LaTeX

A proposal for citation commands in LaTeX3

Pedro J. Aphalo

1 Introduction

There are several recent published and unpublished papers on the problem of what kinds of citations and bibliographies are commonly used in different disciplines (Rhead, 1990, 1993, and unpublished, Wonneberger and Mittelbach, 1991), one proposal of how LaTeX3 might support them (Rhead, 1991, and unpublished), and one implementation called Camel (Bennett, 1996) more in line with the current proposal than with Rhead's.

Rhead describes three citation schemes: citation-by-key,[1] author-date and short-form. Although he accepts that if possible using a single set of commands for the three citation schemes would be preferable, he argues that this is not possible (Rhead, 1991).

I agree in general with David Rhead's description of the different citation schemes, but in contrast to what he assumes in his proposal for citation commands, I think that it is possible to mark the input for the three citation schemes using the same set of commands. Below I describe an alternative proposal for the syntax of citation commands in LaTeX3 which does not rely on different citation commands for different citation schemes. I hope that having two contrasting proposals available will highlight the tradeoffs involved, and help the developers of LaTeX3 design a good user interface for citations. However, I do not expect either proposal to be adopted without substantial changes.

As a user, I strongly object to having three different sets of commands for citations. (In disciplines like biology different schemes are used by different journals and publishing houses, so having to change from one scheme to a different one is a real problem.) Having a user interface with more commands also makes it more difficult for the user to learn how to use them! I do not object so much to having different `\bibitem`[2] commands because I use BibTeX for all

[1] I prefer the more general name 'citation-by-key' to 'reference-by-number' as used by Rhead, because although the key is usually a number this is not always the case as in `alpha.bst`.

[2] To keep this discussion simple I will assume a system which follows the same general design for the generation of citations and lists of references as that implemented in the current versions of LaTeX and BibTeX. This implementation

my manuscripts, but I think that a consistent syntax would also be highly desirable.

Preserving the separation of contents and format would provide for a generic markup of manuscripts that could be easily translated to in-house formats. The normal LaTeX document styles and bibliography styles could use exactly the same syntax as in-house styles and make electronic submission for publication much easier than nowadays — e.g. Elsevier Science Publishers and Kluwer Academic Publishers have each one a single, but different, generic format for submission of manuscripts as LaTeX source files to *many* of their journals (Elsevier Science, 1995; Kluwer Academic Publishers, 1997). It would be much better if different publishing houses used the same standard format.

Another advantage of abstracting formatting issues into style files is that the same document can be formatted differently for different purposes (e.g. TeXinfo). If one considers the possibility of using LaTeX for documents to be printed but also viewed on-line, using different formats on paper and screens, the use of a single and consistent syntax becomes very important. For the development of LaTeX viewers, such as TechExplorer from IBM (see `http://www.ics.raleigh.ibm.com/ics/techexp.htm`) the use of such a consistent syntax would guarantee their compatibility with many LaTeX source files without any need of editing.

These are some examples of the advantages of generic markup, which are behind the main objective of this proposal: to achieve a generic syntax for citations capable of supporting the different citation schemes.

2 Rationale

Citation-by-key is the only scheme currently supported by LaTeX without extensions. Consequently the available citation commands are too limited and citation styles that add new commands have proliferated (e.g. `chicago.sty, harvard.sty, authordate.sty`). This is not good, and LaTeX3 should aim at providing a complete set of commands flexible enough to provide to the needs of different citation schemes (however, it should provide only a few basic examples of their use in citation styles and an interface for easily defining new citation styles).

The information needed in different citation schemes is not very different and could be thought of as different subsets of a citation's 'full' information set. The amount of information required for this superset is certainly finite and in practice the different subsets overlap a lot. Abstracting a superset common to all three citation schemes is not trivial, but my attempt at doing such an abstraction is the basis for the present proposal. Another requirement is to have this information parsed into small enough units (i.e. having enough 'fields' in the `\bibitem` commands) so that the right pieces can be chosen by the different citation styles. This implies a trade-off between bibliographic data entry against ease of document markup, but as bibliographic data entry can be automated by use of a program such as BibTeX I think that easy of document markup should be favoured.

In practice, provided that some discipline is used when typing a piece of text using the citation commands proposed here (e.g. use of `\citeasnoun` and `\citenoname` even for the citation-by-key scheme), no adjustment would be required in most cases when changing a document from one citation scheme to another. Consequently, there is no reason from the user's point of view that justifies breaking one of the design principles of LaTeX: logic structure and format should be kept separate.

