The `bibexport.sh` script

Nicolas Markey

2005/02/27

Abstract

`bibexport.sh` is a small shell script, relying on BibTEX, that extracts entries of one or several `.bib` file(s). It will expand abbreviations and cross-references, except standard month and journal abbreviations. The output is indented as neatly as possible, yielding a readable `.bib` file even if the original one was not.

1 Exporting `.bib` files

1.1 Why and how?

BibTEX aims at allowing for the use of one single `.bib` file, containing many entries, from which BibTEX extracts only the `\cite`d ones. When sending a document to someone else, this requires either to send the whole file, or to extract some entries from the `.bib` file.

BibTEX also has a mechanism for using abbreviations and cross-references. When extracting entries of a large `.bib` file, it can be interesting to develop those abbreviations, in order to get a clean, self-contained `.bib` file. Also, it may be useful to develop cross-references in a `.bib` file, independently of any document.

`bibexport` can either extract entries that are cited in a document, or all the entries of one or several `.bib` files. It will always develop cross-references and abreviations, except standard abbreviations for months or some journals, that are defined in standard BibTEX styles. This script uses BibTEX. This has both pros and cons:

+ it is very simple. Basicaly, the script simply calls BibTEX, and the `.bst` file just outputs the name and the content of each field.

+ since it uses BibTEX, we are sure that it will handle everything "properly", *i.e.* in the same way as they will be handled when cited in a LaTEX document;

= BibTEX has some strict limitations (especially "no more than 78 consecutive non-space characters") that we must be aware of. On the other hand, any such problem occuring within the script would also occur when compiling a document;

- abbreviations and cross-references will *always* be developped. It could be argued that this is also a positive point, but having the choice would have been better.

- Many people seem to find BibTEX's internal language clumsy, and thus the script could be difficult to adapt to special needs. However, this is *really* not that difficult, as will be explained later on. And please, don't hesitate to ask for any improvement or for the addition of some extra (but still relevant) fields.

1.2 Related scripts

Several other tools exist for achieving this task:

- `aux2bib`, written by Ralf Treinen, relies on `bib2bib`, which is a CAML program for selecting some entries in one or several `.bib` files. It does not expand anything, but includes all the necessary definitions and entries.
- `bibextract.sh`, by Nelson Beebe. This script uses AWK for extracting some entries out of a `.bib` file. It is said to be noncompliant with cross-references.
- subset.bst, by David Kotz. export.bst develops the same ideas (but I discovered that only later on). subset.bst does not handle `@preamble`, neither does it "protect" standard abbreviations.
- Probalby some others...

1.3 Some examples

- extracting `\cite`d references of a document, also including cross-references:

  `bibexport -o <result>.bib <file>.aux`

- extracting `\cite`d references of a document, without crossrefs, and using a special `.bst` file:

  `bibexport -b <style>.bst -o <result>.bib <file>.aux`

- export all the entries of two `.bib` files (including crossrefed entries):

  `bibexport -a -o <result>.bib <file1>.bib <file2>.bib`

- export all the entries of two `.bib` files (without crossrefs):

  `bibexport -a -n -o <result>.bib <file1>.bib <file2>.bib`

  In fact, the only difference between this and the previous one is that `crossref` field will be filtered out at the end of the script.

- export all the entries of two `.bib` files, using an extra file containing cross-referenced entries (that should not be included):

  ```
  bibexport -a -e <crossref>.bib -n -o <result>.bib \
                <file1>.bib <file2>.bib
  ```

2 The code

2.1 The shell script

2.1.1 Initialization

usage() We first define how the script should be used:

```
⟨*script⟩
function usage()
{
echo "bibexport: a tool to extract BibTeX entries out of .bib files.
usage: $0 [-h|v] [-n] [-b bst] [-a [-e file]...] [-o file] file...
  -a, --all          export the entire .bib files
  -b, --bst          specifies the .bst style file  [default: export.bst]
  -e, --extra        extra .bib files to be used (for crossrefs)
  -c, --crossref     include entries that are crossref'd   [default: yes]
  -n, --no-crossref  don't include crossref'd entries       [default: no]
  -o file            write output to file        [default: bibexport.bib]
  -h, --help         print this message and exit
  -v, --version      print version number and exit";
exit 0;
}
⟨/script⟩
```

