

Solr Search Reference

Simple queries

Queries are put to CKAN **package_search** using the “main query” (**q**) and “filter query” (**fq**) parameters. Paging of results can be done with **start** and **rows** parameters. The two types of query parameter are very similar. The intended usage is that the filter query (fq) filters the totality of the search index and caches it for further use in a main query. The filter query does not rank search returns by scoring, the main query does.

In the absence of a field specification the fields in a main query are weighted by "**name^4 title^4 tags^2 groups^2 text**" where **text** is a bin which contains all of the dataset and resource metadata not otherwise indexed.

Example:

https://data.humdata.org/api/action/package_search?fq=climada – this returns the 7 Climada datasets

Example:

https://data.humdata.org/api/action/package_search?fq=afghanistan&start=100&rows=100 – this returns 100 dataset records, starting at the 100th record relating to Afghanistan.

Querying metadata fields

Specific metadata fields can be selected in both types of query using a syntax like **fieldname:value**. Spaces are allowed in search strings but they must be escaped with a \ i.e. **dataset_source:ETH\ Zurich\ Climada** returns datasets with dataset_source set to

“ETH Zurich Climada”. The following characters all need escaping because they are Solr search operators: + - && | | ! () { } [] ^ " ~ * ? : \

Alternatively the search text can be quoted which removes the need for escaping:

Example:

`https://data.humdata.org/api/action/package_search?fq=dataset_source:"ETH Zurich Climada"`

Example:

`https://data.humdata.org/api/action/package_search?fq=dataset_source:ETH\ Zurich\ Climada` – this illustrates escaping spaces in a query

Only metadata fields that have been indexed by Solr can be included in a query specification; the list of indexed fields is shown in the following table.

archived	batch	caveats	creator_user_id
data_update_frequency	dataseries_name	dataset_preview	dataset_source
has_geodata	has_quickcharts	has_showcases	id
is_requestdata_type	isopen	last_modified*	license_id
license_title	maintainer	metadata_created*	metadata_modified*
methodology	name	notes	num_of_showcases
num_resources	num_tags	organization	overdue_date*
owner_org	package_creator	pageviews_last_14_days	qa_completed
res_description	res_extras_broken_link	res_extras_in_hapi	res_format
res_name	res_url	review_date*	solr_additions
state	subnational	title	total_res_downloads

type	updated_by_script	url	vocab_Topics
------	-------------------	-----	--------------

Table 1: Indexed Solr fields enabled for search

Those fields prefixed res_ refer to resource metadata, the rest to dataset metadata. Those fields marked with a * are date type fields which support some special search features described below. The approved tags which appear on HDX datasets are actually stored in a field called “vocab_Topics”.

Example:

https://data.humdata.org/api/action/package_search?fq=vocab_Topics:education – returns datasets with the “education” tag.

Querying with wildcards and ranges

The multicharacter (*) and single character (?) wildcard operators are supported but cannot be used in quoted search terms:

Example:

https://data.humdata.org/api/action/package_search?fq=dataset_source:ETH\ Zurich\ Cli?ada - returns 7 Climada datasets with a query that uses the single character wildcard.

Example:

https://data.humdata.org/api/action/package_search?fq=dataset_source:ETH\%20Zurich\%20Cli* - returns 7 Climada datasets with a query that uses the multi-character wildcard.

As well these simple wildcard usages the search API also supports range queries which can include wildcards.

Example:

[https://data.humdata.org/api/action/package_search?fq=num_resources:\[2 TO 5\]](https://data.humdata.org/api/action/package_search?fq=num_resources:[2 TO 5]) – this selects datasets which have between 2 and 5 resources.

Range queries can be used to select values which are not null with the query **fieldname:[* TO *]**, this is the approved way of doing such selections. The query **fieldname:*** includes datasets that have null values as well as all other values.

Querying date fields

CKAN's date search facilities are powerful but not always obvious. You can do an exact search for a datetime with a query like **metadata_created:"2019-12-04T10:23:27.806321Z"**, note that if the trailing Z is omitted the search fails with an Invalid Date String error, however dates are returned from `package_search` without a trailing Z! Date fields can also be queried with a range expression which allows for the special values NOW, DAY, MONTH, YEAR, HOUR, MINUTE, these can be combined with "normal" dates with +, - and / operators (/ is rounding):

Example:

[https://data.humdata.org/api/action/package_search?fq=last_modified:\[NOW-1DAY TO NOW\]](https://data.humdata.org/api/action/package_search?fq=last_modified:[NOW-1DAY TO NOW]) – this query returns datasets modified in the last 24 hours.

Combining query terms

Query terms can be combined with logical operators, the simplest of these is "NOT" or its equivalents "-" and "!"

Example:

https://data.humdata.org/api/action/package_search?fq=NOT res_format:CSV – this query shows datasets with resources that do not have format CSV.

Query terms can be combined with the "AND" operator (or "+"), in fact several terms are added to every query by CKAN (+capacity:public -dataset_type:showcase +state:(active) +site_id:"default").

Example:

[https://data.humdata.org/api/action/package_search?fq=res_format:\(CSV and XLS\)](https://data.humdata.org/api/action/package_search?fq=res_format:(CSV and XLS)) –

this returns datasets whose resources have either CSV or XLS formats.