It would be possible either to have in a file only the subset of the bibliographic information needed for the scheme in use, and to regenerate (probably by means of a program like BibTeX) the `\bibitem` commands when switching to a different scheme, or to have the full set of information always available in the `\bibitem` commands. I think that the second option should be favoured because it makes switching between styles a lot easier and also because one may need to include the name of authors or dates of publication in the text independently of the citation scheme being used. Another reason is that the use of the same syntax for `\bibitem` commands for all the schemes (a simpler user interface) would simplify the implementation of mixed citation schemes.

A reason given for having different commands is that some documents use more than one citation scheme. However, the use of this 'mixed' scheme is not a common situation, and its full support should be addressed by special styles and not by allowing the simultaneous use of more than one citation style with LaTeX's default commands. Styles supporting the 'mixed' scheme could rely on optional arguments of the same standard `\cite` commands to switch between citation schemes, and in this way they could remain compatible with the three simple schemes.

Mittelbach and Rowley (1993, p. 2) state that 'It [LaTeX] was designed to separate content and form as much as possible...', and that one of the

could change in the future, but discussing such possible changes is not an aim of this article.

aims for LaTeX3 is to separate the interfaces used for generic markup by the author of a LaTeX document and the specification of how the document elements will be formatted (Mittelbach and Rowley, 1993, p. 5).

Keeping the command set consistent and fully implemented in all styles is the basis for keeping format and structure separate. Such a command set *allows* generic markup which makes it possible to change the citation style without having to edit the whole document to replace incompatible variations of the `\cite` commands.

3 Commands

I propose the following basic set of commands, to be implemented in *all* citation styles. There is a trade-off: the number of commands is larger than what would be needed to support a single citation scheme, especially the citation-by-key scheme, but this is the price that has to be paid for having a single set of commands which can support well all three citation schemes.

I have tried to keep the syntax of the commands consistent with the rest of LaTeX. The biggest departure from 'normal' LaTeX syntax is the use of optional arguments *within* the curly brackets of the main argument of `\cite`.[3] This seems to me the most logical way of making clear that the optional string arguments remain attached to the citation specified by each citation key (the order of citations within a pair of *citation brackets* is *not* guaranteed to be the same as the order of the citation keys supplied as argument: styles may arrange them either in alphabetical or date order, for the author-date scheme, or in key order for the citation-by-key scheme, or leave them in the argument's order). The syntax also assumes that styles that format first citations differently from later ones automatically detect which ones are first citations (this is possible to achieve, and has been implemented by Peter Williams in the Harvard family of bibliography styles). The commands marked • cannot be replaced with simpler ones, those marked ◦ can be replaced with more basic ones but are included because they are used frequently. The commands marked + should be considered optional: a syntax to be used if implemented. `\authorof` and `\yearof` are not citation commands, but are very useful as they guarantee consistency of spelling for author names and consistency for dates. They could also be very useful for generating documents using templates or 'boiler plates'.

[3] A syntax first proposed by David Rhead.

- • `\cite[`*opt*`]{[`*str*`]`*key*`[`*str*`],...`[4]`}`, where *opt* is a style specific option, *str* is a text string, and *key* is the citation key of a `\bibitem` (or of a BibTeX database entry), generates a citation string, including enclosing brackets or footnote(s). Options, if not supported, should be ignored,[5] with a warning except for `f` and `l` below which should be supported whenever they are meaningful and quietly ignored otherwise.
  `\cite[f]{[`*str*`]`*key*`[`*str*`],...}` ... treat as first citation(s) of the *key*(s) even if they are later instances of the citations.
  `\cite[l]{[`*str*`]`*key*`[`*str*`],...}` ... treat as later citation(s) of the *key*(s) even if they are first instances of the citations. (The options `f` and `l` apply to the whole compound citation, because forcing a special format is related to the context of the citation and not to individual *key*s. Special formatting is needed when some citations are not considered to be part of the sequence of citations in the main body of a document, as is frequently the case for citations within tables. Moreover, the `\cite` commands with `f` or `l` options should be ignored when automatically deciding whether a citation to a given *key* is the first or a later one.)
- ◦ `\citeaffixed[`*opt*`]{`*aff*`}{[`*str*`]`*key*`[`*str*`],...}` generates a citation string with *aff* affixed, including enclosing brackets or footnote(s). `f` and `l`, and other options as in `\cite`.
- ◦ `\citeasnoun[`*opt*`]{`*key*`[`*str*`]}` generates a string with author's name(s) and citation to be used as a noun in a sentence. `f` and `l`, and other options as in `\cite` above. (Rhead would not include this command and the next one, arguing that the authors' names are not part of the citation but rather part of the text. I think that he is only partly right, because at least in the author-date scheme, whether the name list is abbreviated or not depends on whether the names precede the first or a later instance of a citation. In other words, the instructions to authors of several biological journals do implicitly consider them part of the citation.)
- ◦ `\citepossesive[`*opt*`]{`*key*`[`*str*`]}` generates a string with author's name(s) and citation to be

[4] The sequence of *key* plus attached strings can be repeated, using comma as separator. This also applies to the next command, but not to other cite commands described below.

