The `urlbst` package

Norman Gray
`<https://nxg.me.uk>`

Version 0.9.1, 2023 January 30

The `urlbst` package consists of a Perl script which edits BibTEX style files (`.bst`) to add a `webpage` entry type, and which adds a few new fields – notably including `url` – to all other entry types. The distribution includes preconverted versions of the four standard BibTEX `.bst` style files.

It has a different goal from Patrick Daly's `custom-bib` package [2] – that is intended to create a BibTEX style `.bst` file from scratch, and supports `url` and `eprint` fields. This package, on the other hand, is intended for the case where you already have a style file that works (or at least, which you cannot or will not change), and edits it to add the new `webpage` entry type, plus the new fields.

The added fields are:

- `url` and `lastchecked`, to associate a URL with a reference, along with the date at which the URL was last checked to exist;

- `doi`, for a reference's DOI (see `https://doi.org`);

- `eprint`, for an arXiv eprint reference (see `https://arxiv.org`); and

- `pubmed` for a reference's PubMed identifier (PMID, see `http://pubmed.gov`).

The script's home page is `https://purl.org/nxg/dist/urlbst` (**please quote this rather than the URL it redirects to**). It is on CTAN at `https://ctan.org/pkg/urlbst`. The code repository is at `https://heptapod.host/nxg/urlbst`.

Usage

Command line:

```
% urlbst [--setting key=value] [--literal key=value]
    [input-file [output-file]]
```

where the `input-file` is an existing `.bst` file, and the `output-file` is the name of the new style file to be created. If either file name is missing, the default is the standard input or standard output respectively.

There are a number of settings which can be applied on the command-line, to adjust the support in the generated style file. See the following tables for the (current) list, and `urlbst --setting help` and `urlbst --literal help`. Some of these can be alternately set with particular options as shown in Table 3.

setting	description	default
inlinelinks	0=URLs explicit; 1=URLs attached to titles	0
addpubmed	0=no PUBMED resolver; 1=include it	1
addeprints	0=no eprints; 1=include eprints	1
hrefform	0=no crossrefs; 1=hypertex hrefs; 2=hyperref hrefs	0
adddoi	0=no DOI resolver; 1=include it	1
doiform	0=with href; 1=with `\doi{}`	0

Table 1: Defined settings

literal	description	default
urlintro	prefix before URL	URL:
pubmedprefix	prefix before PUBMED ref	PMID:
pubmedurl	prefix to make URL from PMID	
		http://www.ncbi.nlm.nih.gov/pubmed/
eprintprefix	prefix printed before eprint ref	arXiv:
linktextstring	anonymous link text	[link]
citedstring	label in 'lastchecked' remark	cited
eprinturl	prefix to make URL from eprint ref	https://arxiv.org/abs/
doiurl	prefix to make URL from DOI	https://doi.org/
doiprefix	printed text to introduce DOI	doi:
onlinestring	label that a resource is online	online

Table 2: Literal strings for inclusion in the generated file

Each of these options sets a flag in the generated `.bst` file, but you should feel free to adjust/hack these flags by editing the generated file (for example `#0␣'adddoi␣:=` will disable generation of DOI links; don't blame me for the eccentric `.bst` syntax).

For example:

```
% urlbst --eprint bibstyle.bst
```

would convert the style file `bibstyle.bst`, including support for e-prints, and sending the result to the standard output (ie, the screen, so it would more useful if you were to either redirect this to a file or supply the output-file argument).

If the setting `addeprint` is true (ie, 1), then we switch on support for `eprint` fields in the modified .bst file, with a citation format matching that recommended in `https://arxiv.org/help/faq/references`. If the option `adddoi` is true, we include support for a `doi` field, referring to a Digital Object Identifier (DOI) as standardised by `https://doi.org/`. And if `addpubmed` is true, we in-

option	setting
`--[no]eprint`	addeprints
`--[no]doi`	adddoi
`--nohyperlinks`	hrefform=0
`--hyper[tex/ref]`	hrefform=1/2
`--inlinelinks`	inlinelinks

Table 3: Backward-compatibility options

clude support for a `pubmed` field, referring to a PubMed identifier as supported at `http://www.pubmed.gov`.

