

Managing biological literature citations using a BIB_TE_X-based journal name database¹²³

Xuejun Dong¹, Geoffrey W. Patton², Anne C. Nyren¹,
Bob D. Patton¹, and Paul E. Nyren¹

¹North Dakota State University,
Central Grasslands Research Extension Center,
4824 48th AVE SE,
Streeter, ND 58483 USA.

²U.S. Environmental Protection Agency,
1200 Pennsylvania Ave.,
Mailcode 8601P, Washington, DC 20460 USA⁴

Abstract

Today, many research articles in the biological sciences have multiple authors and an increasingly long list of cited papers covering a large number of journal titles. To ensure accuracy in cross-referencing and citation and to meet the specific requirements of different publishers, the average author is spending more and more time organizing and managing the literature database. In spite of these efforts, citation inconsistencies within a manuscript may divert an author's attention from the scientific aspects of the article. To promote efficient scientific writing and reviewing, we have customized, in a BIB_TE_X format, a database for 4387 mostly biological journals published worldwide mainly in the English language. This database is readily usable to meet the requirements for journal name format by many of today's leading scientific publishers. We provide a simple L^AT_EX template so as to facilitate an automated and accurate handling of the expanding biological literature. Also included in this paper is an eco-physiological literature database containing bibliographic information of 835 papers.

Keywords: Biological journals database, name abbreviations, L^AT_EX, automated management of journal citations.

¹ First draft: January 23, 2009 / Last modified: September 22, 2009

²For correspondence: X. Dong (xuejun.dong@ndsu.edu) or G. Patton (gwpatton1@yahoo.com)

³This is a detailed instruction for using the journal name database. An overview article highlighting the main features of this project can be found in the 2009 summer issue of **Issues in Science & Technology Librarianship** (Number 58) at <http://www.istl.org/09-summer/>

⁴Required disclaimer: The views expressed here are those of the author and do not represent the position or endorsement of the U.S. EPA.

1 Introduction

Powered with today's new equipment, advanced analytical tools, as well as high speed information exchange across the globe, scientists are tackling ever more challenging problems. Meanwhile, most published papers are characterized by wider collaborations, multiple authorships, and a long reference list. With the varied requirements of citation formats by different publishers, accurate and consistent literature citation is forcing scientists to spend more and more time organizing and managing the increasingly larger literature database. To cope with these problems, various computer packages, such as EndNote (Combs, 1998; Siegler and Simboli, 2002), ProCite (Rohmann, 1999; Wachtel, 1987), Reference Manager (Neal, 1993; Shmaefsky, 2000; Wachtel, 1987), BIB_TE_X (Harrison and Munson, 1989; Lamport, 1994), etc., have been widely used to aid the management of scientific literature.

Many popular reference management packages serve three main functions: online search tool, reference database, and bibliography maker (Paling, 1998). The commercial packages such as EndNote, ProCite, and Reference Manager are capable of searching literature from huge collections of some national databases, in addition to their powerful literature analysis and management capabilities (Rohmann, 1999; Shmaefsky, 2000). However, all of these packages have to be used with a word processing package in order to obtain desirable output (Neal, 1993; Paling, 1998). BIB_TE_X, as fully integrated with the L^AT_EX document processing system, can be used to generate a consistent local bibliography database and a flexible bibliography output (Patashnik, 1988; Lamport, 1994; Merkel, 2002). Advanced BIB_TE_X analysis and searching capabilities can be achieved through the use of the literature management system JabRef (Lin, 2006), which can search important research databases such as Medline (<http://www.nlm.nih.gov>) and CiteSeer (<http://CiteSeer.ist.psu.edu/>).

In this “informatics” age, with new biological journals being launched every month, the use of any literature citation management system relies on a good database of journal abbreviations, because different journals have different requirements for journal name format, such as full name, abbreviations, or abbreviations with stops. With an easy-to-use mechanism for efficient handling and editing of the existing journal name database, the time biologists spend in literature management may be significantly reduced, allowing more time for the scientific aspects of research and publication. The latest version of EndNote has a very good capability in this respect by making full use of large journal name databases. In addition, the database compiled by Patton (2006) can be used by EndNote users to make their Term Lists (Paling, 1998; Siegler and Simboli, 2002) and thus facilitate efficient handling of the journal name abbreviations. Although the procedures are available for generating BIB_TE_X output from within the EndNote system for uses by L^AT_EX (Macleod, 2005), the fact that additional steps are needed would discount the practical value of this method of literature management. In producing a document with literature citations in L^AT_EX, at least four computer mouse-clicks are usually required in order to produce the final PDF document (Dodson, 2001). The necessity of additional steps in using the EndNote to BIB_TE_X linkage (Macleod, 2005) would affect the efficiency of L^AT_EX-based document processing.

