

# Searching KnowUK via XML Gateway

## Table of Contents

1. [Overview](#)
2. [Access and Authentication](#)
3. [Operations](#)
  1. [explain](#)
  2. [searchRetrieve](#)
4. [Diagnostic Messages](#)
5. [Appendix A: Publication source ids](#)
6. [Appendix B: Category ids](#)
7. [Appendix C: Context set mappings](#)
8. [Appendix D: Example searchRetrieveResponse XML](#)

## Overview

We have implemented this XML Gateway to the [Search/Retrieve via URL](#) (SRU) standard protocol for Internet search queries. Readers should become familiar with the documentation of this standard before reading the rest of this document as we include here only the details of how our interface implements this standard.

## Access and Authentication

### Query Types

The base URL for this operation is:

<http://xml.knowuk.co.uk/xml/advancedSearch.do>

Both HTTP POST and GET requests are supported.

### Authorization and Authentication

Access via the XML Gateway requires authentication. We currently support both IP authentication and username/password access. See [Extra Request Data](#) below for details of how to submit a username and password.

# Operations

There are two different operations that are supported:

- [explain](#)
- [searchRetrieve](#)

For each of these operations, we describe the general purpose of the operation, the request parameters and the format of the response.

## The explain operation

### Purpose

The [explain](#) operation will give details about the service provided, including the indexes (search fields) and schemas supported for querying.

### Request Parameters

Parameter Name	Mandatory or Optional	Values	Default	Description
<b>version</b>	mandatory	1.1	<i>Not applicable</i>	The SRU <a href="#">version</a> .
<b>recordPacking</b>	optional	xml	<i>xml</i>	How the explain record should be escaped in the response. Currently only an xml response is supported.
<b>stylesheet</b>	optional	valid URLs	<i>none</i>	A URL for an xml stylesheet to be included in the response.
<b>operation</b>	mandatory	explain	Not applicable	Indicates the type of operation being requested.
<b>x-info-1-accept</b>	optional	any or anyResponse	<i>none</i>	Indicates that the client will accept extraResponseData in the returned XML. If this parameter is omitted, no information about the extraResponseData will be included in the explain Response.

### Response

The response record will be in XML following the [ZeeRex Schema for SRU](#).

## The searchRetrieve operation

### Purpose

The [searchRetrieve](#) operation is used for retrieving a set of records associated with a particular query.

### Standard Request Parameters

Parameter Name	Supported?	Mandatory or Optional	Values	Default	Description
<b>version</b>	yes	mandatory	1.1	<i>Not applicable</i>	The version of the request, and a statement by the client that it wants the response to conform to a version less than or preferably equal to this version.
<b>query</b>	yes	mandatory	<i>A CQL query.</i>	<i>Not applicable</i>	Contains the query expressed in CQL to be processed by the server. See more below on <a href="#">CQL</a> support.
<b>startRecord</b>	yes	optional	<i>integer greater than 0</i>	1	The position within the sequence of matched records of the first record to be returned. The first position in the sequence is 1. If the value given is larger than the current number of results associated with the result set, an appropriate <a href="#">diagnostic</a> will be returned.
<b>maximumRecords</b>	yes	optional	<i>integer greater than or equal to 0</i>	50	Indicates the maximum number of records to be returned at a time by the search.
<b>numberOfRecords</b>	yes	optional	<i>integer greater than or equal to 0</i>	25	Indicates the desired number of records to be returned at a time by the search.
<b>recordPacking</b>	yes	optional	xml	xml	Indicates how the record should be escaped in the response.
<b>recordSchema</b>	yes	optional	mods	mods	See the explanation from the <a href="#">standard</a> .
<b>recordXPath</b>	no				
<b>recordSetTTL</b>	no				Result set timeout is fixed at 10 minutes.

<b>sortKeys</b>	limited	optional	See <a href="#">below</a> .	<i>none</i>	Contains a sequence of sort keys to be applied to the results. See <a href="#">sorting</a> below.
<b>stylesheet</b>	yes	optional	<i>A URL for an XML stylesheet</i>	<i>none</i>	The client requests that the server simply return this URL in the response. See the description in the <a href="#">standard</a> .
<b>operation</b>	yes	mandatory	searchRetrieve	<i>Not applicable</i>	Indicates the type of operation being requested.