[5] This allows compatibility with other citation styles that do not need the extra information. However, style authors should strive to give unique names to options, unless they have the same function.

used as a possessive in a sentence. `f` and `l`, and other options as in `\cite` above.

- `\citenoname[`*opt*`]{[`*str*`]`*key*`[`*str*`]}` behaves as `\cite` but does not include the authors in the author-date scheme. `f` and `l`, and other options as in `\cite` above. (It is needed for some 'MLA' and 'MHRA' examples in one of Rhead's unpublished papers. Rhead uses the name `\dcite` for this command in his proposal for supporting the author-date scheme.)
- `\citation{`*str*`}` formats the string(s) as a single compound citation — i.e. encloses them in brackets or sets them as a footnote —, and simultaneously disables the generation of additional brackets or footnotes by `\cite` commands used as part of its *str* argument. (The logical markup has a different meaning than a `\cite` with multiple keys, although in some cases the formatting may be the same: the list of keys in a `\cite` command is unordered and the style is allowed to sort them. In contrast, the different `\cite` commands within a citation are ordered, and any connecting text also remains always where it is in the input text, between a given pair of `\cite` commands. This command and the following one make the `\citeNP` family of commands used in `chicago.sty` and `ltugboat.cls` with the `harvardcite` option, as used in the source of this document, redundant.)
- `\nocitation{`*str*`}` does not format the string(s) as a citation — i.e. does not enclose them in brackets or set them as a footnote —, but similarly to `\citation` disables the generation of additional brackets or footnotes by `\cite` commands used as part of its *str* argument. (To be used for example in tables summarizing data from other publications, in which table cells normally contain citations formatted as normal text, without brackets in the case of the author-date scheme.)
- `\nocite{`*key*`,...}` generates no output, but forces inclusion of references to the *key*(s) in the list of references.

+ `\bibref{`*key*`}` generates a full reference at the position in the document where the command appears, not a citation. (To be used in abstracts in which references are usually given in full rather than as a pointer to a list of references, and also for writing commented lists of suggested reading.)
+ `\yearof{`*key*`}` year of *key*; it is not a citation — i.e. a reference to *key* is not included in the list of references.
+ `\authorof[`*opt*`]{`*key*`}` author or authors of *key*; it is not a citation — i.e. a reference to *key* is not included in list of references. `f` and `l`, and other options as in `cite` above.

4 Related problems

4.1 Pinpointing locations within a reference

In the previous section, to keep the description of the `\cite` commands simple, I have ignored the problem of citing specific parts of a publication. The syntax for `\cite` given above does not explicitly support style and language independent pinpointing to pages, sections, chapters, etc. Full support for pinpoints should not only provide the pinpoint prefix and typeface, but also scan their argument to determine whether a plural is needed (e.g. 'p.' or 'pp.' for pages). Styles could differ also in the location of the pinpoint: (Hoff 1992, pp. 143–179) vs. (pages 143–179 of Hoff 1992).

This functionality could be accommodated by the following syntax for the `\cite` command: `\cite{Hoff92<p:143--179>}`; also valid input is `\cite{Hoff92<p:257>[,␣describes␣well␣my␣feelings]}`. In other words `<...>` would be used for pinpoints, and `[...]` for strings. The problem of this approach is that it probably would make the code for `\cite` complicated. The revised syntax of the `\cite` command would become:

- `\cite[`*opt*`]{[`*str*`]`*key*`<`*pinpoint*`>[`*str*`],...}` with arguments enclosed in `[ ]` and `< >` being optional, or
- `\cite[`*opt*`]{`*key*`<`*pinpoint*`>,...}` if the optional *str* arguments are considered redundant.

with a similar syntax for all other `\cite` commands.

For either of these two last variants of the syntax for `\cite` commands the proposed pinpoint arguments are: `vol:`*volume(s)*, `part:`*part(s)*, `ch:`*chapter(s)*, `sec:`*section(s)*, `p:`*page(s)*, `fig:`*figure(s)*, `tab:`*table(s)*, `plate:`*plate(s)*, `eq:`*equation(s)*, `th:`*theorem(s)*, `col:`*column(s)*, `para:`*paragraph(s)*, `line:`*line(s)*.
Compound pinpoints such as `<vol:3,ch:9,p:1012>` are also valid.

