

# Class Outline

## EndNote Filters and Connection Files

### 1 What Is the Difference between a Filter and a Connection File?

In truth filters and connection files are very similar. Connection files contain filters, but in addition to this they also contain the connection setting information that lets EndNote query a z39.50 server for the information to be filtered.

Shown below are the sections or parts of a connection file and a filter. Notice that all the parts available for a filter are also available for a connection file, but for a connection file that there are also two additional parts: the connection settings and the search attributes.

Connection File	Filter
About This Connection	About This Filter
Connection Settings	
Search Attributes	
Templates	Templates
Author Parsing	Author Parsing
Continuation Lines	Continuation Lines
Default Reference Type	Reference Type
Editor Parsing	Editor Parsing
Field Editing	Field Editing
MARC Records	MARC Records
Record Layout	Record Layout
Source Parsing	Source Parsing

Because of these similarities everything said below in reference to connection files can also be applied to filters, unless specified otherwise.

### 2 Building a Connection File: The Pieces

#### 2.1 About This Connection/Filter

Record any limitations, notes, or background information about the filter or connection file here. This is also a very good place to put any instructions about how to download the text file to match the filter, as we do for the filters we produce.

#### 2.2 Connection Settings

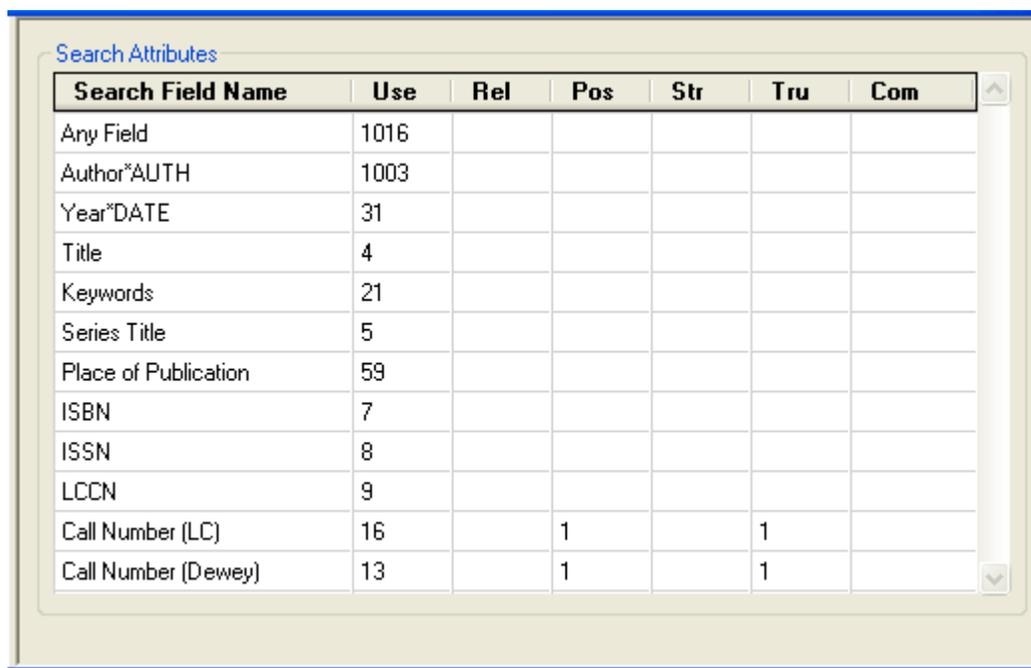
To build a connection file you will need the following information (please note that you will need to contact the library that supports the server to find some of this information, perhaps even a special librarian in charge of the server):

1. Is the database server you are connecting to using the z39.50 protocol? If it is not, you cannot build a connection file for it. If it is, continue to gather the following:

2. The server address. This can either be a URL or an IP address. Note that this will not be the same URL you would go to using your internet browser to access the same database.
3. The database command, called the database name in EndNote. This is the command that will “start up” your interaction with that database on the server.
4. The port that the server uses. Ports 210 and 7090 are among the most common, but it could be another. The server, not EndNote, controls which port must be used for the connection. Note that it will not use port 80, which is generally used for http connections, so it will not be accessible through a proxy server. Users who want to access the database from off-site may need a VPN connection to do so.
5. The record syntax used by the database.
6. The text-encoding method used by the database.

