A Set of BiBTeX File Processing Scripts

Technical Report CLIP X/98.1
Draft printed on: 16 April 2002
Version 0.4#45 (2002/4/10, 18:47:34 CEST)

M. Hermenegildo
clip@clip.dia.fi.upm.es
http://www.clip.dia.fi.upm.es/
The CLIP Group
School of Computer Science
Technical University of Madrid

Copyright ©1990-98 M. Hermenegildo and the CLIP group

This document may be freely read, stored, reproduced, disseminated, translated or quoted by any means and on any medium provided the following conditions are met:

1. Every reader or user of this document acknowledges that is aware that no guarantee is given regarding its contents, on any account, and specifically concerning veracity, accuracy and fitness for any purpose.
2. No modification is made other than cosmetic, change of representation format, translation, correction of obvious syntactic errors, or as permitted by the clauses below.
3. Comments and other additions may be inserted, provided they clearly appear as such; translations or fragments must clearly refer to an original complete version, preferably one that is easily accessed whenever possible.
4. Translations, comments and other additions or modifications must be dated and their author(s) must be identifiable (possibly via an alias).
5. This licence is preserved and applies to the whole document with modifications and additions (except for brief quotes), independently of the representation format.
6. Any reference to the "official version", "original version" or "how to obtain original versions" of the document is preserved verbatim. Any copyright notice in the document is preserved verbatim. Also, the title and author(s) of the original document should be clearly mentioned as such.
7. In the case of translations, verbatim sentences mentioned in (6.) are preserved in the language of the original document accompanied by verbatim translations to the language of the traslated document. All translations state clearly that the author is not responsible for the translated work. This license is included, at least in the language in which it is referenced in the original version.
8. Whatever the mode of storage, reproduction or dissemination, anyone able to access a digitized version of this document must be able to make a digitized copy in a format directly usable, and if possible editable, according to accepted, and publicly documented, public standards.
9. Redistributing this document to a third party requires simultaneous redistribution of this licence, without modification, and in particular without any further condition or restriction, expressed or implied, related or not to this redistribution. In particular, in case of inclusion in a database or collection, the owner or the manager of the database or the collection renounces any right related to this inclusion and concerning the possible uses of the document after extraction from the database or the collection, whether alone or in relation with other documents.

Any incompatibility of the above clauses with legal, contractual or judiciary decisions or constraints implies a corresponding limitation of reading, usage, or redistribution rights for this document, verbatim or modified.

Table of Contents

Summary 1

1 Introduction 3
- 1.1 Overall description 3
- 1.2 Installation 3
  - 1.2.1 Installing the scripts 3
  - 1.2.2 Installing the manuals 3
- 1.3 The actual utilities 4
- 1.4 Some more specific utilities 5
- 1.5 Support libraries used by bibutils 5
- 1.6 Examples 6
- 1.7 Version/Change Log (`bibutils`) 6

2 Conventions for BiBTeX files 7
- 2.1 The `butype` field 7
- 2.2 Recommendations for the `author` field 8
- 2.3 Other bibutils-specific BiBTeX fields 8
- 2.4 Other fields specific to the CLIP group 8
- 2.5 Standard Fields Required in BiBTeX Files 9

Concept Definition Index 11

Summary

This package is a collection of simple utilities and libraries for processing BiBTeX files. It allows generating documents or parts of documents automatically from a BiBTeX database and performing some simple maintenance tasks (such as sorting entries or merging files) on those databases. The libraries include a Prolog parser for references in *BiBTeX* format.

This documentation corresponds to version 0.4#45 (2002/4/10, 18:47:34 CEST).

1 Introduction

This package is a collection of simple utilities and libraries for processing BiBTeX files. It allows generating documents or parts of documents automatically from a BiBTeX database and performing some simple maintenance tasks (such as sorting entries or merging files) on those databases. The libraries include a Prolog parser for references in *BiBTeX* format.

