in the Preview pane, a graphical view of how your concept relates to other terms is shown (as a blue bubble with connecting terms) in Search Explorer to the right. This allows you to select only the precise terms related to the word “pay-off” that should be included in your search. You can continue building and viewing related concepts in the Explorer view by clicking and dragging words. Clicking a word (related concept) in Explorer view allows you to build on that term as a related concept. Clicking Refresh (at the bottom left of the Concept Builder window) shows you how many documents, based on the current selected concept terms, will be found in your results.

   Terms: 9  Documents: 1,883  Refresh
For example, if your document count is too large, and for your case, you are only interested in the related term “profits”, you can refine your search by clicking the word “profits” in the Explorer view. A new (orange) bubble appears, stemmed from your original concept.

Depending on where you want to focus your search, use the play buttons (at the top of the window) to go forward or back through your changes to adjust the total number of documents.

Each time you arrange or adjust your terms, click the Refresh link to update the document count. (The link becomes unavailable if the count is current after the last modification).

3. When you are ready to run the search, click Save Concept. This returns you to the Advanced Search page, where you can click Run Search to view your results.

You can also use Concept Builder to run the same terms as keywords. Clicking Save as Keywords from the Concept Builder window returns you to the Advanced Search page with the terms pre-populated in the Keywords section.

4. After saving and running your search, view your results showing the highlighted terms (as shown in the Common Concept Terms box). This lists the terms selected for the original concept as well as other conceptually related terms.

5. Report on your search results by viewing the Concept section of the Search Report, which displays the original concept term and all common concept terms included in your search.

Additional Notes
• Search Preview displays up to 200 related terms, out of which you can select 20. Plus, any additional concept terms are shown, which Clearwell determines are closely related to the selected concept terms.

• In the search results, the following terms are displayed:
  • The original list of input terms, including
  • Any additional terms you selected in Search Preview and Search Explorer, plus
  • Additional list of ranked terms
  • Concept terms are also highlighted in the document results, indicating the reason a specific document was considered related.
  • Stop words such as “and”, “or”, “the” in your original terms are excluded before searching for related terms or documents. See “Appendix D - Stop Words for Performance-Sensitive Indexes” for a full list of excluded words.
  • Concept searches can be combined with Tag, Folder, Participant and other selections in an Advanced Search.
  • All terms listed in the Common Concept Terms box are shown in order of frequent occurrence near the selected terms.
  • Best practice is to save your concept first, then save your search. If you want to run a Keyword search, click the “edit” icon to re-open Concept Builder. From the Concept Builder window, click Save As Keywords, then save (or run) that search.
  • You can always go from a concept search to a keyword search; however, if a search is saved after running it as a keyword search, your concept information is not saved, and therefore not available to reconstruct the concept.
Freeform Searches

The Freeform Search feature allows you to construct queries using the full power of Clearwell’s underlying search engine. This section describes how to construct effective freeform searches.

About the Freeform Search Page

To open the Freeform Search page, click Advanced Search on the Basic Search bar, and click Freeform. Note that separate text boxes are provided for message queries and file queries. Separate queries are required for messages and files because the data is stored in separate indexes. The query strings entered in each text box are treated as an AND search, along with any other search criteria you specify on the page.

Basic Freeform Queries

Freeform queries can include terms, fields, and logic operators, as described below. Note the following rules:

- All searches are case-insensitive
- Each of the two query fields (message and file) can contain up to 8,000 “tokens”. Tokens are individual query elements, such as terms and fields.
- The maximum text length of a query depends on your browser, but is usually 128K.
- In addition to the Freeform Search page, Clearwell supports basic freeform queries in the Basic search and Advanced search Any of these words fields, including phrase, logic operators, grouping, wildcard, and proximity searches.
- Advanced freeform queries, such as field selection, fuzzy searches and boosting are not supported in the Basic search or Advanced search Any of these words fields.
- Field selection, fuzzy searches and boosting are only supported through the Freeform search page.

