The collref Package*

Niklas Beisert

Institut für Theoretische Physik
Eidgenössische Technische Hochschule Zürich
Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland

nbeisert@itp.phys.ethz.ch

31 August 2014, v2.0b

Abstract

collref is a LaTeX 2_ε package to automatically collect multiple `\bibitem` references which always appear in the same sequence in `\cite` into a single `\bibitem` block.

Contents

1 Introduction

Suppose a manuscript uses the following set of four references:

[1] Reference A

[2] Reference B

[3] Reference C

[4] Reference D

Now if references B and C cover similar or related material, they might always be cited together as in "[. . . , 2, 3, . . .]" throughout the manuscript. In some (physics) journals it is then customary to collect the two references into a single reference

[*] Earlier versions developed at: Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut), Potsdam, Germany; AEI-2009-054

[1] Reference A

[2] Reference B
Reference C

[3] Reference D

and cite it by "[. . . , 2, . . .]". The package collref automates this process by analysing the `\cite` commands and identifying blocks of references which always appear in conjunction. These blocks are collapsed to a single item in the bibliography. Please note that collref requires the sequence of `\bibitem` entries to match with the sequence of `\cite` blocks. This is most easily achieved through the use of BibTeX with any *unsorted* style.

Similar CTAN Packages. The objective and some of the implementation of the collref package is similar to the CTAN packages mcite by Thorsten Ohl and mciteplus by Michael Shell, but the functionality is different is several respects:

- collref is intended to work transparently: LaTeX documents which compile with collref should also compile fine without invoking collref (obviously without collected references). The package decides automatically which references can be collapsed, no further interaction of the author is required.

  mcite and mciteplus leave the decision/duty to collapse certain references using the modified syntax `\cite{A,*B,*C}`.

- mcite and mciteplus are intended to handle punctuations in collapsed references correctly. This requires a specialised BibTeX style.

  No effort is made in collref in this regard. Some minor modification in `collref.sty` together with a modified BibTeX style might achieve basic punctuation features similar to mcite.

2 Usage

Inclusion. To use collref simply add the command

```
\usepackage{collref}
```

to the preamble of your LaTeX document. No further interaction is required.

Punctuation. collref provide basic punctuation between collected references. This is specified through the package options `\usepackage[`*opt*`]{collref}` where *opt* is one of the following

`nosep` (default)	`parsep`	`bulletsep`	`punctsep`
no separator:	separated by `\par`:	separated by ‘•’:	punctuated by ‘;’ and ’.’:
[1] A [2] B C [3] D	[1] A [2] B C [3] D	[1] A [2] B • C [3] D	[1] A . [2] B ; C . [3] D . cf. note on spacing below.

Alternative separators can be specified in the preamble through the command:

```
\collectsep[punctuation]{separator}
```

The *separator* appears between references in a block and the *punctuation* at the end of a block of references.

Bibliography Preparation. Please note that only such blocks of references can be collapsed which appear in the same order for `\cite` commands as for `thebibliography`. It is recommended to prepare the bibliography through BibTeX which does this automatically. You must use a style which does not sort the references but preserves the order in which they were `\cite`'d, e.g. `unsrt.bst`.

Also note that collref suppresses new paragraphs invoked by empty lines in the bibliography. This allows to use standard BibTeX styles which commonly separate reference entries by empty lines. If these empty lines would be expanded to new paragraphs, collref would not be able to join two references properly. Therefore new paragraphs have to be invoked by the command `\par`.

If you wish to use the style `punctsep`, please refer to the following note on spacing and punctuation.

Spacing and Punctuation. References are usually punctuated in some way. Three of the predefined styles – `nosep`, `parsep` and `bulletsep` – preserve the punctuation from the bibliography.

The fourth predefined style – `punctsep` – automatically performs the punctuation. This however requires care in the preparation of the bibliography: The entries have to be provided *without* punctuation. Furthermore, there must not be *whitespaces* at the end of an entry. They can be suppressed with '`%`' or `\ignorespaces` directly following the last word of the entry. See Appendix B for an example. The standard BibTeX styles, e.g. `unsrt.bst`, have to be adjusted to remove the punctuation and whitespaces.