1.1 Overall description

The general idea is that `.bib` files are converted to `.pl` files using `bib2pl` and then the `.pl` (Prolog) files are used to generate several listings, reports, etc. automatically, by writing simple Prolog scripts. A number of sample scripts are provided. Some library files are also available with some predicates which may facilitate writing other, similar scripts.

In order for the `bibutils` package to work optimally, `.bib` files must adhere to some conventions, which are specified in Chapter 2 [Conventions for BiBTeX files], page 7.

In general, automatically generated files produced by the scripts include `_auto` in the name in order to identify them as such.

As mentioned above, these utilities have been written as Prolog scripts. Such scripts are supported directly by certain Prolog systems (such as `CIAO Prolog`) and can be implemented easily on other Prolog systems (e.g., a `prolog_shell` package is available for `SICStus Prolog` from `http://www.clip.dia.fi.upm.es/Software`).

1.2 Installation

1.2.1 Installing the scripts

Because these are Prolog scripts, no actual installation is needed after unpackaging the distribution. However:

- You must make sure that the utilities are accessible for execution. This means either copying them to a common directory (such as `/usr/local/bin`) or including the distribution directory in your *path for executables* (e.g., `PATH`).
- In order to run the scripts, you need to have installed previously a *Prolog system* which can run *Prolog scripts* (e.g., *CIAO* or *SICStus* with the prolog_shell package).

  The Prolog script processor (e.g., `ciao-shell` or `prolog_shell`) must also be accessible for execution, i.e., it must be in your execution path. Alternatively, you can edit the header of the scripts and put absolute file paths.

You may be also be able to run the code of the scripts from the top level of *other Prolog systems* by commenting out the headers and loading it into the Prolog system in the standard way.

1.2.2 Installing the manuals

The package *documentation* is included in the `doc` directory in several formats. You may want to copy it to more accessible places (this can be done by typing `gmake install`).

1.3 The actual utilities

`bib2pl f1.bib f2.bib ...`
: Converts one or more bib files into one `.pl` file (file generated is `f1_f2_..._auto.pl`).

  Caveats/Bugs:

  - Some syntactically incorrect entries that bibtex can read (and that therefore may go undeteced) are not supported (e.g., commas after the body of the last item in an entry).
  - `@string` is not supported yet.

`mergesortbibs f1.bib f2.bib ...`
: Merges and sorts one or more bib files into a single `.bib`. file (file generated is `f1_f2_.._auto.bib`).

  Caveats/Bugs: same as above.

`pl2bib bibfile.pl`
: Converts one `.pl` file (containing a paper database) into a `.bib` file.

`pl2pubsbyyear SName PName bibfile.pl people_urls.pl psdir psdirurl`
: Generates a list of the publications contained in file `bibfile.pl` in which `SName` is an author (if `-` is used then all papers in `bibfile.pl` are selected). The list is produced in latex (in English), classified by *publication type* and *YEAR of publication*.

  - `PName` is a (normally longer) name of `SName`, which will be used to produce the section titles.
  - `people_urls.pl` must contain a predicate `url_db(`*Name*`,`*URL*`,`*email*`)` (all arguments are strings) and is used to obtain *urls for authors*, etc. For example:

    ```
    url_db(""M.Hermenegildo"",
           ""http://www.clip.dia.fi.upm.es/~herme"",
           ""herme@fi.upm.es"").
    ```

  - `psdir` is a directory (which should be accessible via WWW) where postscript (`.ps` / `.ps.gz`) versions of the papers are stored. These `.ps` files must be named *keyword.ps*, where *keyword* is the keyword field in the bib entry.
  - `psdirurl` is the url address of that directory.

  (An example application is `examples/clippubsbyyear`.)

`pl2pubsbytopic SName PName bibfile.pl people_urls.pl topics.pl psdir psdirurl`
: Generates a list of the publications contained in file `bibfile.pl` in which `SName` is an author (if `-` is used then all papers `bibfile.pl` are selected). The list is produced in latex (in English), classified by *publication type* and *TOPIC of the paper*.

