

BASE HTTP Interface

BASE (Bielefeld Academic Search Engine, <https://www.base-search.net/>) indexes web resources and harvests metadata of scientific repositories via OAI-PMH. It is based on the open source search engine software Lucene/SOLR. Currently more than 9000 [content providers](#) and 270 million documents are covered in BASE. The data structure related to the index structure is based on Dublin Core with several extensions.

BASE supplies a HTTP API. It can be used for non-commercial purposes only. The interface is IP controlled and interested users have to register at <https://www.base-search.net/about/en/contact.php>.

The interface is located at:

<https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi>

Once registered (i.e. an ip address or range has been configured) the interface can be easily tested by editing the url in a browser (located in the activated ip range) and looking for the delivered xml responses.

Rate limiting: One request per second (1 QPS).

Currently the following methods are supported:

Method	Description
ListRepositories	This method lists all available repositories or the member repositories of a certain collection. The method delivers an xml list of internal string identifiers for the repository.
ListProfile	This method lists basic information about the requested repository included in BASE (defined by the internal BASE string identifier)
PerformSearch	This is the interface core function. It performs queries and delivers the corresponding results. The BASE interface supports the SOLR query term syntax.

I. ListRepositories

Syntax:

<interface-url>?func=ListRepositories&coll=<collection_name>

Parameter	Value	Status	Description
func	ListRepositories	mandatory	
coll	<collection_name>	optional	Existing pre-defined collections: see Appendix, section „Collection-related queries“. The collection name must exist, otherwise all BASE repositories will be listed.
format	json	optional	Set response format to JSON Default: XML

Examples:

1. Repositories located in Europe:

<https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=ListRepositories&coll=ceu>

Response:

```
<?xml version="1.0" encoding="utf-8"?>
<collection>
  <collection_name>ceu</collection_name>
  <list_repositories>
    <repository>
      <activation_date>2020-08-26</activation_date>
      <name>OPUS – Publikationsserver der Hochschule Augsburg</name>
      <name_en>OPUS – Publication Server of Hochschule Augsburg</name_en>
      <internal_name>fthsaugsburg</internal_name>
    </repository>
    <repository>
      <activation_date>2020-08-26</activation_date>
      <name>Publikationsserver der Hochschule Landshut</name>
      <name_en>Publication Server of Hochschule Landshut</name_en>
      <internal_name>fthslandshut</internal_name>
    </repository>
    (...)
  </list_repositories>
</collection>
```

2. Repositories located in Germany:

<https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=ListRepositories&coll=de>

3. Repositories located in North Rhine-Westphalia, Germany:

<https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=ListRepositories&coll=denw>

II. ListProfile

This method delivers metadata information about a specific repository.

Syntax:

<interface-url>?func=ListProfile&target=<internal_name>

Parameter	Value	Status	Description
func	ListProfile	mandatory	
target	<internal_name>	mandatory	Internal name of a single repository as delivered by ListRepositories.
format	json	optional	Set response format to JSON Default: XML

Example:

1. Basic information of the repository „ftubbi epub“ including number of all documents, Open Access documents, Non-Open Access documents, documents with Creative Commons license and documents with Public Domain license:

<https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=ListProfile&target=ftubbi epub>

Response:

```
<?xml version="1.0" encoding="utf-8"?>
<repository>
<activation_date>2011-10-07</activation_date>
<country>de</country>
<name>PUB - Publikationen an der Universität Bielefeld</name>
<name_en>PUB - Publications at Bielefeld University</name_en>
<num_non_oa_records>56424</num_non_oa_records>
<num_oa_cc_records>1525</num_oa_cc_records>
<num_oa_pd_records>118</num_oa_pd_records>
<num_oa_records>12790</num_oa_records>
<num_records>69223</num_records>
</repository>
```

III. PerformSearch

PerformSearch is the essential method to search and retrieve BASE data. As a new feature the current version supports the usage of facets.

