# bibupdate

Automatically update the entries of a bibtex file

**usage: bibupdate [-h|-H] [-a] [-c] [-f] [-i FIELDS] [-l LOG] [-m | -M] [-q] [-r]** [-w LEN] bibtexfile [outputfile]

This is a command line tool for updating the entries in a BibTeX file using mrlookup. By default bibupdate tries to update the entry for each paper in the BibTeX file unless the entry already has an `mrnumber` field (you can disable future checking of an entry by giving it an empty `mrnumber` field).

**Options**:

```
-a, --all             update or validate ALL BibTeX entries
-c, --check_all       check/verify all bibtex entries against a database
-k, --keep_fonts      do NOT replace fonts \Bbb, \germ and \scr in titles
-h, --help            show this help message and exit
-H, --Help            print full program description
-i FIELDS, --ignored-fields FIELDS
                      a string of bibtex fields to ignore
-l LOG, --log LOG     log messages to specified file (defaults to stdout)
-o  --overwrite       overwrite existing bibtex file
-q, --quieter         print fewer messages
-w LEN --wrap LEN     wrap bibtex fields to specified width

-m, --mrlookup        use mrlookup to update bibtex entries (default)
-M, --mathscinet      use mathscinet to update bibtex entries (less flexible)
```

**Note:** As described below, you should check the new file for errors before deleting the original version of your BibTeX file.

By default, bibupdate does not change your original database file. Instead, it creates a new file with the name *updated_file.bib*, if your original file was *file.bib*. It is also possible to have it replace your current file (use carefully!), or to specify an output file.

BibTeX is widely used by the LaTeX community to maintain publication databases. This script attempts to add missing fields to the papers in a BibTeX database file by querying mrlookup and getting the missing information from there. This is not completely routine because to search on mrlookup you need either the authors or the title of the article and both of these can have non-standard representations. If the article is already published then it is also possible to use the publication year and its page numbers. To search on mrlookup we:

- use the authors (can be problematic because of accents and names with von etc)
- use the page numbers, if they exist
- use the year only if there are no page numbers and this is NOT a preprint
- use the title if there are no page numbers (or this is a book)

If there is a unique (good, non-fuzzy) match from mrlookup then bibupdate replaces all of the current fields with those from mrlookup, except for the citation key. The values of any fields that are not specified by mrlookup, such as `eprint` fields, are retained. By default, a message is printed whenever existing fields in the database are changed. If the title of the retrieved paper does not (fuzzily) match that of the original article then the entry is NOT updated and a warning message is printed.

Although some care is taken to make sure that the new BibTeX entries correspond to the same paper that the original entry referred to there is always a (small?) chance the new entry corresponds to an entirely different paper. In my experience this happens rarely, and mostly with unpublished manuscripts. In any case, before you delete your original BibTeX file *you are strongly advised to check the updated file BibTeX file carefully for errors!*

To help the user to compare the updated fields for each entry in the BibTeX file the program prints a detailed list of all changes that are made to existing BibTeX fields (any new fields that are added are not printed). Once bibupdate has finished running it is recommended that you compare the old and new versions of your database using programs like *diff* and *tkdiff*.

As bibupdate calls mrlookup this program will only be useful if you have papers in your database that are listed in MathSciNet. As described below it is also possible to call MathSciNet directly, however, this is less flexible because the `mrnumber` field for each paper is required.

I wrote this script because I wanted to automatically add links to journals, the arXiv and DOIs to the bibliographies of my papers using hyperref. This script allowed me to add the missing urls and DOI fields to my BibTeX database. As a bonus the script helped me to correct many minor errors that had crept into my BibTeX file over the years (for example, incorrect page numbers and publication years). Now I use the program to automatically update the preprint entries in my database when the papers appear in MathSciNet after they are published.

## Options and defaults

**-a, --all** Update or validate ALL BibTeX entries

By default bibupdate only checks each BibTeX entry with the mrlookup database if the entry does *not* have an `mrnumber` field. With this switch all entries are checked and updated.

-c --check_all Check/validate all bibtex entries against a database

Prints a list of entries in the BibTeX file that have fields different from those given by the corresponding database. The original BibTeX file is not changed.

**-k, --keep_fonts** do NOT replace fonts Bbb, germ and scr in titles

The BibTeX entries generated by mrlookup use \Bbb, \germ and \scr for the \mathbb, \mathfrak and \mathscr fonts. By default, in the *title* fields, these fonts specifications are automatically changed to the following more LaTeX friendly fonts:

- \Bbb X --> \mathbb{X}
- \scr X --> \mathcal{X}
- \germ X --> \mathfrak{X}

By using the -k option the fonts specification used by MathSciNet are used.