Terms

There are two types of terms: single terms and phrases:

- A single term is one word, such as “coffee” or “tea.” A phrase is a group of words enclosed in double quotation marks, such as “grande latte.”
- Multiple terms can be combined together with logic operators to form a more complex query (see “Logic Operators”).
Logic Operators

Individual query elements can be combined together into more complex search requests by using logic operators. Refer to the table for basic Logical Operators in the section “Boolean Searches”.

The following table describes additional logic operators and how they can be used to combined search terms.

<table>
<thead>
<tr>
<th>Operator</th>
<th>Description</th>
<th>Example</th>
<th>Clearwell Query Syntax</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Includes only documents that contain terms after the + symbol (but only the word immediately following the symbol)</td>
<td>Search for “mocha” (but may contain “beans”)</td>
<td>+mocha beans</td>
</tr>
<tr>
<td>–</td>
<td>Excludes documents that contain the term after the – symbol.</td>
<td>Search for “bagel” but not contain “cream cheese”</td>
<td>bagel -&quot;cream cheese&quot;</td>
</tr>
</tbody>
</table>

When performing searches, Clearwell treats messages and attachments as separate documents. With an AND search, a match occurs for a message only if all of the words are in the message or in an attachment. A match does not occur if the words are split between the message and an attachment.

Wildcard Searches

Clearwell supports the use of ? and * for single- and multiple-character wildcard searches, respectively.

The single-character wildcard indicates that a match occurs on any character in the wildcard position. For example, to search for “text” or “test,” enter: te?t

Multiple-character wildcard searches look for 0 or more characters. For example, to search for “test”, “tests” or “tester”, enter: test*

You can also use the wildcard searches at the beginning or middle of a term. te*t
You can perform wildcard searches in any of the following Advanced search fields: Any of these words, All of these words, phrase, None of these words, Source name and location, Subject, or Attachment/file – Any of the words, and in Basic search. You can use wildcards in phrase and proximity searches in Basic search or Advanced search Any of these words fields. Wildcards in phrase or proximity searches are not supported in any other fields.

Specifically, wildcard queries can be done in Freeform search, however, Clearwell does not support wildcard searches when used in phrase or proximity queries. For example, the following query will find hits with "flaming", "flamingo" or "flamingopink" in the body content:

```
+u_body:flaming*
```

However, the following query will ignore the wildcard and is essentially a simple search for "flaming lawn ornament" and will not find a document with "flamingo lawn ornament" in the body:

```
+u_body:"flaming* lawn ornament"
```

In this example, the letter "o" completely changes the meaning of the phrase.

**Note:** Searches containing non-ASCII characters and wildcards could return an error due to too many results. If this error occurs, group the non-ASCII characters and wildcards in angle brackets. This prevents the wildcard from running as a separate search.

**Grouping**

Clearwell supports using parentheses to group clauses to form sub-queries. This can be very useful if you want to control the boolean logic for a query. To search for either “coffee” or “tea” and “milk” in a document, use the query:

```
(coffee OR tea) AND milk
```

Parentheses can also group multiple clauses to a single field. To search for messages that contains both the word “latte” and the phrase “espresso machine” use the query:

```
(+latte +"espresso machine")
```
Proximity Searches

Proximity searches find words that have a specific number of intervening words. When performing proximity searches, the word order in the phrase does not matter. Clearwell supports proximity searches containing two or more terms. You can perform a proximity search two ways:

- Separate search terms with `w/n`.
  Example: `budget w/10 issues`
  
  **Note:** Because `w/n` is now an operator, searches containing the string, `w/n`, are interpreted as proximity searches. Verify that the saved searches of upgraded cases are not impacted. Upgraded cases containing saved searches with the string, `w/n`, and result in an error. Saved searches with the string, `NOT w/n`, are run as a proximity search.

- Add a tilde (~) at the end of a phrase (quoted string) followed by the total number of other words that are allowed to come between the words in the phrase.
  Example: `"budget issues"~10`

Both searches will find documents where there are 10 or fewer intervening words between “budget” and “issues” or where there are 10 or fewer intervening words between “issues” and “budget.”

**Note:** Wildcard characters (* or ?) can be used within proximity searches only in Basic search and the Advanced search **Any of these words** fields.