L^AT_EX differs from the above-mentioned commercial packages (such as EndNote) in that it is free and has widely available web-based tutorials. Patashnik (1988) proposed an idea of automating the use of journal and literature databases so that repeated manual modifications of the literature are reduced. However, to our best knowledge, Patashnik’s idea has not been turned into a full-scale application for the management of citations of rapidly expanding biological journals using L^AT_EX. The objective of this paper is to translate the biological journals database of Patton (2006) into a BIB_TE_X format in order to facilitate L^AT_EX-based document writing in biology. The basics of how to use L^AT_EX are explained by authors such as Greenberg (2005), Williams (2001), and Dodson (2001), and more detailed information is available from Lamport (1994). This paper assumes that the reader has a preliminary experience with L^AT_EX and BIB_TE_X.

2 A journal name database and its usage

Assume that L^AT_EX (for example, Mik_TE_X 2.5 bundled with BIB_TE_X) and WinEdt (a word processor by default configured with Mik_TE_X) are installed in a personal computer running on the Windows XP operating system. The handling of the literature listings and citations involves the use of an appropriate bibliography style file (with .bst extension) stored under the directory C:\Program Files\Mik_TE_X 2.5\BIB_TE_X\bst and both a journal name database file and a literature database file (both with .bib extension) stored under the directory C:\Program Files\Mik_TE_X 2.5\BIB_TE_X\bib. A typical L^AT_EX source file for producing a technical document can be something like: List 1:

List 1:

```

\documentclass[a4paper,12pt]{article}           %Line 1
\usepackage{natbib}                            %Line 2
\begin{document}                               %Line 3
In grasslands, surface soil dryness can greatly influence %Line 4
respiration~\citep{Redmann:Aa1978a, Redmann:Aa1978b}. %Line 5
\bibliographystyle{kluwer}                     %Line 6
\bibliography{Jshort,MyRef}                   %Line 7
\end{document}                                %Line 8

```

In List 1, line 1 tells L^AT_EX to use “article” document class with text size of 12 pt on A4 size paper; line 2 evokes the “natbib” package, which is a L^AT_EX macro able to handle both the author-year and the numerical styles of literature citations. Lines 3 and 8 mark the beginning and end of the document. In line 5, ~\citep is a L^AT_EX command to cite the two papers (enclosed within curly braces with unique nicknames) parenthetically. These two cited papers are stored in the literature database file MyRef.bib (which is stored under C:\Program Files Mik_TE_X 2.5\BIB_TE_X\bib; see [Supplementary Material 1](#); also note that in line 7 the file extension .bib is omitted). Line 6 instructs BIB_TE_X to use a popular author-year bibliography style of kluwer.bst (which is part of a standard Mik_TE_X installation and stored under C:\Program Files\Mik_TE_X 2.5\BIB_TE_X\bst). For

an author, line 7 is especially important because it instructs BIB_TE_X to use an abbreviated version of the journal name database (`Jshort.bib`, see [Supplementary Material 2](#)) and a literature database file (`MyRef.bib`). One important thing to note here is that the journal name database file should always be followed by the literature database file. The L^AT_EX output is shown in List 2.

List 2:

In grasslands, surface soil dryness can greatly influence respiration (Redmann, 1978a, 1978b).

References

Redmann, R. E.: 1978a, Seasonal dynamics of carbon dioxide exchange in a mixed grassland ecosystem, *Can J Bot* 56, 1999–2005.

Redmann, R. E.: 1978b, Soil respiration in a mixed grassland ecosystem, *Can J Soil Sci* 58, 119–124.

In List 2, we find that the two cited papers are listed at the end of the document automatically with only one instruction: `\bibliography{Jshort,MyRef}` from List 1. Also note that the journal names are abbreviated. When necessary, the author can also use two other versions of the journal names: the abbreviated-dotted version (`Jshort_dot.bib` as listed in [Supplementary Material 3](#)) or the full name version (`JLong.bib` as listed in [Supplementary Material 4](#)). In the latter two cases, the listed journal names will be changed accordingly with the rest of the citation remaining the same. Now let's extract a portion of the bibliography file `Jshort.bib` to see what it looks like in the database file.