## 2.3 Search Attributes

Search attributes control the query that is sent to the server. They are independent of the information that will be sent back. Think of these as being *how* you ask the question, but remember that the answer may not give you exactly the response you expect, so you may need to experiment with these settings. In many ways, the search attributes can be the most difficult part of building a connection file if you want to customize it. However, frequently using the default settings will fit your needs, and to use those you can simply leave most of the fields blank. Below is an example of a search attributes screen, followed by descriptions of each of the attributes. All attributes are in numerical values.



Search Field Name	Use	Rel	Pos	Str	Tru	Com
Any Field	1016					
Author*AUTH	1003					
Year*DATE	31					
Title	4					
Keywords	21					
Series Title	5					
Place of Publication	59					
ISBN	7					
ISSN	8					
LCCN	9					
Call Number (LC)	16		1		1	
Call Number (Dewey)	13		1		1	

Attribute	Description
Use	This attribute describes what type of information is being searched for. It's easy to think of this as just the field being searched, but it is more detailed than that. For example, Use value 1 equals a personal name, so you might use it when searching for a person's name but not a corporation's name. Use attribute 2 equals a corporate name. Not all use attributes will be supported for all databases. When you create a new, blank connection file it will be prepopulated with the most commonly-used use attributes, but can then be edited.
Relation	EndNote does not support editing this field.
Position	<p>The position attribute specifies the location of the search term within the field or subfield in which it appears. The options are—</p> <ol style="list-style-type: none"> <li data-bbox="586 709 1333 772">1. <b>First in field:</b> The search term must be the first data in the field.</li> <li data-bbox="586 779 1276 842">2. <b>First in subfield:</b> The search term may appear in any subfield but must be the first data in the subfield.</li> <li data-bbox="586 848 1360 911">3. <b>Any position in field:</b> The search term may appear anywhere in the field.</li> </ol>
Structure	<p>The structure attribute specifies the type of search term, for example a single word or a phrase.</p> <ol style="list-style-type: none"> <li data-bbox="586 1010 1297 1073">1. <b>Phrase:</b> One or more groups of characters separated by blanks, in exactly the order they are entered.</li> <li data-bbox="586 1079 1179 1104">2. <b>Word:</b> A group of characters with no blanks.</li> <li data-bbox="586 1110 1312 1205">3. <b>String:</b> The entire term is to be treated as a string, rather than a sequence or set of individual words. An example might be a credit card number with blanks within it.</li> <li data-bbox="586 1211 1344 1274">4. <b>Word List:</b> One or more groups of characters separated by blanks, in which there is no implied order.</li> </ol>

Attribute	Description
Truncation	<p>Truncation determines if one or more characters may be omitted when matching the search term. The effects of truncation vary depending on what the structure attribute is set to as explained below.</p> <ol style="list-style-type: none"> <li>1. <b>Right Truncation:</b>  <b>Word or Phrase:</b> Last word of the term is right truncated.  <b>String:</b> Entire term is right truncated.  <b>Word List:</b> Each word is right truncated.</li> <li>2. <b>Left Truncation:</b>  <b>Word or Phrase:</b> First word of the term is left truncated.  <b>String:</b> Entire term is left truncated.  <b>Word List:</b> Each word is left truncated.</li> <li>3. <b>Left and Right Truncation:</b>  <b>Word or Phrase:</b> First word of term is left truncated and last word of term is right truncated.  <b>String:</b> Entire term is left and right truncated.  <b>Word List:</b> Each word is left and right truncated.</li> </ol> <p>100. <b>Do Not Truncate:</b> No truncation is applied.</p>
Completeness	<ol style="list-style-type: none"> <li>1. <b>Incomplete subfield:</b> Words other than those in the search term may appear in the subfield or field in which the term appears.</li> <li>2. <b>Complete Subfield:</b> No words other than those in the search term should appear in the entire subfield in which the term appears, but additional words may appear in other subfields in the field.</li> <li>3. <b>Complete Field:</b> No words other than those in the search term should appear in the entire field in which the term appears.</li> </ol>

After you have set your Connection Settings and Search Attributes, close and save the connection file, then close EndNote to clear the connect log of any previous connection attempts. Reopen EndNote and do a search using your new connection file. Do not be alarmed if you import a lot of “garbage.” That is normal because you have not yet created the filter portion of the file. By performing a search EndNote creates a log file in your EndNote Documents and Setting folder. This file is called “connect.log.” Locate that file, open it in a text editor and save it as a text file. You can now use this text file to create the filter portion of your connection file.