The genereated .bst file may include support for hyperlinks where appropriate, in the generated entries in the bibliography. If the `hrefform` setting is zero, this is suppressed, and it if is 1 or 2, the support is include using HyperTEX (see `https://arxiv.org/hypertex/#implementation`), supported by xdvi, dvips and others, or the `hyperref` package, respectively. When URLs are included in the bibliography, they are written out using the `\url{...}` command. The `hyperref` support is more generic, and more generally supported, and so you should choose this unless you have a particular need for the HyperTEX support. The `--nohyperlinks` option, which is present by default, suppresses all hyperlinking.

By default, any URL field is displayed as part of the bibliography entry, linked to the corresponding URL. If the `inlinelinks` setting is true, however, then the URL is not displayed in the printed entry, but instead a hyperlink is created, linked to suitable text within the bibliography entry, such as the citation title. This option does not affect the display of eprints, DOI or PubMed fields. It makes no sense to specify `inlinelinks` with `hrefform=0`, and the script warns you if you do that, for example by failing to specify one of the link-style options. This option is (primarily) useful if you're preparing a version of a document which will be read on-screen; the point of it is that the resulting bibliography is substantially more compact than it would otherwise be.

The support for all the above behaviours is always included in the output file. The options instead control only whether the various behaviours are enabled or disabled, and if you need to alter these, you may do so by editing the generated .bst file and adjusting values in the `{init.urlbst.variables}` function, where indicated.

The generated references have URLs inside `\url{...}`. The best way to format this this is with the `url` package (see [5] for pointers), but as a poor alternative, you can try `\newcommand{\url}[1]{\texttt{#1}}`. The `hyperref` package automatically processes `\url{...}` in the correct way to include a hyperlink, and if you have selected `hyperref` output, then nothing more need be done. If you selected HyperTEX output, however, then the script encloses the `\url` command in an appropriate HyperTEX special.

When the style file generates a link for DOIs, it does so by prepending the string `https://doi.org/` to the DOI. This is generally reasonable, but some DOIs have characters in them which are illegal in URLs, with the result that the resulting `doi.org` URL doesn't work. The only real way of resolving this is to write a URL-encoding function in the style-file implementation language, but while that would doubtless be feasible in principle, it would be hard and very, very, ugly. The only advice I can offer in this case is to rely on the fact that the DOI will still appear in the typeset bibliography, and that users who would want to take advantage of the DOI will frequently (or usually?) know how to resolve the DOI when then need to. As a workaround, you could include a URL-encoded DOI URL in the url field of the entry (thanks to Eric Chamberland for this suggestion). A slightly more generic approach is to apply the setting `doiform=1`, which will generate DOIs wrapped in the macro `\doi{...}`. This allows you to supply a generic `\doi` macro to format them as you desire.

The `doi` field may include the `https://doi.org/`, or it may omit it; the stylefile adds or removes the prefix as appropriate. Note that this parsing is

3

rudimentary: it won't detect any of the variants of this prefix, meaning `http:` or `dx.doi.org`, both of which are now deprecated.

The `urlbst` script works by spotting patterns and characteristic function names in the input `.bst` file. It works as-is in the case of the four standard BibTeX style files `plain.bst`, `unsrt.bst`, `alpha.bst` and `abbrv.bst`. It also works straightforwardly for many other style files – since many of these are derived from, or at least closely inspired by, the standard ones – but it does not pretend that it can do so for all of them. In some cases, such as the style files for the `refer` or `koma-script` packages, the style files are not intended to be used for formatting; others are sufficiently different from the standard files that a meaningful edit becomes impossible. For the large remainder, however, the result of this script should need only relatively minor edits before being perfectly usable.

New `.bib` entry and field types

The new entry type webpage has required fields title and url, and optional fields author, editor, note, year, month and lastchecked. The url and lastchecked fields are new, and are valid in other entry types as well: the first, obviously, is the URL which is being cited, or which is being quoted as an auxiliary source for an article perhaps; the second is the date when you last checked that the URL was there, in the state you quoted it; this is necessary since some people, heedless of the archival importance of preserving the validity of URLs, indulge in the vicious practice of reorganising web pages and destroying links. For the case of the webpage entry type, the editor field should be used for the 'maintainer' of a web page.