List 3:

```
...
...
@string{CanJBot = {Can J Bot}}
...
@string{CanJSoilSci = {Can J Soil Sci}}
...
```

Actually, this `Jshort.bib` file is so simple that it just consists of many entries with a common format: `@string{A={B}}`, where A is a unique nickname for the particular journal, and B is for the abbreviated name for this journal. In fact, a user can freely edit the abbreviation if they know that what is listed in this journal name database for this particular journal is not authoritative, or is not acceptable with a particular publisher. At the same time, one may add a new journal entry using the same format as shown above and save the file. Actually, in the building of these three journal name database files, we have added nearly 100 journal entries that are not covered in the original listing of biological journals in Patton (2006). Some of these new entries are for non-biological journals, multidisciplinary journals, or the most recently launched journals.

Next, let's show how the database `MyRef.bib` looks in List 4.

List 4:

```

...
@ARTICLE{Redmann:Aa1978a,
  AUTHOR      = {R. E. Redmann},
  TITLE       = {Seasonal dynamics of carbon dioxide exchange in a mixed
    grassland ecosystem},
  YEAR        = {1978},
  JOURNAL      = CanJBot,
  volume      = {56},
  number      = {},
  pages       = {1999--2005},
  month       = {},
}

@ARTICLE{Redmann:Aa1978b,
  AUTHOR      = {R. E. Redmann},
  TITLE       = {Soil respiration in a mixed grassland ecosystem},
  YEAR        = {1978},
  JOURNAL      = CanJSoilSci,
  volume      = {58},
  number      = {},
  pages       = {119--124},
  month       = {},
}
...

```

List 4 lists only two articles that are cited in this sample document fragment of List 2, however, one can have hundreds of articles and other different reference materials listed in this same file (see [Supplementary Material 5](#)). After “LaTeXing” the source file, only those papers (or other materials) that are actually cited in the main text will be listed under the reference section of the document. While most of the entries are self-explanatory, two points in particular are important here. First, each article has a unique “key” that is to be referred to in the main text. These keys help L^AT_EX to find the correct papers from the literature database. Here the keys for the two articles are: “Redmann:Aa1978a” and “Redmann:Aa1978b”, respectively. A downside of this approach to literature citation “automation” is that users must first spend time building their own literature database manually, using a key for each of the literature entries. However, once the literature database is built, one can use it many times without major editing. This is where using this L^AT_EX/BIBT_EX literature management method benefits authors in the long run. This method also has some advantages over the usual situation in which one has to edit the format of the literature listing manually every time one switches to a different publisher. This method also solves the problem encountered in revisions, where new articles are now cited in the main text, but are not in the listing, or citations have been deleted from the

text but not from the reference listing.

The second important entry in List 4 are the journals' nicknames, namely, "Journal = CanJBot" and "Journal = CanJSoilSci". Those unique nicknames (unique for each particular journal) are where the literature database and the journal name database are linked. The statement of "Journal = CanJSoilSci" provides all the information L^AT_EX/BIBT_EX needs in order to find the real name of the journal, whether it is a full name, abbreviated, or abbreviated-dotted, in the journal name database. The reason the literature database should be preceded by the journal name database in List 1 line 7 (namely, first `Jshort`, then `MyRef`) is that L^AT_EX first needs a definition of which journal each of the journal nicknames represents before it can find the complete information for each of the cited papers as listed in List 4. Most of the nicknames are simply the concatenated version of the abbreviated journal names as used by Patton (2006). However, when the journal names are very long, the nicknames are shortened, but still uniquely defined.

For authors who would like to have their documents in a form readable in MS Word or WordPerfect, the L^AT_EX source (with `.tex` extension) can be translated into a `.rtf` format through the use of some translation software such as `l2r`, provided that the document does not contain a lot of graphics and other illustrations.

The bibliography style file used in List 1 is "kluwer"⁵. The style file determines the appearance of all the bibliography items in an article (including journal names as discussed in this paper) in both the in-text citation and the bibliography listing. As both the literature citation mode (author-year or numerical) and listing format may differ with different journals, it is necessary have ready the appropriate style file when one writes for a particular publisher. Fortunately, the style files can be obtained or generated easily. See [Supplementary Material 6](#) for websites where one can find the bibliography style files for selected journals, as well as for detailed instructions of how to make a new bibliography style file for one's favorite journal.

3 Discussion

Currently, EndNote is well known for efficiently handling automatic literature citations, including journal name abbreviations. However, EndNote is not shareware. The L^AT_EX compatible journal name databases, as proposed in this study, provide an alternative way of handling journal name citations (functionally similar to the Term List in EndNote) for biologists. Using either method, one can handle a large amount of literature with ease. Here, we, as well as our colleagues, must be aware of the possible misuse of literature citations. Blindly citing those easily obtained references (from today's powerful reference management packages) without careful study and understanding may pose some dangers in the credibility of the research. One good thing is that the proposed literature management method in this paper requires that researchers periodically manually enter and maintain the databases. Through actually entering the data, one can become familiar

⁵Visit <http://jo.irisson.free.fr/bstdatabase/faq.html> for explanation of how to use different types of style files.

with the authors' names and the contents of the papers.

