

The copac-clean utility

Version 0.7

Guido Milanese
guido.milanese@unicatt.it

10th January 2014

Abstract

This standalone utility edits the Copac/BibTeX records, improving some details and adding some useful fields. The programme is written using Snobol for editing, tck/tk for a simple GUI and a bash script as framework. The utility is only available for use under Linux, but could easily be ported to Windows.

Contents

| | | |
|----------|--|----------|
| 1 | Copac-clean: description and features | 2 |
| 1.1 | Editing | 2 |
| 1.2 | Hyphenation | 3 |
| 1.3 | Location/Address | 3 |
| 1.4 | New fields | 3 |
| 1.5 | Final result | 4 |
| 2 | Installation | 4 |
| 2.1 | Dependencies | 4 |
| 2.2 | xclip | 4 |
| 2.3 | Tcl/tk | 4 |
| 2.4 | Snobol | 5 |
| 3 | Installing Copac-clean | 5 |
| 4 | Usage | 6 |
| 4.1 | With GUI | 6 |
| 4.2 | Without GUI | 6 |
| 4.3 | Paste the new version | 9 |
| 5 | Bugs and TODO | 9 |
| 6 | Acknowledgements | 9 |

1 Copac-clean: description and features

Copac (<http://copac.ac.uk>) is one of the best resources for academics, students, librarians: it unifies more the 70 catalogues of libraries of the UK, featuring around «40 million records in 452 languages» (Copac information from <http://copac.ac.uk/faq/>). Copac records can be exported in several formats, including BibTeX. However, some minor details of the Copac/BibTeX format can be improved, and – particularly for Biblatex users – some more fields can be easily added.

1.1 Editing

This is a typical COPAC/BibTeX record:

```
@proceedings{Giannantoni:1996,
author = {Giannantoni, Gabriele. and Gigante, Marcello,},
title= {Epicureismo greco e romano : atti del congresso
internazionale, Napoli, 19-26 maggio 1995},
series = {Elenchos ; 25 },
series = {Elenchos (series) ; 25. },
address = {[Napoli] : },
publisher = {Bibliopolis,},
year = {c1996.},
size = {3 v. (1128 p.) ; 23 cm.},
language = {English ; French ; German ; Italian},
isbn = {ISBN: 8870883620},
note = {Papers from an international congress
held in Naples, 19-26 May 1993.},
note = {Includes bibliographical references (v. 3, p. [1005]-1056)
and indexes.},
note = {Italian, English, French and German contributions.},
keywords = {Epicurus -- Congresses.},
keywords = {Epicurus -- Influence -- Congresses.},
keywords = {Philosophy, Ancient -- Congresses.},
location = {King's College London ; University of London - ULRLS},
}
```

The first problem that any BibTeX user can detect is useless punctuation. For example:

```
author = {Giannantoni, Gabriele. and Gigante, Marcello,},
```

should be improved as follows:

```
author = {Giannantoni, Gabriele and Gigante, Marcello},
```

The same applies to all the other entries (see e.g. year, address, etc.). Improving this inaccurate, useless punctuation is the first purpose of this utility.

See now the note field(s):

```

note = {Papers from an international congress
        held in Naples, 19-26 May 1993.},
note = {Includes bibliographical references (v. 3, p. [1005]-1056)
        and indexes.},

```

In most BibTeX / Biblatex styles the contents of note field is printed (e.g. «Reprint of the edition Oxford, Oxford U.P. 1976»). The note field is used by Copac for comments (e.g. if the book offers a bibliography); for this reason the utility changes note to annote.

Another change is author if the record type is proceedings (see the example above). In this case author is changed to editor.

1.2 Hyphenation

This feature is useful only to Biblatex users. If the field language features one and only one language, the field is transformed from language to langid, and the language name is transformed according to table 2, page 23 of Philipp Lehman's Biblatex manual, version 2.8a, 25.11.2013. If more than one language is listed, or if the language is not listed in Lehman's table, the field language is left unchanged.