### Extra Request Data

The following table outlines the [extra request parameters](#) implemented by the KnowUK XML API.

Parameter Name	Mandatory or Optional	Values	Default	Description
<b>instit1, instit2</b>	optional	Username and password pair as <i>instit1=username&amp;instit2=password</i>	<i>none</i>	If a user is recognised by IP, this field may be ignored. If authentication fails, an appropriate <a href="#">diagnostic</a> will be returned.
<b>x-sortType</b>	optional	"relevance" or "alphabetical"	per search	Pre-defined sort orders suited to the relevant datasets

### Common Query Language Support

The Common Query Language ([CQL](#)) is a formal language for representing queries to information retrieval systems. We currently support [Level 2 conformance](#) to the standard.

A query consists of one or more search clauses joined by boolean operators. The service currently supports:

- Term-only queries (e.g., war or "world war").
- Queries with search clauses using the 'index relation searchTerm' syntax (e.g., title any war), for supported relations (see [below](#)).
- Either of the above types of clauses combined with boolean operators (e.g. title any war and author any smith, e.g. war or conflict).

For unsupported queries, an appropriate [diagnostic](#) will be returned. The following table provides more details about the specific parts of the language that are supported.

Note that the default **Context Set** for indexes is Chadwyck-Healey (ch="http://www.chadwyck.co.uk/"). See Appendix C for mappings to Dublin Core and SerialsSolutions Central Search context sets.

Index name	Sort Key?	Description
keyword (default)	no	Keyword search entire record, including full text
title	yes	Article title
sourceID	yes	Publication ids (see Appendix A)
type	yes	Category (see Appendix B)
cql.resultSetId	n/a	When the cql.resultSetId index is used, it should be the only index in the query
x-sortType	n/a	"relevance" (default) (sort by relevance descending, article title ascending) "alphabetical" (sort by article title ascending, journal title ascending )

Query Feature	Values	Default	Description
<b>Relation</b>	<ul style="list-style-type: none"> <li>▪ cql.any</li> <li>▪ cql.all</li> <li>▪ cql.=</li> </ul>	cql.all	cql.any means find any of the terms; cql.all means find all of them; cql.= means find the phrase given within the field indicated by the index. The qualifier cql need not be given.
<b>Boolean Operators</b>	<ul style="list-style-type: none"> <li>▪ and</li> <li>▪ or</li> <li>▪ not</li> </ul>	<i>Not applicable</i>	These have their usual meaning, except that not means "and not" and is thus a binary operator, not a unary one.

All queries are **case-insensitive**. The cql.exact relation, relational modifiers and boolean modifiers are not supported. Search terms must be enclosed in double quotes when they include any of these characters: <, >, =, /, (, ), and whitespace. (See the [CQL documentation](#).)

## Sorting

Sorting of result sets can be requested one of two ways:

1. Using the `x-sortType` parameter. These trigger tailored sort algorithms which match those used in the Web interface.
2. Using the `sortKeys` parameter. This parameter can be supplied either as part of the original request or as a subsequent request in which the result set identifier is supplied in the `query` parameter in this way: `query=cql.resultSetId=[id]`. The sorting syntax will follow the [syntax](#) given in the standard with the following exceptions.
  - The `sortKeys` values are specified by index name rather than XPath.
  - We do not currently support specifying multiple result set identifiers for use in sorting.
  - Case sensitivity in sorting is not supported.

Quick Search results are always returned in relevance order.

The appropriate [diagnostic message](#) will be returned for unsupported or invalid queries.