  - `PName`, `people_urls.pl`, `psdir`, and `psdirurl` are as above.
  - `topics.pl`, the topics database, must contain a predicate `topic(`*TopicKeyword*`,`*TopicText*`)` (first argument is an atom, the second a string) which relates topic keywords with some text describing the topic. The former is what is included in the entries in `bibfile.bib` database, in the `butopics` field. The latter is the text used in section headings in the generated document. For example:

    ```
    topic(anal,""Program Analysis"").
    ```

  (An example application is `examples/clippubsbytopic`. Note for CLIP group: the CLIP topics database is in `/home/clip/PlDbs/clip_topics.pl`.)

`pl2pubsbytopiclist SName PName bibfile.pl people_urls.pl topics.pl psdir psdirurl topicid1 ... topicidn`
: Same as bove, but a number of topics can be given and a single list is produced, merging the papers in all those topics.

`pl2engcur bibfile.pl Name people_urls.pl`
: Generates the list of publications for person `Name`, in latex (English), classified by *type of publication*. `people_urls.pl` is as above.

  (An example application is `examples/engcur`.)

`pl2tramoscur bibfile.pl Name people_urls.pl`
: Generates the list of publications for person `Name`, in latex (Spanish), classified by *type of publication*. `people_urls.pl` is as above.

  (An example application is `examples/tramoscur`.)

`incversion file`
: Simple script which takes as argument a filename which containing an integer followed by a period. Increments the integer by one in the file and returns the new value in stdout. This may be useful for for example for updating the version numbers of automatically generated documents.

  (An example application is `examples/clippubsbyyear`.)

1.4 Some more specific utilities

These utilities are more specific to the CLIP group's environment, but (parts of them) may be useful to other people or may serve as a starting point for developing other, similar scripts:

`pl2cicytcur bibfile.pl Name`
: Generate "publicaciones", "otras publicaciones", etc. for CV in CICYT (Spanish funding agency) format for person `Name`.

  (An example application is `examples/cicytcur`.)

`pl2upmcurbooks bibfile.pl Name`
: Generate "libros" part for CV in UPM/MEC format for person `Name`.

  (An example application is `examples/upmcur`.)

`pl2upmcurpub bibfile.pl Name`
: Generate "publicaciones" part for CV in UPM/MEC format for person `Name`.

  (An example application is `examples/upmcur`.)

`pl2clip_dia_annual bibfile.pl year`
: Generate skeleton of a group or person's contribution to the UPM DIA department's annual report for `year`.

1.5 Support libraries used by bibutils

These libraries include predicates which are used by several of the scripts. They may also be useful for writing other, related scripts:

`support.pl`
: Miscellaneous library predicates.

`bibparser.pl`
: bibtex file parser.

`pp_papers_bib.pl`
: Output papers in bibtex format.

`pp_papers_latex.pl`
: Output papers in latex itemize format (language can be selected).

`pp_papers_pl.pl`
: Output papers in Prolog format.

1.6 Examples

The directory `examples` contains directories or links to directories with *example applications*.

1.7 Version/Change Log (`bibutils`)

Version 0.4#42 (2002/4/7, 13:44:20 CEST)
Created new `pl2tramoscur` script to generate CV publications as in `pl2engcur` but in Spanish. This is suitable for example for the CNEAI 6-year evaluations. (Manuel Hermenegildo)

Version 0.4#19 (1998/11/6, 17:35:39 MET)
Added automatic generation of readmes and brief installation instructions. (Manuel Hermenegildo)

Version 0.4#17 (1998/10/29, 13:12:40 GMT)
Moved documentation to `lpdoc`. (Manuel Hermenegildo)

2 Conventions for BiBTeX files

This is a description of the conventions that `.bib` files must adhere to in order for the `bibutils` package to work optimally. Basically, all the standard types of entry (`article`, `inproceedings`, `book`, etc.) and all the standard fields (`author`, `title`, `publisher`, etc.) are supported by `bibutils`. Some of these fields are *required* for the correct operation of `bibutils` (see Section Section 2.5 [Standard Fields Required in BiBTeX Files], page 9). In addition, some new fields which particular to `bibutils` can (and, in some cases, must) be added to the BiBTeX entries for full operation of `bibutils`. These new fields are described below. In particular, the main restriction is that typically *every entry in the* `.bib` *file which is elligible to appear in the generated documents must contain the field* `butype`.