1.3 Location/Address

Copac/BibTeX records use address for the place of publication and location for library names. This choice causes problems of BibTeX/Biblatex compatibility: location is used by Biblatex in place of the traditional address field, and library is used for library names. Copac-clean leaves address unchanged: this gives no problems because BibTeX uses this field normally and Biblatex interprets it as an alias for location. To avoid conflicts, location is changed to library, that is ignored (= not printed), but can be parsed, by BibTeX, and is standard in Biblatex. See the discussion in <http://tex.stackexchange.com/questions/57720/bibtex-handling-manuscripts>.

1.4 New fields

The Copac/BibTeX format knows only the title fields, but for some bibliographic styles (e.g. author-title styles, or the traditional “long” citations still common in the humanities), a shorttitle field is necessary. This utility outputs a shorttitle identical to title, because each user/journal/publisher may have preferences difficult to standardise. For BibTeX (not Biblatex) bibliographic styles, a booktitle field is necessary in the case of proceedings, *Festschriften*, edited books etc. For this reason, if and only if the record type is *proceedings* a booktitle field is automatically added – needless to say, it is identical to title¹.

¹These features may be possibly transformed to options: shorttitle is useless for author-date styles, and booktitle is useless for BibTeX users.

1.5 Final result

After running this utility, the example printed at page 2 is transformed as follows:

```
@proceedings{Giannantoni:1996,  
  editor = {Giannantoni, Gabriele and Gigante, Marcello},  
  title = {Epicureismo greco e romano: atti del congresso  
    internazionale, Napoli, 19-26 maggio 1995},  
  booktitle = {Epicureismo greco e romano: atti del congresso  
    internazionale, Napoli, 19-26 maggio 1995},  
  shorttitle = {Epicureismo greco e romano: atti del congresso  
    internazionale, Napoli, 19-26 maggio 1995},  
  series = {Elenchos ; 25},  
  series = {Elenchos (series) ; 25},  
  address = {[Napoli]},  
  publisher = {Bibliopolis},  
  year = {c1996},  
  size = {3 v. (1128 p.) ; 23 cm},  
  langid = {italian},  
  isbn = {ISBN: 8870883620},  
  annote = {Papers from an international congress  
    held in Naples, 19-26 May 1993},  
  annote = {Includes bibliographical references  
    (v. 3, p. [1005]-1056) and indexes},  
  annote = {Italian, English, French and German contributions},  
  keywords = {Epicurus -- Congresses},  
  keywords = {Epicurus -- Influence -- Congresses},  
  keywords = {Philosophy, Ancient -- Congresses},  
  library = {King's College London ; University of London - ULRLS}}
```

2 Installation

2.1 Dependencies

The programme requires `xclip`, a modern Snobol interpreter, and optionally a `tcl/tk` interpreter.

2.2 `xclip`

`xclip` can be normally installed from the software manager provided by Linux standard distributions. It is a tiny and efficient command line utility that can read from clipboard or output to clipboard. The homepage is <http://sourceforge.net/projects/xclip/>.

2.3 `Tcl/tk`

`Tcl` and its graphical toolkit, `tk`, are offered by the software installation manager of standard Linux distributions. `Tcl/tk` is highly portable and its scripts run with no or very little adaptation on Linux, Windows, and Mac. The homepage is [http:](http://)

[//www.tcl.tk/](http://www.tcl.tk/). I recommend to install the GUI of Copac-clean because, if you are using COPAC, running this utility without opening a text shell is much more comfortable.

2.4 Snobol

Snobol is a powerful language, with unparalleled string processing abilities. It is a niche language, but it is very useful to linguists and handy for text transformations as in the present case: if interested in Snobol, refer to Phil Budne's page (<http://www.snobol4.org/>). I strongly recommend to install the recent CSNOBOL4 release 1.5 (October 1, 2013) by Phil Budne. Unfortunately, Snobol is not offered by standard Linux distributions, but the installation is very easy (the installation detects if you are on a 32 or 64 bit system):

1. download `ftp://ftp.ultimate.com/snobol/snobol4-1.5.tar.gz`
2. `tar zxfv snobol4-1.5`
3. `cd snobol4-1.5`
4. (a) if you plan to use the tcl/tk GUI run:
`configure -with-tcl=/path/to/tclConfig.sh`
The path must be explicit; in my xubuntu 13.10 system it is `/usr/lib/tclConfig.sh`. Obviously you must install tcl/tk before Snobol.
- (b) if you do not plan to use the tcl/tk GUI run:
`configure`
`make`
5. as root (or sudo), run
`make install`
6. test if it's all right typing `snobol4 -v` in a shell. You should read a message containing the version number.