Syntax:

<interface-url>?func=PerformSearch&coll=<collection>&query=<queryterm>&(…)
<interface-url>?func=PerformSearch&target=<internal_name>&query=<queryterm>&(…)

Parameter	Value	Status	Description
func	PerformSearch	mandatory	
query	<queryterm>	mandatory	Search term For syntax details see Appendix, section „Query syntax“. Max. 1000 characters are allowed.
coll	<collection>	optional	For existing, pre-defined collections see Appendix, section „Collection-related queries“
target	<internal_name>	optional	Internal name of a single repository as delivered in ListRepositories
hits	<number>	optional	Maximum number of records per response (0 is allowed) Default: 10; max.: 120
offset	<number>	optional	Starting position in the result set Default: 0; max.: 999

sortby	<field+asc desc>	optional	Sorting A sort ordering must include a single field name (see Appendix, section „Fields“, table column „Sorting“.), followed by a whitespace (escaped as + or %20 in URL strings), followed by sort direction (asc or desc). Default: sort by relevance
boost	oa	optional	Push open access documents upwards in the result list. Default: without boosting
format	json	optional	Set response format to JSON Default: XML
fields	<field1,field2,...>	optional	The result records only contain fields listed in the comma-separated field list. For existing, pre-defined fields see Appendix, section „Fields“.
facets	<field1,field2,...>	optional	The response contains an extra section "facet_counts/facet_fields" with fields from the comma-separated facets list. This section provides a breakdown or summary of the results. From the user's perspective, faceted search breaks up search results into multiple categories, typically showing counts for each, and allows the user to "drill down" or further restrict their search results based on those facets. Use of faceting does not affect the results section of a search response. For existing, pre-defined facet fields see Appendix, section „Fields“, table column „Facet“.
facet_limit	<number>	optional	Maximum number of constraint counts that should be returned for the facet fields. Default: 100; min:1; max: 500
facet_sort	count index	optional	Ordering of the facet field constraints: count - sort by count (highest count first) index - alphabetical sorting Default: count
f_<field> e.g. f_dcsubject	<search_term>	optional	Search term for the facet field. The results can be refined by certain criteria (facets), such as subject. For existing, pre-defined facet fields see Appendix, section „Fields“, table column „Facet“.

If the arguments „target“ or „coll“ are missing or the value for „target“ does not exist, the query addresses all BASE repositories.

Examples:

For query syntax details see Appendix, section „Query syntax“.

1. Documents from the repository „ftubbiiepub“ containing the terms „lossau“ and „summann“ (search in the whole document):

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&target=ftubbiiepub&query=lossau+summann](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&target=ftubbiiepub&query=lossau+summann)

2. Documents from Italian repositories containing the term „manghi“ in the dcreator field (author). The flag „boost“ pushes open access documents upwards in the result list:

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&coll=it&query=dcreator:manghi&boost=oa](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&coll=it&query=dcreator:manghi&boost=oa)

3. Documents containing the terms „algebra“ and „linear“ or „algebra“ and „numerical“ or „algebra“ and „linear“ and „numerical“:

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&query=algebra+AND+\(linear+OR+numerical\)](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=algebra+AND+(linear+OR+numerical))

4. Documents containing the terms „operator“ or „algorithms“ and „linear“ or „numerical“:

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&query=\(operator+OR+algorithms\)+AND+\(linear+OR+numerical\)](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=(operator+OR+algorithms)+AND+(linear+OR+numerical))

5. Documents containing the terms „linear“ and „algebra“ but not „numerical“:

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&query=linear+algebra+NOT+numerical](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=linear+algebra+NOT+numerical)

6. Documents containing the terms „schmidt“ in dcreator field (author) and „biology“ in dtitle. The response starts after record 5 (offset=5) and contains max. 5 hits with the fields dtitle, dcreator and dyear:

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&query=dcreator:schmidt+dtitle:biology&hits=5&offset=5&field
s=dtitle,dcreator,dyear](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=dcreator:schmidt+dtitle:biology&hits=5&offset=5&fields=dtitle,dcreator,dyear)

7. Documents containing the term „unix“ and published between 1983 and 2009, sorted by year of publication (dyear) in descending order:

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&query=unix+dyear:\[1983+T0+2009\]&sortby=dyear+desc](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=unix+dyear:[1983+T0+2009]&sortby=dyear+desc)

8. Documents containing the term „unix“ with facets dsubject and dyear each with max. 10 hits (default facet sorting „count“):

[https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?
func=PerformSearch&query=unix&facets=dsubject,dyear&facet_limit=10](https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=unix&facets=dsubject,dyear&facet_limit=10)

9. Documents containing the term „unix“ and the exact terms "computer science" in dcs subject and „2008“ in dcyear; with facets „dcs subject“ and „dcyear“ each with max. 10 hits:

https://api.base-search.net/cgi-bin/BaseHttpSearchInterface.fcgi?func=PerformSearch&query=unix&facets=dcs subject,dcyear&f_dcs subject=%22computer%20science%22&f_dcyear=2008&facet_limit=10

Please note:

Results of the search „query=unix+dcs subject:education“ can differ from results of the faceted search „query=unix&f_dcs subject=education“. Facets search for the exact term „education“ in the dcs subject field, while the pure query search delivers records with dcs subject fields containing the term „education“ also as a single keyword of a field's contents, e.g. "higher education".