The concept for this database originated with the `@string {}` command of BIB_TE_X (Patashnik, 1988; Lamport, 1994, pages 158-159). Our purpose for publishing the three forms of the database is to share this information world-wide with biologists who use L^AT_EX and to promote more consistent, error-free, and efficient scientific writing. In terms of the versatility of the BIB_TE_X system (Merkel, 2002), we are confident that most of the output requirements by today's publishers can be met through the use of this database with the combination of L^AT_EX and BIB_TE_X packages. The journal name database mentioned in this paper will grow to incorporate new journals. As stated previously, it is extremely easy to edit the journal name database; however, it is a good habit to edit all three versions simultaneously. Some abbreviations may not be trusted as stated in the original database maintained by Patton (2006). Some international journals are not abbreviated. For example, Japanese journals such as *Idengaku Zasshi* are not abbreviated. Users of these databases should be aware of these potential problems. The literature database `MyRef.bib` can also be easily updated following rules as detailed in [Supplementary Material 7](#).

In conclusion, we believe that the L^AT_EX-compatible journal name database with 4387 journals in its current three forms can significantly aid biologists who use L^AT_EX to manage today's increasingly larger literature database more efficiently, and thus enhance professional writing and publishing.

Acknowledgments

X.D. thanks Dr. C. T. J. Dodson at University of Manchester, UK, for advice in using L^AT_EX. Also, we thank J. Patton, Streeter ND, for suggestions. The two papers authored by R. E. Redmann were used only to illustrate the nature of the literature and journal name databases discussed in this paper.

References

1. Combs, Jr., J. (1998). Endnote 3: Reference database management, bibliography generation, and Z39.50 search and retrieval software in one package from Niles Software, Inc. *Library Software Review*, 17(2), 149-156.
2. Dodson, C.T.J. (2001). PDF LaTeX tutorial for scientific word processing. Retrieved October 6, 2008, from <http://www.maths.manchester.ac.uk/~kd/latexut/pdfbyex.htm>.
3. Greenberg, H. J. (2005). A simplified introduction to LaTeX. Retrieved October 6, 2008, from <http://www.ctan.org/tex-archive/info/simplified-latex/>.
4. Harrison, M. A. and Munson, E. V. (1989). On integrated bibliography processing. *Electronic Publishing*, 2(4), 193-209.
5. Lamport, L. (1994). *LaTeX: A Document Preparation System*. 2nd ed. Reading MA: Addison-Wesley Publishing Company.

6. Lin, R. (2006). An introduction to BIBTeX and JabRef. Retrieved October 6, 2008 from www.biostat.jhsph.edu/bit/compintro/jabref-20060224.pdf.
7. Macleod, R. (2008). Using EndNote with LaTeX/BIBTeX. Retrieved October 6, 2008 from <http://www.cvrtri.utah.edu/macleod/latex/endnote.html>.
8. Merkel, S. (2002). Reference sheet for Natbib usage. Retrieved October 6, 2008 from <http://merkel.zoneo.net/Latex/natbib.php>.
9. Neal, P. R. (1993). Personal bibliographic softwares- a comparative review. *BioScience*, 43(1), 44-51.
10. Paling, S. (1998). EndNote. *The Journal of Academic Librarianship*, 24, 430-431.
11. Patashnik, O. (1988). BIBTEXing. Retrieved January 9, 2009 from <http://amath.colorado.edu/documentation/LaTeX/reference/faq/BIBTeX.pdf>.
12. Patton, G. (2006). Biological journals and abbreviations. Retrieved October 6, 2008 from <http://home.ncifcrf.gov/research/bja/>.
13. Rohmann, G. (1999). ProCite and EndNote: bibliographic management software. *Information Technology, Learning and Performance Journal*, 17(1), 51-53.
14. Shmaefsky, B. R. (2000). e-reference tool. *Science*, 288(5473), 1981.
15. Siegler, S. and Simboli, B. (2002). EndNote at Lehigh. *Issues in Science and Technology Librarianship*, 34, Spring, 2002.
16. Wachtel, R. E. (1987). Personal bibliographic database. *Science*, 235(4792), 1093-1095.
17. Williams, J. (2001). How to use BIBTeX. Retrieved October 6, 2008 from <http://www.hep.manchester.ac.uk/u/jenny/jcwdocs/latex/BIBTeXbasics.html>