Bibcheck

Manual for Version 1.4 – last updated on 18 January 2024

Bibcheck is a Lua-based script that checks each \bibitem of a .tex file against MathSciNet, zbMATH, Crossref, and arXiv, and creates the bibliography based on a .bst file.
Bibcheck is available at GitHub. For bug reports, comments and suggestions contact Tamas Bori.

Contents

1 Installation

1.1 Install Lua v5.1 (or higher) on Windows

1. Download LuaForWindows_v5.1.5-52.exe (27.8 MB).
2. Run the .exe file and always click 'accept/next'.

1.2 Install Wget (v1.14 or higher) on Windows

1. Download the .exe (!) file of Wget (32-bit or 64-bit). (Tested with Version 1.21.)
2. Copy `wget.exe` to, say, `C:\Program Files\wget-1.21.4-win64\` or any other folder.
3. Add Wget to the Windows PATH:
   (a) Open the Start Search, type in 'env' and choose 'Edit the system environment variables'.
   (b) Click the 'Environment Variables' button.
   (c) Under 'System Variables' find the row with 'Path' in the first column and click edit.
   (d) Click 'New' and type in the new path, say, `C:\Program Files\wget-1.21.4-win64`
   (e) Dismiss all of the dialogs by choosing OK. Your changes are saved.
4. To check whether the installation was successful, open a command terminal (by typing 'cmd' in the search menu) and type

```
wget -qO-
"https://mathscinet.ams.org/mathscinet-mref?dataType=bibtex&ref=J. Whitehead,
On 2-spheres in 3-manifolds. Bull. Amer. Math. Soc. 64 (1958), 161--166"
```

or

```
wget -qO-
"https://zbmath.org/citationmatching/match?f=latex&q=J. Whitehead,
On 2-spheres in 3-manifolds. Bull. Amer. Math. Soc. 64 (1958), 161--166"
```

in both cases without any line break. The return should be MR 103473 resp. Zbl 0084.19103.

If this does not work, try again with `wget --no-check-certificate -qO- ...`.

1.3 Install Lua and Wget on Mac/Linux

1. We recommend to use the package manager Homebrew.
2. See Method #1 on how to install Homebrew and Wget.
3. After installing Lua, the two Lua libraries 'LuaFileSystem' and 'Penlight' might be missing. Then run:

```
brew install luarocks
luarocks install luafilesystem
luarocks install penlight
```

1.4 Install Bibcheck

1. Create a new folder for Bibcheck.
2. Copy all eight .lua files and the .csv into the Bibcheck folder: `bibcheck.lua`, `config.lua`, `functions.lua`, `journal_abbreviations_mathematics.csv`, `dkjson.lua`, `ftcsv.lua`, `xml2lua.lua`, `xmlhtree.lua`, `XmlParser.lua`. **NEW**
3. It could be (e.g., under Windows 11) that instead of '`wget -qO-`' in `bibcheck.lua` and `functions.lua` you need to use '`wget --no-check-certificate -qO-`'; see the test described in §1.2.4. Bibcheck is doing this from version 1.4 onwards automatically (i.e., adds the `--no-check-certificate` parameter to `wget` if there was no output without it, assuming it was due to an SSL error). **NEW**

2 How to use Bibcheck

2.1 Case 1: \bibitem's

If your .tex file contains a `thebibliography` environment,[1] proceed as follows:

1. Open the command terminal, go to the paper's directory and write

   ```
   lua path\bibcheck.lua main.tex bstfilename
   ```

   e.g., on Windows:

   ```
   lua C:\tools\bibcheck\bibcheck.lua main.tex ems
   ```

   Note: The spaces in the original tex filename will be replaced by underscores.

2. The script runs for a few seconds. What happens is the following:
   - Every `\bibitem` in `main.tex` is compared with MathSciNet.
   - If there is a match, this **match** is written into a .bib file.
   - Preprint entries are compared with arXiv.
   - If there is a match, then this **match** is written into the .bib file.
   - If there is no match in any of MathSciNet and arXiv, then the **original entry** is written into the .bib file.
3. Next, an automatic run of `latex` and `bibtex` creates a .bbl file. This .bbl file is pasted into the original .tex file `main.tex`., which is then renamed `main_bibchecked.tex`.
4. Now you can start editing the new ('bibchecked') .tex file. For this, *it is essential to check each* `\bibitem`:

[1] This is the default case for Bibcheck.

