The copac-clean utility
Version 0.8

Guido Milanese <`guido.milanese@unicatt.it`>

28th August 2015

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

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 25 of Philipp Lehman's Biblatex manual, version 3.0, 20.04.2015. 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 title[1].

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 the zenity GUI tool.

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/.

---

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

2.3 zenity

Zenity is a GTK+ lightweight tool – standard in most Linux distributions – that outputs message windows, error windows and the like. See `https://help.gnome.org/users/zenity/stable/`.

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 2.0 (January 1, 2015) 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/snobol4-2.0.tar.gz`

2. `tar zxfv snobol4-2.0`

3. `cd snobol4-2.0`

4. `configure`

5. `make`

6. as root (or sudo), run `make install`

7. 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: I would recommend /home/yourname/bin/copaclean. 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.

4 Usage

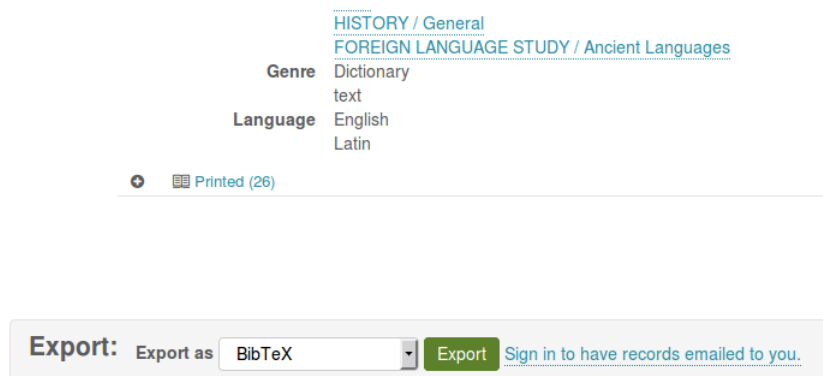1. select BibTeX in the Export drop down menu (at bottom of the page): see fig. 1.

Figure 1: Selecting BibTeX export from Copac

2. select the whole record, and save it to the clipboard (Ctrl-C is the standard shortcut): see fig. 2.

In Ubuntu systems, use Alt-F2: see fig. 3. Better, copy the provided copac-clean.desktop file to your desktop: the programme should run without any further adaptation.

If you prefer, open a terminal and type

```
copac-clean.sh
```

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 the clipboard is empty or does not hold 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).

4.1 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 window[2]. If you use Jabref, click on the name of the database you want to use, and paste the clipboard contents with a simple `Ctrl-V` or through the Modify – Copy menu entry.

5 Bugs and TODO

List of features that I would like to add:

1. A Windows version, if requested by users.

---

[2]Documentation on Jabref: `http://jabref.sourceforge.net/`.

```
@book{Maltby:1991,
author = {Maltby, Robert.},
title= {A lexicon of ancient Latin etymologies},
series = {ARCA, classical and medieval texts, papers, and monographs },
series = {25},
address = {Leeds },
publisher = {Cairns},
year = {1991},
size = {xviii, 669 p. ; 22 cm.},
language = {English ; Latin},
isbn = {ISBN: 090520574X},
isbn = {ISSN: 0306-8269},
note = {bibl.},
note = {includes Latin text, with English introduction},
keywords = {Latin language -- Etymology -- Dictionaries.},
keywords = {Latin language -- Glossaries, vocabularies, etc.},
keywords = {Latin language -- Etymology -- Dictionaries -- Latin.},
keywords = {Latin language.},
keywords = {Dictionary -- Latin -- Etymology.},
keywords = {DICTIONARIES. Latin: Etymological.},
keywords = {Latin language -- Etymology},
location = {British Library ; British School at Rome Library ; Durham University Libraries ; King's College London Library ; National Library
of Scotland ; National Library of Wales / Llyfrgell Genedlaethol Cymru ; Queen's University Belfast ; Senate House Libraries, University of
London ; St. Andrews University Libraries ; The London Library ; Trinity College Dublin Library ; University College London Library ;
University of Aberdeen Libraries ; University of Birmingham Libraries ; University of Bristol Libraries ; University of Cambridge Libraries ;
University of Edinburgh Libraries ; University of Exeter Libraries ; University of Glasgow Libraries ; University of Leeds Libraries ;
University of Liverpool Libraries ; University of Manchester Libraries ; University of Nottingham Libraries ; University of Oxford Libraries ;
University of York Libraries},
}
```