## Search examples

Parameters	Description
"ernest rutherford cambridge"	Find all of the terms "ernest", "rutherford" and "cambridge" in the default (keyword) field.
keyword all "ernest rutherford cambridge"	Find all of the terms "ernest", "rutherford" and "cambridge" in the keyword field. (Note: This is the equivalent of a Web interface search on "All of these terms").
keyword = "ernest rutherford"	Find the exact phrase "ernest rutherford" in the keyword field. (Note: This is the equivalent of a Web interface search on "Exact phrase")
keyword any "ernest rutherford cambridge"	Find any of the terms "ernest", "rutherford" or "cambridge" in the keyword field. (Note: This is the equivalent of a Web interface search on "Any of these terms")
title = "Cambridge Regional College"	Find the exact phrase "Cambridge Regional College" in the item title field.
keyword all "ernest rutherford cambridge" and type = "BIOG"	Find all the terms "ernest", "rutherford" and "cambridge" in records from biography sources only.
keyword all "ernest rutherford cambridge" and sourceID = "WHOWASWHO"	Find all the terms "ernest", "rutherford" and "cambridge" in records from "Who Was Who" only.

## Response

The [response](#) will conform to the searchRetrieveResponseType defined within [SRU/SRW schema](#). The citation metadata, outlined in the [table below](#), will be represented using [mods data elements](#).

Field	XPath	Comments
Document title	mods:titleinfo/mods:title	Document / article title
Author	mods:name[@type="personal"]/mods:namePart	Author where present
Publication	mods:relatedItem[@type="host"]/mods:titleInfo/mods:title	Name of publication
Publisher	mods:originInfo/mods:publisher	Publisher
Date	mods:relatedItem[@type="host"]/mods:part/mods:date	Date of publication
Category	mods:relatedItem[@type="series"]/mods:titleInfo/mods:title	Content type
Identifier	mods:identifier	Unique internal identifier for the item
Location	mods:extension/localUrl	Product URL for the item Note that URLs in KnowUK are not durable in the long term, as publications may be updated using different unique identifiers.
Relevance score	mods:extension/relevanceScore	Relevance score as returned by the RetrievalWare search
Publication ID	mods:extension/sourceID	Publication ID (see Appendix A)
Hit in context	mods:extension/hitInContext	A keyword hit and surrounding text (may include encoded HTML markup)

There is a small [example](#) of response XML. Notice that not all data fields are present in every record.

## Search examples

### Initiating a search

This is an example query:

```
?version=1.1&query=title+any+tree&operation=searchRetrieve
```

This initiates a search and retrieves the total number of matches for the query "title contains tree" for each database as well as a subset of those matching records for each database. The [response](#) contains the `resultSetId` for this set (in this example, `result_set_id_string`), the status information about the search, and the records retrieved.

### Continuing a search

If more records than were retrieved with the initial search are wanted, a client may request that more be retrieved by using the `cql.resultSetId` with a new `startRecord`:

```
?version=1.1&operation=searchRetrieve&query=cql.resultSetId=R1234567890&startRecord=51
```

## Diagnostic Messages

Diagnostic messages will be returned when, for example, the query provided is not of the proper syntax, or the application's server is not responding, or the version requested is not supported. The diagnostic messages will follow the [SRU standard diagnostics](#), augmented where necessary to provide specific messages from our application.

## Appendix A: Publication source ids

ch.sourceID	Publication title
TG_*	Life Events Topic Guides
LOCINF	KnowUK Local Information
ASKHOLLIS	A.S.K. Hollis
ACTORS_YB	Actors' Yearbook
AIRFORCE	Air Force List
PM_BRIT	All-Time Greats of British and Irish Sport
ARMYLIST	Army List
NHS_MANAGE	Binley's Directory of NHS Management
BLOOMSBURY_GRG	Bloomsbury Good Reading Guide
BGCHANISLE	Blue Guide Channel Islands
BG_HOUSES	Blue Guide Country Houses of England
BG_ENGLAND	Blue Guide England
BGIRELAND	Blue Guide Ireland
BG_LONDON	Blue Guide London
BG_MGL	Blue Guide Museums & Galleries of London
BG_OXCAM	Blue Guide Oxford and Cambridge
BG_SCOT	Blue Guide Scotland
BGWALES	Blue Guide Wales
BRITARCH	British Archives
BRITQUAL	British Qualifications
CAREHOME	Care Homes for Older People

