

# Search Services

## Archive.org Search API

This is documentation for version v1 of the Archive.org Search APIs.

To view beta search API documentation, visit <https://archive.org/services/search/beta/>

### meta : Information about search results

GET

/fields

#### Implementation Notes

Returns the list of fields that can be requested

#### Response Class (Status 200)

List of fields

Model

Model Schema

```
{
  "count": 0,
  "fields": [
    "string"
  ]
}
```

Response Content Type

application/json

#### Parameters

Parameter	Value	Description	Parameter Type	Data Type
callback	<input type="text"/>	JSONP callback function name	query	string

#### Response Messages

HTTP Status Code	Reason	Response Model	Headers
default	Unexpected error	<div><div>Model</div><div>Model Schema</div></div> <div><pre>{   "code": 0,</pre></div>	

```
"message": "string"
}
```

Try it out!

search : Search and scraping services

GET /scrape

Implementation Notes

See POST description

Response Class (Status 200)

Scaping API

Model | Model Schema

```
{
  "cursor": "string",
  "previous": "string",
  "count": 0,
  "remaining": 0,
  "total": 0,
  "items": [
    {}
  ]
}
```

Response Content Type application/json

Parameters

Parameter	Value	Description	Parameter Type	Data Type
secret	<input type="text"/>	S3 secret	formData	string
access	<input type="text"/>	S3 access key	formData	string
callback	<input type="text"/>	JSONP callback function name	query	string
REQUIRE_AUTH		Whether authorization	query	

	<input type="text"/>	is required. Must be set, along with s3 keys, to access 'all' scope		boolean
debug	<input type="text" value="false (default)"/>	If this value is set, information for debugging may be returned	query	boolean
xvar	<input type="text" value="production"/>	The experimental variant to use (setting this may return unexpected results)	query	string
total_only	<input type="text" value="false (default)"/>	Whether to return the total number of results only (if a cursor is given, this will return the total only from the point of the cursor).	query	boolean
scope	<input type="text"/>	The scope of the query. Possible values are 'standard' or 'all' The 'all' scope requires authorization	query	string
cursor	<input type="text"/>	Cursor for scrolling (used for subsequent calls)	query	string
count	<input type="text" value="100"/>	Number of query results to return. The minimum is 100, the maximum is 10,000. The default is 5,000. It's not bad to return a large number.	query	integer
sorts	<input type="text" value="title asc"/>	sort collations, comma delimited. Any of the sort fields from the <code>fields</code> endpoint can be specified. If the <code>identifier</code> is	query	string

specified as a sort field, it must be the last sort. The format is `<field> [asc|desc]`, where `asc` means ascending, and `desc` means decending. The default is `identifier asc` (and is always the implied last collation). Example: `title asc, publicdata desc`.

fields	identifier, title	Metadata fields, comma delimited. The list of requestable fields can be found at the <code>fields</code> endpoint. The default is to return the identifier only.	query	string
q	mediatype:(movie)	Lucene-type search query. The query language is described on <a href="#">Internet Archive advanced search page</a> .	query	string

Response Messages

HTTP Status Code	Reason	Response Model		Headers
default	Unexpected error	Model	Model Schema	
<pre>{   "code": 0,   "message": "string" }</pre>				

Try it out!

Hide Response

Curl

```
curl -X GET --header 'Accept: application/json' 'https://archive.org/services/search/v1/
```

## Request URL

```
https://archive.org/services/search/v1/scrape?debug=false&xvar=production&total_only=fal
```

## Response Body

```
{
  "items": [
    {
      "identifier": "The_Worm_in_the_Apple_Animation_Test",
      "title": "\"The Worm in the Apple\" Animation Test"
    },
    {
      "identifier": "RainerCicatka1WienerHochquelllleitung",
      "title": [
        "1. Wiener Hochquelllleitung",
        "1. Wiener Hochquelllleitung"
      ]
    },
    {
      "identifier": "169_audiovisual__Logo_animation",
      "title": "16:9 audiovisual - Logo animation"
    },
    {
      "identifier": "1929FleischerScreenSongAfterTheBall",
      "title": "1929 Fleischer Screen Song \"After the Ball\""
```

## Response Code

```
200
```

## Response Headers

```
{
  "server": "nginx/1.4.6 (Ubuntu)",
  "transfer-encoding": "Identity",
  "content-type": "application/json; charset=UTF-8",
  "date": "Wed, 27 Sep 2017 22:24:41 GMT",
  "connection": "keep-alive",
  "x-powered-by": "PHP/5.5.9-1ubuntu4.21",
  "cache-control": "no-cache"
}
```

POST

/scrape

### Implementation Notes

The Scrape API allows efficient queries to scrape (that is find all) documents (typically for IA items) meeting a search query. The metadata `fields` desired can be specified, as well as how the results should be sorted (via `sorts`). The results are returned in batches specified by `count` (which defaults to 5,000). A `cursor` is returned, which can be used to retrieve the next batch of results. If you want to know just the total number of results, specify `total_only`.

### Response Class (Status 200)

Scaping API

Model

Model Schema

```
{
  "cursor": "string",
  "previous": "string",
  "count": 0,
  "remaining": 0,
  "total": 0,
  "items": [
    {}
  ]
}
```

Response Content Type 

application/json

### Parameters

Parameter	Value	Description	Parameter Type	Data Type
secret	<input type="text"/>	S3 secret	formData	string
access	<input type="text"/>	S3 access key	formData	string
callback	<input type="text"/>	JSONP callback function name	query	string
REQUIRE_AUTH	<div><input type="checkbox"/></div>	Whether authorization is required. Must be set, along with s3 keys, to access 'all' scope	query	boolean

debug	<input type="text" value="false (default)"/>	If this value is set, information for debugging may be returned	query	boolean
xvar	<input type="text" value="production"/>	The experimental variant to use (setting this may return unexpected results)	query	string
total_only	<input type="text" value="false (default)"/>	Whether to return the total number of results only (if a cursor is given, this will return the total only from the point of the cursor).	query	boolean
scope	<input type="text"/>	The scope of the query. Possible values are 'standard' or 'all' The 'all' scope requires authorization	query	string
cursor	<input type="text"/>	Cursor for scrolling (used for subsequent calls)	query	string
count	<input type="text" value="1000"/>	Number of query results to return. The minimum is 100, the maximum is 10,000. The default is 5,000. It's not bad to return a large number.	query	integer
sorts	<input type="text"/>	sort collations, comma delimited. Any of the sort fields from the <code>fields</code> endpoint can be specified. If the <code>identifier</code> is specified as a sort field, it must be the last sort. The format is <code>&lt;field&gt; [asc desc]</code> , where	query	string

asc means ascending, and desc means decending. The default is identifier asc (and is always the implied last collation). Example: title asc, publicdata desc.

fields	<div>identifier</div>	Metadata fields, comma delimited. The list of requestable fields can be found at the fields endpoint. The default is to return the identifier only.	query	string
q	<div>(required)</div>	Lucene-type search query. The query language is described on <a href="#">Internet Archive advanced search page</a> .	query	string

Response Messages

HTTP Status Code	Reason	Response Model		Headers
default	Unexpected error	Model	Model Schema	
<div><pre>{   "code": 0,   "message": "string" }</pre></div>				

Try it out!