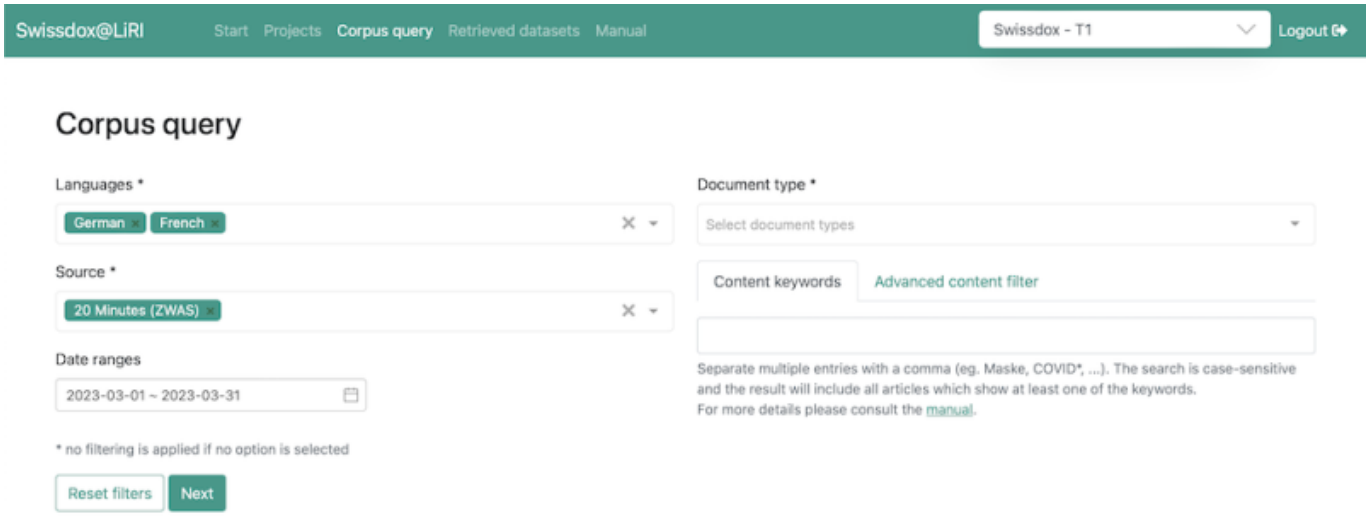
Searching the corpus

On the **Corpus query** page, you define a query according to your particular research interests. The query comprises of filters for various attributes as shown in the image below.



The following attributes are available.

Field	Description
Languages	Most articles are available in German and French, fewer articles are available in Italian, Romansh and English.
Document type	PDLN identifiers describe the type of the respective source.
Source	Media articles come from a large list of sources.
Time intervals	Most available articles date from the last 25 years, but some date back to the beginning of the last century.

Content keywords are a comma-separated list of terms provided by the user, of which at least one must be present in an article. Keywords are case-sensitive and asterisks (*) can be used as placeholders in words.

- `finden` will match the exact word (e.g. *finden*, but **not** *auffinden* or *findig*)
- `find*` will match all words starting with *find* (e.g. *finden* and *findig*, but **not** match *auffinden* or *auffindbar*)
- `*finden` will match words ending with *finden* (e.g. *finden*, *auffinden*, but **not** *auffindbar* or *findig*)
- `ge* Partei*` will match a word starting with *ge*, followed immediatly by a word starting with *Partei* (e.g. *gesamtschweizerische Partei*, *gelernte Parteiparolen*, but **not** *gegenerische, neue Partei*)

The **Advanced content filter** (see below) allows you to build a logical tree structure with AND, OR and NOT operators. Keywords are case-sensitive and asterisks (*) can be used as placeholders in words.

The **Content keywords** search terms are only applied to the actual textual content of the article - this excludes textual elements like captions or legends, author names, table

content etc.

Simple search using content keywords

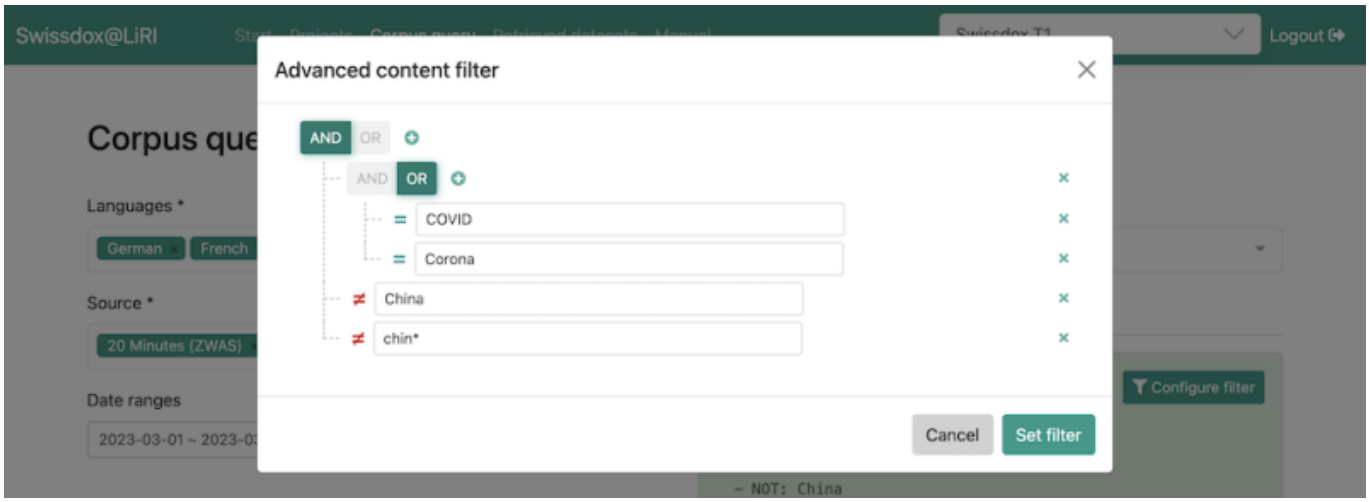
Assuming you are interested in articles in German that contain the word “Covid” and have been published in January 2022. In this scenario, you would set the filters in query builder like this:

