

**NAME**

citesub – substitute standardized BibTeX citation labels

**SYNOPSIS**

**citesub** [ **-f** *substitution-file* ] [ **-v** ] [ file(s) ] >outfile

**DESCRIPTION**

**citesub** applies citation label substitutions to input BIB<sub>T</sub><sub>E</sub>X, L<sup>A</sup>T<sub>E</sub>X, and T<sub>E</sub>X files. The substitutions will usually be generated automatically by the companion **biblabel** program (see below), which can be used to standardize the form of BIB<sub>T</sub><sub>E</sub>X citation labels to the conventions adopted for the BibNet Project.

Most existing BIB<sub>T</sub><sub>E</sub>X bibliography files have been found to have rather haphazardly-chosen, and unsystematic, citation labels that are very likely to conflict with labels in other bibliography files; **biblabel** and **citesub** provide an automatic way to rectify this.

To avoid confusion between labels with common prefixes, such as *Smith80* and *Smith80a*, **citesub** checks for leading context of a left brace, quote, comma, whitespace, or beginning of line and trailing context of a right brace, comma, quote, percent, whitespace, or end of line so as to match these styles:

```
@Book{Smith:1980:ABC,

crossref = "Smith:1980:ABC",

crossref = {Smith:1980:ABC},

\cite{Smith:1980:ABC}

\cite{Smith:1980:ABC,Jones:1994:DEF}

\cite{%
    Smith:1980:ABC,%
    Jones:1994:DEF%
}
```

Although one might expect that simple application of standard software tools like the UNIX **awk**(1) and **sed**(1) utilities could do the string substitution job, this is not the case. For one thing, the required context sensitivity complicates the regular-expression patterns that are needed. For another, most UNIX **sed**(1) implementations have a built-in limit of about 100 substitutions, which is far too few for typical bibliographies. Finally, simple application of **awk**(1) and **awk**(1) involves matching every input line with every substitution pattern, which results in quadratic run-time behavior that proves impossibly slow for large bibliographies.

**citesub** provides an efficient solution whose run time is essentially proportional to the size of the input files, and *independent* of the number of substitutions to be carried out. This is achieved by tokenizing the input lines, and then looking up each token in a constant-access time (hash) table of substitutions. An initial prototype programmed in the *awk* language led to a final version in C that ran about 50 times faster, processing about 4000 input lines/sec on an entry-level Sun SPARCstation LX workstation.

**OPTIONS**

Except for the option described below, all command-line words are assumed to be input files. Should such a filename begin with a hyphen, it must be disguised by a leading absolute or relative directory path, e.g. */tmp/-foo.bib* or *./-foo.bib*.

**-f** *substitution-file* This option specifies the name of a file containing pairs of old and new citation labels, one pair per line, surrounded by arbitrary amounts of whitespace. This file is most easily generated by the companion program **biblabel**(1).

If this option is omitted, then the substitution filename will be derived from that of the first input file by replacing its extension by *.sub*. Thus, the commands  
*citesub -ffoo.sub foo.bib >foo.bib-new*

and  
`citesub foo.bib >foo.bib-new`  
 are equivalent.

If the substitution file is named "-", then **citesub** follows the common UNIX convention and interprets it to mean standard input, allowing the substitutions to be provided from a pipeline, such as

`biblabel foo.bib | citesub -f - >foo.new`

**-v** Display the program version number, and possibly installer, location, and compile-time information, on *stderr*.

## WARNING AND ERROR MESSAGES

**citesub** will issue warning messages in the following cases:

- *Incorrect number of fields in substitution file.* The line is ignored.
- *Invalid citation label character.* **citesub** expects that citation labels follow the requirements of the  $\text{\LaTeX}$  grammar described in the paper

Nelson H. F. Beebe, *Bibliography prettyprinting and syntax checking*, TUGboat 14(3), 222, October (1993) and TUGboat 14(4), 395--419, December (1993).

Citation labels must contain only these characters:

ABCDEFGHIJKLMNOPQRSTUVWXYZ

abcdefghijklmnopqrstuvwxyz

0123456789

:-+/.?\_

**citesub** will continue processing, but since only input tokens containing the above set of characters are candidates for substitution, such erroneous labels will not be substituted.

- *No label substitutions found in file.* This is not necessarily an error, but might be. **citesub** will then simply copy its input to its output.

## SEE ALSO

**awk**(1), **bibcheck**(1), **bibclean**(1), **bibextract**(1), **bibjoin**(1), **biblabel**(1), **biblex**(1), **biborder**(1), **bibparse**(1), **bibsort**(1), **bibtex**(1), **bibunlex**(1), **sed**(1).

## AUTHOR

Nelson H. F. Beebe, Ph.D.

Center for Scientific Computing

University of Utah

Department of Mathematics, 322 INSCC

155 S 1400 E RM 233

Salt Lake City, UT 84112-0090

USA

Email: [beebe@math.utah.edu](mailto:beebe@math.utah.edu), [beebe@acm.org](mailto:beebe@acm.org),

[beebe@computer.org](mailto:beebe@computer.org), [beebe@ieee.org](mailto:beebe@ieee.org) (Internet)

WWW URL: <http://www.math.utah.edu/~beebe>

Telephone: +1 801 581 5254

FAX: +1 801 585 1640, +1 801 581 4148

## AVAILABILITY

**citesub** is freely available; its master distribution can be found in the **biblabel** package at

`ftp://ftp.math.utah.edu/pub/tex/bib/`

`http://www.math.utah.edu/pub/tex/bib/`

in the files

`biblabel-x.yy.jar`

`biblabel-x.yy.tar.gz`

`biblabel-x.yy.zip`

*biblabel-x.yy.zoo*

where *x.yy* is the current version. Each of the popular archive format unpacks into an identical distribution tree.

That site is mirrored to several other Internet archives, so you may also be able to find it elsewhere on the Internet; try searching for the string *biblabel* at one or more of the popular Web search sites, such as

```
http://search.microsoft.com/
http://www.altavista.com/
http://www.dejanews.com/
http://www.dogpile.com/
http://www.euroseek.net/
http://www.excite.com/
http://www.go2net.com/
http://www.google.com/
http://www.hotbot.com/
http://www.infoseek.com/
http://www.inktomi.com/
http://www.lycos.com/
http://www.northernlight.com/
http://www.snap.com/
http://www.stpt.com/
http://www.websmostlinked.com/
http://www.yahoo.com/
```

## COPYRIGHT

```
#####
#####
#####
###
###      citesub: substitute standardized BibTeX citation labels      ###
###
###      Copyright (C) 1994, 1996, 1997, 2001 Nelson H. F. Beebe      ###
###
### This program is covered by the GNU General Public License (GPL), ###
### version 2 or later, available as the file COPYING in the program ###
### source distribution, and on the Internet at                        ###
###
###      ftp://ftp.gnu.org/gnu/GPL                                     ###
###
###      http://www.gnu.org/copyleft/gpl.html                         ###
###
### This program is free software; you can redistribute it and/or    ###
### modify it under the terms of the GNU General Public License as   ###
### published by the Free Software Foundation; either version 2 of   ###
### the License, or (at your option) any later version.              ###
###
### This program is distributed in the hope that it will be useful,   ###
### but WITHOUT ANY WARRANTY; without even the implied warranty of   ###
### MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the    ###
### GNU General Public License for more details.                     ###
###
### You should have received a copy of the GNU General Public        ###
### License along with this program; if not, write to the Free       ###
### Software Foundation, Inc., 59 Temple Place, Suite 330, Boston,    ###
### MA 02111-1307 USA                                                 ###
#####
```

#####  
#####  
#####