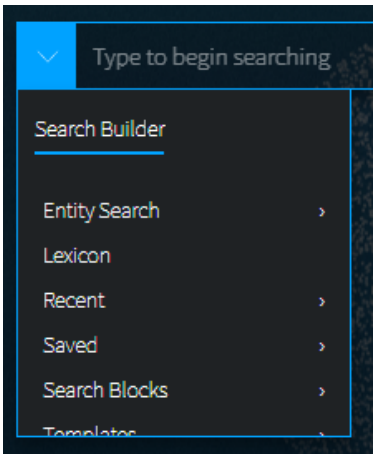
Searching in DarkOwl Vision

Using the Search Bar

1. The search bar works like most search engines; simply type words, phrases, numbers, or characters. The Search Techniques section goes into more detail and options for searching, but here a few quick start tips:
    - Use Boolean operators when searching more than one keyword. See: Searching with Booleans.
    - Use quotations (**"Jane Doe"**) to send the query as one phrase, *Jane Doe*.
    - Use the **exact:** search operator (**exact:fullz**) to prevent word stemming, and search for exact matches of that term. See: Stemming and Searching for Exact Terms.
2. Use the left drop-down menu to open **Search Builders**, tools to help you create queries.



Entity Search is the best way to search for Emails, Credit Cards, Cryptocurrency Addresses, IP Addresses, and Social Security Numbers in our system. You can also create Search Blocks from these builders.

The **Lexicon** includes curated lists of commonly searched items.

Templates are pre-populated search templates to help you get started quickly.

Pre-built **Search Blocks,** as well as any custom search blocks that you create, are accessible from this menu for easy access.

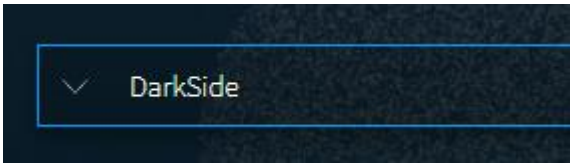
Once you've started searching, **Recent Searches** and **Saved Searches** will also appear in this menu, for easy access.

3. Click on the right Filter menu to refine your search, create targeted queries, and select options. More information about these is provided in this section: Filters, Dates, Advanced Search Options.

Search Techniques

Single Terms

To find any document containing a specific keyword, simply put that keyword into the search bar.



Phrase Searching

To find two or more keywords in a specific order, place the keywords within double quotes:

- **"AES 256"**

Without the double quotes, the search would be sent as: *AES OR 256.* This search is an inclusive search and may return results that only include the term AES, that only include the term 256, that include both terms but not next to each other, and that include both terms next to each other.

Boolean Searching

Use Boolean operators **AND, OR,** and **NOT** to specify inclusions, alternate terms, or exclusions. (You can substitute AND, OR, and NOT with **&&, ||,** and **!** respectively.) Keywords and field searches can be separated by any of the above in order to fine-tune your results.

- **drugs OR crime –** find documents with either 'drugs' or 'crime'
- **drugs AND crime –** find documents with both 'drugs' and 'crime'
- **DarkOwl AND (hack OR ddos OR 0day) –** find documents with DarkOwl and any one of three hacking keywords

Note: Boolean operators must be in all caps. If they aren't in all caps, DarkOwl Vision will consider the word "and", "or", etc as keywords.

Using Subqueries/Boolean Order of Operations

You can group together phrases to form subqueries, using parentheses ( ) to indicate each clause. This is especially important when both ANDs and ORs are used, to designate the correct order of operations for your subqueries.