Nested Proximity Searches

Nested proximity searches combine two query types, proximity and grouping. Examples of nested proximity searches include:

- “apple pie” w/5 (“strawberry cheesecake” w/10 “lemon tart”)
- NOT (“apple pie” w/10 “lemon tart”)
- “maple scone” NOT (“apple pie” w/10 “lemon tart”)

Advanced Freeform Search Features

The following types of freeform searches are supported:

- Fuzzy Searches
- Fields
- Boosting Terms
Fuzzy Searches

Clearwell supports fuzzy searches based on the Levenshtein Distance, or Edit Distance algorithm. To perform a fuzzy search, add a tilde (~) at the end of a one-word term.

For example, to find terms like “foam” and “roams” in the subject of an email, enter the following fuzzy search:

\[ \text{u_subject:roam~} \]

**Fields**

Fields let you search specific parts of an email, such as the subject, body, or recipient list. Fields are unstemmed, which means that a match occurs only on the exact text specified in the query.

The following table describes the message query fields.

**Note:** All field names are case sensitive. You must enter all names exactly as shown in the following tables.

**Fields Available in Message Queries**

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>fromListIndexed</td>
<td>The sender of an email. Normally this is a single participant, but in some cases (such as when a message is sent “on behalf of” someone else) there can be multiple senders in the index.</td>
</tr>
<tr>
<td>toListIndexed</td>
<td>The recipients of the email as specified on the To: line.</td>
</tr>
<tr>
<td>ccListIndexed</td>
<td>The recipients of the email as specified on the cc: line.</td>
</tr>
<tr>
<td>bccListIndexed</td>
<td>The recipients of the email as specified on the bcc: line.</td>
</tr>
<tr>
<td>containedSenderListIndexed</td>
<td>List of senders identified in forwarded emails contained within an original email.</td>
</tr>
<tr>
<td>containedRecipientsListIndexed</td>
<td>List of recipients identified in forwarded emails contained within an original email.</td>
</tr>
<tr>
<td>ID:&lt;document_ID&gt;</td>
<td>The document ID number of a specific document. For example: ID:0.7.87.2171</td>
</tr>
<tr>
<td>importance</td>
<td>The importance of the email. Valid values are:</td>
</tr>
<tr>
<td></td>
<td>• 0. Low importance</td>
</tr>
<tr>
<td></td>
<td>• 1. Normal importance</td>
</tr>
<tr>
<td></td>
<td>• 2. High importance</td>
</tr>
<tr>
<td></td>
<td>For example, to search only messages with normal or high importance, add the following to the query: importance: (1 OR 2)</td>
</tr>
</tbody>
</table>
### Field Name

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>scope</td>
<td>The scope of the email. Valid values are:</td>
</tr>
<tr>
<td></td>
<td>• 0. Internal (sent between internal participants)</td>
</tr>
<tr>
<td></td>
<td>• 1. Inbound (sent from an external participant to an internal participant)</td>
</tr>
<tr>
<td></td>
<td>• 2. Outbound (sent from an internal participant to one or more external participants)</td>
</tr>
<tr>
<td></td>
<td>For example, to search only internal or inbound messages, add the following to the query:</td>
</tr>
<tr>
<td></td>
<td><code>scope:(0 OR 1)</code></td>
</tr>
<tr>
<td>sendersDept</td>
<td>The group(s) of the senders</td>
</tr>
<tr>
<td>recipientsDepts</td>
<td>The group(s) of the recipients</td>
</tr>
<tr>
<td>topicNounPhrase</td>
<td>The most important phrases in the email, as determined by Clearwell topic classification.</td>
</tr>
<tr>
<td>u_subject</td>
<td>Unstemmed subject.</td>
</tr>
<tr>
<td>u_body</td>
<td>Unstemmed message text.</td>
</tr>
<tr>
<td>u_quotedTextN</td>
<td>Unstemmed quoted text regions.</td>
</tr>
<tr>
<td>nonEmailAttachmentNames</td>
<td>Attachment names found within an email.</td>
</tr>
</tbody>
</table>

The following table describes the file query fields.