Example Response Format:

```
<?xml version="1.0" encoding="UTF-8"?>
<response>
  <lst name="responseHeader">
    <bool name="zkConnected">true</bool>
    <int name="status">0</int>
    <int name="QTime">128</int>
  </lst>
  <lst name="params">
    <str name="q">creator:summann</str>
    <str name="facet">true</str>
    <arr name="facet.field">
      <str>f_dcs subject</str>
      <str>f_dcyear</str>
    </arr>
    <str name="facet.limit">10</str>
    <str name="facet.mincount">1</str>
    <str name="facet.sort">count</str>
    <str name="fl">
      dccollection,dccontenttype,dcontinent,dccountry,dccreator,dccdate,dc
      description,dcdocid,dcdoi,dcformat,dcidentifier,dclang,dclanguage,dc
      link,dcperson,dcpublisher,dcrights,dcs source,dcs subject,dctitle,dcyear
      ,dctype,dcclasscode,dctypenorm,dc deweyfull,dc deweyhuns,dc deweytens,d
      cdeweyones,dcautoclasscode,dcrelation,dccontributor,dccoverage,dchda
      te,dcoa,dcrightsnorm</str>
    <str name="fq">collection:ftubbi epub</str>
  </lst>
</lst>
<result name="response" numFound="70" start="0" maxScore="8.910765">
<doc>
  <date name="dchdate">2019-09-23T12:46:37Z</date>
  <str name="dc docid">
a3e98a979171fc179135968f2e5bbf5d7e8696f2ff964b6f02050bb479bcd8b
</str>
  <str name="dcontinent">ceu</str>
  <str name="dccountry">de</str>
  <str name="dccollection">ftubbi epub</str>
```

```
<str name="dcprovider">PUB - Publications at Bielefeld
University</str>
<str name="dctitle">10 years BASE: A contribution to the worldwide
development of repositories</str>
<arr name="dccreator">
<str>Pieper, Dirk</str>
<str>Summann, Friedrich</str>
</arr>
<arr name="dcperson">
<str>Pieper, Dirk</str>
<str>Summann, Friedrich</str>
</arr>
<arr name="dcauthorid">
<str>Pieper, Dirk | orcid:0000-0002-6083-9348</str>
<str>Summann, Friedrich | orcid:0000-0002-6297-3348</str>
</arr>
<arr name="dcsubject">
<str>Bielefeld Academic Search Engine</str>
<str>ddc:020</str>
</arr>
<str name="dcdate">2014</str>
<int name="dcyear">2014</int>
<arr name="dctype">
<str>info:eu-repo/semantics/conferenceObject</str>
<str>doc-type:conferenceObject</str>
<str>text</str>
</arr>
<arr name="dctypenorm">
<str>13</str>
</arr>
<arr name="dcidentifier">
<str>https://pub.uni-bielefeld.de/record/2710028</str>
</arr>
<str name="dclink">https://pub.uni-bielefeld.de/record/2710028</str>
<str name="dcsource">Pieper D, Summann F. 10 years BASE: A
contribution to the worldwide development of repositories. Presented
the China IR Seminar 2014.</str>
<arr name="dclanguage">
<str>eng</str>
</arr>
<str name="dcrights">info:eu-repo/semantics/openAccess</str>
<arr name="dcdeweyfull">
<str>020</str>
</arr>
<arr name="dcdeweyhuns">
<str>0</str>
</arr>
<arr name="dcdeweytens">
<str>02</str>
</arr>
<arr name="dcdeweyones">
<str>020</str>
```

