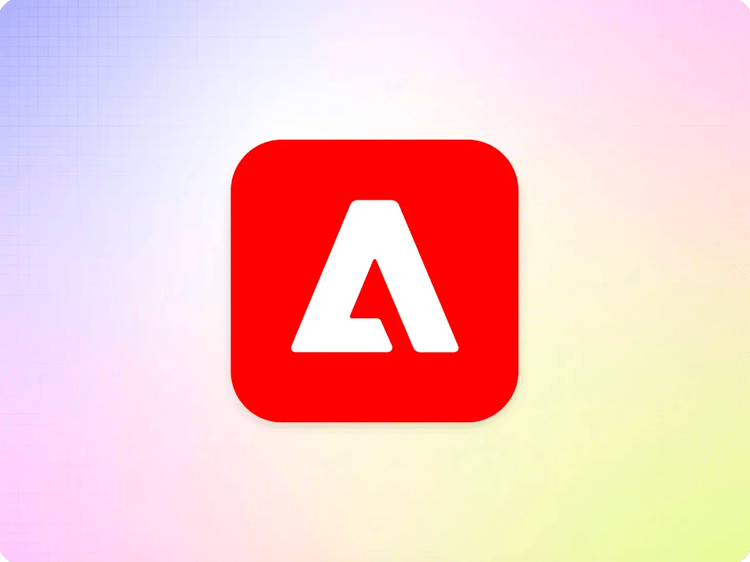
| Section Metadata | |
| --- | --- |
| style | content |



# Indexing reference

In your helix-query.yaml#, you can define one or more index definitions. A sample index definition looks as follows: <https://gist.github.com/dominique-pfister/92cb67b6f95e1edee6a7d6508b124039>

The include and exclude section dictates what documents get indexed. Everything that is included but not excluded gets indexed. The double asterisk \*\* matches everything under a prefix, **including** the prefix, so in the example above, the path /documents gets indexed as well. If you leave out that section entirely, *everything* gets indexed.

The select property is a CSS selector that grabs *all* matching HTML elements out of your document. The indexer will apply your selectors on the HTML markup, not on the rendered DOM, so you must write them accordingly. (right click -> View Page Source, on the page you want to extract information from, to see the exact HTML the indexer will work on).

If you just want the *first* matching element to be returned, use selectFirst instead of select:

first-img:  
 selectFirst: img  
 value: attribute(el, "src")

To verify that a CSS selector entered is selecting what you expect, you can use the [aem cli](/developer/indexing#debugging-your-index-configuration) - aem up -–print-index, navigate to the page where the selector should extract a meaningful value and check the console. The cli will use the helix-query.yaml file from your local filesystem and will print the extracted values, or an empty string if it cannot find the information it is looking for.

aem up --print-index  
...  
info: Index information for /my/test/page  
info: Index: mysite  
info: author: "John Smith"

Note, that not all CSS selectors are supported. Internally, we use a library called hast-util-select, and the list of supported selectors can be found here: <https://github.com/syntax-tree/hast-util-select#support>

The value or values property contains an expression to apply to all HTML elements selected. The property name value is preferred when you need a string, values on the other hand provides you with an array of all the matches found. The expression can contain a combination of functions and variables:

## innerHTML(el)

Returns the HTML content of an element.

## textContent(el)

Returns the text content of the selected element, and all its descendents.

## attribute(el, name)

Returns the value of the attribute with the specified name of an element.

## match(el, re)

Matches a regular expression containing parentheses to capture items in the passed element. In the author example above, the actual contents of the <p> element selected might contain by John Smith, so it would capture everything following by .

## characters(el, start, end)

Returns the substring from start to end of the given element or text. If start or end are negative, they address the position counted from the end of the text. end is optional, and defaults to the length of the text.

## words(el, start, end)

Useful for teasers, this selects a range of words out of an HTML element.

## replace(el, substr, newSubstr)

Replaces the first occurrence of a substring in a text with a replacement.

## replaceAll(el, substr, newSubstr)

Replaces all occurrences of a substring in a text with a replacement.

## parseTimestamp(el, format)

Parses a timestamp given as string in a [custom format](https://day.js.org/docs/en/parse/string-format#list-of-all-available-parsing-tokens), and returns its value as number of seconds since 1 Jan 1970.

## dateValue(el, format)

Parses a timestamp given as string, and returns its value as serial number, where January 1, 1900 is serial number 1. For more information see [DATEVALUE function](https://support.microsoft.com/en-gb/office/datevalue-function-df8b07d4-7761-4a93-bc33-b7471bbff252)

## el

Returns the HTML elements selected by the select property.

## path

Returns the path of the HTML document being indexed.

## headers[name]

Returns the value of the HTTP response header with the specified name, at the time the HTML document was fetched.

| Footnotes |
| --- |
| The full definition of the helix-query.yaml is available here: <https://github.com/adobe/helix-shared/blob/main/docs/indexconfig.md> |

| Metadata | |
| --- | --- |
| name | footnotes |