NAME

cattobib – convert Z39.50 library catalog server data to BibTeX markup

SYNOPSIS

cattobib [**−?**] [**−by***xxx*] [**−help**] [**−keep-files**] [**−logfile**] [**−quiet**] [**−server** [**−test**] [**−version**]
        search-key-1 search-key-2 search-key-3 . . . >*BIBTEX-file*

DESCRIPTION

cattobib converts Z39.50 library-catalog server data to BIBTEX markup, using one or more search keys provided on the command line. This allows convenient re-use of publication data, and eases the tedious and error-prone task of creating BIBTEX entries for books and other cataloged publications.

The library catalog server can be specified by a command-line option, with the default server being the world's largest library catalog, the US Library of Congress.

ANSI/NISO Standard Z39.50-1995 and ISO Standard 23950:1998 ''Information and documentation — Information retrieval (Z39.50) — Application service definition and protocol specification'' define a library catalog protocol that allows client programs to communicate with library catalog servers around the world, and retrieve data in a small number of different formats, notably USMARC (*United States MAchine-Readable Cataloging*) and SUTRS (*Simple Unstructured Text Record Syntax*).

OPTIONS

Command-line options may be abbreviated to a unique leading prefix, and may begin with either one or two leading hyphens:

−?   Give a brief help message on *stdout*, and exit immediately with a successful status code.

−by*xxx* Pass this sort-order option to **bibsort**(1). The suffix *xxx* is one of *day*, *label*, *number*, *pages*, *seriesvolume*, *volume*, or *year*, or any other suffix supported by future versions of that program.

−help   Give a help message on *stdout* describing the options and known Z39.50 servers, and exit immediately with a successful status code.

−keep-files
           Preserve intermediate scratch files in */tmp* (or wherever the environment variable *TMPDIR* points). They are named *cattobib.bib.nnn* (raw BIBTEX data before several cleanup steps), *cattobib.fifo.nnn* (raw data from Z39.50 server for the last item searched), and *cattobib.label.nnn* (citation-label substitution file), where *nnn* is the process number. Normally, these files are not of interest and are deleted on exit, unless this option is specified.

−logfile *filename*
           Log the output on the specified filename, which must be a new file, instead of on *stdout*.

−quiet   Suppress search status messages.

−server *name-or-path*
           Specify a Z39.50 catalog server. The *name-or-path* value is either a path to a particular Z39.50 server, usually in the form *hostname:portnumber/databasename*, or an abbreviation for such a server as shown below. A regularly-updated directory of Z39.50 servers can be found on the Web at

                   http://www.indexdata.dk/targettest/

           In this list, vertical bars separate alternatives, and asterisk matches any word with that prefix:

```
alberta | ab       University of Alberta
amherst | umass    University of Massachusetts, Amherst
boulder | co       University of Colorado, Boulder
british | br       British Library
calgary            University of Calgary
congress | lc      US Library of Congress
copac | uk         COPAC (union of 24 research-university
```

```
                              catalogs in the UK and Ireland)
           denmark | dk        Royal Library of Denmark
           dsb | dsl           Danish State Library
           duke                Duke University
           florida | fl        Florida Center for Library Automation
           marriott | ut       University of Utah Marriott Library
           melbourne           University of Melbourne
           melvyl | ca         University of California MELVYL catalog
           minn* | mn          University of Minnesota
           newyork | ny        New York University
           nla | au            National Library of Australia
           nlnz | nz           National Library of New Zealand
           norway | no         National Library of Norway
           nsw | unsw          University of New South Wales
           odense | sdu        University of Southern Denmark
           oxford | ox*        Oxford University
           sweden | se         National Library of Sweden
           texas | tx          University of Texas at Austin
           toronto             University of Toronto
           usc                 University of Southern California
```

−test Run in test mode: library-catalog data is supplied on *stdin*, instead of being fetched from a Z39.50 catalog server. This option is primarily intended for the installation-time validation test suite, but can be also use for local testing and tuning of the format-conversion software.

−version
Display the version number and revision date on *stdout*, and exit immediately with a successful status code.