- **DarkOwl AND (drugs OR crime)** – find documents with DarkOwl and either drugs or crime
- **("AES-256" OR "AES 256") AND ("RSA-4096" OR "RSA 4096") –** find documents with one of AES-256 or AES 256 and one of RSA-4096 or RSA 4096

Stemming and Searching for Exact Terms

DarkOwl Vision supports a process called *stemming*, which tries to reduce a word to an approximation of its stem or root form. Usually, terms are stemmed to plural/singular versions or different tenses. This means that searching will return matches on related forms of a word, unless you specify otherwise, for instance:

- **Hack** may return **Hacked, Hacker, Hacking,** in addition to **Hack**

When you want to search for a specific term, including special characters and punctuation, use the **exact:** operator to prevent word stemming:

- **exact:hack –** will return only documents containing the word hack

Excluding Items from a Search

Keywords can be excluded in multiple ways:

- Using the 'NOT' Boolean operator
- Prefacing the term with a hyphen
- Prefacing the term with an exclamation mark

For example, the three searches below are equivalent and will find documents that contain 'DarkOwl' but not 'drugs'. Note that when excluding a keyword via hyphen or exclamation mark, it must be placed directly before the keyword with no space in between.

- DarkOwl NOT drugs
- DarkOwl -drugs
- DarkOwl !drugs

You can also exclude values in other fields in the same way:

- DarkOwl NOT domain:drugs.onion
- DarkOwl -domain:drugs.onion
- DarkOwl !domain:drugs.onion

Searching for Entities in the Search Bar

When searching for specific entities, such as an email address or credit card number, directly in the search bar, always use the designated search operator. The Search Builders (in the drop-down on the left-side of the search bar) are a shortcut and automatically convert your query to the correct syntax.

- email:first.last@company.com
- ccn:1111111111111111
- cryptocurrency:15ivMrk8VzaK9TEN85XYssVbU3Yd6tLzb9
- ipAddress:127.0.0.1\/24
- ssn:123-45-6789

When searching for multiple entities, use the search operator and a Boolean OR, as follows:

- email:(first.last@company.com OR last.first@company.com)
- ccn:(1111111111111111 OR 2222222222222222)

Searching for both Keywords and Entities in the Search Bar

When searching for both keywords and specific entities (such as an email address or credit card number) directly in the search bar, use the following format:

- ("First Last" OR Nickname) AND email:first.last@company.com

Using Wildcards

Wildcards (***** or **?**) are currently allowed *in limited usage*, in the middle or end of terms only. (*****) is used to find *zero or more* unknown characters; (**?**) is used to find *any one* unknown character. Examples:

- **dar*** – will find "dar", "dart", "darkowl", "daredevil", etc
- **d?rk** – will find "dark", "dork", "dirk", etc; will not find "drk" (however, **d*rk** would)

DarkOwl Vision does not support *leading wildcards*. In other words, a search term cannot begin with either one of the wildcard characters.

Using Proximity Searches

You can find words in proximity to each other by using quotations and selecting a maximum distance allowed: **"password hack"~2**. We support a maximum distance of 9.

Using Pattern Matching / Regular Expressions

Lucene-based regular expressions are allowed and should be wrapped by forward slashes (/). Not all functionality you may be familiar with may be supported. Additionally:

- ***These queries may time out***, particularly when searching for a high volume of unknown characters. Regex searching is computationally heavy and will result in slower, less performant searches.

To use a regular expression in Vision, place the expression between two forward slash characters:

- **/r[0-9a-zA-Z]{24,34}/** – to find results matching the pattern of a Ripple cryptocurrency address (which starts with 'r', then has anywhere between 24 to 34 alphanumeric characters)

Note: Not all regex functionality you may be used to is supported.

Using Special Characters

The following characters are reserved:

+ - = && | | > < ! ( ) { } [ ] ^ " ~ * ? : \ /

If any of the above characters are in a keyword or phrase being searched, you can escape the character with a backslash: \. For example, to search for mentions of a URL within a document, such as *https://darkowl.com/darkint-blog*, you must escape the colon, forward slashes, and hyphen, otherwise the search will return an error.