Control. The package collref provides one command to control which references (not) to collect:

`\nocollect{`*label*`}`

It ensures that the label *label* starts a new block of references. It is not collapsed with earlier references. Later references, however, can still be collapsed to the end of *label.*

Labels for Blocks of References. While collref aims to automatically collect similar references into a single block, it is often convenient for the author to refer to such blocks with a single citation label. Standard TeX/LaTeX commands can be used to define such a block:

`\newcommand{`\blocklabel`}{`*label1, label2, . . .*`}`

Subsequently this block can be referenced with `\cite{`*. . . ,* \blocklabel, . . .`}`.

Interaction with CTAN Packages. The package collref has been tested with other CTAN packages concerned with citations and the bibliography:

- cite: collref works in conjunction with cite. Note that you must load cite *before* collref so that the latter can pass the correctly reduced list of references down to cite. Tested with v5.1 and v5.3 (2010/09/10).
- hyperref: collref works in conjunction with hyperref. The two packages can be loaded in any sequence. Tested with v6.78s and v6.83m (2012/11/06).

3 Revision History

v2.0b: 2014/08/31

- updated author addresses
- minor internal changes

v2.0: 2009/09/07

- proper punctuation added
- blocks of references enabled
- manual extended

v1.0: 2009/06/09

- streamlined detection of chains
- manual and installation package added
- renamed package to collref due to name clash on CTAN
- first version published on CTAN

v0.9:

- package named collect; unpublished

4 Acknowledgements

Thanks to Oleg Zhirov for suggesting proper punctuation and labels for blocks of references.

A Files and Installation

The package consists of the files

`README`	readme file
`collref.ins`	installation file
`collref.dtx`	source file
`collref.sty`	package file
`collsamp.tex`	sample file
`collref.pdf`	manual

The distribution consists of the files `README`, `collref.ins` and `collref.dtx`.

- Run (pdf)LaTeX on `collref.dtx` to compile the manual `collref.pdf` (this file).
- Run LaTeX on `collref.ins` to create the package `collref.sty` and the sample `collsamp.tex`. Copy the file `collref.sty` to an appropriate directory of your LaTeX distribution, e.g. *texmf-root*`/tex/latex/collref`. Alternatively, you may copy `collref.sty` to the local directories of manuscripts for which you wish to use collref.

B Sample File

In this section we provide a sample file.

```
\documentclass{article}
%\usepackage{cite}
\usepackage[punctsep]{collref}
%\usepackage{hyperref}
```

```

\begin{document}

\def\tworef{c8,c9}

\cite{c1,c2,c3,c4}
\nocollect{c3}
\cite{c5,c6,c7,\tworef}
\cite{c5,c6,c7}
\cite{c7,\tworef}

\begin{thebibliography}{11}
\bibitem{c1} reference 1%
\bibitem{c2} reference 2%
\bibitem{c3} reference 3%
\bibitem{c4} reference 4%
\bibitem{c5} reference 5
\bibitem{c6} reference 6
\bibitem{c7} reference 7 %
\bibitem{c8} reference 8\ignorespaces
\bibitem{c9} reference 9\ignorespaces
\end{thebibliography}

\end{document}
```

It produces the output:

```
[1, 2] [3, 4, 5] [3, 4] [4, 5]

  [1] reference 1; reference 2.
  [2] reference 3; reference 4.
  [3] reference 5 ; reference 6 .
  [4] reference 7 .
  [5] reference 8; reference 9.
```

Note the different behaviour for references 5, 6 and 7 for which trailing whitespaces were not removed.

C Implementation

In this section we describe the package `collref.sty`.

Internal Lists. For each bibliography label *label* the package maintains a predecessor `\nc@p@`*label* and a successor `\nc@s@`*label*. These are initially undefined. When a label *label* is first cited these labels are set to the *predecessor* and *successor* labels, respectively, in `\cite{`*. . . , predecessor, label, successor, . . .*`}`. An empty `\nc@p@`*label* or `\nc@s@`*label* refers to the beginning and end of a block, respectively. Whenever `\cite` finds conflicting blocks (non-matching predecessors or successors in two `\cite`'s), it terminates the blocks to the maximum common overlap.