- If there was a **MathSciNet or arXiv match**, then the original entry appears as a comment (%) below the match. It is important to compare the original entry and the match because sometimes MathSciNet or arXiv may deliver a wrong result. However, a quick comparison of, e.g., the page range provides clarity. Or, if the original entry has an MR or arXiv number, you can compare it with that of the match. Aiming to help the critical comparison, in the comment field some additional text information from MathSciNet is added by Bibcheck from v1.4.
- If there was **neither a MathSciNet nor an arXiv match**, then the entry is now at the beginning of the bibliography. This means, you have to (a) sort this entry and (b) format it by hand. This is a disadvantage of the tool; but it happens to only a few entries.

5. In addition, each `\bibitem` is compared with zbMATH too. If there is a match with high enough relevance score, then its zbMATH number is added to the reference. Should there be a match with too low relevance score, this match and its relevance score are also listed as a comment under the corresponding entry. Again: you should compare the original entry and the zbMATH match in both directions: to ensure that all inserted zbMATH numbers correspond to the right publication, indeed, and conversely, that zbMATH numbers omitted (because of their low relevance score) are mismatches, indeed. As a result, you may like to re-insert some omitted zbMATH numbers from the comment manually.

6. That's it!

2.2 Case 2: BibTeX

If there is a .bib file, say `bibsource.bib`, proceed as follows:

1. In the main .tex file write:

```
\bibliographystyle{ems}
\bibliography{bibsource}
```

2. Run `bibtex` to create `bibsource.bbl`.

3. Copy the content of `bibsource.bbl` into the .tex file.

4. Proceed as in Case 1.[2]

2.3 Case 3: amsrefs

If the bibliography is prepared using the `amsrefs` package, proceed as follows:

1. Copy all `\bib` entries into a new .ltb file, say `source.ltb`.

2. Create a temporary .tex file, say `temp.tex`, consisting of the following lines:[3]

```
\documentclass{article}
\usepackage{amsrefs,ltb2bib}
\begin{document}
  \begin{bibdiv}
    \begin{biblist}
      \bibselect*{source}
    \end{biblist}
  \end{bibdiv}
  \writebib{source}{bibsource}
\end{document}
```

3. Compile `temp.tex`. This will create a .bib file `bibsource.bib` containing all `\bib` entries from `source.ltb`.

   Warning: The resulting .bib file is not perfect: wrong brackets around author names, wrong structure of `@incollection` entries, capitalization of proper names is lost (due to missing brackets), and maybe more. However, these conversion errors should have little effect on the hit rate in MathSciNet and zbMATH.

4. In the *original* .tex file (say, `main.tex`), replace the bibliography by

[2] Isn't this procedure totally cumbersome? Yes, it is. Maybe someday a new version of Bibcheck will make it better.

[3] The reason we create a new .tex file is that there is a conflict with the `\title` command: if the .tex file contains `\title`, then all article and book titles disappear when converting ltb to bib.

```
\bibliographystyle{ems}
\bibliography{bibsource}
```

5. Run `bibtex`. But before you do this, you must remove `\usepackage{amsrefs}` (otherwise the .bib file will be in amsrefs format).
6. Copy the content of `main.bbl` into the .tex file.
7. Now you can finally use Bibcheck as in Case 1.
8. **Warning:** What Bibcheck considers to be the "original entry" is of course only the entry from the defective .bib file created in step 2.3.3. Therefore, in cases of doubt, the original *manuscript* should be consulted for comparison.[4]

2.4 Use a .bat file on Windows

Instead of using the command terminal, you can create and use a batch file (.bat):

1. Create, using any text editor, a file `bibcheck.bat` containing the following four lines:

   ```
   @echo off
   chcp 65001
   lua "C :\...\ bibcheck.lua" "%~f1" ems
   pause
   ```

   Here `C :\...\ bibcheck.lua` is the full path of `bibcheck.lua`. Keep the quotation marks!
2. Create a desktop shortcut of that .bat file.
3. Now drag and drop your .tex file onto the desktop shortcut. Easy!