For example, in Figure 1 we illustrate two potential `.bib` file entries. The `@webpage` entry type is the new type provided by this package, and provides reference information for a webpage; it includes the new `url` and `lastchecked` fields. There is also an example of the standard `@book` entry type, which now includes the `url` and `lastchecked` fields as well. The difference between the two references is that in the `@book` case it is the book being cited, so that the `url` provides extra information; in the `@webpage` case it is the page itself which is of interest. You use the new `eprint`, `doi` and `pubmed` fields similarly, if the bibliographic item in question has an e-print, DOI or PubMed reference.

How do you use this in a document? To use the the `alphaurl.bst` style – which is a pre-converted version of the standard `alpha.bst` style, included in the `urlbst` distribution – you simply make sure that `alphaurl.bst` is in your BibTeX search path (use the command `kpsepath bst` to find this path and `kpsewhich alphaurl.bst` to confirm that BibTeX can find it) and add `\bibliographystyle{alphaurl}` to your LaTeX document.

Sources

There are various sources which suggest how to format references to web pages. I have followed none specifically, but fortunately they do not appear to materially disagree.

ISO-690 [3] is a formal standard for this stuff. Walker and Taylor's *Columbia*

4

```
@Manual{w3chome,
  url =           {http://www.w3.org},
  title =         {The World Wide Web Consortium},
  year =          2009,
  lastchecked = {26 August 2009}}

@Book{schutz,
  author =        {Bernard Schutz},
  title =         {Gravity from the GroundUp},
  publisher =     {Cambridge University Press},
  year =          {2003},
  url =           {http://www.gravityfromthegroundup.org/},
  lastchecked = {2008 June 16}}
```

Figure 1: The new `@webpage` entry type, and the `url` field in action

Guide to Online Style [6] provides extensive coverage (but is only available on dead trees). There are two style guides associated with the APA, namely the published APA style guide [1] (a paper-only publication, so should be ignored by all, if there's any justice in the world), and what appears to be the 1998 web-citation proposal for that [4], which also includes some useful links. The TeX FAQ [5] has both practical advice and pointers to other sources.[1]

Hints

If you use Emacs' BibTeX mode, you can insert the following in your `.emacs` file to add knowledge of the new webpage entry:

```
(defun my-bibtex-hook ()
  (setq bibtex-mode-user-optional-fields '("location" "issn")) ; e.g.
  (setq bibtex-entry-field-alist
        (cons
         '("Webpage"
           ((("url" "The URL of the page")
             ("title" "The title of the resource"))
            (("author" "The author of the webpage")
             ("editor" "The editor/maintainer of the webpage")
             ("year" "Year of publication of the page")
             ("month" "Month of publication of the page")
             ("lastchecked" "Date when you last verified the page was there")
             ("note" "Remarks to be put at the end of the entry"))))
         bibtex-entry-field-alist)))
(add-hook 'bibtex-mode-hook 'my-bibtex-hook)
```

After that, you can add a webpage entry by typing `C-c␣C-b␣webpage` (or `M-x␣bibtex-entry`).

It is a *very* good idea to use the `url` package: it deals with the problem of line-breaking long URLs, and with the problem that BibTeX creates, of occasionally inserting %-signs into URLs in generated bibliographies.

See also the URL entry in the UK TeX FAQ [5], and references therein.

---

[1]Emory University's Goizueta Business Library once had a collection of useful links on this topic, but they've whimsically changed the URL at least twice since I first distributed `urlbst`, and I've got fed up fixing their broken link.

Acknowledgements, and release notes

Thanks are due to many people for suggestions and requests: to Jason Eisner for suggesting the `--inlinelinks` option; to 'ijvm' for code contributions in the `urlbst` script; to Paweł Widera for the suggestion to use `\path` when formatting DOIs; to Michael Giffin for the suggestion to include PubMed URLs; to Katrin Leinweber for the pull request which fixed the format of DOI references.

0.9.1, 2023 January 30 Added code to spot and wrangle the `https://doi.org/` URL prefix – the code now behaves correctly whether this is present or not.

0.9, 2022 December 1
- Changed repository location to heptapod[2] (when Bitbucket dropped support for Mercurial). The issue links below, pointing to Bitbucket, will therefore no longer work.
- Refactoring of the mechanism for configuring the generated `.bst` file, specifically adding the `--setting` option.
- Added the `doiform` setting to generate DOIs wrapped in `\doi{...}`.

0.8, 2019 July 1
- The presence of a preexisting `format.doi`, `format.eprint` or `format.pubmed` function is now detected, and warned about. The resulting `.bst` file might still require some manual editing. Resolves issue 8[3].
- Clarified licences (I hope).
- Adjust format of DOI resolver. Resolves issue 11[4], thanks to code contributed by Katrin Leinweber[5].