## 2.4 Templates

In creating the template, place the field tags in the left column and the fields the data will be imported into in the right column. Be sure to include any spaces that are to the left of the field tag; they are considered part of the tag. You do not need to worry about spaces between the tag and the beginning of the field information.

You can use the Insert Field button to select fields to insert. This minimizes the risk of typos and also shows the fields that are available for the specific Reference Type.

Ideally there will be one tag per field, but in reality there may be multiple fields within a single tag. For example, below are variations of the source line for the OCLC ERIC database for journal article references. Notice that in general the more complex version of the source lines are first, with simpler source lines following. EndNote will try to match the punctuation and letters that are not parts of field names for the contents of the record's data. If the punctuation and field names do not match for any given line, EndNote will then try to match to the next line. The more consistent the data is, the less source lines are usually necessary.

Tag	Fields
Source:	Journal  Volume, Issue, Pages, Date
Source:	Journal vVolume nIssue pPages Date  Year
Source:	Journal vVolume nIssue pPages Year
Source:	Journal nIssue pPages Date  Year
Source:	Journal nIssue pPages Year
Source:	Journal nIssue Date  Year
Source:	Journal nIssue Year
Source:	Journal vVolume pPages Date  Year
Source:	Journal vVolume pPages Year
Source:	Journal vVolume Date  Year
Source:	Journal vVolume Year
Source:	Journal pPages Date  Year
Source:	Journal pPages Year
Source:	Notes

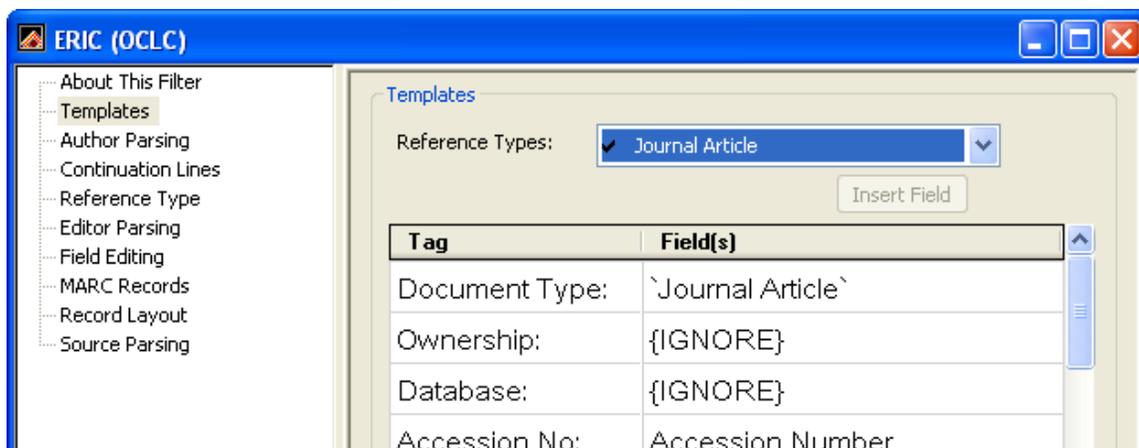
Note the vertical line between some fields. This tells EndNote to put in a field all the data preceding the last occurrence of the separator that follows the line in the field. For example, in the following line anything that follows the Pages field up until a four-digit number at the end of a line will be imported into the Date field. EndNote expects the four-digit number to be a year, so it can tell that anything not a four-digit number at the end of the line should be part of the Date field.

```
Source:      Journal vVolume pPages Date| Year
```

This means that in either of the following cases, the date information would be correctly entered in the Date field.

```
Source:      JAMA v10 p25-28 December 31 1933
Source:      New England Journal of Medicine v3 p68-89 June 1998
```

Another special character that is used in the templates is the accent grave mark (`). This character lets EndNote identify which reference type a record should be imported as. In the example below the selected Reference Type is Journal Article and the information for the Document Type tag reads as “`Journal Article`.” This indicates that if EndNote finds the exact text string “Journal Article” in the Document Type field, that reference will be imported as a Journal Article and the Journal Article template in the filter will be used to parse it.



