NAME

bibextract − extract BibTeX entries from a list of .bib files

SYNOPSIS

bibextract *keyword-regexp value-regexp bibfile(s)*

DESCRIPTION

bibextract extracts from a list of BIBTEX *.bib* files those bibliography entries that match a pair of specified regular expressions, sending them to *stdout*, together with all BIBTEX "@Preamble{...}" commands, and just those "@String{...}" commands that are actually used by the matched entries.

If no bibliography files are specified on the command line, then *stdin* is read instead, so that **bibextract** can be used in a UNIX pipeline.

The order of entries, and spacing within "@Name{...}" text, is preserved exactly. Successive entries are separated by a single blank line.

The first regular-expression pattern, *keyword-regexp*, is used to select which *"keyword = "value""* pairs to examine further; it matches against the keyword part only. It may include alternate keywords separated by vertical bar, such as "author|editor". If it is an empty string, then the entire bibliographic entry text, including the entry type name, is examined.

The second regular-expression pattern, *value-regexp*, is used to further select from the value strings of *"keyword = "value""* pairs the bibliography entries to be output. It too may contain alternates separated by vertical bar, such as "brown|smith". The selection algorithm therefore consists of the logical AND of match successes against the keyword and value strings.

Letter case is ignored in regular-expression matches, so that "Brown|Smith", "BROWN|smith", and "brown|smith" are equivalent. The original letter case of the output entries is always preserved.

If the input BIBTEX data comes from files named on the command line, each output entry will contain a final key/value pair of the form:
  bibsource = "file://hostname/FILENAME",
The value string is a World-Wide Web Uniform Resource Locator, where FILENAME is the full path name of the source file in which the entry was found. Such lines are silently ignored by standard BIBTEX styles, so they are harmless, but they help to track the origin of bibliography entries.

If you don't want the *bibsource* lines to be added, simply supply the BIBTEX file from *stdin*.

bibextract can be used to extract from a large BIBTEX bibliography data base just those bibliography entries that match a particular pair of regular expressions.

bibextract expects the bibliography file(s) to be consistently formatted in the style produced by **bibclean**(1), which allows use of simple pattern matching to recognize the required entries.

EXAMPLES

Here are some examples:

Extract all entries mentioning chaos in any field:

> **bibextract** *"" "chaos" bibfile(s) >new-bibtex-file"*

Extract entries with names Brown or Smith occurring in either of the author or editor fields:

> **bibextract** *"author|editor" "brown|smith" bibfile(s) >new-bibtex-file*

Extract entries for titles containing the letter 'z' anywhere after a vowel; note that single quotes are necessary to provide the necessary protection from shell expansion:

> **bibextract** *"title" '[aeiou].*z' bibfile(s) >new-bibtex-file*

Extract all conference proceedings entries:

> **bibextract** *"" '@proceedings' bibfile(s) >new-bibtex-file*

BUGS

bibextract is not smart enough to incorporate BibTeX cross references unless they are themselves matched by the specified regular expression.

That feature should be added.

SEE ALSO

bibcheck(1), **bibclean**(1), **bibdup**(1), **bibjoin**(1), **biblabel**(1), **biblex**(1), **biborder**(1), **bibparse**(1), **bibsort**(1), **bibtex**(1), **bibunlex**(1), **citesub**(1), **citetags**(1), **latex**(1), **gawk**(1), **nawk**(1), **tex**(1).

FILES

/usr/local/share/lib/bibextract/bibextract.awk	**nawk**(1) program for tag extraction.
/usr/local/bin/bibextract	user-callable shell script to invoke **nawk**(1).

AUTHOR

Nelson H. F. Beebe
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
Tel: +1 801 581 5254
FAX: +1 801 585 1640, +1 801 581 4148
Email: beebe@math.utah.edu, beebe@acm.org, beebe@ieee.org (Internet)
WWW: http://www.math.utah.edu/~beebe