### Fields Available in File Queries

<table>
<thead>
<tr>
<th>Field name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>u_NEACContent</td>
<td>Unstemmed file content. Use this field to find an exact match on the specified file text.</td>
</tr>
<tr>
<td>NEAName</td>
<td>The filename.</td>
</tr>
<tr>
<td>u_NEAMetadata</td>
<td>Location where file metadata (such as camera type for a photo) is indexed (in newer versions).</td>
</tr>
</tbody>
</table>

### Boosting Terms

Clearwell allows you to boost certain terms in your search relative to other terms. To boost a term, add a caret after the term, followed by a boost factor (a number). The higher the boost factor, the more relevant the term will be considered when ranking results.
For example, to search for both “breakfast” and “donuts,” but “breakfast” is much more relevant than “donuts,” you can enter:

```
breakfast^4 donuts
```

By default, the boost factor of all terms and phrases is 1. Although the boost factor must be positive, you can use a value less than 1 (such as 0.2) to decrease a term’s relevance.

**Common Freeform Searches**

The following examples show how Freeform Search can be used to satisfy common E-discovery requests.

**Finding traffic between two groups**

While the Dashboard can be used to monitor group-to-group communication, in some cases you may want to carry out more detailed searches using the full power of Clearwell Advanced search.

To include a constraint in a Freeform query that restricts the result set to messages that were sent between two groups, use the following query:

```
(sendersDept:"<Group 1>" AND recipientsDepts:"<Group 2>") OR 
(sendersDept:"<Group 2>" AND recipientsDepts:"<Group 1>")
```

This logic can be made more complex as necessary; such as to track interactions between more than two groups, or to find documents sent from one of several groups to another group.

**Searching for files of a particular type**

Freeform search allows you to distinguish between searches on file content and the file name, so that you can limit your searches to files of a particular type. For example, to find loose XLS files and messages that have XLS attachments that contain the word “budget”, use the following file query:

```
+NEAName:(xls) +NEAContent:(budget)
```

**Finding the blind copy messages that a user received**

In a standard advanced search, the “Recipient” field does not distinguish between the To, Cc, or BCC lines. Using Freeform Search, however, you can easily distinguish between these three fields. For example, to find all messages that were grouped using bcc to someone named Smith, add the following to your message query:

```
+bccListIndexed:(smith)
```
Non-English Language Searches

Clearwell supports searches in all common languages. When performing searches with languages that use characters, such as Chinese, Japanese and Korean, note the following:

- If you enter characters with no spaces, such as:

  北京中国
  (Beijing China)

  Clearwell will interpret this as a phrase search and will find documents containing these characters in the exact order you specify.

- To search for documents containing ANY of these characters, enter the characters with spaces or using explicit OR operators. For example:

  北京中国, or 北京 OR 中国

  will search for Beijing OR China.

- To search for documents containing ALL of these characters but in no particular order, enter the characters using explicit AND operators. For example:

  北京 AND 中国

  will search for Beijing AND China.

- If you do a wildcard search (using ? or *) with Kanji style multi-character sets, you may have mixed, or no results. These conditions are more complex. For example:

  重大收支机会

  this is broken into three tokens:

  重大
  收支
  机会

  To search for any of the tokens in wildcard form, supply only that token with a wildcard.

  Further, for accurate interpretation of wildcards in Chinese, Japanese and Korean languages, Clearwell requires enclosing the phrase with angle brackets: < and >. This enables proper language boundary detection and identification. In the above example, the last of the three tokens and its wildcard variations can be searched using:

  <收*>

**Note:** Clearwell does not currently provide any translation functionality. For more information and examples on multi-character handling, and how to search in languages other than English, refer to the *Clearwell Multi-Language Support Guide.*
Punctuation Searches

Clearwell indexes some punctuation characters but does not index others. In order to maximize searchability, Clearwell also indexes punctuation characters differently depending on where they are found. For example, To, From, Cc, Bcc and filename content is indexed differently from other content. Clearwell also alternatively indexes different types of terms, such as email addresses or terms containing numbers. See the Appendices for details on how punctuation characters are indexed.