ch.sourceID	Publication title
INFORMAGP	Informa Database of Primary Care
PRESSURE	Interest Groups
PM_INT	International Who's Who of Sport
LAWSOCDIR	Law Society's Directory of Solicitors
LIBSDIR	Libraries Directory
MAM_SPO	Mammoth Book of Sports & Games of the World
MYB	Municipal Year Book
MUSICIANS_YB	Musicians' and Songwriters' Yearbook
NTGUIDE	National Trust Guide
NAVYLIST	Navy List
NIBIOGS	Northern Ireland Assembly Biographies
POETRY_WRITERS_YB	Poetry Writers' Yearbook
BEST	ProQuest Book of the Best
PUBBODIES	Public Bodies
REGION	Regional Trends
MFD	Religions in the UK: A Multi-Faith Directory
WW_SCH	Schools' Who's Who
SCOTLAW	Scottish Law Directory
SLQA	Scottish Legal Questions Answered
SCOTBIOGS	Scottish Parliamentary Biographies
EAC	Sheltered Housing for Older People

CATHSCOT	Catholic Directory for Scotland
COFE	Church of England Year Book
CIVBIOG	Civil Service Biographies
CIVISERVYB	Civil Service Year Book
CDNB	Concise Dictionary of National Biography
ECCTIS	Course Discover: Courses
INSTS	Course Discover: Universities and Further Education Colleges
CROCK	Crockford's Clerical Directory
POT	Debrett's People of Today
DOMUS	Digest of Museum Statistics (DOMUS)
DIPLOMLIST	Diplomatic Service List
SELF_HELP	Directory of British Self Help Groups and Support Organisations
GMTRUST	Directory of Grants and Funds for Scotland
VOLORG	Directory of National Voluntary Organisations in Scotland
EDU_GRANTS	Educational Grants Directory
EUBIOGS	European Parliamentary Biographies
FILMMAKERS_YB	Filmmakers' Yearbook
FTWDR	Financial Times World Desk Reference
YEARAHEAD	Foresight
GAB_SPEC	Gabbitas Guide to Schools for Special Needs

SOCIAL	Social Trends
TITLEFORMS	Titles and Forms of Address
UKONS	UK 2005 - The Official Yearbook of the United Kingdom of Great Britain and Northern Ireland
UKAF	UK Advice Finder
EXPERT	UK Register of Expert Witnesses
VAD	Voluntary Agencies Directory
WELFARE	Welfare Benefits and Tax Credits Handbook, Abridged
WELBIOGS	Welsh Assembly Biographies
WDIES	What to Do When Someone Dies
WHICHLEGAL	Which? 450 Legal Problems Solved
WHICHCREER	Which? Guide to Choosing a Career
WHICHCOMED	Which? Guide to Complementary Therapies
WDIVORCE	Which? Guide to Divorce
WHICHHOTEL	Which? Guide to Good Hotels
WLIVETOG	Which? Guide to Living Together
WMONEY	Which? Guide to Money
WPENSION	Which? Guide to Planning Your Pension
WRENTLET	Which? Guide to Renting and Letting
WHICHSCOT	Which? Guide to Scotland
WTOURIST	Which? Guide to Tourist Attractions

WHICHBED	Good Bed & Breakfast Guide
WHICHFOOD	Good Food Guide
GSG	Good Schools Guide and Directory
GD_WLKS	Good Walks Guide
GLABIOGS	Greater London Assembly Biographies
IND_NEEDS	Guide to Grants for Individuals in Need
COMBIOGS	House of Commons Biographies
LORBIOGS	House of Lords Biographies
HUTCHINSON	Hutchinson Almanac
HUTCH_SYMBOLS	Hutchinson Dictionary of Symbols in Art
HUTCH_MYTHOLOGY	Hutchinson Dictionary of World Mythology
ENCY	Hutchinson Encyclopedia of Britain
ENCYBIOG	Hutchinson Encyclopedia of Britain - Biographies
ENCYWEB	Hutchinson Encyclopedia of Britain - Websites
UKGAZ	Hutchinson UK Gazetteer