4.2 Signals and other terms

What Rhead calls "signals" are very often used in texts about law, and less frequently in other disciplines. For these terms, the typefaces and abbreviations used depend on house styles. An optional style could define them, but it is arguable whether they differ from the more general problem of using abbreviations and symbols, except for the fact that

they are widely used. The advantage of providing an optional style file with their default definitions as part of LaTeX3 would be the standardisation of the names used for this group of commands. Rhead proposes a list of such commands as used in texts about law: `\accord`, `\Accord`, `\and`, `\butcf`, `\Butcf`, `\butsee`, `\Butsee`, `\cf`, `\Cf`, `\compare`, `\Compare`, `\contra`, `\Contra`, `\eg`, `\Eg`, `\etseq`, `\ibid`, `\Ibid`, `\id`, `\Id`, `\infra`, `\loccit`, `\opcit`, `\see`, `\See`, `\seealso`, `\Seealso`, `\seegenerally`, `\Seegenerally`, `\supra`, `\re`, `\Re`, `\versus`, `\with`. Only some of them are used in texts unrelated to law; the full list should be implemented only in law-specific styles. Some of these "signals" can be used together with citations (a) as explanatory text: `\citeaffixed{\see␣}{Hoff92}` would print as '(see Hoff 1992)', or (b) as pinpoints: `\cite{Hoff92[,␣\loccit]}` would print as '(Hoff 1992, *loc. cit.*)'. I think that pinpoints like *loc. cit.* and *op. cit.* should be handled automatically by citation styles because they are a formatting issue and have no intrinsic meaning — e.g. (Hoff, *op. cit.*) in the right context has exactly the same meaning as (Hoff 1992). In contrast *cf.*, *see*, *versus*, etc. should be specified by the author because they alter the meaning of citations.

5 Examples

A few simple examples of the use of these commands and of how the output might look for the different citation schemes are provided below in the following order: (i) author-date, (ii) citation-by-key using numeric keys, (iii) citation-by-key using alphanumeric keys, and (iv) short-form. In the cases in which the citations within a single `\cite` command could be automatically sorted in either alphabetical or chronological order, both possibilities are shown. For the short-form scheme fake footnotes are given at the end of each item in the list of examples, the numbers for numeric keys in the examples are also faked, but not the authors and titles.

▷ `\cite{Borges78,Hudson18}`
(Borges, 1978; Hudson, 1918) or (Hudson, 1918; Borges, 1978)
(1,2)
[Bor78, Hud18] or [Hud18, Bor78]
[1,2]

[1] Borges, J. L., *El Libro de los Seres Imaginarios.*
[2] Hudson, W. H., *Far Away and Long Ago.*

▷ `\citeaffixed{see}{Borges78,Hudson18}`
(see Borges, 1978; Hudson, 1918) or (see Hudson, 1918; Borges, 1978)
(see 1,2) or even (1,2)
[see Bor78, Hud18] or [see Hud18, Bor78]
[1,2]

[1] See Borges, J. L., *El Libro de los Seres Imaginarios.*
[2] See Hudson, W. H., *Far Away and Long Ago.*

▷ `\citeasnoun{Borges78}`
Borges (1978)
Borges (1)
Borges [Bor78]
Borges[1]

[1] Borges, J. L., *El Libro de los Seres Imaginarios.*

▷ `\citepossesive{Borges78}`
Borges' (1978)
Borges' (1)
Borges' [Bor78]
Borges'[1]

[1] Borges, J. L., *El Libro de los Seres Imaginarios.*

▷ `\citenoname{Borges78}`
(1978)
(1)
[Bor78]
[1]

[1] Borges, J. L., *El Libro de los Seres Imaginarios.*

▷ `\cite{Borges78<p:45-46>,`
`Hudson18<ch:3>}`
(Hudson, 1918, chapter 3; Borges, 1978, pp. 45–46) or (Borges, 1978, pp. 45–46; Hudson, 1918, chapter 3)
(1, pp. 45–46, 2 chapter 3) or (1 chapter 3, 2 pp. 45–46)
[Bor78 pp. 45–46, Hud18 chapter 3] or [Hud18 chapter 3, Bor78 pp. 45–46]
[1,2]

[1] Borges, J. L., *El Libro de los Seres Imaginarios*, pp. 45–46.
[2] Hudson, W. H., *Far Away and Long Ago*, chapter 3.

