bibneat

Making your BibTeX files nice and neat!

Lorenzo Peri

June 8, 2025

Abstract

`bibneat` is a command-line tool for cleaning, merging, and supercharging your BibTeX files. It ensures your bibliographies are consistent, beautiful, and always compile. This document describes its features, usage, and options in detail, and introduces its speedy sibling, `bibgrab`.

1 Introduction

`bibneat` is here to help you:

- Format your .bib files beautifully and consistently
- Merge multiple bibfiles with ease (no more duplicate headaches)
- Shorten and clean up entries by filtering out fields not used by BibTeX
- Handle Unicode like a pro:
  - Canonical (NFC) and compatibility-optimized (NFKC) Unicode normalization
  - Convert Unicode characters (accents, symbols, etc.) to their equivalent LaTeX encoding
- Connect to arXiv.org and doi.org APIs to:
  - Check and validate bibliography entries
  - Fill in missing fields automatically
  - Detect if an arXiv entry has been published and (optionally) replace it with the published version

2 Guarantees

`bibneat` makes two pinkie promises:

- If your .bib file compiled before, it will compile after bibneat touches it. Guaranteed.
- Your bibliography keys are sacred. bibneat will **never** change them, so every `\cite{}` in your documents will keep working, even if entries are modified or replaced.

3 Quick Start

Basic Usage

```
bibneat [options] -i input1.bib -i input2.bib ... -o output.bib
```

4 Options for `bibneat`

`-h, --help` Displays this message.

`-i` Input .bib files for bibneat.

`-o` Output .bib file (default `bibneat.bib`). Can be the same as one of the input files.

`-a, --append`
Append to output file rather than overwriting it.

`-kb, --keep-bibtex`
Keep only fields that BibTeX knows about (delete the others).

`-xc, --arxiv-check`
Check the arXiv API for each arXiv entry (validates the arXiv ID and checks whether a related publication exists). Do not alter the entry.

`-xr, --arxiv-replace`
Replace each arXiv entry with the BibTeX received from the arXiv API. Do not follow known DOIs for publications.

`-xf, --arxiv-follow`
Replace each arXiv entry with the BibTeX received from the arXiv API. If there is a known publication DOI, replace the arXiv entry with the published version.

`-dc, --doi-check`
Check the doi.org API for each entry with known DOI (validates the DOI). Do not alter the entry.

`-dr, --arxiv-replace`
Replace each entry with known DOI with the BibTeX received from the doi.org API.

`-t, --timeout`
Timeout (in seconds) for each API request. Default 20s.

`--connection-timeout`
Timeout (in seconds) for the CURL connection stage (resolution+handshake) for each API request. Default 15s. Must be > timeout to have effect.

`-p, --add-encoding-preamble`
Add UTF8 encoding as BibTeX preamble (inputenc and fontenc packages must be present in latex document preamble to have effect).

`-uc, --unicode-canon-normalize`
Perform canonical unicode normalization of each entry (may fix compilation errors).

`-uk, --unicode-compat-normalize`
: Perform compatibility-optimized unicode normalization of each entry (should fix most compilation errors, but may result in loss of details in non-latin character sets).

`-u2l, --unicode2latex`
: Transforms unicode special characters (accents, symbols, etc.) in their equivalent latex encoding (may fix compilation errors).

5 Examples

- Merge and clean two bibfiles:

  ```
  bibneat -i refs1.bib -i refs2.bib -o merged.bib -kb
  ```

- Normalize unicode and convert to LaTeX:

  ```
  bibneat -i messy.bib -o clean.bib -uc -u2l
  ```

- Check arXiv and DOI entries:

  ```
  bibneat -i myrefs.bib -xc -dc
  ```

6 bibgrab: The Little Sibling

`bibgrab` is a quick tool to fetch BibTeX entries from arXiv or DOI identifiers. Just give it an arXiv ID, arXiv URL, or DOI, and it will print a ready-to-cite BibTeX entry.

6.1 Usage

Basic Usage

```
bibgrab [arxiv-id|arxiv.org url|doi|doi.org url]
```

6.2 Options for `bibgrab`

`-h, --help` Displays this message.

`-x, --arxiv` ArXiv entry to look up. Accepts arXiv IDs (with or without the `arXiv:` prefix) and arxiv.org URLs.