0.7, 2011 July 20 Add `--nodoi`, `--noeprints` and `--nopubmed` options (which defaulted on, and couldn't otherwise be turned off)

0.7b1, 2011 March 17 Allow parameterisation of literal strings, with option `--literal`.

0.6-5, 2011 March 8 Adjust support for inline links (should now work for arXiv, DOI and Pubmed)

0.6-4, 2009 April 28 Work around BibTeX linebreaking bug (thanks to Andras Salamon for the bug report).

0.6-3, 2009 April 19 Fix inline link generation (thanks to Eric Chamberland for the bug report).

0.6-2, 2008 November 17 We now turn on inlinelinks when we spot format.vol.num.pages, which means we include links for those styles which don't include a title in the citation (common for articles in physical science styles, such as aip.sty).

---

[2] `https://heptapod.host/nxg/urlbst`

[3] `https://bitbucket.org/nxg/urlbst/issue/8`

[4] `https://bitbucket.org/nxg/urlbst/issue/11`

[5] `https://bitbucket.org/nxg/urlbst/pull-requests/1/`

0.6-1, 2008 June 16 Fixed some broken links to the various citation standards (I think in this context this should probably *not* be happening, yes?). The distributed `*url.bst` no longer have the `--inlinelinks` option turned on by default.

0.6, 2007 March 26
- Added the option `--inlinelinks`, which adds inline hyperlinks to any bibliography entries which have URLs, but does so inline, rather than printing the URL explicitly in the bibliography. This is (primarily) useful if you're preparing a version of a document which will be read on-screen. Thanks to Jason Eisner for the suggestion, and much testing.
- Incorporate hyperref bugfixes from Paweł Widera.
- Further reworkings of the inlinelinks support, so that it's now fired by a `format.title` (or `format.btitle`) line, with a fallback in `fin.entry`. This should be more robust, and allows me to delete some of the previous version's gymnastics.
- Reworked `inlinelinks` support; should now be more robust. Incorporate hyperref bugfixes from Paweł Widera.
- Added the option `inlinelinks`, which adds inline hyperlinks to any bibliography entries which have URLs, but does so inline, rather than printing the URL explicitly in the bibliography. This is (only) useful if you're preparing a version of a document which will be read on-screen.

0.5.2, 2006 September 6 Another set of documentation-only changes, hopefully clarifying installation.

0.5.1, 2006 January 10 No functionality changes. Documentation and webpage changes only, hopefully clarifying usage and configuration

0.5, 2005 June 3 Added support for Digital Object Identifiers (DOI) fields in bibliographies.

0.4-1, 2005 April 12 Documentation improvements – there are now examples in the help text!

0.4, 2004 December 1 Bug fixes: now compatible with mla.bst and friends. Now uses `./configure` (optionally). Assorted reorganisation.

0.3, 2003 June 4 Added `--eprint`, `--hypertex` and `--hyperref` options.

0.2, 2002 October 23 The 'editor' field is now supported in the webpage entry type. Basic documentation added.

0.1, 2002 April Initial version

References

[1] American Psychological Association. *Publication Manual of the American Psychological Association*, 6th edition, 2009. URL: `http://www.apastyle.org/manual/index.aspx` [cited 17 March 2011].

[2] Patrick W Daly. Index of custom-bib [online]. 2003. URL: `https://ctan.org/pkg/custom-bib` [cited 17 March 2011].

[3] International Standards Organisation. ISO 690-2 [online]. URL: `http://en.wikipedia.org/wiki/ISO_690` [cited 17 March 2011].

[4] B Land. Web extension to American Psychological Association style (WEA-PAS) (rev.2.0) [online]. July 2001. URL: `http://www.beadsland.com/weapas/` [cited 17 March 2011].

[5] Robin Fairbairns, editor. UK TeX FAQ: URLs in BibTeX bibliographies [online]. 2002. URL: `https://texfaq.org/FAQ-citeURL` [cited 17 March 2011].

[6] Janice R. Walker and Todd Taylor. *The Columbia Guide to Online Style*. Columbia University Press, 2nd edition, 2006. URL: `http://cup.columbia.edu/book/978-0-231-13210-7/` [cited 17 March 2011].