3 What to pay attention to

1. Before you bibcheck a .tex file, you should
   - replace each `\bysame` by the respective authors.
2. Be aware of **mismatches**. They mainly appear when
   - the original entry is a preprint (e.g., arXiv) and MathSciNet returns a journal article with the same title;
   - the original entry is the original version (e.g., in Russian) and MathSciNet returns the English translation, or the other way around;
   - the original entry and the MathSciNet match have a very similar title, maybe even the same author, but are different;
   - the original entry is a preprint without arXiv ID, and arXiv returns another paper's data. **NEW**

   In such cases please check the comments below the `\bibitem` carefully: the original entry, the NOTE entry of the MathSciNet response, the data returned by zbMATH, and the MathSciNet TEX type response. **NEW**
3. The bibliography in `main_bibchecked.tex` is based mostly on MathSciNet. Here are some known problems with MathSciNet:
   - Book **series** titles are often not abbreviated. Although Bibcheck from version 1.4 is prepared for handling this, it is recommended to check, and if necessary, abbreviate these titles according the MathSciNet standard.
   - Since our style asks to write 'Springer' instead of 'Springer-Verlag', and similarly, 'Birkhäuser' instead of 'Birkhäuser Verlag', Bibcheck version 1.4 and above takes care of removing '**Verlag**' automatically, but there is still a possibility that you may need to remove it manually in some cases. (Also consider removing unnecessary Press, Publishing, Co., GmbH, Verlag, Editions, Editore, etc. from the publisher's name.) **NEW**

[4] Instead of using this ltb2bib approach, it might make more sense to use the `\bib` entries from amsrefs as the input for Bibcheck (i.e. without the detour via BIBTEX). One probably only has to remove the fields names (author=, title=, etc.) and replace `\bib` with `\bibitem`. Maybe Bibcheck will be able to do that in one of the next versions.

- **Publisher** names (for book-like entries) are often abbreviated. Although according to our new style these can be kept in the abbreviated or full format MathSciNet responds, it is worth checking these however, and replace them in either direction, for consistency (at least within an article). NEW
- Some **Astérisque** papers are of entry type `@incollection` containing a field `journal`:

```
@incollection {Shelstad1989,
     AUTHOR = {Shelstad, D.},
      TITLE = {A formula for regular unipotent germs},
       NOTE = {Orbites unipotentes et repr{\'{e}}sentations, II},
    JOURNAL = {Ast{\'{e}}risque},
   FJOURNAL = {Ast{\'{e}}risque},
     NUMBER = {171-172},
       YEAR = {1989},
      PAGES = {275--277},
       ISSN = {0303-1179},
    MRCLASS = {22E35 (11F70 11R39 11S37)},
  MRNUMBER = {1021506},
MRREVIEWER = {Joe Repka},
}
```

  Officially, `journal` is not an allowed field for `@incollection`. Thus, .bst files (such as `ems.bst`) will not print `journal`, they rather expect a field `booktitle`. This is solved by Bibcheck above version 1.4 by converting the `@incollection` type to `@article`. The book titles may be listed in the comment below the `\bibitem`, after `%__ MathSciNet Bib NOTE:`. NEW

4. **zbMATH**: The hit rate with zbMATH is not as high as with MathSciNet. This means: Not all publications listed in zbMATH are found, and when they are, there is a possibility that the Zbl number returned by zbMATH belongs to an another paper. In this case, a manual search using the zbMATH Citation Matcher often leads to success. The hit probability increases if you only enter *parts* of the `\bibitem`. Whether or not a Zbl number is included by Bibcheck (from v1.4 onwards) in a `\bibitem` entry is decided by the 'relevance score' returned by zbMATH together with every match. It is an indicator measuring how likely the match is correct. If the score is lower than a certain limit, the `\Zbl{number}` (besides the score) will be included in the prefix of the zbMATH comment below the `\bibitem`.