EXAMPLES

Search the default Z39.50 server for a book by its ISBN (*International Standard Book Number***):**

```
% cattobib 1-57586-011-2
%% Searching [z3950.loc.gov:7090/Voyager] for [1575860112]: flags = [@attr 1=7]
@Book{Knuth:1999:DT,
  author =        "Donald Ervin Knuth",
  title =         "Digital typography",
  volume =        "78",
  publisher =     "CSLI Publications",
  address =       "Stanford, Calif.",
  pages =         "xv + 685",
  year =          "1999",
  ISBN =          "1-57586-011-2 (cloth), 1-57586-010-4 (paperback)",
  LCCN =          "Z249.3 .K59 1999",
  bibdate =       "Fri Nov 19 07:23:50 MST 2004",
  bibsource =     "z3950.loc.gov:7090/Voyager",
  series =        "CSLI lecture notes",
  URL =           "ftp://uiarchive.cso.uiuc.edu/pub/etext/gutenberg/;
                  http://www.loc.gov/catdir/description/cam029/98027331.html;
                  http://www.loc.gov/catdir/toc/cam022/98027331.html",
  acknowledgement = ack-nhfb,
  subject =       "Printing; Data processing; Computerized typesetting;
                  Computer fonts; TeX (Computer file); METAFONT",
}
```

Remark: The ISBN is a *unique* identifier assigned to books published throughout the world since about 1972. It consists of ten decimal digits, the last of which may also be the letter X, divided into four hyphen- (or rarely, space-) separated parts: country or language, publisher, book number within the publisher, and a final check digit that can be used to detect invalid ISBNs.

Country/language groups 0 and 1 are English, 2 is French, 3 is German, 4 is Japanese, 5 is Russian, and so on. The Republic of Srpska is 99938.

Large publishers have small numbers (e.g., Collins is 00, McGraw-Hill is 07, and Prentice-Hall is 13), and small publishers have big numbers (e.g., Peachpit Press is 938151 and Personal TeX is 9631044).

When a publisher exhausts its range of book numbers, it gets a new publisher number: O'Reilly and Associates is assigned numbers 937175, 56592, and 596.

Search the default Z39.50 server for a book by its title:

```
% cattobib 'Digital Typography Sourcebook'
%% Searching [z3950.loc.gov:7090/Voyager] for [Digital Typography Sourcebook]: flags = []
@Book{Bryan:1996:DTS,
  author =        "Marvin Bryan",
  title =         "The digital typography sourcebook",
  publisher =     "Wiley",
  address =       "New York",
  pages =         "xxiv + 384, 3",
  year =          "1996",
  ISBN =          "0-471-14811-3 (paper/CD-ROM)",
  LCCN =          "Z250.7 .B79 1996",
  bibdate =       "Fri Nov 19 07:25:19 MST 2004",
  bibsource =     "z3950.loc.gov:7090/Voyager",
  URL =           "ftp://uiarchive.cso.uiuc.edu/pub/etext/gutenberg/;
                   http://www.loc.gov/catdir/bios/wiley047/96013161.html;
                   http://www.loc.gov/catdir/description/wiley033/96013161.html;
                   http://www.loc.gov/catdir/toc/onix04/96013161.html",
  acknowledgement = ack-nhfb,
  subject =       "Computer fonts",
}
```

Search the British Library for the same book:

```
% cattobib --server br 'Digital Typography Sourcebook'
%% Searching [z3950cat.bl.uk:9909/BLAC] for [Digital Typography Sourcebook]: flags = []
%% IGNORED:  Number of hits: 1, setno 1
...
@Book{Bryan:1997:DTS,
  author =        "Marvin Bryan",
  title =         "The digital typography sourcebook",
  publisher =     "Wiley",
  address =       "New York ; Chichester",
  pages =         "xxiv + 384",
  year =          "1997",
  ISBN =          "0-471-14811-3 (paperback)",
  bibdate =       "Fri Nov 19 07:26:13 MST 2004",
  acknowledgement = ack-nhfb,
  subject =       "Computer fonts",
}
```