2.1 The `butype` field

In order for the `bibutils` to work properly, publications must be correctly classified in the source `.bib` file using appropriate values of the `butype` field. The following general considerations apply:

- The `butype` field will typically be the same as the BiBTeX type, i.e., Article, Book, InProceedings, etc. (except that it must always be written all in *lowercase*). However, there are some cases in which, in order to classify more clearly the type of publication, the value of `butype` to be used may differ from the BiBTeX type (see below for more details).
- It is often the case that the BiBTeX file contains *multiple entries for a given paper*. For example, a short and a long version, or two entries with different keywords, for compatibility. In these cases, it is necessary to identify the entry whose data will appear in the output of `bibutils`. This version is identified in the bibtex file by the presence of a `butype` field. Entries without this field will be ignored by the automatic processing tools. Generally the entry with the most detailed information is the one in which this field is included. It is important to avoid including the `butype` field in more than one entry for a given publication, since otherwise the publication will appear repeated in the output.

The following are the possible values for the `butype` field and their meaning:

1. *Refereed/serious publications* (the "real stuff"):

   `article` Only refereed articles published in a journal.

   `inproceedings`
   Only refereed articles published in conference proceedings.

   `book` Book *with editor*, including formal proceedings.

2. *Unrefereed, but important publications*:

   `mastersthesis`
   Master's thesis.

   `phdthesis`
   Ph.D. thesis.

   `incollection`
   Paper in book *with editor*.

   `proceedings`
   No need for publisher, etc. (e.g., workshop proceedings?).

   `workshop` Workshop papers. These have "inproceedings" bibtex type but "workshop" butype.

   `invited` Invited papers. These have "inproceedings" bibtex type but "invited" butype.

`invitedtalk`
: Invited talks and tutorials at major conferences, when there is no paper associated with them (in that case they are "invited"). These have "inproceedings" bibtex type but "invitedtalk" butype.

3. *Unrefereed, miscellaneous publications*:

   `techreport`
   : Technical Report.

   `manual` A user's manual.

4. *Other publications*:

   `misc` Accepts anything (e.g., internet postings).

   `unpublished`
   : Note required! For manuscripts...

2.2 Recommendations for the `author` field

In order to ensure correct parsing of authors and sorting of papers:

- Use braces around *compound author names* such as `{Garc\\'{\\i}a de la Banda}`.
- *author names* Use "M. Hermenegildo and F. Bueno" rather than "M. Hermenegildo, F. Bueno" or "Manuel Hermenegildo and Francisco Bueno".

2.3 Other bibutils-specific BiBTeX fields

`butopics` = `<` list of topics`>`
: This field is needed in order to be able generate lists of publications classified by topics. It contains one or more atoms (separated by commas) describing topics. These atoms are related to the textual descritpion of the topics themselves by a topics database. See the description of the `pl2pubsbytopic` script for details.

  Note for CLIP group: the CLIP topics database is in `/home/clip/PlDbs/clip_topics.pl`.

`url` = `<url>`
: An *online copy of the paper* (or other information on it) is available at location `<url>`.

  Note that it is normally not required to include this field: if *postscript file for the paper* named `keyword.ps` or `keyword.ps.gz`, is available in a certain *directory of papers* (given to the script upon startup). `keyword` is the keyword used in the bib file (i.e., the same used in `@cite{}`/`\\cite{}` commands). Thus, url normally only needs to be used for exceptions to this rule. A page describing the paper may be available in `keyword.html`.

`bookurl` = `<url>`
: Contains a pointer to some WWW accessible document related with the collection (e.g., online proceedings, description of conference, etc.).

2.4 Other fields specific to the CLIP group

This section lists some other conventions that BiBTeX files that are local to the CLIP group should adhere to.

`papercopy` = `<number>`
: Means that a paper copy of the paper exists in the CLIP lab library, filed under that number.