You can perform this search multiple ways:

1. Escaping the special characters: **https\:\/\/darkowl.com\/darkint-blog**
2. Putting the whole keyword in quotes: **"https://darkowl.com/darkint-blog"**

Without escaping the special characters, this search will be interpreted as:

- Searching within a field called 'https' (which doesn't exist) for:
    - An empty regular expression (// signifies the start and end of a regex with no content)
    - The keyword 'darkowl.com'

- The start of a regular expression starting with 'darkint-blog'
- No end to the regular expression (will return an error)

Field Searching (Search Operators for Metadata Searches)

Every search performed will look in one or more fields for the keyword(s) being searched. By default, the search bar will search both the 'title' and 'body' fields of documents. This means that results will be returned if the keywords you're looking for are found in either the body of the document or the title (or both). For example, a search of just **the word '*drugs*' in the search bar** is equivalent to: ***title:drugs OR body:drugs***.

Most searches will not require specifying a field name, since title and body are automatically searched. However, other metadata fields can be searched in addition to title and body, for example:

- **title:alphabay**
- **hackishness:1**
- **domain:drugs.onion**

The list of metadata fields is below. When searching within these fields, type the following search operators in the search bar, and then the query content:

- inUrl:
- contentType:
- headers.server:
- headers.last-modified:
- title: (to search within this field exclusively)
- domain:
- leak:
- network:
- hackishness:

Multiple values within the same field can be searched in a number of ways. The following examples are equivalent:

- **domain:(drugs.onion OR crime.onion)**
- **domain:drugs.onion OR domain:crime.onion**

You can also look for phrases within specific fields using double quotes:

- **title:"Forum rules"**

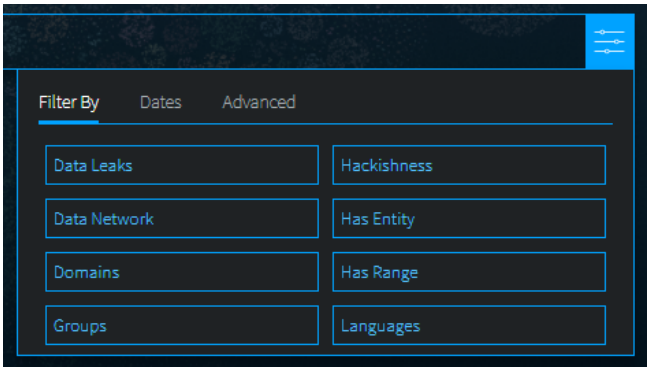
Subqueries within fields are supported:

- **title:(darkowl AND (drugs OR crime))**

Note: when searching using fields, there should not be a space after the ':' character.

Filters, Dates, and Advanced Search Options

Refine your search with various filter and date options; click on the Filters icon, on the right side of the search bar, to get started.



Data Leaks

Filter to content from known breaches or other leak data, tagged by name in DarkOwl Vision. See the Lexicon for descriptions of the Data Leaks in this list. Type or scroll to find the desired Data Leak. ***Click once to include, click twice to exclude*.**

- For free text searching, type *leak:leakname* in the search bar, prefixing with - to exclude.

Data Network

Filter to content from a particular DarkOwl data collection network. Options include: Discord, FTP, I2P, IRC, Onion, OpenNIC, Telegram, Zeronet. ***Click once to include, click twice to exclude*.** More information: Data Networks.

- For free text searching, type *network:networkname* in the search bar, prefixing with - to exclude.

Domains

Filter to content from one or more domains, or exclude a particular domain by typing a hyphen in front of the domain. Type only the domain portion (such as, **arch3rsecgjqcmjb.onion**; no need for the www or http:// prefix) in the filter box. Be sure to remove any trailing slashes or paths from the domain.

- For free text searching, type *domain:domain.com* in the search bar, prefixing with - to exclude.

Groups

Groups are combined filters that narrow your search to specific categories; ***click to include*:**

- *Authenticated Sites:* Filter to content from sites requiring credentials or other challenges.
- *Blogs:* Filter to content from sites identified as blogs.
- *Chans:* Filter to content from a curated set of chan/imageboard forums selected by our analysts.
- *Darknet:* Filter to content from the Tor, I2P, and Zeronet darknets.
- *Forums:* Filter to content from sites identified as forums.
- *Paste Sites:* Filter to content from a curated set of paste sites selected by our analysts.