CREF
FILE
EXT
EXTRA
EXTRABIB
SPACE
BST

We define the default value of some variables:

- `$CREF`: the value of `-min-crossrefs`,
- `$FILE`: the input file(s),
- `$EXT`: the extension (`.aux` or `.bib`) of input files,
- `$EXTRA`: list of possible extra `.bib` files without extension,
- `$EXTRABIB`: list of possible extra `.bib` files with extension,
- `$SPACE`: file name separator (can be ␣, comma or empty),
- `$BST`: the `.bst` file to be used.

```
⟨*script⟩
CROSSREF="1";
FILE="";
EXT=".aux";
EXTRA="";
EXTRABIB="";
SPACE=" ";
BST="export";
⟨/script⟩
```

2.1.2 Handling arguments

If no argument have been supplied, we call `usage()`.

```
⟨*script⟩
if [ $# -eq 0 ]; then
  usage;
```

```
fi
```
⟨/script⟩

Otherwise, we enter a `while`-loop for handling the whole list of arguments:

⟨∗script⟩
```
while [ $# != 0 ]; do
        case $1 in
```
⟨/script⟩

- `-a` or `--all`: export all the bibliography. This means that we input `.bib` files.

⟨∗script⟩
```
                -a|--all)
                  EXT=""; SPACE="";
                  shift ;;
```
⟨/script⟩

- `-b` or `--bst`: specifies the style file. It seems that BibTEX does not like the `./style.bst` syntax, and we have to handle that case separately.

⟨∗script⟩
```
            -b|--bst)
                if [ "`dirname $2`" = "." ]; then
                    DOLLARTWO="`basename $2 .bst`";
                else
                    DOLLARTWO="`dirname $2`/`basename $2 .bst`";
                fi
                BST="${DOLLARTWO}";
                shift 2;;
```
⟨/script⟩

- `-e` or `--extra`: when we want to export all the entries of a `.bib` file, we can specify an extra `.bib` file that would contain entries that we don't want to export, but that are needed, *e.g.* for crossrefs.

⟨∗script⟩
```
            -e|--extra)
                if [ "`dirname $2`" = "." ]; then
                    DOLLARTWO="`basename $2 .bib`";
                else
                    DOLLARTWO="`dirname $2`/`basename $2 .bib`";
                fi
                EXTRA="${EXTRA}${DOLLARTWO},";
                EXTRABIB="${EXTRABIB},${DOLLARTWO}.bib";
                shift 2;;
```
⟨/script⟩

- `-o` or `--output`: the name of the output file.

⟨∗script⟩
```
            -o|--output-file)
                if [ "`dirname $2`" = "." ]; then
                    DOLLARTWO="`basename $2 .bib`";
                else
                    DOLLARTWO="`dirname $2`/`basename $2 .bib`";
```

```
                    fi
                    OUTPUT="${DOLLARTWO}.bib";
                    shift 2 ;;
⟨/script⟩
```

- `-c` or `--crossref` (or others): this options means that we want crossrefs to be included. Note that for any entry, field inheritage will be performed.

```
⟨*script⟩
            -c|--crossref|--crossrefs|--with-crossref|--with-crossrefs)
                CREF="1" ;
                shift ;;
⟨/script⟩
```

- `-n` or `--no-crossref`: don't include crossref'ed entries.

```
⟨*script⟩
            -n|--no-crossref|--without-crossref)
                CREF="20000" ;
                shift ;;
⟨/script⟩
```

- `-v` or `--version` for version number:

```
⟨*script⟩
            -v|--version)
                echo "2.10"; exit 0;;
⟨/script⟩
```