Field	Value
Languages	German
Document type	*(leave blank)*
Source	*(leave blank)*
Content keywords	COVID
Date ranges	2022-01-01 ~ 2022-01-31

Note that if no option is selected, no filtering is performed. In this example field **Source** is left blank, so all sources will be taken into consideration. If you are interested in filtering articles using more keywords, multiple entries of keywords can be separated by commas (eg. aske, COVID*, ...). Search is case-sensitive and results will include articles with at least one of the keywords (OR operator). In combination with keywords, wildcards (*) can be used.

Advanced search using content filter

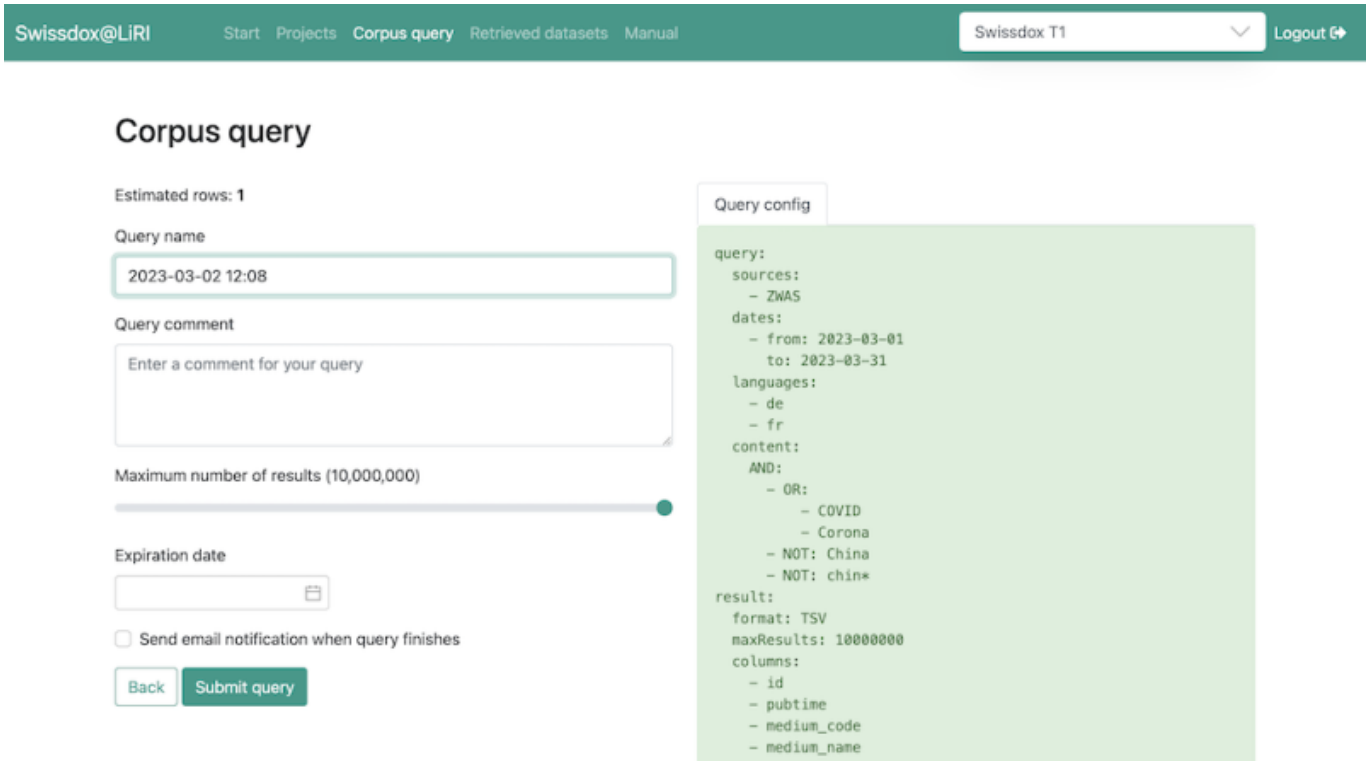
If you are interested in articles comprising the words “COVID” or “Corona”, but not mentioning “China” or “chinesisch” or any word form of that adjective, the **Advanced content filter** option allows you to use a combination of the logical operators `AND`, `OR` and `NOT`. The matching expression would look like this:



Submitting the query

On the next page, you can provide a meaningful name for your query to later be able to relate it to

the dataset that will be compiled.



If you want to receive a notification by email when a dataset is ready, check the option on this page. An expiry date will make the dataset disappear after the indicated date.

Datasets compiled this way can be downloaded from the application itself or retrieved via our API.