5. **Crossref**: If the MathSciNet match does not contain any DOI, the original entry is sent to the Crossref API.

   Crossref always returns *one* DOI, which will be added to `main_bibchecked.tex` as a comment together with the so-called 'relevance score' (again, an indicator of how likely the DOI provided by Crossref really is the DOI you are looking for) and the title of the paper.

   Treat the DOIs provided by Crossref with caution: they must not be adopted without verification.

   Only if the DOI is the correct one, copy the complete URL around the title:

   `\href{https://doi.org/10.4171/ggd/702}{On the Basilica operation}`

   If neither the MathSciNet API nor the Crossref API provide a matching DOI, then it is quite likely that the publication has no DOI. However, you can still manually search for one, for example using the Crossref Metadata Search.

6. **arXiv**: If MathSciNet API does not return a match, and the `\bibitem` seems to be a preprint with an arXiv ID given, Bibcheck queries the arXiv API for that ID. If there is a match, the original entry will be replaced by the data returned (with author, title, publication year information, the latter completed with the first and the current – or the requested – version and IDs). NEW

   If MathSciNet API does not return a match, and the `\bibitem` seems to be a preprint without an arXiv ID given, Bibcheck queries the arXiv API for the full text of the entry, and the original entry will be replaced by the data of the first match returned (with author, title, publication year information, the latter completed with the first – and the current, if there are more – version and ID). In this case there is a greater chance that the paper returned by the arXiv API is different from the original one, so it must be checked carefully.

   Matched arXiv entries will also be properly formatted and sorted. One should check the title of these entries as, e.g., some proper names could be typeset in lowercase (despite that Bibcheck tries to figure out whether some initial letters should be capitalized).

   Occurrences of `\eprint` are replaced by `\arXiv`, and the definition of the latter is also included in the preamble of the bibliography.

4 The config file

1. When running `bibcheck`, you may skip the last argument (BSTFILENAME). This will activate the default bibliography style as defined in `config.lua`:

   `M.bibstyle = 'ems'`

2. The temporary .bib file contains the MR and zbMATH numbers:

   `zblnumber = {1460.22007}, mrnumber = {4028458}`

   If your .bst file knows how to deal with such entries (e.g., `ems.bst` does), the final `\bibitem` will contain

   `\Zbl{1460.22007} \MR{4028458}`

   The .tex/.cls/.sty file, in turn, must have a definition for the `\Zbl` and `\MR` commands.

   This means, if you do not need any zbMATH numbers, you may either

   - use a .bst file that ignores the field `zblnumber`, or
   - define `\newcommand\Zbl[1]{}`, or
   - change the respective line in `config.lua` to

     `M.checkzbMATH = false`

   The latter option disables the comparison with zbMATH and makes Bibcheck run much faster!

3. If you are not satisfied with the threshold for zbMATH relevance score used by Bibcheck to decide which `\Zbl`'s are included in the `\bibitem`'s automatically, you can adjust it by changing the value of the following setting in `config.lua`: **NEW**

   `M.ZblRelevanceScoreThreshold = 5.9`

   Based on the experience so far, the most reliable result is obtained with a value between 5 and 6, but one still need to check the `%__ zbMATH` comment below the `\bibitem` entry carefully, especially with relevance scores lower than 7.

4. If you do not need DOIs,

   - use a .bst file that ignores the field `doi`, *and*
   - change the respective line in `config.lua` to

     `M.checkCrossref = false`

5. By default the resulting .tex file has the suffix `_bibchecked`. This value can be changed in `config.lua`:

   `M.suffix = '_bibchecked'`

6. If you want to *keep* the temporary .bib file, remove the line `'.bib'` from the `M.remove_files` list in `config.lua` (or add the comment sign `--`):

   ```
   M.remove_files = {
     --  '.bib',
     '.bbl',
     '.aux',
     '.log',
     '.dvi',
     '.blg',
   }
   ```

   However, we strongly recommend *not* to continue working with the .bib file, but always with the .tex file (`main_bibchecked.tex`) since the .tex file contains important comments (`%`), which are missing in the .bib file.