Interface. The package provides two public commands, described above:

```
\newcommand{\collectsep}[2][]{\def\nc@punct{#1}\def\nc@sep{#2}}
\newcommand{\nocollect}[1]{\nc@breakbefore{#1}\ignorespaces}
```

Package Options. The package provides four predefined separators described above:

```
\DeclareOption{nosep}{\collectsep{}}
\DeclareOption{parsep}{\collectsep{\par}}
\DeclareOption{bulletsep}{\collectsep{\textbullet{} }}
\DeclareOption{punctsep}{\collectsep[.]{; }}
\ExecuteOptions{nosep}
\ProcessOptions
```

Internal Commands. Some internal commands for abbreviation:

```
\newcommand{\nc@getcsname}[1]{\csname #1\endcsname}
\newcommand{\nc@setcsname}[2]{\expandafter\xdef\csname #1\endcsname{#2}}
```

Command to terminate the chain before a label: The predecessor of the label is terminated. If the predecessor was active, its successor is also terminated.

```
\newcommand{\nc@breakbefore}[1]{%
  \edef\nc@citepred{\@ifundefined{nc@p@#1}{}{\nc@getcsname{nc@p@#1}}}%
  \ifx\nc@citepred\@empty\else\nc@setcsname{nc@s@\nc@citepred}{}\fi%
  \nc@setcsname{nc@p@#1}{}%
}
```

Command to terminate the chain after a label. Similar to the above command.

```
\newcommand{\nc@breakafter}[1]{%
  \edef\nc@citesucc{\@ifundefined{nc@s@#1}{}{\nc@getcsname{nc@s@#1}}}%
  \ifx\nc@citesucc\@empty\else\nc@setcsname{nc@p@\nc@citesucc}{}\fi%
  \nc@setcsname{nc@s@#1}{}%
}
```

Citations. Hack for `\@citex`: It is assumed that (as in $\LaTeX\,2_\varepsilon$) `\cite` eventually passes down to `\@citex`.

```
\let\nc@old@citex\@citex

\def\@citex[#1]#2{%
  \let\nc@citecomma\@empty%
  \let\nc@citestring\@empty%
  \let\nc@citelast\@empty%
  \edef\nc@citelist{#2}%
```

Main loop to process the arguments of `\cite`. The current label is stored in `\nc@citethis`.

```
  \@for\nc@citethis:={\nc@citelist}\do{%
    \edef\nc@citethis{\expandafter\@firstofone\nc@citethis\@empty}%
```

The first entry has no predecessor, terminate the chain.

```
    \ifx\nc@citelast\@empty%
      \nc@breakbefore{\nc@citethis}%
    \else%
```

Non-first entry: Fill undefined successor and predecessors entries with the current chain sequence.

```
      \@ifundefined{nc@s@\nc@citelast}%
                   {\nc@setcsname{nc@s@\nc@citelast}{\nc@citethis}}{}%
      \@ifundefined{nc@p@\nc@citethis}%
                   {\nc@setcsname{nc@p@\nc@citethis}{\nc@citelast}}{}%
```

Get the successor and predecessors for the last and current entry, respectively.

```
   \edef\nc@citesucc{\nc@getcsname{nc@s@\nc@citelast}}%
   \edef\nc@citepred{\nc@getcsname{nc@p@\nc@citethis}}%
```

In case of mismatching chains: terminate all links.

```
   \ifx\nc@citesucc\nc@citethis%
    \ifx\nc@citepred\nc@citelast%
    \else%
     \nc@breakafter{\nc@citelast}%
     \nc@breakbefore{\nc@citethis}%
    \fi%
   \else%
    \nc@breakafter{\nc@citelast}%
    \nc@breakbefore{\nc@citethis}%
   \fi%
  \fi%
```