**-i FIELDS, --ignored-fields=FIELDS** A string of BibTeX fields to ignore when writing the updated file

By default bibupdate removes the following fields from each BibTeX entry:

- coden
- mrreviewer
- fjournal
- issn

This list can be changed using the -i command line option:

```
bibupdate -i "coden fjournal" file.bib   # ignore coden and fjournal
bibupdate -i coden -i fjournal file.bib  # ignore coden and fjournal
bibupdate -i "" file.bib                 # do not ignore any fields
```

**-l LOG, --log LOG** Log output to file (defaults to stdout)

Specify a log filename to use for the bibupdate messages.

-m --mrlookup Use mrlookup to update bibtex entries (default)
-M --mathscinet Use mathscinet to update bibtex entries

By default mrlookup is used to update the BibTeX entries in the database. This has the advantage of being a free service provided by the American Mathematical Society. A second advantage is the more flexible searching is possible when mrlookup is used. It is also possible to update BibTeX entries using MathSciNet, however, these searches are currently only possible using the `mrnumber` field (so this option only does something if combined with the --all option or the -check-all-option).

**-o** **--overwrite** Overwrite the existing bibtex file with the updated version

Replace the existing BibTeX file with the updated file. A backup version of the original BibTeX is made with a .bak extension. it is also possible to specify the output filename as the last argument to bibupdate.

**-q, --quieter** Print fewer messages

There are three levels of verbosity in how bibupdate describes the changes that it is making. These are determined by the q-option as follows:

```
bibupdate     bibfile.bib    (Defalt) Report all changes
bibupdate -q  bibfile.bib    (Warning mode) Only print entries that are changed
bibupdate -qq bibfile.bib    (Quiet mode) Only printer error messages
```

By default all changes are printed (to stdout, although a log file can be specified by the -l option). In the default mode bibupdate will tell you what entries it changes and when it *is not* able to find the paper on the database (either because there are no matches or because there are too many). If it is not able to find the paper and bibupdate thinks that the paper is not a preprint then it will mark the missing entry with an exclamation mark, to highlight that it thinks that it should have found the entry in mrlookup but failed. Here is some sample output:

```
------------------------------
? did not find Webster:CanonicalBasesHigherRep=Canonical bases and higher repre
++++++++++++++++++++++++++++++
+ updating Weyl=
+ publisher: Princeton University Press
+          -> Princeton University Press, Princeton, NJ
------------------------------
? did not find Williamson:JamesLusztig=Schubert calculus and torsion
------------------------------
! did not find QSAII=On Quantitative Substitutional Analysis
```

Each bibtex entry is identified by the citation key and the (first 50 characters of the sanitised) document title, as specified by your database. Of the three missed entries above, bibupdate thinks that the first and third are preprints (they are not marked with an !) and that the final article should already have been published. With the entry that bibupdate found, only the publisher field was changed to include the city of publication.

In *warning mode*, with the -q option, you are "warned" whenever changes are made to an entry or when the paper is not found in the external datbase. That is, when papers are found (with changes) or when they are missed and bibupdate thinks that they are not preprints. In *quiet mode*, with the -qq option, the program only reports when something goes wrong.

-w LEN --wrap LEN Wrap bibtex fields to specified width

Limits the maximum line length in the output BibTeX file. In theory this is supposed to make it easier to compare the updated BibTeX file with the original one, however, in practise this doesn't always work.

## Known issues

bibupdate_ reads BibTeX files using a small number of regular expressions so there may be be some corner cases where it fails to extract all of the field entries.

There are a small number of cases where bibupdate fails to correctly identify papers that are listed in MathSciNet. These failures occur for the following reasons:

- Apostrophes: Searching for a title that contains, for example, "James's Conjecture" confuses mrlookup.
- Ambiguous spelling: Issues arise when there are multiple ways to spell a given author's name. This can often happen if the surname involves accents (such as Koenig and K\"onig). Most of the time accents themselves are not a problem because the AMS is LaTeX aware.
- Pages numbers: electronic journals, in particular, often have strange page numbers (for example "Art. ID rnm032, 24"). bibupdate assumes that page numbers are always given in the format like 4--42.
- Occasionally MathReviews combines two or more closely related articles. This makes it difficult to search for them.

All of these problems are due to idiosyncrasies with mrlookup so there is not much that we can do about them.

### Installation

You need to have Python installed. In principle, this program should work on any system that supports Python, however, I only promise that it will work on an up-to-date mac or Linux system. In the event that it does not install I may not be able to help you as I will not have access to your system.

From the command line type:

```
pip install bibupdate
```

Instead of pip, you should also be able to use easy_install. The program should run on python 2.7 and 2.8...I haven't tried python3. You can also clone or download the git repository and work directly with the source.

### Support

This program is being made available primarily on the basis that it might be useful to others. I wrote the program in my spare time and I will support it in my spare time, to the extent that I will fix what I consider to be serious problems and I may implement feature requests.

### To do

- More intelligent searches using MathSciNet.
- Add lookup using MRef and, when an entry is not found, allow additional searches
- Add an rc file?
- Fix the wrapping of bibtex fields.
- Add interface to the arXiv using `http://arxiv.org/help/api` or `http://arxiv.org/help/oa`.

### Author

Andrew Mathas