

The Culture Grid search service

The culture grid search service utilises the SOLR (<http://lucene.apache.org/solr/>) enterprise search platform. It offers access to search services via the SOLR interface directly or via an SRU (<http://www.loc.gov/standards/sru/>) service which acts as a bridge between the SRU protocol and the SOLR platform. Both are equally usable, there is a small response overhead when using SRU, however this is offset by the fact that using SRU insulates systems from changes to the SOLR service.

Search access points/hit list metadata fields

SOLR	SRU	Description
authority	authority	The person or organisation responsible for the item
dc.description	dc.description	The description of the item
dc.identifier	dc.identifier	The identifier of the item
dc.subject	dc.subject	The subject for the resource
dc.title	dc.title	The title of the resources
dcmi.type	dcmi.type	The type of the resource
dcterms.isPartOf	collection	The collection to which the resource belongs
dcterms.spatial	dcterms.spatial	The spatial characteristics of the resource
dcterms.temporal	dcterms.temporal	The temporal characteristics of the resource
have_thumbnail	have_thumbnail	Whether the resource has a thumbnail cached in CG or not
institution_sector	institution_sector	Sector (present in institution records only)
postcode	postcode	Postcode (present in institution records only)
pndsterms.thumbnail	pndsterms.thumbnail	Not an access point even though it is in both access methods. This is a field meant to be used in the returned hit list to allow the display of a thumbnail if present
record_type	record_type	The type of the record (item, collection, institution)
subject.combined	subject.combined	The combination of the following three fields
subject.event	subject.event	Any events associated with the resource
subject.organisation	subject.organisation	Any organisations associated with the resource
subject.person	subject.person	Any people (agents) associated with the resource

text	fullText	All of the text content concatenated into a single text field
unparsed.place	place	The text name of the place (only in institutions)
	location	Only on SRU. This is parsed into latitude and longitude by the SRU service and used as a filter for the SOLR query to specify distance ordering.

Sample queries

SRU

The SRU/CQL query syntax is described as part of the specification. However there are some non-standard features that have been implemented as extensions within the culture grid implementation.

[http://culturegrid.org.uk/discover/xsl/default?operation=searchRetrieve&startRecord=1&query=dc.title%20all%20"Scree"%20OR%20fullText%20all%20"Scree"&x-facet=default&stylesheet=@/xslt/sruSearchResults.xsl&maximumRecords=10](http://culturegrid.org.uk/discover/xsl/default?operation=searchRetrieve&startRecord=1&query=dc.title%20all%20)

parameter	value
operation	searchRetrieve
query	dc.title all "Scree" OR fullText all "Scree"
x-facet	default
startRecord	1
stylesheet	@/xslt/sruSearchResults.xsl
maximumRecords	10

The above query (from the front page of the discover service) is an example of executing a query using SRU across the full text field and title. This will return all results which match in either field (and will by default give weight to exact matches in title).

Support for facets has been added within the SRU implementation as it is not yet part of the base SRU spec. Facets can be accessed in two different ways:

- By specifying &x-facet=default as part of any query.
 - This will return the default facets as used by the culture grid service.
- By specifying the required facets directly
 - &x-facet=true together with &x-facet.field=dc.subject (for example)

SRU Spatial searching

Spatial searching based on a distance sort from a postcode is only available from SRU (From SOLR the latitude and longitude must be specified directly)

http://peoplesnetwork.gov.uk/discover/xsl/default?operation=searchRetrieve&startRecord=1&query=dcterms.isPartOf%20adj%20%22MLAInstitutions%22%20AND%20location%20adj%20%22S3%208PZ%22&radius=5&sortBy=geo_distance/sort.ascending&stylesheet=@/xslt/sruFindALibrarySearchResults.xsl

The above query can be described as follows:

parameter	value
operation	searchRetrieve
Query	dcterms.isPartOf adj "MLAInstitutions" AND location adj "S3 8PZ"
radius	5
sortBy	geo_distance/sort.ascending
startRecord	1
stylesheet	@/xslt/sruSearchResults.xsl
maximumRecords	10

SOLR searching

The SOLR search base url is <http://culture-grid.org.uk/index/>.

The SOLR syntax is fully described in the SOLR documentation. However, to facilitate spatial searching a local SOLR extension has been added. An example of the use is as follows:

This means that proximity searching cannot be carried out directly it requires a set of results to filter. So if you wish to find the closest item to you would need to filter on a query which returns records (for example dcterms.isPartOf:MLAInstitutions).

The following return all records in the MLAInstitutions collection <http://culture-grid.org.uk/index/select/?q=dcterms.isPartOf%3AMLAInstitutions&version=2.2&start=0&rows=10&indent=on>

With additional parameters to order by proximity the latitude and longitude of the postcode must be known and supplied as follows:

http://culture-grid.org.uk/index/select/?q=dcterms.isPartOf%3AMLAInstitutions&version=2.2&start=0&rows=10&indent=on&lat=53.3813152288178&long=-2.4811235692305&radius=5&q=geo&sort=geo_distance asc

Supported features

SRU

The SRU search access point only supports exact single phrase searching in that it doesn't support an exact query of the form

"blue john mine"OR "mam tor" OR castleton.

Multiple phrase searching is treated as individual tokens

e.g.

blue OR john OR mine OR mam OR tor OR castleton.

Exact phrase searching is limited to a single phrase e.g. "blue john mine"