▷ `\citation{see␣\cite{Borges78}␣or`
`\cite{Hudson18}}`
(see Borges, 1978 or Hudson, 1918)
(see 1 or 2) or (see 2 or 1), where the first number always refers to `Borges78`
[see Bor78 or Hud18]
[1]

[1] See Borges, J. L., *El Libro de los Seres Imaginarios* or Hudson, W. H., *Far Away and Long Ago.*

6 Caveats

The command set proposed cannot support the needs of all legal texts. References to cases and tables of cases require special commands, but citations of books, articles in periodicals, etc., can most probably be supported by the commands given above. In one of his unpublished manuscripts, Rhead gives an example of how the necessary extensions could be supported by a law-specific style. A basic example could be provided as part of LaTeX3 so as to provide a guideline for programmers of styles for legal texts.

While working on this proposal I have deliberately 'forgotten' all problems that have to do with the implementation of the commands. My philosophy is that first we should have clear what we want, and only afterwards worry about the implementation. Only as a last resort should we change the syntax to suit the limitations of TeX. My idea is that we should be very open minded about implementation issues, and even consider *heretical* alternatives such as the use of a preprocessor that reads not only the `.aux` file but also the `.tex` file or maybe even generates the file to be processed by TeX replacing the `\cite` commands with something else that is easier for TeX to process.[6] However, we should not forget that using preprocessors has serious drawbacks: (1) the preprocessors should be as portable as TeX itself, and ports should exist for all platforms and operating systems for which TeX is available. (2) preprocessing or additional passes through TeX itself slow down document formatting.

7 Acknowledgements

This paper is a revised version of my contribution to task 04 of the volunteer work for the LaTeX3 project (Mittelbach, Rowley, and Downes, 1993). This proposal owes much to the extensive work of David Rhead (mostly unpublished, but kindly made available by him) and to the authors of the many citation styles that are available in the CTAN archives. David Rhead (UK), Peter Williams (Australia), John Grace (UK), John Wells (Canada) and Mario Natiello (Sweden) kindly answered my questions and made useful suggestions. A suggestion of Frank Bennett (UK) plus his efforts in implementing Camel were the main incentive behind the revision of this manuscript for *TUGboat*. Finally, the detailed comments from an anonymous referee helped me improve this document.

[6] As long as the syntax of the commands remains consistent, support for different schemes could even rely on different preprocessors!

References

Bennett, F. G., Jr. "Camel: kicking over the bibliographic traces in BibTeX". *TUGboat* **17**(1), 22–28, 1996.

Borges, J. L. *El Libro de los Seres Imaginarios*. Emecé Editores, Buenos Aires, 1978. This edition first published in 1967 by Kier, Buenos Aires.

Elsevier Science. "Preparing Articles with LaTeX. Instructions to Authors for Preparing Compuscripts". 1995. This is file `instraut.dvi`, included in the ESP-LaTeX package archived in the CTAN archives in the directory `macros/latex/contrib/supported/elsevier`.

Hoff, B. *The Te of Piglet*. Dutton, Penguin Books, New York, 1992.

Hudson, W. H. *Far Away and Long Ago*. J. M. Dent and Sons, London, 1918. Reprinted 1982, Eland Books, London.

Kluwer Academic Publishers. "User manual for `kluwer.cls`, version 1997/05/30. Instructions for authors". 1997. This is file `usrman.tex`, included with the LaTeX class archived in the CTAN archives in the directory `macros/latex/contrib/supported/kluwer`.

Mittelbach, F. and C. Rowley. "The LaTeX3 Project". Public, official document, LaTeX3 Project, London, 1993. File `l3d001.tex` archived in CTAN.

Mittelbach, F., C. Rowley, and M. Downes. "Volunteer work for the LaTeX3 project (Version 6.2a)". Public document, LaTeX3 Project, 1993. File `vol-task.tex` archived in CTAN.

Rhead, D. "Towards BibTeX style-files that implement principal standards". *TeXline* (10), 2–8, 1990.

Rhead, D. "How might LaTeX 3.0 deal with citations and reference lists?". *TeXline* (13), 13–20, 1991.

Rhead, D. "The "operational requirement" for support of bibliographic references by LaTeX 3". *TUGboat* **14**(4), 425–433, 1993.

Wonneberger, R. and F. Mittelbach. "BibTeX Reconsidered". *TUGboat* **12**(1), 111–124, 1991. TeX 90 Conference Proceedings.

◇ Pedro J. Aphalo
Faculty of Forestry
University of Joensuu
P.O. Box 111
FIN-80101 Joensuu
Finland
`pedro.aphalo@joensuu.fi`
`http://cc.joensuu.fi/~aphalo/`