7. If you do not want to check preprints against arXiv, change the respective line in `config.lua` to **NEW**

   `M.checkArXiv = false`

   (It could make Bibcheck run a bit faster only for bibliographies containing quite a few preprints.)

8. The `note` field of the .bib entry from MathSciNet, and the TEX type result of MathSciNet are added as comments under the matched entries. Both may hold additional information potentially missing in the `\bibitem`. If you do not want to receive these comments, change the respective lines in `config.lua` to `false`: NEW

   ```
   M.saveMRNote = false
   M.checkMRTeX = false
   ```

   The latter setting could speed up Bibcheck slightly.

5 Changelog

- Version 1.1
  - `split_at_bibitem`: removes blank lines inside any `\bibitem`.
  - `make_bib` and `space_warning`: removes spaces in `\bibitem` labels and adds a warning in `main_bibchecked.tex`.
  - `make_bib` and `undress`: curly braces and tex commands of the form `\XYZ` removed before the `\bibitem` is sent to the zbMATH Citation Matcher.
  - `zbl_ID`: Bug fixed.
  - `escape_percents`: Function added; otherwise strings such as `\%3` (e.g., in URLs) would not work in Lua 5.1.
- Version 1.2
  - `make_bib` and `undress`: further simplify each `\bibitem` before it is sent to the zbMATH Citation Matcher. This way we try to increase the zbMATH hit rate.
  - `make_bib`: DOIs added. If the MathSciNet match does not contain any DOI, the original entry is sent to the Crossref API.
- Version 1.3
  - `make_bib`: Exit immediately if two `\bibitem`'s have the same label (modulo capitalization).
- Version 1.4
  - `crossref_info`: Fixed a bug resulting script termination if Crossref returned a match without title.
  - `make_bib`, `get_crossref`: David Kolf's `dkjson.lua` included in the package.
  - `make_bib`, `get_arxiv` and `make_tex`: Implementing the query of arXiv API, processing its result and related modifications (also using `@manoelcampos`'s `xml2lua`, included).
  - `make_bib`: Fixing .bib entries (undefined or ancient TEX commands, font styles and accents corrected, improper element values removed [translations in `series`, doubled `year` values, Verlag from `publisher`] or fixed [page prefixes for e-prints replaced by 'article no.', ProQuest theses converted to `@phdtheses` with `school`, `@incollections` with *Astérisque* as `journal` converted to `@article`, No. removed from the beginning of `volume`'s and `number`'s, `mrnumber` with less than 7 digits filled up with leading zeros, one-letter abbreviation of the first or last word of `journal` is glued respectively to the next or previous word with non-breaking space (~)]).
  - `zbl_info`, `zbl_ID`: Include `zblnumber` value in .bib entry based on the relevance score returned by the zbMATH response, threshold can be adjusted in config, score (and if it is lower than the threshold, the `\Zbl` number) put into the zbMATH comment prefix below the `\bibitem`.
  - `make_bib`, `add_comments`, `mr_note_info`, `mr_tex_info`: The `note` entry of the MathSciNet response, as well as the MathSciNet TEX type response (requested based on a setting in `config.lua`) are saved (by default) as comments below the `\bibitem`.
  - `make_bib`: Book series are checked against the AMS abbreviated series list CSV (included in the Bibcheck package [source: JabRef], also downloaded automatically if the local file doesn't exist or was modified more than 180 days ago) (also using `@FourierTransformer`'s `ftcsv`, included), and if matched (case-insensitively), the series' full titles are replaced by their abbreviated versions.
  - `execute`: If `wget` produces no output, Bibcheck will add the `--no-check-certificate` parameter to it and then try to download the same URL, assuming the lack of output was due to an SSL error, hopefully fixing an issue automatically that arises sometimes on certain platforms/installations.

Registered company: European Mathematical Society – EMS – Publishing House GmbH
Trade register number: HRB 204966 B · Registration court: Amtsgericht Charlottenburg · Managing director: Dr. André Gaul