Frequently Asked Questions About Punctuation Searches

How does Clearwell handle punctuation searches?

The ability to find documents containing terms that include punctuation characters depends on whether the characters are indexed. If a character is not indexed, then it will, typically, be replaced by a space during indexing. Such a character will also not be searchable and will be replaced by a space when included in a search.

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>If Not Indexed:</td>
</tr>
<tr>
<td>If # is not indexed</td>
</tr>
</tbody>
</table>

In this example, this search will find all documents that originally contained #revenue and find all the documents containing revenue on its own or associated with other non-indexed characters, such as (revenue).

Why does Clearwell remove punctuation?

Clearwell removes punctuation characters in this way in order to improve search results. Without this, keyword searches may not find documents in which a word (instead of punctuation) occurred at the end of a sentence, or was enclosed in parentheses. However, it is possible to adjust how Clearwell indexes punctuation characters. Refer to Appendix A for more information.

If a search contains a term with a punctuation character that has been indexed, only documents containing the term that includes the punctuation character will be found.

<table>
<thead>
<tr>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search:</td>
</tr>
<tr>
<td>Documents containing 10% (and % is indexed)</td>
</tr>
</tbody>
</table>
Can Clearwell index more than one version of a term?

In order to maximize searchability, Clearwell will sometimes index two versions of a term: one with punctuation characters indexed and one or more terms in which punctuation characters are removed. This makes it possible to construct searches to find documents containing these characters, if desired.

Example

<table>
<thead>
<tr>
<th>If document contains term:</th>
<th>Then search for either:</th>
</tr>
</thead>
<tbody>
<tr>
<td>risk/reward</td>
<td>A. Search: risk</td>
</tr>
<tr>
<td>(and / is a searchable</td>
<td>then search for:</td>
</tr>
<tr>
<td>and non-searchable</td>
<td>reward</td>
</tr>
<tr>
<td>character – both indexed</td>
<td>B. Search: &lt;risk/reward&gt;</td>
</tr>
<tr>
<td>and not indexed)</td>
<td>(finds only documents</td>
</tr>
<tr>
<td></td>
<td>that precisely contain</td>
</tr>
<tr>
<td></td>
<td>this term)</td>
</tr>
</tbody>
</table>

Note: The use of the “less than” and “greater than” signs <> alerts Clearwell to not remove any punctuation characters when searching the index. It is possible to modify which characters will have this behavior. Clearwell indexes email addresses and numbers in a similar manner. The original address or number containing the term will be indexed and the terms after removal of punctuation characters will also be indexed. (See Appendix A for more information.)

How does Clearwell use punctuation as query syntax?

Clearwell’s search engine uses certain punctuation characters as part of the syntax for constructing search queries. For example, parentheses are used to group terms, and quotes are used for phrase and proximity searches. As a result, searching for these punctuation characters requires instructing Clearwell to not use these characters as part of the query syntax but instead to search for these characters. There are two ways to instruct Clearwell to search for these characters:

- Use the back slash (\) escape character: For example, to search for +10 use \\+10
- Use quotes: For example, to search for +10, use “+10”.
- Note: Use this only if the phrase to be searched for does not contain quotes itself.

The following characters need to be escaped or quoted: + · & \ | ! ( ) { } [ ] ^ ~ : ".

Wildcard characters * and ? are not currently searchable.
## Search Examples

### Leading wildcard searches

<table>
<thead>
<tr>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for words containing &quot;inflate&quot;</td>
<td><em>inflate</em></td>
</tr>
</tbody>
</table>

### Proximity searches

<table>
<thead>
<tr>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for inflation and profit with 10 or fewer intervening terms in either direction</td>
<td>inflation w/10 profit &quot;inflation profit&quot;~10</td>
</tr>
</tbody>
</table>

### Proximity searches containing wildcards

<table>
<thead>
<tr>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for inflat* profit* with 10 intervening terms</td>
<td>inflat* w/10 profit* &quot;inflat* profit*&quot;~10</td>
</tr>
</tbody>
</table>