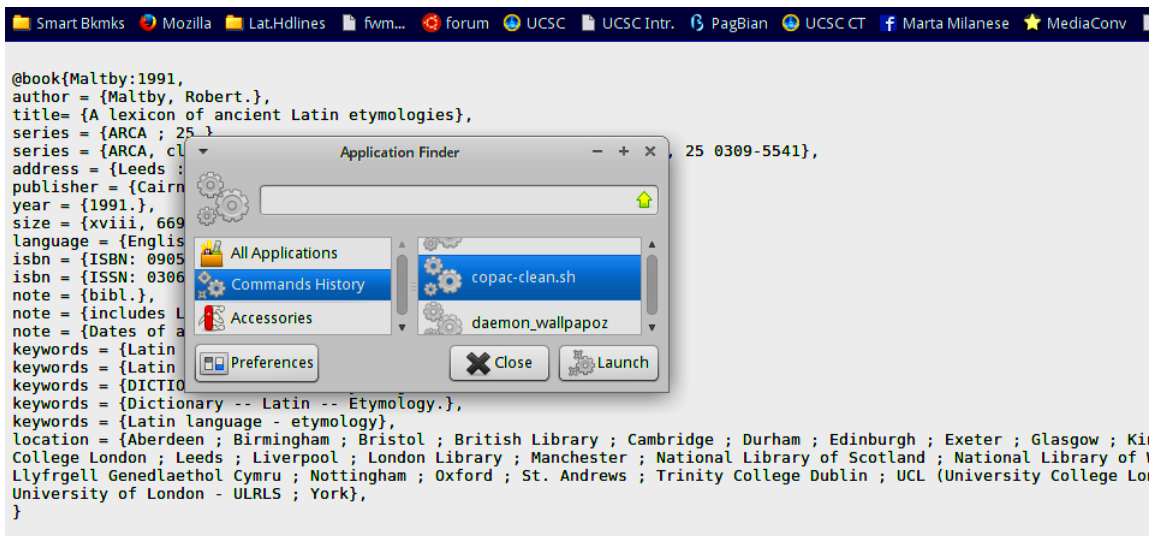Figure 2: A BibTeX record from Copac
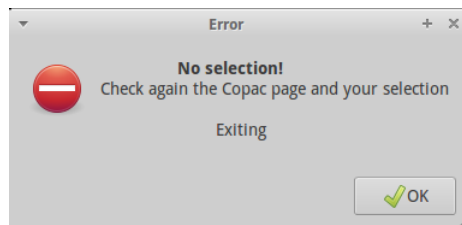

Figure 3: Running Copac-clean
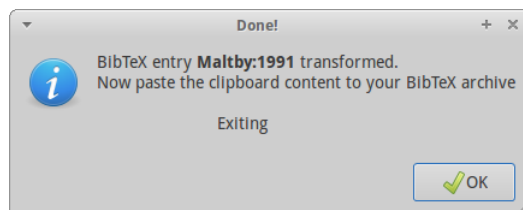

Figure 4: Error message: wrong clipboard contents

Figure 5: End of transformation

2. Choice among Spitbol and Snobol.

3. More flexibility, e.g. shorttitle yes/no and booktitle yes/no.

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.

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
28.08.2015	version 0.8	changed GUI from tcl/tk to zenity and improved documentation accordingly; added desktop configuration file

8   Author, copyright, license, disclaimer

COPAC-CLEAN. Automatic editing of Copac/BibTeX records.
Version 0.8
Copyright (C) 2015 Guido Milanese <guido.milanese@unicatt.it>
    License: LaTeX Project Public License (LPPL 1.3c)
    This work may be distributed and/or modified under the conditions of the LaTeX Project Public License, either version 1.3 of this license or (at your option) any later version.
    The latest version of this license is in

```
http://www.latex-project.org/lppl.txt
```

and version 1.3 or later is part of all distributions of LaTeX version 2005/12/01 or later.

This work has the LPPL maintenance status 'maintained'.

The Current Maintainer of this work is Guido Milanese.

This work consists of the files copac-clean.desktop, copac-clean.pdf, copac-clean.sno, copac-clean-lib.inc, copac-clean.sh, copac-clean.tex