3 Installing Copac-clean

Now copy (1) `copac-clean.sh` (2) `copac-clean.sno` (3) `copac-clean-lib.inc` in the same directory. This directory must be in the PATH of your system. Type `echo $PATH` in a terminal to know the Path of your system. The script `copac-clean.sh` must be executable: if not, type `chmod +x copac-clean.sh`. If you do not have writing permissions on that particular folder as normal user, type `sudo chmod +x copac-clean.sh` and enter your password, or, on some systems, login as root and type the command.



Figure 1: Selecting BibTeX export from Copac

4 Usage

The first part is common to the GUI/no-GUI options:

1. select BibTeX in the Export drop down menu (at bottom of the page): see fig. 1.
2. select the whole record, and save it to the clipboard (Ctrl-C is the standard shortcut): see fig. 2.

4.1 With GUI

Run this utility from the GUI. In Ubuntu systems, use Alt-F2: see fig. 3.

The Copac record must be in your clipboard! To be sure of this, a clipboard utility as glipper, that controls your clipboard selections, can be useful. If in the clipboard there is not a copy of the Copac-BibTeX record, you'll get an error message (see fig. 4.)

If everything goes fine, you'll get a message confirming that the record has been edited (see fig. 5). Go now to section 4.3.

4.2 Without GUI

First, follow the instruction common to GUI/no-GUI (see the beginning of section 4). Then, open a terminal and type

```
copac-clean.sh --notk
```

If in the clipboard there is not a Copac record, you'll get an error message (see fig. 6). If the clipboard is used for a Copac record, and if the programme exits correctly, the record is transformed and you'll get a confirmation message (see fig. 7). Go now to section 4.3.

```

copac.ac.uk/search?author=Maltby+Robert&m=4&format=BibTeX&action=Export
Smart Bkms Mozilla Lat.Hdlines fvm... forum UCSC UCSC Intr. PagBian UCSC CT Marta Milanese MediaConv

@book{Maltby:1991,
author = {Maltby, Robert.},
title= {A lexicon of ancient Latin etymologies},
series = {ARCA ; 25 },
series = {ARCA, classical and medieval texts, papers, and monographs, 25 0309-5541},
address = {Leeds : },
publisher = {Cairns,},
year = {1991.},
size = {xviii, 669 p. ; 22 cm.},
language = {English ; Latin ; Multiple languages},
isbn = {ISBN: 090520574X},
isbn = {ISSN: 0306-8269},
note = {bibl.},
note = {includes Latin text, with English introduction},
note = {Dates of available copies: 1990, 1991, 2006.},
keywords = {Latin language -- Glossaries, vocabularies, etc.},
keywords = {Latin language -- Etymology -- Dictionaries -- Latin.},
keywords = {DICTIONARIES. Latin: Etymological.},
keywords = {Dictionary -- Latin -- Etymology.},
keywords = {Latin language - etymology},
location = {Aberdeen ; Birmingham ; Bristol ; British Library ; Cambridge ; Durham ; Edinburgh ; Exeter ; Glasgow ; Ki
College London ; Leeds ; Liverpool ; London Library ; Manchester ; National Library of Scotland ; National Library of
Llyfrgell Genedlaethol Cymru ; Nottingham ; Oxford ; St. Andrews ; Trinity College Dublin ; UCL (University College Lo
University of London - ULRLS ; York},
}

