

Bibulus — a Perl/XML replacement for BibTeX

Thomas Widmann

3/2, 54 Mavisbank Gardens

Glasgow G51 1HL

Scotland, EU

thomas@widmann.uklinux.net

<http://www.widmann.uklinux.net>

Abstract

In this paper Bibulus will be introduced. Bibulus is a very flexible, multilingual system for formatting bibliographies in several formats, including L^AT_EX. Bibliographies are stored in XML, but Bibulus includes a program to convert old BibTeX databases.

Bibulus is being developed as open-source software at <http://savannah.nongnu.org/projects/bibulus/>.

This paper will first summarise the current BibTeX situation. Then, the main features of Bibulus will be introduced, followed by a guide to switching to Bibulus from BibTeX.

Résumé

Dans cet article nous présentons Bibulus, un système multilingue et très flexible pour composer des bibliographies dans plusieurs formats, y compris L^AT_EX. Les bibliographies sont stockées en XML, mais Bibulus propose également un programme pour convertir des vieilles banques de données BibTeX.

Bibulus est développé comme logiciel *Open Source* dans <http://savannah.nongnu.org/projects/bibulus/>.

Dans la première partie de l'article nous présentons l'existant. Puis, les caractéristiques principales de Bibulus seront présentées, suivies d'un guide du passage de BibTeX à Bibulus.

Background

BibTeX is a great tool — something which is demonstrated by the fact that it is still being used today, almost twenty years after it was created — but it has some problems. Some of these have been solved by additional packages:

Natbib and other L^AT_EX packages add real support for author–date citations.

Custom-bib makes it possible for ordinary human beings to construct .bst files according to their needs.

Jurabib and other packages add support for citations in footnotes and for use of *ibid.* citations.

Chapterbib and bibunits allow the use of multiple bibliographies in a single document.

amsrefs among other things allows the definition of bibliographies within the L^AT_EX file and adds many more entry types.

However, to the best of my knowledge, there are still problems which are not satisfactorily handled by any existing programs:

- Sorting can be very difficult for non-English bibliographies.

- Sharing a single bibliography file between documents in several languages is very difficult, because notes, dates and types are stored as ordinary text in the bibliography file.
- Creating bibliographies in multiple writing systems is not handled at all (except for solutions that allow just one additional alphabet, such as `hellas.bst` for Greek).
- Interacting with programs other than L^AT_EX is problematic, because many features are hard-coded in BibTeX.
- Although it is not too difficult to define a .bst file thanks to Custom-bib, it would be nice to be able to define the appearance of the bibliography within the L^AT_EX file.¹
- Like TeX, BibTeX is compiled with a certain memory limit, which means that an expanded version is sometimes needed, something which is likely to cause a newcomer immense problems.

¹. To be fair, this is handled by `amsrefs`, but this is not a part of BibTeX.

Enter Bibulus

Bibulus is an attempt to address the problems listed above. In the following some of its main features will be introduced.

XML Bibulus requires its databases to be in Bibulus XML (specified in a DTD² bundled with Bibulus). This is not as problematic as it sounds for two reasons.

First, Bibulus comes with a program which will convert old BibTeX databases to XML.

Second, data in XML is generally very easy to convert to another kind of XML, for instance by means of an XSLT script. Although no such scripts are included in the Bibulus module at the moment, it could easily be done if there was a need for them, for instance in order to use XML bibliographies conforming to other bibliographic DTDs with Bibulus.

Unicode Bibulus is truly multilingual. It uses Unicode internally, but it can both read and write other character sets.

Perl Bibulus is written in pure Perl. This means it is very portable, as Perl is available on most platforms.

Not just L^AT_EX To Bibulus, L^AT_EX is just yet another input/output format. If you want, you can also get your bibliography in pure ASCII or in HTML, and other formats are easily added.³

Specifying the style in your document Bibliography styles can be defined in the L^AT_EX file. While definitions via `\bibliographystyle` are still supported, one can also use this new format:

```
\bibulus{citationstyle=numerical,
          surname=comes-first,
          givennames=initials,
          blockpunctuation=.}
```

Of course, many more options are possible.

Entries can also be specified or modified within the L^AT_EX document. For instance, the following command will add a note with the text “Great!” to the entry called `sample`.

```
\bibulusadd{sample}{note}{Great!}
```

Open Source Bibulus is released under the GNU Public Licence which among other things means you get the source code and are free to make any changes you want. Bibulus is very much work in progress, both in the sense that many features have not been implemented yet and that there is a good chance your requests will be implemented.

2. The Bibulus DTD has been inspired by the one described in *The L^AT_EX Web Companion* [1] and by various bibliographic DTDs that can be found on the Internet.