### Proximity searches containing exact phrases

<table>
<thead>
<tr>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for the exact phrase “stock option” with 2 intervening words between &quot;stock option&quot; and backdate</td>
<td>&quot;stock option&quot; w/2 backdate</td>
</tr>
</tbody>
</table>

### Nested proximity searches

<table>
<thead>
<tr>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find “inflate” within 5 terms of “profit” within 10 terms of “options” within 5 terms of “backdating”</td>
<td>(inflate w/5 profit) w/10 (options w/5 backdating)</td>
</tr>
</tbody>
</table>

### Proximity and NOT searches

<table>
<thead>
<tr>
<th>Example</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Search for “stock” except when there are 20 intervening words between stock and &quot;option&quot;</td>
<td>stock NOT w/20 option</td>
</tr>
<tr>
<td>Search for all documents not containing &quot;stock&quot; and “option” within 20 terms</td>
<td>NOT (stock w/20 option)</td>
</tr>
</tbody>
</table>
Frequently Asked Questions

Does Clearwell perform in-text character searches?

By default, Clearwell performs term based searching and does not perform an in-text character search. For example, searching for "ch" will not find the word "searches". If it is required to find in-text characters, a leading/trailing wildcard search, such as "*ch*", can be performed. You can also use Search Preview in order to analyze the results of these wildcard searches to evaluate which terms are relevant and which terms are not.

How do I know when to use a stemmed vs. literal search?

Clearwell provides the ability to search with both stemmed variations as well as literal, without requiring re-processing of data. The Basic Search field always performs stemmed searches. In the Advanced Search screen, you can choose whether to run a stemmed search or a literal search by using the Search All Variations of the Keyword Terms (Stemmed Search) checkbox.

Stemmed searches find variations of words such as plurals or alternative verb forms based on a set of linguistic rules. Wildcard searches will find all words that match the characters defined in the wildcard search, so a search for hir*, which is intended to find documents related to hiring, will find all documents containing words whose first three letters are hir. By finding variations of the specified keyword, both stemmed and wildcards searches can find more relevant documents containing these variations that otherwise might have been missed. However, each of these technologies have tradeoffs. In addition to finding more relevant documents, they can find non-relevant documents or false positives. For example, the search hir* might find documents containing the word "hirl", which could be someone's last name, which likely is not relevant to hiring.

In general, the use of stemming vs. wildcards depends on a cost-benefit analysis that weighs the value of finding more relevant documents versus the cost of finding more false positives. Wildcard searches will tend to find more relevant documents but also more false positive documents. Stemmed searches have been designed to find fewer false positives, but they may not find some relevant documents that a wildcard search might find. For example, stemmed searches will typically not find misspelled words that wildcard searches might find. With Clearwell’s Transparent search, you can dramatically reduce the number of false positive documents by excluding irrelevant variations in wildcard or stemmed searches. Users should choose the search method that best matches their search objectives.

How do I search for all emails to or from another person and perform privilege searches containing names?

Searches for email to or from people can be conducted using the sender and recipient fields within advanced search. As part of this approach, the participant picker (which can be accessed by clicking on the icon to the right of these fields) can be used to identify the participants whose emails you wish to find by using searches like *[lastname]* or *{firstname}* In Clearwell, a participant is a unique email address and/or display name.
With a search for potentially privileged documents, it is typically necessary to find emails or files that reference the designated people, such as attorney names, anywhere within a document not just the sender and recipient fields. In these situations, it is recommended to use the Search Preview to identify all the terms that contain part or all of the person’s name anywhere within the document in addition to running a search using the participant picker.
Appendix A – Treatment of Punctuations for Cases Started with or after V4.5

- The treatment of punctuation characters has changed in version 4.5 as part of the addition of multiple language support. For cases started in version 4.0, punctuation characters will be handled as they were in version 4.0. Please refer to Appendix B for information on this behavior.

- This Appendix covers how punctuation characters are treated for characters in the Latin script. For information on the treatment of characters written in other scripts including Chinese, Japanese and Korean, please refer to the Clearwell Multiple Language Guide. All of the following rules apply to the Keywords, Email – Subject, and File/Attachment – Any Of The Words fields within Advanced search.