The {IGNORE} options that you see in the example above are also available from the Insert Field button. Use this option if you do not want that data imported into any field.

## 2.5 Author Parsing

The Author Parsing options allow you to customize how EndNote will recognize authors' names. Surprisingly often the default "Smart" option works remarkably well, but if you notice that authors' names are not importing properly, you can adjust how they are recognized and parsed here.

## 2.6 Continuation Lines

This setting controls how EndNote recognizes if a new line is part of the preceding field or should be treated as a new field. If the second and following lines for a tag are indented, you can set the number of spaces indented here. When EndNote sees that number of spaces at the beginning of a new line it will consider that line part of the same field. When EndNote does not see that number of spaces at the beginning of a new line it knows that line should be treated as a new tag. You can set the indentation for all fields at once or on a field-by-field basis if the indentation is not consistent for all fields.

## 2.7 Default Reference Type

If a record does not contain any of the text strings identified within the templates using the accent grave ( ` ) mark, it will be treated as the reference type set here. As explained above, that mark can be used to specify reference types for different records, and such records will be sorted into their appropriate reference types. All records without such identifying text will be imported using the default reference type and its template.

## 2.8 Editor Parsing

Editor Parsing works the same way Author Parsing does, but it applies to the author-type fields other than the primary authors, such as editors or series editors. For some databases author and editor names may be entered differently, so it is necessary to have two different setting options.

## 2.9 Field Editing

Field Editing will allow you to set capitalization options for data that was originally entered in all caps, and you can do it on a field-by-field basis. This will work *only* with data that was originally all caps. If there is even a single lowercase letter in the field, EndNote will not be able to change the case.

You can also use these settings to eliminate unwanted characters within a field. You can select the field in the Field Editing options, and then create a list of the characters you want removed from the field during import. For example, some databases may place hyphens between each word in the title or notes fields. If you add hyphens to your list of characters to be edited out for that field, your final data will have spaces instead of hyphens in it.

## 2.10 MARC Records

MARC stands for Machine Readable Cataloging. The MARC Records options let you identify if the records for a given database use a numerical coding system such as MARC records. An example of MARC tags and fields is shown below.

Tag	Field(s)
008	Year
020	\$aISBN
022	\$aISBN
100	\$aAuthor
110	\$aAuthor\$bAuthor\$dAuthor\$nAuthor\$cAuthor\$qNotes\$kNotes
111	\$aNotes\$cNotes\$dNotes\$eNotes\$nNotes\$fNotes\$gNotes\$kNotes\$pNotes...
130	\$aNotes\$dNotes\$kNotes\$lNotes\$sNotes\$fNotes\$mNotes\$pNotes
222	\$aNotes\$bNotes
245	\$aTitle\$bTitle\$cNotes\$fTitle\$gTitle\$hTitle\$kTitle\$nTitle\$pTitle\$sTitle

A three-digit number is used for the tag, with subfield delimiters separating possible variations in data for that tag. In the example above the subfield delimiter is a dollar sign followed by a single letter. Each MARC field tag may have several subfields associated with it, and each subfield may have several variations. The biggest difference in building a filter with MARC coding is that if you have multiple variations of a source line, all of the variations go in the same cell in the template. This means you may have several subfields in a cell in your template, and within each subfield you may have several variations of the source line. For example, in the 773 tag below there are six variations on the \$g subfield. If you were working with SUTRS (Simple Unstructured Text Record Syntax) data such as you had seen above, where words or letters rather than numbers were used for the tags, each variation would be on a separate line in the template.

```

773      $cNotes $dNotes $tJournal (Notes)$tJournal $gVolume, no.
Issue (Date): Pages $gno. Issue (Date): Pages $gVolume
(Date): Pages $gVolume (Date) $g (Date) $gNotes $xNotes
$yNotes
798      $tJournal. $xNotes $dNotes. $gDate Year. v. Volume (Issue)
$Year. (Issue) $gYear. v. Volume (Issue) $gYear $xISSN
$zNotes

```

In this screen you can also set which column the data starts in. We don't think in columns much these days for text files, but it simply means how many characters and spaces there are, counted from the left margin, before the data begins. You can also set if the first character at that point is data or a subfield delimiter.