at

```
</arr>
<arr name="dcclasscode">
<str>020</str>
</arr>
<int name="dcoa">1</int>
<arr name="dclang">
<str>eng</str>
</arr>
</doc>
(...)
</result>
<lst name="facet_counts">
<lst name="facet_queries"/>
<lst name="facet_fields">
<lst name="f_dcsubject">
<int name="ddc:020">42</int>
<int name="Bielefeld Academic Search Engine">12</int>
<int name="ddc:004">9</int>
<int name="Repositories">3</int>
<int name="Bibliothek">2</int>
<int name="Communication">2</int>
<int name="Comunicazione">2</int>
<int name="Informatica">2</int>
<int name="Informatics">2</int>
<int name="Informatik">2</int>
</lst>
<lst name="f_dcyear">
<int name="1994">5</int>
<int name="2009">5</int>
<int name="2014">5</int>
<int name="2017">5</int>
<int name="2004">4</int>
<int name="2005">4</int>
<int name="2006">4</int>
<int name="2013">4</int>
<int name="2015">4</int>
<int name="2016">4</int>
</lst>
</lst>
<lst name="facet_ranges"/>
<lst name="facet_intervals"/>
<lst name="facet_heatmaps"/>
</lst>
</response>
```


Appendix

1. Collection-related queries:

Supported values for the request parameter „coll“ and as contents of response fields „dccontinent“ and „dccountry“.

Continents	
caf	Africa
cas	Asia
cau	Australia/Oceania
ceu	Europe
cna	North America
csa	South America
cww	Web server without geographic relation (org)
Countries (Country code ISO 3166)	
de	German repositories
nl	Dutch repositories
uk	British repositories
etc.	(...)
German Country states	
debw	Baden-Württemberg
deby	Bayern
debe	Berlin
debb	Brandenburg
dehb	Bremen
dehh	Hamburg
dehe	Hessen
demv	Mecklenburg-Vorpommern
deni	Niedersachsen
denw	Nordrhein-Westfalen
derp	Rheinland-Pfalz
desl	Saarland
desn	Sachsen
dest	Sachsen-Anhalt
desh	Schleswig-Holstein
deth	Thüringen
Austrian Country states	

BASE Interface Guide – Version 1.23 (October 2021)

atbgld	Burgenland
atktn	Kärnten
atnö	Niederösterreich
atoö	Oberösterreich
atsbg	Salzburg
atstmk	Steiermark
att	Tirol
atvbg	Vorarlberg
atw	Wien
Swiss Country states	
chag	Aargau
chai	Appenzell Innerrhoden
char	Appenzell Ausserrhoden
chbe	Bern
chbl	Basel-Landschaft
chbs	Basel-Stadt
chfr	Freiburg
chge	Genf
chgl	Glarus
chgr	Graubünden
chju	Jura
chlu	Luzern
chne	Neuenburg
chnw	Nidwalden
chow	Obwalden
chsg	St. Gallen
chsh	Schaffhausen
chso	Solothurn
chsz	Schwyz
chtg	Thurgau
chti	Tessin
chur	Uri
chvd	Waadt
chvs	Wallis
chzg	Zug
chzh	Zürich

2. Fields (used for search and response)

Used as request parameters in „query“, „fields“, „facets“, „sortby“ and as contents of response fields. Please note that only fields marked with „x“ are allowed in the respective parameters.

Internal fields					
Field	Description	Status	Query	Facet	Sorting
dccollection	internal BASE repository name	single	x	x	
dchdate	date of harvesting/updating	single	x		x
Bibliographic DC fields (as defined in the Dublin Core definition)					
Field	Description	Status	Query	Facet	Sorting
dccontributor	contributor	multi	x		
dccoverage	coverage	single	x		
dccreator	author	multi	x		x
dcddate	date of publication	single	x		
dcdescription	abstract	single	x		
dcformat	format	multi	x		
dcidentifier	urls	multi	x		
dclanguage	language	multi	x		
dcpublisher	publisher	multi	x		
dcrelation	relation	multi	x		
dcrights	rights	single	x		
dcsource	source	single	x		
dcsubject	subject headings	multi	x	x	
dctitle	title	single	x		x
dctype	publication and document type	multi	x		
Dublin Core similar extra fields					
Field	Description	Status	Query	Facet	Sorting
dcauthorid	author and his identifiers	multi	x		
dcautoclasscode	DDC class code (internal automatic categorization)	multi	x		
dcclasscode	DDC class code derived from the document metadata	multi	x		
dccontenttype	type of content	multi	x	x	
dccontinent	continent (internal code, see section „Collections“)	single	x		
dccountry	country code (ISO 3166)	single	x		
dcdeweyfull	DDC class code (original + automatic)	multi	x		
dcdeweyhuns	DDC class code (original + automatic): the hundreds value	multi	x	x	