```

Figure 2: A BibTeX record from Copac

```

Smart Bkms Mozilla Lat.Hdlines fvm... forum UCSC UCSC Intr. PagBian UCSC CT Marta Milanese MediaConv

@book{Maltby:1991,
author = {Maltby, Robert.},
title= {A lexicon of ancient Latin etymologies},
series = {ARCA ; 25 }
series = {ARCA, cl
address = {Leeds :
publisher = {Cairn
year = {1991.},
size = {xviii, 669
language = {Englis
isbn = {ISBN: 0905
isbn = {ISSN: 0306
note = {bibl.},
note = {includes L
note = {Dates of a
keywords = {Latin
keywords = {Latin
keywords = {DICTIO
keywords = {Dictionary -- Latin -- Etymology.},
keywords = {Latin language - etymology},
location = {Aberdeen ; Birmingham ; Bristol ; British Library ; Cambridge ; Durham ; Edinburgh ; Exeter ; Glasgow ; Ki
College London ; Leeds ; Liverpool ; London Library ; Manchester ; National Library of Scotland ; National Library of
Llyfrgell Genedlaethol Cymru ; Nottingham ; Oxford ; St. Andrews ; Trinity College Dublin ; UCL (University College Lo
University of London - ULRLS ; York},
}

```

Figure 3: Running Copac-clean

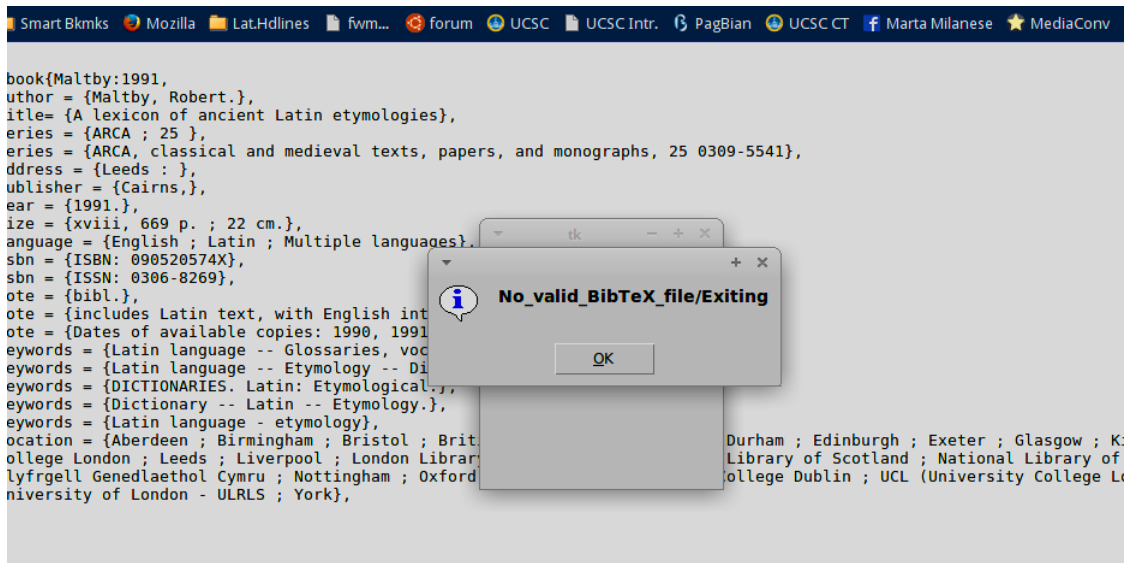


Figure 4: Error message: wrong clipboard contents

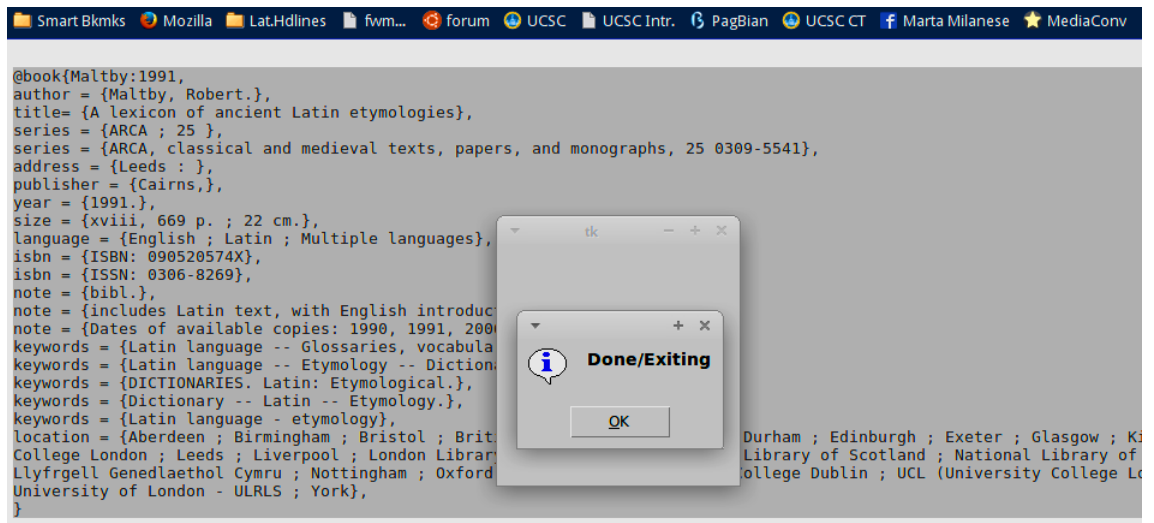


Figure 5: End of transformation

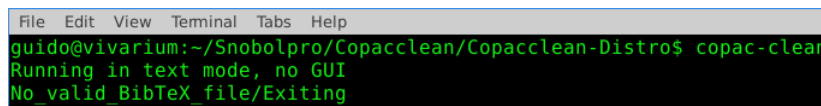


Figure 6: Error message: wrong clipboard contents (no GUI)


```
File Edit View Terminal Tabs Help
guido@vivarium:~/Snobolpro/Copacclean/Copacclean-Distro$ copac-clean.sh --notk
Running in text mode, no GUI
Done/Exiting
```

Figure 7: End of transformation (no GUI)

4.3 Paste the new version

Now your clipboard is used by the transformed BibTeX record, not by the old version (you can check it with glipper or similar utilities). Paste the new, transformed record in your BibTeX file or in a Jabref window². If you use Jabref, select a new record, open the BibTeX window, and paste the clipboard contents. Make sure to remove the empty record that Jabref outputs, that has normally this form:

```
@Book { ,
      }
```

5 Bugs and TODO

List of features that I would like to add:

1. A Windows version, if requested by users.
2. Choice among Spitbol and Snobol.
3. More flexibility, e.g. shorttitle yes/no and booktitle yes/no.
4. Automate the importation of edited records to Jabref.

6 Acknowledgements

My warmest thanks to PHIL BUDNE (phil@ultimate.com) for making his excellent CSNOBOL available and for his generous help. Many thanks to the community of Snobol users, and to RAFAL M. SULEJMAN (<http://rms.republika.pl/>) whose vim syntax files are a daily blessing.

²Documentation on Jabref: <http://jabref.sourceforge.net/>.

7 History

| | | |
|------------|-------------|---|
| 10.09.2013 | version 0.1 | |
| 15.09.2013 | version 0.2 | first usable |
| 30.10.2013 | version 0.3 | some fields corrected; comments edited |
| 27.12.2013 | version 0.4 | changed author to editor in proceedings |
| 28.12.2013 | version 0.5 | general editing of functions |
| 07.01.2014 | version 0.6 | improved check - using langid and library |
| 10.01.2014 | version 0.7 | improved license and documentation |

Author, copyright, license, disclaimer

COPAC-CLEAN. Automatic editing of Copac/BibTeX records.

Version 0.7

Copyright (C) 2014 Guido Milanese <guido.milanese@unicatt.it>

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 3 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, see <<http://www.gnu.org/licenses/>>. 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.

If you do not have a copy of the GNU General Public License write to the Free Software Foundation, Inc., 675 Mass Ave, Cambridge, MA 02139, USA.

If the author of this software was too lazy to include the full GPL text along with the code, you can find it at: <http://www.gnu.org/copyleft/gpl.html>.