- other dash-options are erroneous (except `-h`, but...):

```
⟨*script⟩
            -*)
                usage;;
⟨/script⟩
```

- there should only remain file names: we add those names to the list of files.

```
⟨*script⟩
            *)
                if [ "`dirname $1`" = "." ]; then
                    DOLLARONE="`basename $1 ${EXT}`";
                else
                    DOLLARONE="`dirname $1`/`basename $1 ${EXT}`";
                fi
                FILE="${FILE}${SPACE}${DOLLARONE}${EXT}";
                if [ -z "${SPACE}" ]; then
                    SPACE=",";
                fi;
                shift;;
⟨/script⟩
```

That's all folks:

```
⟨*script⟩
        esac
done
⟨/script⟩
```

2.1.3 The core of the script

We first set the name of the result and intermediary files:

```
⟨*script⟩
FINALFILE=${OUTPUT};
if [ ! "${FINALFILE}" ]; then
    FINALFILE="bibexport.bib";
fi
TMPFILE="bibexp.‘date +%s‘";
⟨/script⟩
```

We then create the `.aux` file for the main run of BibTEX. Note that this could call BibTEX, with the expkeys.bst file, in the case where we want to export all entries of a `.bib` file but not crossrefs. Note how, in that case, we trick BibTEXfor inputing extra files twice: we include then with their short name first (with no extension), and then with the full name. We *need* to do that, since `string` abbreviations must be defined first, while crossrefs must occur after having been referenced.

```
⟨*script⟩
if [ -z "${EXT}" ]; then
    if [ -z "${EXTRA}" ]; then
        cat > ${TMPFILE}.aux <<EOF
\citation{*}
\bibdata{${FILE}}
\bibstyle{${BST}}
EOF
    else
        cat > ${TMPFILE}.aux <<EOF
\citation{*}
\bibdata{${FILE}}
\bibstyle{expkeys}
EOF
        bibtex -min-crossrefs=${CREF} ${TMPFILE} >/dev/null 2>&1;
        mv -f ${TMPFILE}.bbl ${TMPFILE}.aux;
        cat >> ${TMPFILE}.aux <<EOF
\bibdata{${EXTRA}${FILE}${EXTRABIB}}
\bibstyle{${BST}}
EOF
    fi
else
    cat ${FILE} | sed -e "s/bibstyle{.*}/bibstyle{export}/" > ${TMPFILE}.aux;
fi
⟨/script⟩
```

This was the hard part. We now call BibTEX, clean and rename the output file, and remove intermediary files:

```
⟨*script⟩
bibtex -min-crossrefs=${CREF} ${TMPFILE}
echo -e "Generated on ‘date +\"%c (%s).\"‘\n" > ${FINALFILE};
if [ ! "x${CREF}" = "x1" ]; then
    sed -r -e \
        "/^ *[cC][rR][oO][sS][sS][rR][eE][fF] *= *[^,]+,?$/d" \
        ${TMPFILE}.bbl >> ${FINALFILE};
else
```

```
    cat ${TMPFILE}.bbl >> ${FINALFILE};
fi
rm -f ${TMPFILE}.bbl ${TMPFILE}.aux ${TMPFILE}.blg
⟨/script⟩
```

2.2 The **expkeys.bst** file

The only role of that file is to export the list of entries to be exported. It is used when we export all the entries of `.bib` files, except those of *extra* `.bib` files. Thus:

```
⟨*expkeys⟩
ENTRY{}{}{}
READ
FUNCTION{export.key}
{
  "\citation{" cite$ "}" * * write$ newline$
}
ITERATE{export.key}
⟨/expkeys⟩
```

2.3 The **export.bst** file

2.3.1 Some configuration values

`left.width` We define here the indentation values, and the field delimiters. *short* width are
`right.width` used for `@preamble`.

```
⟨*export⟩
FUNCTION{left.width}{#18}
FUNCTION{right.width}{#55}
FUNCTION{url.right.width}{#61}
FUNCTION{left.short.width}{#10} %% for @preamble
FUNCTION{right.long.width}{#63}
FUNCTION{left.delim}{"{"}
FUNCTION{right.delim}{"}"}
%FUNCTION{left.delim}{quote$}
%FUNCTION{right.delim}{quote$}
⟨/export⟩
```

2.3.2 Entries

We use standard entries here. Of course, more entries could be added for special `.bib` files. Those extra entries will also have to be added in the main exporting function.