3. For instance, as suggested by a reviewer, it would be easy to make Bibulus output BibTeX databases.

Name “Bibulus” means *fond of drink, thirsty* in Latin. Furthermore, M. Calpurnius Bibulus was consul in Rome together with C. Iulius Cæsar in the year 59 BC.

The name was given in the hope that the program will be fond of “drinking” many books, and that it will rule together with the best typesetting systems.⁴

Tour of Bibulus

In the following we shall have a closer look at various aspects of Bibulus.

Converting BibTeX databases Bibulus comes with a conversion program called `bib2xml` which will convert a BibTeX database to Bibulus XML. As an example, let us convert the file `xampl.bib` which comes with BibTeX.

```
> bib2xml xampl
This is BibTeX, Version 0.99c (Web2C 7.3.1)
The top-level auxiliary file: tmp6600.aux
The style file: bib2xml.bst
Database file #1: xampl.bib
Odd edition number: Silver
```

It should be clear from the above that `bib2xml` calls BibTeX to do its job, thus ensuring it can parse all documents that BibTeX can. It produced one warning, since Bibulus assumes that editions can only be numbers, but `xampl.bib` contains a “silver edition”.

The result of this is a file, `xampl.xml`, which conforms to the Bibulus DTD.

For instance, in the original file there is an entry which looks like this:

```
@PHDTHESIS{phdthesis-full,
  author = "F. Phidias Phony-Baloney",
  title = "Fighting Fire with Fire:
          Festooning {F}rench Phrases",
  school = "Fanstord University",
  type = "{PhD} Dissertation",
  address = "Department of French",
  month = jun # "-" # aug,
  year = 1988,
  note = "This is a full PHDTHESIS entry",
}
```

In Bibulus XML, this has become:

```
<thesis id="phdthesis-full" type="phd">
  <author>
    <name gender="unknown"
           nametype="familylast">
      <given>F. Phidias</given>
      <family>Phony-Baloney</family>
    </name>
  </author>
  <title>Fighting fire with fire:
        Festooning French
        phrases</title>
  <institution>Fanstord
                University</institution>
```

4. Hopefully more happily than its ancient namesake.

```

<place>Department of French</place>
<year month="8">1988</year>
<note>This is a full
  PHDTHESIS entry</note>
</thesis>
  Some notes:
  • The BibTeX entry types @MASTERSTHESIS and
    @PHDTHESIS have been unified.
  • The type field is gone; instead, there is a type at-
    tribute which can contain certain predefined types.
    (Allowing free-form text makes it impossible to out-
    put the entry in another language.)
  • The month field has become an attribute to year;
    furthermore, it now only allows a single month, not
    a range of months.
  • The title has been down-cased. The BibTeX prac-
    tice of down-casing on the fly is problematic, not
    least because it requires an escape mechanism for
    words that should remain in upper case. Instead,
    Bibulus follows amsrefs in storing the lower-case
    version in the file and up-casing on the fly instead
    (this applies only to English, of course).
  • Two attributes to name have been added. The first,
    nametype, allows the handling of names in Chinese
    and other languages where the family names comes
    before the given name, even when used in Western
    languages (e.g., “Deng Xiaopeng” would normally
    be written in this way, not as “Deng, Xiaopeng”
    or “X. Deng”).5 The second attribute, gender, is
    necessary to certain languages where other words in
    the sentence inflect according to the gender of the
    author of a reference.

```

Let us regard a further example from the same file.

```

@ARTICLE{article-full,
  author = {L[eslie] A. Aamport},
  title = {The Gnats and Gnus Document
    Preparation System},
  journal = {\mbox{G-Animal's} Journal},
  year = 1986,
  volume = 41,
  number = 7,
  pages = "73+",
  month = jul,
  note = "This is a full ARTICLE entry",
}

```

This is transformed by bib2xml to the following:

```

<article id="article-full">
  <crossref id="article-full-PART2"/>
  <author>
    <name gender="unknown"

```

5. It is thus not to be used for, e.g., Hungarian names that behave as other Western names when used in English.

```

    nametype="familylast">
      <given>L[eslie] A.</given>
      <family>Aamport</family>
    </name>
  </author>
  <title>The gnats and gnus document
    preparation system</title>
  <pages>73+</pages>
  <note>This is a full ARTICLE
    entry</note>
</article>
<magazine id="article-full-PART2">
  <journal>G-Animal's Journal</journal>
  <volume>41</volume>
  <number>7</number>
  <year month="7">1986</year>
</magazine>

```

Most of this is hardly surprising by now, except for the fact that the entry has been split into two. This does not affect the output since Bibulus (like BibTeX) will in-line cross-references that are only used a limited number of times (specified by the user). This allows for a significant simplification of the DTD.

