

SimSage Rules Engine

Every SimSage source has an associated rules engine. This Rules Engine enables the creation or removal of hashTags to documents/records.

Admin UI

The rules engine provides an interface for adding simple logic. Navigate to your sources in the admin UI, and create a new source, or edit an existing source.

Edit Source: rock-test

General Web Crawler Metadata ACLs MimeTypes **Rules** Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

rule description #hashtag Delete

AND + Rule + Group

Close Test Save

the rules tab

The rules tab (shown above) shows the rule engine editor page. To start, click + Add Rule.

You can have multiple rules per source. Each rule is responsible for adding or removing a single hashTag to or from a document. Each rule needs a description (rule description), a hashTag, and a set of logic rules.

Edit Source: rock-test

General Web Crawler Metadata ACLs MimeTypes **Rules** Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret #topsecret Delete

AND AND OR

+ Rule + Group

Close Test Save

Top Secret classification rule example

We can add a series of logic rules that are either AND (all rules must succeed for success) or OR (any of the rules must succeed for success).

Proceed by clicking the + Rule button next to the AND/OR drop down box. This adds a single expression line for creating a rule.

Edit Source: rock-test

General Web Crawler Metadata ACLs MimeTypes **Rules** Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret #topsecret Delete

OR

Document Content contains word

Document Content
metadata: Title
metadata: URL
metadata: Author
Last Modified date
Created date

+ Rule + Group

Close Test Save

the first rule

We get to select what sort of a rule we want. The options are:

Operator	Description
Document Content	The rule looks at the body content of a document / record
metadata: Title	The rule looks at the title of a document / record (NB. not all documents have titles)

metadata: URL	The rule looks at the URL of a document / record (usually its path in a file-system or its http address depending on the type of source)
metadata: Author	The rule looks at the author of a document / record (NB. not all documents have authors)
Last modified date	The rule looks at the date-time stamp of the document / record's last modified date
Created date	The rule looks at the date-time stamp of the document / record's creation date

Selecting any of the Document Content / metadata: values brings up a set of text operators.

Edit Source: rock-test

General

Web Crawler

Metadata

ACLs

MimeTypes

Rules

Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret

#topsecret

Delete

OR

+ Rule

+ Group

Document Content

contains word

contains word

starts with

ends with

is exactly

does not contain word

does not start with

does not end with

is not equal to

Close

Test

Save

text operators

These text operators are:

Text operator	Description
contains word	The select field (see above) contains the word you're about to supply. Lowercase words match words of mixed case (e.g., "dog" will match with "dog" and "Dog"). Mixed case words will only match themselves (e.g, "Dog" will only match "Dog", and not "dog", or "DOG").

starts with	The selected field starts with exactly the word you're about to supply. The case sensitivity rules are the same as above.
ends with	The selected field ends with exactly the word you're about to supply. The case sensitivity rules are the same as above.
is exactly	The selected field must equal the word you're about to supply. The case sensitivity rules are the same as above.
does not contain word	The opposite of contains word. This value inverts the logical meaning and succeeds if the word is absent.
does not start with	The opposite of starts with
does not end with	The opposite of ends with
is not equal to	The opposite of is exactly

Edit Source: rock-test

General Web Crawler Metadata ACLs MimeTypes **Rules** Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret #topsecret Delete

OR

Document Content contains word top secret

Close Test Save

our first rule

We've just added our first rule. It reads:

If the document's content contains the case insensitive word "top secret" anywhere, then add a new hashTag to the document "#topsecret" (or remove that hashTag if it is present and this rule does not succeed).

Now, suppose that we also want to check the title, and perhaps the folder where the document is stored for similar values. Click +Rule again to add more options.

Edit Source: rock-test

General

Web Crawler

Metadata

ACLs

MimeTypes

Rules

Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret

#topsecret

Delete

OR

+ Rule

+ Group

Document Content

contains word

top secret

metadata: Title

contains word

top secret

metadata: URL

contains word

top-secret

Close

Test

Save

top secret classification looking at various part of a document / record

We can add many more rules. Let’s continue our example, and add another rule that adds a #deprecated hashTag to each document that is over 2 years old. Click + Add Rule at the top.

Edit Source: rock-test

General

Web Crawler

Metadata

ACLs

MimeTypes

Rules

Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret

#topsecret

Delete

OR

+ Rule

+ Group

Document Content

contains word

top secret

metadata: Title

contains word

top secret

metadata: URL

more than this many days old

top-secret

determine if a document is too old

Delete

OR

+ Rule

+ Group

Last Modified date

more than this many days old

Close

Test

Save

a new rule for document / record age tagging

Note that the operators for Last Modified date and Created date are different. This is because we’re dealing with dates instead of text. The options are:

Date operator	Description
---------------	-------------

more than this many days old	The value about to be supplied is a number of days. Any document older than this many days will be tagged
more than this many weeks old	The value is a number of weeks. Any document older than this many weeks will be tagged
more than this many months old	The value is a number of months. A month in this context is always take to be a 30 day period. Any document older than this many months will be tagged
more than this many years old	The value is a number of years. A year is always taken as 365 days. Any document older than this many years will be tagged
less than this many days old	These are identical to their “more than this many...” operators, but take the document to be newer than the stated value.
less than this many weeks old	These are identical to their “more than this many...” operators, but take the document to be newer than the stated value.
less than this many months old	These are identical to their “more than this many...” operators, but take the document to be newer than the stated value.
less than this many years old	These are identical to their “more than this many...” operators, but take the document to be newer than the stated value.

Edit Source: rock-test

General Web Crawler Metadata ACLs MimeTypes **Rules** Check Document

Schedule

HashTags start with #, one hashtag per rule.
Days, weeks, months, and years must be numbers.
Word values must be small, not sentences (~ 100 characters).
Lowercase word values match all cases, mixed-case words only match exact.

+ Add Rule

determine if a document is Top Secret #topsecret Delete

OR

Document Content	contains word	top secret	
metadata: Title	contains word	top secret	
metadata: URL	contains word	top-secret	

determine if a document is too old #deprecated Delete

OR

Last Modified date	more than this many years old	2	
Created date	more than this many years old	2	

Close Test Save

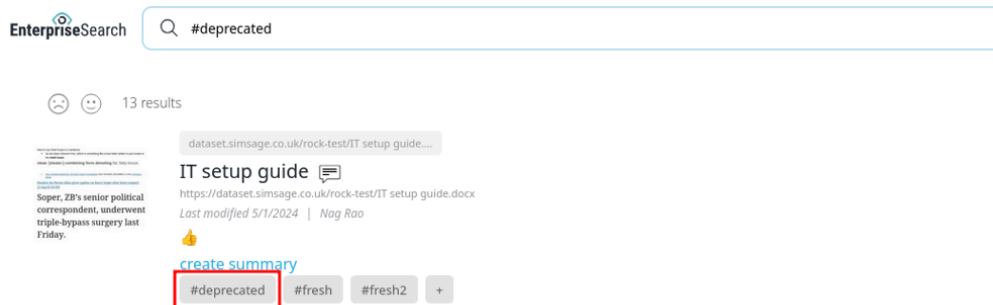
A source with two rules

We've just set up an entirely new rule in addition to our top secret classifier. This rule adds a #deprecated to each document where either the Last modified date, or Created date of that document is more than two years (more than 2 x 365 days).

Usage

You will need to re-process documents every-time you change "the rules". Changing the value of a hashTag for an existing rule will not remove the old hashTag / previous hashTags from documents already processed.

After re-processing, wait for SimSage to finish indexing, at which stage your hashTags become searchable.



example hashTag #deprecated search