`ENTRY`

```
⟨*export⟩
ENTRY{
% Standard fields:
    address
    author
    booktitle
    chapter
    edition
    editor
```

```
    howpublished
    institution
    journal
    key
    month
    note
    number
    organization
    pages
    publisher
    school
    series
    title
    type
    volume
    year
% Special (but still somewhat standard) fields (natbib, germbib, ...):
    abstract
    doi
    eid
    isbn
    issn
    language
    url
}{}{}
⟨/export⟩
```

2.3.3 Basic functions

No comment.

or
and
not

```
⟨*export⟩
FUNCTION{not}
{
    {#0}
    {#1}
  if$
}
FUNCTION{and}
{
    'skip$
    {pop$ #0}
  if$
}
FUNCTION{or}
{
    {pop$ #1}
    'skip$
  if$
}
⟨/export⟩
```

2.3.4 Splitting strings

We design functions for splitting strings, so that the final `.bib` file will be cleanly indented.

space.complete
split.string
split.url
split name

```
⟨*export⟩
INTEGERS{left.length right.length}
STRINGS{ s t }
FUNCTION{space.complete}
{
  'left.length :=
  duplicate$ text.length$ left.length swap$ -
  {duplicate$ #0 >}
    {
      swap$ " " * swap$ #1 -
    }
  while$
  pop$
}
FUNCTION{split.string}
{
  'right.length :=
  duplicate$ right.length #1 + #1 substring$ "" =
    {""}
    {
      's :=
      right.length
      {duplicate$ duplicate$ s swap$ #1 substring$ " " = not and}
        {#1 -}
      while$
      duplicate$ #2 <
        {
          pop$ "    " s * ""
        }
        {
          duplicate$ s swap$ #1 swap$ substring$
          swap$
          s swap$ global.max$ substring$
        }
      if$
    }
  if$
}
FUNCTION{split.url}
{
  'right.length :=
  duplicate$ right.length #1 + #1 substring$ "" =
    {""}
    {
      's :=
      right.length
      {duplicate$ duplicate$ s swap$ #1 substring$ "/" = not and}
        {#1 -}
```

```
      while$
      duplicate$ #2 <
        {
          pop$ "    " s * ""
        }
        {
          duplicate$ s swap$ #1 swap$ substring$
          swap$ #1 +
          s swap$ global.max$ substring$
        }
      if$
    }
  if$
}
FUNCTION{split.name}
{
  'right.length :=
  duplicate$ right.length #1 + #1 substring$ "" =
    {""}
    {
      's :=
      right.length
      {duplicate$ duplicate$ s swap$ #5 substring$ " and " = not and}
        {#1 -}
      while$
      duplicate$ #2 <
        {
          pop$ "  " s * ""
        }
        {
          #4 + duplicate$ s swap$ #1 swap$ substring$
          swap$
          s swap$ global.max$ substring$
        }
      if$
    }
  if$
}
```
⟨/export⟩