`-d, --doi` DOI or doi.org URL to look up.

`-f, --arxiv-follow`
: For each arXiv entry, if there is a known publication DOI, replace the arXiv entry with the published version.

`-o` Output .bib file (default `bibneat.bib`). Can be the same as one of the input files.

`-no, --no-output`
: Just perform request, do not write .bib file.

`-a, --append`
: Append to output file rather than overwriting it.

`-kb, --keep-bibtex`
: Keep only fields that BibTeX knows about (delete the others).

`-t, --timeout`
: Timeout (in seconds) for each API request. Default 20s.

`--connection-timeout`
: Timeout (in seconds) for the CURL connection stage (resolution+handshake) for each API request. Default 15s. Must be > timeout to have effect.

`-p, --add-encoding-preamble`
: Add UTF8 encoding as BibTeX preamble (inputenc and fontenc packages must be present in latex document preamble to have effect).

`-uc, --unicode-canon-normalize`
: Perform canonical unicode normalization of each entry (may fix compilation errors).

`-uk, --unicode-compat-normalize`
: Perform compatibility-optimized unicode normalization of each entry (should fix most compilation errors, but may result in loss of details in non-latin character sets).

`-u2l, --unicode2latex`
: Transforms unicode special characters (accents, symbols, etc.) in their equivalent latex encoding (may fix compilation errors).

7 Installation

`bibneat` depends on:

- The C++ standard library
- libcurl (for web requests)
- ICU4C (for Unicode and normalization magic)

If all dependencies are set up, just run:

```
make
```

For more details, see `INSTALL.md`.

8 Why bibneat?

Because life's too short for ugly bibliographies.

9 Further Help

For more details, usage instructions, and advanced options, run:

```
bibneat --help
bibgrab --help
```

What does the encoding preamble do?

When you use the `-p` or `--add-encoding-preamble` option, bibneat adds the following to the top of your .bib file:

```
@preamble{"\makeatletter\@ifpackageloaded{inputenc}{\inputencoding{utf8}}{}\@ifpackageloaded
```

This ensures that, if your LaTeX document loads the `inputenc` and `fontenc` packages, your bibliography will be interpreted as UTF-8 and typeset with the T1 font encoding. This helps avoid weird character issues and makes your bibliography more robust to Unicode content.

Example: Upgrading an arXiv entry to a published version

Suppose you have the following entry in `bibliography.bib`:

```
@misc{peri2024polarimetryspinssolidstate,
  title={Polarimetry With Spins in the Solid State},
  author={Lorenzo Peri and Felix-Ekkehard von Horstig and Sylvain Barraud and Christopher J.
  year={2024},
  primaryclass={cond-mat.mes-hall},
  url={https://arxiv.org/abs/2410.17867},
  archiveprefix={arXiv},
  eprint={2410.17867},
}
```

If you run:

```
bibneat -xc -i bibliography.bib -o bibliography.bib
```

`bibneat` will not change the entry, but will print to the terminal:

```
ArXiv entry peri2024polarimetryspinssolidstate has been published with DOI
https://doi.org/10.1021/acs.nanolett.5c01511
Please consider replacing with published version
```

If you run:

```
bibneat -xf -i bibliography.bib -o bibliography.bib
```

`bibneat` will automatically replace the entry with the published version from the DOI API:

```
@article{peri2024polarimetryspinssolidstate,
  title={Polarimetry with Spins in the Solid State},
  author={Peri, Lorenzo and von Horstig, Felix-Ekkehard and Barraud, Sylvain and Ford, Chris
  year={2025},
  journal={Nano Letters},
  month={may },
  doi={10.1021/acs.nanolett.5c01511},
  url={http://dx.doi.org/10.1021/acs.nanolett.5c01511},
  issn={1530-6992},
}
```

Example: Adding a new arXiv reference to your bibliography

Suppose you are writing a paper and all your references are in `bibliography.bib`. You are now reading an AMAZING paper on the arXiv. Just copy the URL and type:

```
bibgrab -x https://arxiv.org/pdf/2410.17867 -a -o bibliography.bib
```

Now you can start citing it!

Example: Checking if an arXiv preprint has a published version

You bumped into this great paper and want to check if there is a published version. Just type:

```
bibgrab -x 2410.17867 -no
```

See if bibgrab can find a published DOI!

Happy (Bib)TeXing!