WCH_WKBK	Which? Guide to Weekend Breaks in Britain
WCH_MEDIC	Which? Medicine
WCH_PUBS	Which? Pub Guide
WFLAT	Which? Way to Buy, Own and Sell a Flat
WHOUSE	Which? Way to Buy, Sell and Move House
WHITALM	Whitaker's Almanack
WHIT_OLYMPIC	Whitaker's Olympic Almanack
WHIT_SCOT	Whitaker's Scottish Almanack
WHOWASWHO	Who Was Who
WHOSWHO	Who's Who
WHOEDUC	Who's Who in Education
WHONIRE	Who's Who in Northern Ireland
WHOSCOT	Who's Who in Scotland
WILLINGS	Willings Press Guide
WRITERS	Writers' and Artists' Yearbook

## Appendix B: Category ids

ch.type	Category
ARTS	Arts and Media
BIOG	Biography
COUR	Courses and Careers
EDUC	Education
EVNT	Events
KNOW	General Knowledge
GOV	Government and Armed Services
GRNT	Grants and Funding

ch.type	Category
WELF	Health and Welfare
LAW	Law
LIBM	Libraries, Museums and Archives
ORGASSCH	Organisations, Associations and Charities
FINP	Personal Finance
RELIG	Religion
TOUR	Tourism and Leisure

## Appendix C: Context set mappings

Chadwyck-Healey context set (default)	Dublin Core context set	Central Search context set	Description
keyword	<i>n/a</i>	cs.keyword	Keyword search entire record, including full text
sourceID	<i>n/a</i>	<i>n/a</i>	Chadwyck-Healey or ProQuest attributed publication ID, when used for searching instead of the publication title (for which use dc.source)
title	dc.title	cs.title	Article or item title
type	dc.type	<i>n/a</i>	Resource type/Category

## Appendix D: Example searchRetrieveResponse XML

```
<srw:searchRetrieveResponse xmlns:srw=http://www.loc.gov/zing/srw/
  xmlns:diag="http://www.loc.gov/zing/srw/diagnostic/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:mods="http://www.loc.gov/mods/v3"
```

```
xsi:schemaLocation="http://www.loc.gov/zing/srw/
    http://www.loc.gov/standards/sru/xml-files/srw-types.xsd
    http://www.loc.gov/zing/srw/diagnostic/
    http://www.loc.gov/standards/sru/xml-files/diagnostics.xsd
    http://www.loc.gov/mods/v3
    http://www.loc.gov/standards/mods/v3/mods-3-0.xsd">
<srw:version>1.1</srw:version>
<srw:numberOfRecords>23</srw:numberOfRecords>
<srw:resultSetId>R10ED7AAB1C11</srw:resultSetId>
<srw:resultSetIdleTime>1800</srw:resultSetIdleTime>
<srw:records>
  <srw:record>
    <srw:recordSchema>info:srw/schema/1/mods-v3.0</srw:recordSchema>
    <srw:recordPacking>xml</srw:recordPacking>
    <srw:recordData>
      <mods:mods>
        <mods:titleInfo>
          <mods:title>
            <![CDATA[RUTHERFORD, 1st Baron, created 1931, of Nelson, New Zealand, and
              Cambridge (1871-1937)]]>
          </mods:title>
        </mods:titleInfo>
        <mods:relatedItem type="series">
          <mods:titleInfo>
            <mods:title>
              Biography
            </mods:title>
          </mods:titleInfo>
        </mods:relatedItem>
      </mods:mods>
    </srw:recordData>
  </srw:record>
</srw:records>
```

```
<mods:relatedItem type="host">
  <mods:titleInfo>
    <mods:title>
      <![CDATA[Who Was Who]]>
    </mods:title>
  </mods:titleInfo>
  <mods:part>
    <mods:date>
      <![CDATA[January 2006]]>
    </mods:date>
  </mods:part>
</mods:relatedItem>
<mods:originInfo>
  <mods:place>
    <mods:placeTerm>
      <![CDATA[A&C Black]]>
    </mods:placeTerm>
  </mods:place>
</mods:originInfo>
<mods:identifier>
  037/0065972
</mods:identifier>
<mods:extension>
  <relevanceScore>
    49
  </relevanceScore>
  <localUrl>
    /displayCategItem.do?ItemID=037/0065972
  </localUrl>
</mods:extension>
</mods:relatedItem>
```

```

        <hitInContext>
            <![CDATA[
                ... New Zealand, and <font
                color=#ff0000><strong>Cambridge</strong></font>;
                <font color=#ff0000><strong>Ernest</strong></font>
                <font
                color=#ff0000><strong>Rutherford</strong></font>; Kt,
                created 1914;...<BR>]]>
        </hitInContext>
        <sourceID>
            <![CDATA[WHOWASWHO]]>
        </sourceID>
    </mods:extension>
</mods:mods>
</srw:recordData>
<srw:recordPosition>1</srw:recordPosition>
</srw:record>
<srw:record>
    <srw:recordSchema>info:srw/schema/1/mods-v3.0</srw:recordSchema>
    <srw:recordPacking>xml</srw:recordPacking>
    <srw:recordData>
        <mods:mods>
            <mods:titleInfo>
                <mods:title>
                    <![CDATA[Rutherford, Ernest (1871-1937)]]>
                </mods:title>
            </mods:titleInfo>
            <mods:relatedItem type="series">
                <mods:titleInfo>
                    <mods:title>
                        General Knowledge
                    </mods:title>
                </mods:titleInfo>
            </mods:relatedItem>
        </mods:mods>
    </srw:recordData>
</srw:record>