2.3.5 Exporting fields

Here, we have four exporting functions, since we also have to deal with abbreviations:

field.export
abbrv.export
name.export
url.export

```
⟨*export⟩
FUNCTION{field.export}
{
  duplicate$ missing$
    'skip$
    {
      left.delim swap$ * right.delim *
      swap$
```

```
      "  " swap$ * " = " * left.width space.complete
      swap$ "," *
      {duplicate$ "" = not}
        {
          right.width split.string 't :=
          *
          write$ newline$
          "" left.width space.complete t
        }
      while$
    }
  if$
  pop$ pop$
}
FUNCTION{abbrv.export}
{
  duplicate$ missing$
    'skip$
    {
      swap$
      "  " swap$ * " = " * left.width space.complete
      swap$ "," *
      {duplicate$ "" = not}
        {
          right.width split.string 't :=
          *
          write$ newline$
          "" left.width space.complete t
        }
      while$
    }
  if$
  pop$ pop$
}
FUNCTION{name.export}
{
  duplicate$ missing$
    'skip$
    {
      left.delim swap$ * right.delim *
      swap$
      "  " swap$ * " = " * left.width space.complete
      swap$ "," *
      {duplicate$ "" = not}
        {
          right.width split.name 't :=
          *
          write$ newline$
          "" left.width space.complete t
        }
      while$
    }
  if$
  pop$ pop$
```

```
}
FUNCTION{url.export}
{
  duplicate$ missing$
    'skip$
    {
      left.delim swap$ * right.delim *
      swap$
      "  " swap$ * " = " * left.width space.complete
      swap$ "," *
      {duplicate$ "" = not}
        {
          url.right.width split.url 't :=
          *
          write$ newline$
          "" left.width space.complete t
        }
      while$
    }
  if$
  pop$ pop$
}
```
⟨/export⟩

2.3.6 Handling abbreviations

Abbreviations are difficult to deal with if we wish to still use them, since BibTEXwill expand them before we can do anything. All we can do is to define them in a special way, in order to be able to get back to the abbreviations later on. This is precisely what we do:

jan-dec
acmcs-tcs
remove.exports.from.months
remove.export.from.journal

```
⟨*export⟩
MACRO{jan}{"export-jan"}
MACRO{feb}{"export-feb"}
MACRO{mar}{"export-mar"}
MACRO{apr}{"export-apr"}
MACRO{may}{"export-may"}
MACRO{jun}{"export-jun"}
MACRO{jul}{"export-jul"}
MACRO{aug}{"export-aug"}
MACRO{sep}{"export-sep"}
MACRO{oct}{"export-oct"}
MACRO{nov}{"export-nov"}
MACRO{dec}{"export-dec"}
MACRO{acmcs}{"export-acmcs"}
MACRO{acta}{"export-acta"}
MACRO{cacm}{"export-cacm"}
MACRO{ibmjrd}{"export-ibmjrd"}
MACRO{ibmsj}{"export-ibmsj"}
MACRO{ieeese}{"export-ieeese"}
MACRO{ieeetc}{"export-ieeetc"}
MACRO{ieeetcad}{"export-ieeetcad"}
MACRO{ipl}{"export-ipl"}
```

```
MACRO{jacm}{"export-jacm"}
MACRO{jcss}{"export-jcss"}
MACRO{scp}{"export-scp"}
MACRO{sicomp}{"export-sicomp"}
MACRO{tocs}{"export-tocs"}
MACRO{tods}{"export-tods"}
MACRO{tog}{"export-tog"}
MACRO{toms}{"export-toms"}
MACRO{toois}{"export-poois"}
MACRO{toplas}{"export-toplas"}
MACRO{tcs}{"export-tcs"}
INTEGERS{ intxt }
FUNCTION{remove.exports.from.months}
{
  #0 'intxt :=
  duplicate$ missing$
    'skip$
    {'t :=
    ""
    {t #1 #1 substring$ "" = not}
      {
      t #1 #7 substring$ "export-" =
        {intxt
           {right.delim * #0 'intxt :=}
           'skip$
         if$
         duplicate$ "" =
           'skip$
           {" # " *}
         if$
         t #8 #3 substring$ *
         t #11 global.max$ substring$ 't :=}
        {intxt
           'skip$
           {duplicate$ "" =
              {}
              {" # " *}
            if$
            left.delim * #1 'intxt :=}
         if$
         t #1 #1 substring$ *
         t #2 global.max$ substring$ 't :=}
      if$
      }
    while$
    intxt
      {right.delim *}
      'skip$
    if$
    }
  if$
}
FUNCTION{remove.export.from.journals}
{
```

```
  duplicate$ missing$
    'skip$
    {
      duplicate$ #1 #7 substring$ "export-" =
        {#8 global.max$ substring$}
        {left.delim swap$
         right.delim * *}
      if$
    }
  if$
}
```
⟨/export⟩