`projects` = <project names>
: Projects acknowledged in the paper. Used in the generation of publication lists for project reports. (Only bother including for "good" pubs?)

`comment` = <text>
: Since CLIP `.bib` files are processed by several utilities, any comments should be inside a `comment` field and should not appear anywhere else in the `.bib` file.

The following are necessary for the somewhat bizarre format of the department annual research report, and some weird CV formats:

`publisher_location` = <city>
: For `article` butype. City where the publishing ocurred (!).

`paper_presentation_city` = <city>
: For `inproceedings` butype. City where the conference was held.

`paper_presentation_country` = <country>
: For `inproceedings` butype. Country where the conference was held.

The following BiBTeX types do not appear as useful as others and they are not used in the CLIP `.bib` files:

`booklet` Brochure. Small book. Editor not needed.

`conference`
: This seems to mean a talk (?).

`inbook` Paper in book (w/chapter), but with no editor (?).

2.5 Standard Fields Required in BiBTeX Files

The following is a (probably incomplete) table of the (traditional) fields that are *required* for each type of BiBTeX entry so that most of the bibutils will work well:

```
A=article       - I=inproceedings - B=book          - H=mastersthesis -
H=phdthesis     - C=incollection  - P=proceedings   - T=techreport    -
M=manual
*R=required     - *U=useful       - *O=optional     - *N=Not applicable

                A       I       B       H       C       P       T       M
-------------------------------------------------------------------------
*author         R       R       R       R       R       *N      R       R
*title          R       R       R       R       R       R       R       R

booktitle       N       R       N       N       R       N       N       N
editor          O       U       R       N       U       U       N       N
journal         R       N       O       N       N       N       N       N
organization    O       O       N       O       O       O       R       O

*pages          R       R       N       N       R       N       N       N
*npages         O       O       R       R       O       R       R       O

*publisher      R       R       R       O       R       R       O       O
*institution    N       N       N       R       N       N       R       O
series          O       O       O       O       O       O       O       O
edition         R       R       O       N       O       O       N       N
volume          R       N       O       N       O       O       N       N
```

number	R	O	O	N	O	O	R	O
-address	O	O	O	O	O	O	O	O
-note	O	O	O	O	O	O	O	O
*year	R	R	R	R	R	R	R	R
*month	U	U	U	U	U	U	U	U

Concept Definition Index

A

article 7
author names 8
auto (_auto) files 3

B

BiBTeX 1, 3
BiBTeX files, conventions 7
book 7
bookurl 8
butopics 8
butype field 7

C

CIAO 3
comment 9
compound author names 8
Conventions, for BiBTeX files 7
Converting, bib to Prolog 4
Converting, Prolog to bib 4

D

directory of papers 8
documentation 3

E

example applications 6

I

incollection 7
inproceedings 7
installation 3
invited 7
invitedtalk 8

L

Listing publications in one or more topics, merged . . 4
Listing publications, by topic 4
Listing publications, by type of publication 5
Listing publications, by type of publication, spanish 5
Listing publications, by year 4
Listing publications, for UPM Annual report 5
Listing publications, in CICYT CV format 5
Listing publications, in UPM CV format 5

M

manual 8
mastersthesis 7
Merging bibliographies 4
misc 8
multiple entries for a given paper 7

O

online copy of the paper 8
other Prolog systems 3
Other publications 8

P

paper_presentation_city 9
paper_presentation_country 9
papercopy 8
Parsing BiBTeX files 5
path for executables 3
phdthesis 7
postscript file for the paper 8
postscript versions of papers 4
proceedings 7
projects 9
Prolog scripts 3
Prolog system 3
publisher_location 9

R

Refereed/serious publications 7
repetitions in the output 7
required BiBTeX fields 9

S

SICStus 3
Sorting bibliographies 4

T

techreport 8

U

unpublished 8
Unrefereed, but important publications 7
Unrefereed, miscellaneous publications 8
url 8
urls for authors 4

V

Version of a document, incrementing 5

W

workshop 7