Editing There is no Bibulus editor (for the time being), but there exist many XML editors, all of which ought to work well with Bibulus XML. However, Bibulus XML is really not any more complicated than BibTeX databases, so it is also quite feasible to edit the files in a plain text editor.

The same situation holds for validation, i. e., checking that an XML file conforms to the definitions in the DTD: There is no Bibulus validator, but many standard tools can be used, and it is highly recommended to validate Bibulus bibliographic databases in this way instead of relying on built-in error handling.

Notes and annotations in the text BibTeX requires us to write notes and annotations in the bibliographic database, but there are problems with this approach. Annotations are typically unique to each bibliography (this is often true for notes, too). The bibliographic database is therefore the wrong place to specify them — it should be done in the main text instead. Furthermore, these fields require translation when the document is translated, something which is much easier if they are kept together with the main text. Bibulus allows both for backwards compatibility.

Transliterations and translations One of the most important *raison d'être* for a bibliography formatting system is to make it possible to define an entry once and then extract it in many different formats. To achieve this, Bibulus is able to transliterate names and titles automatically, and it is possible to add translations of titles, either in the XML database or in the L^AT_EX source.

Migrating to Bibulus

In the following we shall see how a L^AT_EX user can move from BibTeX to Bibulus.

Getting started The very first step is to convert the BibTeX databases to Bibulus XML, as described on page 469.

Without making any changes to the L^AT_EX document, one can then start to use `bibulustex` instead of `bibtex`. If a standard bibliography style is used (e.g., `plain`), this should produce equivalent output.

However, only a few bibliography styles are defined, so this is likely to be less than needed.

Farewell to `\bibliographystyle` The second step is to use the `\bibulus` command in L^AT_EX to define the style of the bibliography.

The default is a style close to BibTeX `plain`, so only options that differ need to be defined. For instance, if one wants `alpha` labels (first letters of the last name + last two digits of the year) instead of numerical labels and furthermore wants author names to be written in small caps, one can just write the following in the L^AT_EX document:

```
\bibulus{citationstyle=alpha,
          namefont=sc}
```

More Bibulus commands in L^AT_EX The next step is to start to use the L^AT_EX commands for manipulating the bibliography, e. g., by adding notes, annotations or translations of titles.

It is also possible to create an alias for a title if one is not happy with its label in the XML file. For instance, if *The T_EXbook* is stored with the ID `knuth86`, one might want to issue the following command:

```
\bibalias{knuth86}{texbook}
```

After this, citing `knuth86` and `texbook` will be fully equivalent.

Goodbye to `bibulustex` All that `bibulustex` really does is to get the filename from the command line and then do the following:

```
my $bib = new Bibulus::LaTeX;
$bib->procaux($filename);
open (BBL, ">$filename.bbl");
  or die "Could not write $filename.bbl.\n";
print BBL $bib->getbib;
close BBL;
```

If more functionality is needed, one can thus make a personalised version of `bibulustex` with extra functionality.

For instance, to output all years *ab urbe condita* (after the founding of Rome), add the following code after the first line:

```
$bib->whenparsing('year',
                  sub {
                    return $_[0] + 754;
                  });
```

Most of this is just Perl code. What matters is the following: The code within the `sub` is executed when a `<year>` is encountered; the contents of this XML chunk is passed to the `sub` in `$_[0]`; and, whatever the `sub` returns replaces the old contents of the chunk.

This is a very simple and silly example, but the possibilities are endless, especially as the full power of Perl is available.

Extending Bibulus If the built-in extension hooks do not provide enough freedom, one can extend Bibulus quite easily.

Create a file called, say, `myBibulus.pm` and put the following into it:

```
package myBibulus;
use Bibulus::LaTeX;
our @ISA = qw(Bibulus::LaTeX);

sub newblock {
  return "\\par ";
}
```

Now replace the line

```
my $bib = new Bibulus::LaTeX;
```

with

```
my $bib = new myBibulus;
```

in your personalised version of `bibulustex`, and the bibliographies produced will now have the blocks separated by `\par` instead of `\newblock`.

Any internal Bibulus function can be overridden in this way.

Final words

This has been a brief introduction to the main features of Bibulus. As the program is being developed continuously, there may be features available now that have not been described in this paper, and the syntax of certain commands might have changed slightly.

For more information, please visit the project's website, <http://savannah.nongnu.org/projects/bibulus/>. There are also two mailing lists, one for developers and one for users — please consider joining one of them.

Bibulus still has a some way to go, but with the help of the user community, we can do it!

References

- [1] Michel Goossens and Sebastian Rahtz. *The L^AT_EX Web Companion*. Addison-Wesley, Reading, Massachusetts, 1999.