2.3.7 Now, we export...

We gather everything. This is were special fields must be added for being exported:

entry.export.standard
entry.export.extra
entry.export
export

```
⟨*export⟩
FUNCTION{entry.export.standard}
{
  "address" address field.export
  "author" author name.export
  "booktitle" booktitle field.export
  "chapter" chapter field.export
  "crossref" crossref field.export
  "edition" edition field.export
  "editor" editor name.export
  "howpublished" howpublished field.export
  "institution" institution field.export
  "journal" journal remove.export.from.journals abbrv.export
  "key" key field.export
  "month" month remove.exports.from.months abbrv.export
  "note" note field.export
  "number" number field.export
  "organization" organization field.export
  "pages" pages field.export
  "publisher" publisher field.export
  "school" school field.export
  "series" series field.export
  "type" type field.export
  "title" title field.export
  "volume" volume field.export
  "year" year field.export
}
FUNCTION{entry.export.extra}
{
  "abstract" abstract field.export
  "doi" doi field.export
  "eid" eid field.export
  "isbn" isbn field.export
  "issn" issn field.export
  "language" language field.export
  "url" url url.export
```

```
}
FUNCTION{entry.export}
{
  entry.export.standard
  entry.export.extra
}
FUNCTION{export}
{
  "@" type$ * "{" * cite$ * "," * write$ newline$
  entry.export
  "}" write$ newline$ newline$
}
```
⟨/export⟩

2.3.8 Miscellanea

We also have to handle preamble, and to define functions for each entry type (we won't use them but otherwise, BibTEXwould complain).

preamble
header
entries.headers
article-unpublished

```
⟨*export⟩
FUNCTION{preamble}
{
preamble$ duplicate$ "" =
  'pop$
  {
    ",--------------------." write$ newline$
    "|     PREAMBLE      |" write$ newline$
    "`--------------------'" write$ newline$ newline$
    "@preamble{ "  swap$
    quote$ swap$ * quote$ *
    {duplicate$ "" = not}
      {
        right.long.width split.string 't :=
        *
        write$ newline$
        "" left.short.width space.complete t
      }
    while$
    "}" write$ newline$ newline$
    pop$ pop$
  }
if$
}
FUNCTION{header}
{
%"** This file has been automatically generated by bibexport **"
%write$ newline$
%"** See   http://www.lsv.ens-cachan.fr/~markey/bibla.php    **"
%write$ newline$
%"** for more informations about bibexport.                  **"
%write$ newline$
newline$
}
```

```
FUNCTION{entries.header}
{
    ",--------------------." write$ newline$
    "|  BIBTEX ENTRIES   |" write$ newline$
    "`--------------------'" write$ newline$ newline$
}
FUNCTION{article}{export}
FUNCTION{book}{export}
FUNCTION{booklet}{export}
FUNCTION{conference}{export}
FUNCTION{habthesis}{export}
FUNCTION{inbook}{export}
FUNCTION{incollection}{export}
FUNCTION{inproceedings}{export}
FUNCTION{journals}{export}
FUNCTION{manual}{export}
FUNCTION{mastersthesis}{export}
FUNCTION{misc}{export}
FUNCTION{phdthesis}{export}
FUNCTION{proceedings}{export}
FUNCTION{techreport}{export}
FUNCTION{unpublished}{export}
⟨/export⟩
```

2.3.9 Main program

We now can execute and iterate those functions:

```
⟨*export⟩
READ
EXECUTE{header}
EXECUTE{preamble}
EXECUTE{entries.header}
ITERATE{export}
⟨/export⟩
```