Hackishness

Hackishness assigns a rating to every piece of content collected, indicating the likelihood to which the information could be used for criminal activity. The lower bound of hackishness is .01 and the upper bound is 1.0; the UI shows these as percentages on search results. You can quickly filter to results with hackishness by ***using the slider*** on the Hackishness filter to select a desired hackishness threshold.

You can also filter to hackish results using hackishness: in the search bar, which supports searching as range. This means, you can narrow down to values between two parameters, *inclusive or exclusive*, for example:

- **hackishness:[.01 TO 1]**
- **hackishness:{.01 TO 1}**

Note the '[' and '{' characters used above. In Lucene range queries, '[' and ']' are inclusive so the first query above would return values from .01 to 1, including both .01 and 1. The second example would return values from .01 to 1 not including .01 or 1. '[' and '{' can be combined:

- **hackishness:{.5 TO 1]**

The above will find values greater than .5 and up to and including 1.

Has Entity (Credit Cards, Cryptocurrencies, Email, IPs, Social Security Numbers)

Filter to content that have at least one selected Entity. ***Click next to the Entity name to select***.

Has Range (Credit Cards, Cryptocurrencies, Email, IPs, Social Security Numbers)

Filter to content that have a certain number of selected Entities. This filter is helpful in finding "dumps," as many threat actors will post multiple instances of PII on a singular site or document. ***Type values next to a selected Entity***: Enter a lower bound (minimum 1), upper bound (maximum 999999), or use both fields to form a range (50 to 1000).

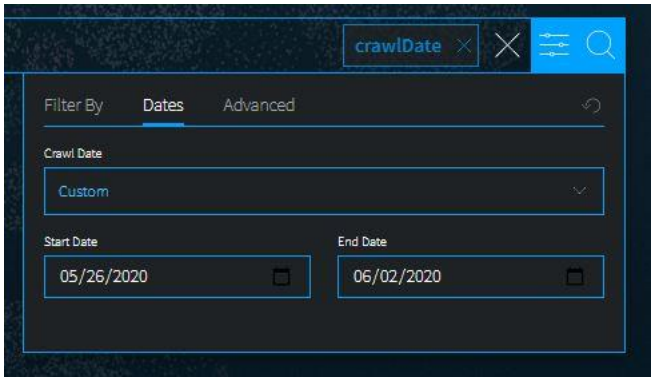
Language

Filter to content in a particular language. Languages are detected by DarkOwl Vision at the time of ingestion, using natural language processing. ***Click once to include***.

- For free text searching, type *language:languagevalue* in the search bar, prefixing with - to exclude.

Dates

Use the drop-down to quickly select a time range for search results, or select *Custom* to choose a specific start/end date.



Crawl dates can be searched in the search bar as a range using the following date format: YYYY-MM-DDTHH-MM-SSZ:

- **crawlDate:[2021-07-01T00:00:00Z TO 2021-07-10T23:59:59Z]**

As crawlDate supports range searching, you can narrow down to values between two parameters, *inclusive or exclusive*, as discussed in the hackishness section. In Lucene range queries, '[' and ']' are inclusive, and '{' and '}' are exclusive.

Advanced Options (Sort By, Show Similar, Empty Bodies)

Use Advanced Options to select a Sort option, or to show all results (including duplicates).

- **Sort options.** Use the drop-down to sort your results by ***Relevance*** (default), By ***Hackishness***, By ***Crawl Date***
- **De-duplicate your results.** You can choose to de-duplicate your result sets or see all results, including similar results. The default is to de-duplicate result sets; you can toggle this on or off.
- **Empty bodies.** Our collection includes documents collected that do not contain any text characters; text content is stored in the Body field. Use the drop-down to select Any document (documents can include text in the body, or no text in the body); Results must have body field (to only return documents that had text content on them); or Results must not have body (to see only documents without text).