## 2.11 Record Layout

The Record Layout options let you set what characters EndNote should look for to determine where the record begins and ends. There are several options here, and which one you use will depend on the data. Try using the Smart option first, it will look for obvious things like extra line breaks in between solid blocks of text. For data with a lot of extra and unnecessary characters between records you may want to select a first and last tag, but you must be sure the same first and last tags appear in all records. You can also use nontag text strings to identify the beginning and end of a record. For example "Start of record #" in the option would indicate that each record began with the text "Start of record" followed by any number.

## 2.12 Source Parsing

The Source Parsing option applies only to multiple source lines for the Journal Article reference type. You can have many fields that may have multiple variations in how the data is entered. For example, for the Publisher field for the book reference type you might have lines like this—

```

Publisher:      Doubleday, New York, 2003
Publisher:      Houghton Mifflin, 2003

```

Each of these variations would have a separate line in the template, from the complex to the simple, as described above. However, if none of the options fits, EndNote will just leave the line unparsed. For this reason it is always good to have one extremely simple option at the end of the list that puts the entire contents of the tag in one field. For the examples above I would write a series like this in the template—

```

Publisher:      Publisher, City, Year
Publisher:      Publisher, Year
Publisher:      Publisher

```

This way if none of the detailed options fit, everything would be dumped in the Publisher field and I could sort it out later.

For journal articles though, EndNote will let you automate this option, and also offer an option to try to "guess" at a match if it can't find one.

In the Source Parsing screen, select the tag that marks the journal's source line, then select if you wish for EndNote to 1) "guess" at which fields things belong in,

otherwise known as “Smart Parsing,” or 2) import the entire contents of the tag into a field of your choice, generally the Notes field. It is your choice, but I would suggest having EndNote put the data in a field such as Notes so that you can parse it manually later. With Smart Parsing you face the risk of EndNote making a mistake and losing some of your data.

### **3 After Building Your Filter or Connection File**

When you are finished putting together your options, close and save your filter and then test it on your data. If you have built a connection file you can test the filter portion of the file by using the Use Connection File option in the Import window. This way you do not have to connect to an online database and recreate your search again and again to test your file. You may find that many records import correctly, but some do not. Filters and connection files usually take some fine tuning before they work satisfactorily. You may find there are reference types in the original data that you have not yet created templates for, or variations in source lines that you did not anticipate. In some instances you may find that it is impossible to create a completely reliable filter. Usually this is because of inconsistencies in data entry or incomplete data. You may even find that there are some databases that are so poorly tagged that it is not possible to build a filter for them at all. Although this can be a frustrating experience when it happens, in most cases you will find that it can be a rewarding challenge to write your own customized filters and connection files.

### **4 Resources for Creating Filters and Connection Files**

Below are some resources that I have found helpful when working with connection files and filters.

1. For information on USMARC coding:  
<http://www.loc.gov/marc/bibliographic/ecbdhome.html>
2. For information on the Bib-1 attribute set used in connection files:  
<http://lcweb.loc.gov/z3950/agency/defns/bib1.html>
3. OCLC Configuration Information  
[http://www.oclc.org/support/documentation/firstsearch/z3950/z3950\\_databases/](http://www.oclc.org/support/documentation/firstsearch/z3950/z3950_databases/)
4. Ovid  
[http://www.ovid.com/site/support/z3950\\_interop.jsp](http://www.ovid.com/site/support/z3950_interop.jsp)
5. Wilson Web  
<http://www.hwwilson.com/Documentation/WilsonWeb/z3950.htm>
6. IFLA Universal Bibliographic Control and International MARC Core Programme (UBCIM) (AKA—UNIMARC)  
<http://www.ifla.org/VI/3/p1996-1/sectn1.htm>
7. Z39.50 International Standards Maintenance Agency  
<http://www.loc.gov/z3950/agency/>
8. Register of Z39.50 Implementers  
<http://www.loc.gov/z3950/agency/register/entries.html>