- Punctuation characters are treated differently for characters written in Latin scripts and characters written in other scripts. Clearwell will always split words when they contain characters in more than one script when the script change occurs.

- For most terms, Clearwell will treat punctuation characters in four different ways.
  - **Searchable characters** are indexed as-is during processing and can be searched within Clearwell.
  - **Non-searchable characters** (also referred to as Delimiters) are treated as spaces during processing and will normally be removed from search queries.
  - **Trim characters** are removed if they are the first or last character of a term.
  - **Searchable and non-searchable characters** are both indexed and treated as spaces during indexing. These characters will normally be removed from search queries but can be searched for by surrounding a search term with less than and greater than signs <>.

- During indexing, for most terms, Clearwell will find an original term, remove trim characters and treat non-searchable characters as spaces and index the resulting token or tokens.
  - For example, the terms *The quick, brown fox.* will be indexed as: *the quick brown fox*
  - The comma and period are removed because these comma and period are non-searchable characters.

- Terms containing searchable and non-searchable characters however will be indexed multiple times.
  - For example, the term *well-received* will be indexed with the following tokens: *well received well-received*

- Clearwell treats certain terms including email addresses and terms containing numbers, differently from an indexing and search perspective in order to maximize the searchable information contained within these terms.
Email addresses will be indexed in multiple ways. For example: The email address jdoe@sales.company.com will be indexed into the following tokens: jdoe@sales.company.com jdoe company.com sales.company.com

Terms containing numbers will also be indexed multiple times. When indexing terms containing numbers, Clearwell will first trim the original term using the characters designated as Trim characters and index the resulting token. Clearwell will then re-index the original term using the searchable, non-searchable, trim and searchable/non-searchable rules described above. Here are some examples using the default character designations described below.

The term 123.45.6789 will be indexed into the following tokens: 123.45.6789 123 45 6789

The term 123.45.6789 will also be indexed into the same tokens as the comma will be trimmed: 123.45.6789 123 45 6789

As described in the punctuation search section, angle brackets should be used to search for email addresses or numbers that otherwise would not be searchable. For example to search for social security numbers that use hyphens, use the following searches: <???-??-????> Without enclosing this in the angle brackets, Clearwell will interpret this search as ??? OR ?? OR ????.

The following characters cause content to be indexed two ways. Words on either side of the character are indexed both separately and as a compound.

Hyphens ( - , ‐ ) and En dash ( – )

Forward and back slashes ( / \ )
• The following characters are non-searchable:
  o Colon (:) and semi-colon (;)
  o Figure Dash (‐), Em dash (—), horizontal bar (—) and Non-breaking hyphen (‐)
  o Exclamation marks (!) and question marks (?)
  o Parentheses and brackets ( () [] {} ﹙﹚ ﹛﹜ ﹝﹞ )
  o Single Quotes, Double Quotes or guillemets or angle brackets ( ’ ” ‘ ” ‘ ” )
  o Less than or greater than signs (< >)
  o Inverted exclamation marks (¡) and inverted question marks (¿)
  o Generic currency marks (¤)
  o Interpuncts (·) and bullets (• · • ‧ • ‧ ‧)
  o Ellipses (…)
  o Daggers († ‡)
  o Asterisk (*)
  o Vertical pipes (| ; ||)
  o Equals sign (=)
  o Pilcrow (¶)
  o Number sign (#)
  o Underscore (_)
  o At signs (@ @ ®)
  o Greek semi-colon and tonos (; ‹ ‼)