Get content of \b@*label* entry to find out whether the \bibitem{label} entry exists. We need to take special care of extended label definitions in hyperref.

```
  {\def\hyper@@link[##1]##2##3##4{##4}%
   \xdef\nc@citelabel{\nc@getcsname{b@\nc@citethis}}}%
```

Only add those labels which actually exist to the pass-on string. This removes collaped references from the citation marks.

```
  \ifx\nc@citelabel\@empty\else%
   \edef\nc@citestring{\nc@citestring\nc@citecomma\nc@citethis}%
  \fi%
```

Write \citation tag to .aux file in original order. Some duplicate \citation's will be written by the original \citex code, but these will have no impact.

```
  \if@filesw\immediate\write\@auxout{\string\citation{\nc@citethis}}\fi%
```

Continue to next label.

```
  \edef\nc@citelast{\nc@citethis}%
  \def\nc@citecomma{,}%
 }%
```

The last entry has no successor, terminate the chain.

```
 \nc@breakafter{\nc@citelast}%
```

Pass on to original LaTeX code.

```
 \nc@old@citex[#1]{\nc@citestring}%
}
```

Bibliography. Enhance the thebibliography environment to a) set the \nc@biblast label to something, and empty \nc@nextpunct (no predecessor for the first entry), b) convert linebreaks into whitespaces (avoid implicit \par's), and c) put the final punctuation for the last entry.

```
\let\nc@old@thebibliography\thebibliography
\let\nc@old@endthebibliography\endthebibliography

\def\thebibliography{%
 \xdef\nc@biblast{asldjfhasklfh}%
```

```
  \xdef\nc@nextpunct{}%
  \catcode`\^^M=10%
\nc@old@thebibliography}

\def\endthebibliography{%
  \nc@nextpunct%
\nc@old@endthebibliography}
```

Overwrite `\bibitem`: It is assumed that the native LaTeX 2_ε code is equivalent but with the LaTeX internals `\@lbibitem` and `\@bibitem`. Some other packages may also redefine `\bibitem` and this will inevitable cause compatibility issues. This implementation is safe with current versions of hyperref.

```
\def\bibitem{\@ifnextchar[\nc@lbibitem\nc@bibitem}
```

`\nc@noitem` is invoked in place of the original `\@bibitem` or `\@lbibitem` for collapsed references:

```
\def\nc@noitem#1{%
  \if@filesw\immediate\write\@auxout{\string\bibcite{#1}{}}\fi%
\ignorespaces}
```

The hack for `\@bibitem`: It checks whether this reference is part of a block. If so, put separator and collect by `\nc@noitem`. Otherwise put punctuation and pass down to `\@bibitem`. Finally let `\nc@biblast` point to current item, and fill the punctuation `\nc@nextpunct` for the next entry.

```
\def\nc@bibitem#1{%
  \edef\nc@bibpred{\@ifundefined{nc@p@#1}{}{\nc@getcsname{nc@p@#1}}}%
  \ifx\nc@biblast\nc@bibpred\nc@sep\nc@noitem{#1}%
    \else\nc@nextpunct\@bibitem{#1}\fi%
  \xdef\nc@biblast{#1}%
  \xdef\nc@nextpunct{\nc@punct}%
\ignorespaces}
```

Similar hack for `@lbibitem`:

```
\def\nc@lbibitem[#1]#2{%
  \edef\nc@bibpred{\@ifundefined{nc@p@#2}{}{\nc@getcsname{nc@p@#2}}}%
  \ifx\nc@biblast\nc@bibpred\nc@sep\nc@noitem{#2}%
    \else\nc@nextpunct\@lbibitem[#1]{#2}\fi%
  \xdef\nc@biblast{#2}%
  \xdef\nc@nextpunct{\nc@punct}%
\ignorespaces}
```

D Copyright

This work has the LPPL maintenance status ‘maintained’.

The Current Maintainer of this work is Niklas Beisert.

This work consists of the files `README`, `collref.dtx` and `collref.ins` as well as the derived files `collref.sty`, `collsamp.tex` and `collref.pdf`.