Search the National Library of Australia for two books by ISBN:

```
% cattobib -q --server au 0-06-621285-5 0-19-860702-4
@Book{Winchester:2003:KDW,
```

```
        author =        "Simon Winchester",
        title =         "Krakatoa: the day the world exploded, 27 August 1883",
        publisher =     "HarperCollins Publishers",
        address =       "New York",
        pages =         "xvi + 416",
        year =          "2003",
        ISBN =          "0-06-621285-5",
        bibdate =       "Fri Nov 19 07:37:12 MST 2004",
        bibsource =     "catalogue.nla.gov.au:7090/Voyager",
        acknowledgement = ack-nhfb,
        remark =        "Includes bibliographical references and index.",
        subject =       "Natural disasters; Indonesia; Krakatoa; Social
                         aspects; Volcanoes; Indonesia; Krakatoa; Krakatoa
                         (Indonesia); Eruption, 1883",
        usmarc-019 =    "019 1 $a 24669279",
        usmarc-043 =    "043 $a a-io---",
        usmarc-250 =    "250 $a 1st U.S. ed.",
        usmarc-984 =    "984 $a ANL $c YY 551.2109598 W759",
}

@Book{Winchester:2003:MES,
        author =        "Simon Winchester",
        title =         "The meaning of everything: the story of the Oxford
                         English Dictionary",
        publisher =     "Oxford University Press",
        address =       "Oxford",
        pages =         "xxv + 260",
        year =          "2003",
        ISBN =          "0-19-860702-4 (hbk.), 0-19-860702-4 (hbk.)",
        bibdate =       "Fri Nov 19 07:37:23 MST 2004",
        bibsource =     "catalogue.nla.gov.au:7090/Voyager",
        price =         "No price",
        acknowledgement = ack-nhfb,
        remark =        "Includes ndex.",
        subject =       "Oxford English dictionary; Lexicology; History",
        usmarc-019 =    "019 1 $a 25073662",
}
```

BUGS

No matter which server is selected, library-catalog data tends to be rife with errors like these:

- completely wrong author lists;

- duplicated records, sometimes with minor variations;

- faulty title capitalization;

- incomplete, inaccurate, or missing page numbers;

- incorrect author order;

- mangled and missing accents;

- off-by-one copyright years;

- truncated author lists and titles;

- …

The best advice to the user is to search three or more catalogs for the same data, and then merge the results, using a majority vote to resolve discrepancies.

When multiple catalogs provide the same data, it may indicate that the data are likely to be reliable. However, the user is warned that libraries around the world share cataloging data, so there may not be as much data independence as might appear from geographically-distant catalogs.

While the conversion of USMARC and SUTRS markup to BIBTEX works reasonably well, there are many catalog record types that are not converted. When they are known not to be useful in BIBTEX entries, they are silently discarded. Otherwise, **cattobib** preserves them as additional key/value pairs, such as the *usmarc-nnn* keys in the BIBTEX output in the **EXAMPLES** section, or else complains about them in diagnostic messages.

cattobib produces only BIBTEX `@Book{...}` entries, even for conference proceedings, for which a `@Proceedings{...}` entry is required. Library catalog information often does not distinguish between these document types, so the user must convert such entries.

A certain amount of manual cleanup of the BIBTEX output is almost always necessary.

ENVIRONMENT VARIABLES

TESTSHRLIBDIR Directory where format-conversion software is stored. This variable is primarily intended for the installation-time validation test suite, but can also be used for testing alternate versions of the software.

TMPDIR Directory where temporary files are stored (default: */tmp*).

SEE ALSO

bibclean(1), **biblabel**(1), **biborder**(1), **bibsort**(1), **bibtex**(1), **citesub**(1), **yaz-client**(1).

AUTHOR

Nelson H. F. Beebe
University of Utah
Department of Mathematics, 110 LCB
155 S 1400 E RM 233
Salt Lake City, UT 84112-0090
Tel: +1 801 581 5254
FAX: +1 801 581 4148
Email: beebe@math.utah.edu, beebe@acm.org, beebe@computer.org
WWW URL: http://www.math.utah.edu/~beebe