```

```
<mods:relatedItem type="host">
  <mods:titleInfo>
    <mods:title>
      <![CDATA[Hutchinson Encyclopedia of Britain - Biographies]]>
    </mods:title>
  </mods:titleInfo>
  <mods:part>
    <mods:date>
      <![CDATA[March 2005]]>
    </mods:date>
  </mods:part>
</mods:relatedItem>
<mods:originInfo>
  <mods:place>
    <mods:placeTerm>
      <![CDATA[Helicon Publishing]]>
    </mods:placeTerm>
  </mods:place>
</mods:originInfo>
<mods:identifier>
  059/2010433
</mods:identifier>
<mods:extension>
  <relevanceScore>
    42
  </relevanceScore>
  <localUrl>
    /displayCategItem.do?ItemID=059/2010433
  </localUrl>
</mods:extension>
</mods:relatedItem>
```

```
<hitInContext>
  <![CDATA[&lt;font
    color=#ff0000&gt;&lt;strong&gt;Rutherford&lt;/strong&gt;&lt;/font&gt; at
    Manchester]-&lt;font
    color=#ff0000&gt;&lt;strong&gt;Ernest&lt;/strong&gt;&lt;/font&gt; &lt;font
    color=#ff0000&gt;&lt;strong&gt;Rutherford&lt;/strong&gt;&lt;/font&gt; New
    Zealand...&lt;BR&gt;]]>
</hitInContext>
<sourceID>
  <![CDATA[ENCYBIOG]]>
</sourceID>
</mods:extension>
</mods:mods>
</srw:recordData>
<srw:recordPosition>2</srw:recordPosition>
</srw:record>
</srw:records>
<srw:nextRecordPosition>3</srw:nextRecordPosition>
<srw:echoedSearchRetrieveRequest>
  <srw:version>1.1</srw:version>
  <srw:query><![CDATA[&#034;ernest rutherford cambridge&#034;]]></srw:query>
  <srw:xQuery xmlns="http://www.loc.gov/zing/cql/xcql/">
    <searchClause>
      <index>
        srw.serverChoice
      </index>
      <relation>
        <value>
          scr
        </value>
      </relation>
    </searchClause>
  </srw:xQuery>
</srw:echoedSearchRetrieveRequest>
```

```
        <term>
            ernest rutherford cambridge
        </term>
    </searchClause>
</srw:xQuery>
<srw:startRecord>1</srw:startRecord>
<srw:maximumRecords>2</srw:maximumRecords>
<srw:extraRequestData xmlns:ch="http://www.chadwyck.co.uk/extradata/">
    <ch:x-sortType>relevance</ch:x-sortType>
</srw:extraRequestData>
</srw:echoedSearchRetrieveRequest>
</srw:searchRetrieveResponse>
```

**©2006 Copyright ProQuest Information and Learning. All rights reserved.**