• The following characters are Trim characters
  o Apostrophe (’)
  o Ampersand (&)
  o Period (.) and Comma (,)
  o Colon (:) and semi-colon (;)
  o Figure Dash (‐), Em dash (—), horizontal bar (—) and Non-breaking hyphen (‐)
  o Exclamation marks (!) and question marks (?)
  o Parentheses and brackets ( () [] {} ﹙﹚ ﹛﹜ ﹝﹞ )
• Less than or greater than signs (< >)
• Hyphens (‐ ‐ ) and En dash ( – )
• Forward and back slashes (/ \)
• Inverted exclamation marks ( ¡ ) and inverted question marks ( ¿ )
• Interpuncts ( · ) and bullets ( • • • • • • )
• Ellipses ( … )
• Asterisk ( * )
• Vertical pipes ( | ||)
• Equals sign ( = )
• Pilcrow ( ¶ )
• Underscore ( _ )
• Greek semi-colon and tonos ( ; ΄ ΅ · )
• Fullwidth & small versions of commas, exclamation points, period, colon, semi-colons, quotes, reverse solidus’ ( ﹐ ﹑ ﹒ ﹓ ﹔ ﹕ ﹔ ﹐ ? ! " , . / \ )

• All other punctuation characters are searchable. Some examples of these include the following:
  • Percent ( % % )
  • Currency ( $ ¥ £ € etc.)
  • Section mark ( § )
  • Tilde ( ~ )
  • Mathematical symbols such as the plus sign ( + ), division slash (/), and minus sign ( − )

• It is possible to change how characters are treated within Clearwell. Please contact Clearwell Support for more information. If you have a significant number of documents containing foreign language documents, you may want to consider changing some of the character treatment. For example, you may want to consider changing the treatment of apostrophes for cases containing significant amounts of French or Spanish documents.
Appendix B – Treatment of Punctuations for Cases Started Prior to V4.5

- For cases started prior to version 4.5, when searching any fields via the Keywords, Email – Subject, and File/Attachment – Any Of The Words fields, Clearwell will treat punctuation characters as spaces except in the following cases:

<table>
<thead>
<tr>
<th>Cases</th>
<th>Indexed punctuation characters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word without numbers in email and file content</td>
<td>Period when not followed by whitespace ( . )</td>
</tr>
<tr>
<td></td>
<td>At symbol ( @ )</td>
</tr>
<tr>
<td></td>
<td>Apostrophe ( ’ )</td>
</tr>
<tr>
<td></td>
<td>Ampersand ( &amp; )</td>
</tr>
<tr>
<td>Words containing numbers in email and file content</td>
<td>Period ( . )</td>
</tr>
<tr>
<td></td>
<td>Hyphen ( - )</td>
</tr>
<tr>
<td></td>
<td>Forward slash ( / )</td>
</tr>
<tr>
<td></td>
<td>Underscore ( _ )</td>
</tr>
<tr>
<td></td>
<td>Comma ( , )</td>
</tr>
</tbody>
</table>

- When searching the To, From, cc, bcc fields of email via the Any of These Senders or the Any of These Recipients fields, most punctuation and special characters are indexed and not ignored.

- When searching for filenames via the Any of These File Names or Extensions, the following characters will not be indexed and will be treated as spaces. All other punctuation characters will be indexed.
  - Period ( . )
  - Forward and back slashes ( / \ )
  - Hyphens ( - )
  - Underscores ( _ )
  - Commas ( , )
  - Semi-colons ( ; )
  - Quotes ( ’ ”)‘
  - Asterisks ( * )
  - Question marks ( ? )
  - Pipes ( | )
  - Brackets ( [ ] )
Appendix C - Stop Words for Cases Started Prior to V4.5

- In cases started prior to version 4.5, Clearwell did not index stop words. As of 4.5 and beyond, all words are indexed.

- Stop words are ignored in all searches except for phrase searches. For example, the search "the energy policy" will search for documents containing "the", "energy", and "policy" in that order. Stop words are not supported within proximity queries. For example, you cannot search for “the energy policy”~10. The word “the” will be ignored in the search and should be removed. Note also that stop words in the documents that are being searched are counted as intervening words in proximity searches.

- The following default stop words apply to versions prior to 4.5:

  a came him much still way
about can himself must such we
after come how my take well
all could however never than were
also did i not that what
an do if now the when
and each in of their where
another even indeed on them which
any for into only then while
are from is or there who
as further it other therefore will
at furthermore its our they with
be get just out this would
because got like over those you
been had made said through your
before has many same thus
being have me see to
between he might she too
both her more should under
but here moreover since up
by hi most some was