BASE Interface Guide – Version 1.23 (October 2021)

dc Dewey tens	DDC class code (original + automatic): the tens value	multi	x	x	
dc Dewey ones	DDC class code (original + automatic): the ones value	multi	x	x	
dc docid	unique record identifier	single	x		
dc doi	DOI	multi	x		
dc lang	normalized form of the language derived from dclanguage (ISO 639-2/B: three-letter codes)	multi	x	x	
dc link	preferred url for linking	single	x		
dc oa	open access document (0=no, 1=yes, 2=unknown)	single	x	x	x
dc orcid	ORCID (Open Researcher and Contributor ID)	multi	x		
dc person	creator + contributor	multi	x	x	
dc provider	content provider	single			
dc rightsnorm	normalized form of the rights derived from dcrights	multi	x	x	
dc typenorm	normalized form of the publication and document type derived from dc type (internal code, see section „Document types“)	multi	x	x	
dc year	normalized form of the year derived from dc date	single	x	x	x

3. Document types

Codes used in the query aspect and the response field „dctypenorm“

Code	Description
1	Text
11	Book
111	Book part
12	Journal/Newspaper
121	Article contribution to journal/newspaper
122	Other non-article part of journal/newspaper
13	Conference object
14	Report
15	Review
16	Course material
17	Lecture

Code	Description
18	Thesis
181	Bachelor thesis
182	Master thesis
183	Doctoral and postdoctoral thesis
19	Manuscript
1A	Patent
2	Musical notation
3	Map
4	Audio
5	Image/Video
51	Still image
52	Moving image/Video
6	Software
7	Dataset
F	Other/Unknown Material

4. Query syntax

Entering one or more search terms (separated by blanks) will produce hits, which include all the entered terms in one document (AND conjunction). There are, however, more options to combine terms and to search efficiently.

You search ...	Example	Hits (ca.)
A and B	linear algebra	77,000
A and B as a phrase (directly adjoined)	"linear algebra"	34,000
A and/or B	linear OR algebra	1,100,000
A and B or A and C or A, B and C	algebra AND (linear OR numerical)	98,000
A and not B	algebra NOT linear	400,000
from A to B (only for numbers)	dcyear:[1983 TO 2009]	34,000,000
Verbatim search	textus:"linear algebra"	29,000

Plural, genitive case and other word forms will be searched automatically.

The * (asterisk) substitutes any number of characters whereas the ? (question mark) only substitutes one character. They are used to find different spellings or wordings even in different languages in one search. If the asterisk or question mark are used, the automatic search for different word forms will be disabled. The asterisk and the question mark can not be applied in phrase searches.

Examples:

A search for ho*e delivers records containing „home“, „horse“, „horticulture“ etc.

A search for ho?e delivers records containing „home“, „hole“, „hope“ etc.

//F. Summann, R. Mitrenga, 12.10.2021

Changelog

2021-10-12 [Version 1.23]

* Changed:

- Limit for hits

2021-09-27 [Version 1.22]

* Changed:

- Introduction
- Limit for hits
- Limit for offset

* Added:

- Character limit for query

2021-07-07 [Version 1.21]

* Added:

- Parameter target for ListRepositories
- Parameter target for ListProfile
- New examples for PerformSearch

2020-08-28 [Version 1.20]

* Changed:

- Example 1 for ListRepositories
- Example 1 for ListProfile

2020-05-12 [Version 1.19]

* Added:

- "dcorcid" in Dublin Core similar extra fields

2019-10-30 [Version 1.18]

* Added:

- "dcauthorid" in Dublin Core similar extra fields

* Changed:

- Example Response Format

2019-06-06 [Version 1.17]

* Changed:

- "dcrightsnorm" to multi-value field

2018-10-19 [Version 1.16]

* Added:

- "num_oa_cc_records", "num_oa_pd_records", "num_non_oa_records" in ListProfile method

2018-03-08 [Version 1.15]

* Added:

- "activation_date", "num_records", "num_oa_records" in ListProfile method
- "activation_date" in ListRepositories method
- Changelog