
AUTHIDX: An Author/Editor Indexing Package

Nelson H. F. Beebe

Contents

1	Introduction	1001
2	The book bibliography problem	1001
3	Preparing to solve the problem	1002
4	The user interface	1003
5	Running the programs	1004
6	Portability considerations	1004
7	Availability	1005
8	Conclusions	1005

Abstract

This document describes AUTHIDX, a package for automatic generation of author/editor indexes, and citation page references, with AmSTEX, L^ATEX, e-plain TEX, and plain TEX.

1 Introduction

Sadly, the bibliographic data provided in many technical publications is sparse, inaccurate, and hard to use as a resource in itself. Some of this can be attributed to the pain that authors and typists suffer in preparing bibliographic data for publication, and some is also certainly due to human egos: most authors would rather write about what *they* have done, than about what *others* have produced.

The availability of large bibliographic databases, such as those in the TEX User Group-bibliography archive at the author's FTP site, `ftp://ftp.math.utah.edu/pub/tex/bib`, and the huge Computer Science archive at `ftp://ftp.ira.uka.de/pub/bibliography` in Karlsruhe, Germany, offers interesting possibilities for substantially improving the traditional citation bibliographies in technical documents.

First of all, when literate references are readily available in marked-up form that can be converted automatically to the style required by the publisher or organization, much of the tedium of citing references in documents is eliminated. For example, with my favorite bibliographic system, BIBTEX [6, pp. 155–164], all that I require in the running text of this article for the citation in this sentence is the text `\cite[pp.~155--164]{Lamport:1994:LDP}`, plus two simple `\bibliography` and `\bibliographystyle` declarations near

the end of this document, and an invocation of the BIBTEX program between L^ATEX runs.

Second, when a fast and powerful search engine is used to access a large bibliographic archive, such as the BIBSEARCH [12] system at my site (a simple interface to the MG database system [12]), authors can readily locate many references that they might otherwise have overlooked, and they are therefore much more likely to provide more extensive bibliographies in their documents.

Third, when bibliographic data access is easy, and the collection is large, authors are likely to turn up many interesting references that they had not previously encountered, thereby enriching their own ideas, and promoting intellectual progress by building on the work of others.

2 The book bibliography problem

During 1995–1996, at the University of Utah, with U.S. National Science Foundation support, we hosted a *Special Year in Mathematical Biology*. This brought together eighty-six outside researchers in the field (graduate students, post-doctoral fellows, and visiting faculty) from six countries, for visits ranging from a few days to several months, at our Mathematics and Biology departments. At the conclusion of this activity, we set about to prepare a book [8] to document the research carried out during the Special Year.

I was the consulting technical editor on the book, and its production took almost six months of extremely intense work on the part of the contributing authors, editors, and technical support people. Work began in June, and the day before Thanksgiving in late November, I uploaded the final completed POSTSCRIPT files to the publisher's printer shop. Just five weeks later, the book was available in bookstores.

The book has 573 literature references, every one of which was checked against library catalogs and journal databases. Though tedious, this checking was extremely worthwhile, since it uncovered scores of errors and omissions which were repaired before publication.

Since the book documents forefront research in the relatively young field of mathematical biology, most of its references were to work of the last two decades, and researchers and students who use the book will very likely want to read many of those publications.

With conventional bibliographies, the bibliography is of little use in itself, since one has no

easy way to find out where in the text a particular publication was cited, and why, unless, of course, one has an author who is as conscientious as Donald E. Knuth is about indexing cited authors.

I therefore decided that the bibliography in our book would be different: every bibliographic entry would contain a list of page numbers in the book where the publication was cited, and every author of every publication would be indexed. This latter index is important, because good scientific research often requires reading all of the publications of a prominent author in a particular area. In addition, book entries would contain International Standard Book Number (ISBN) values, providing unique identifiers that can be helpful in locating books in libraries, or ordering them through bookstores.

The problem with this objective is that with hundreds of citations, and over a thousand authors in the bibliography, accurate manual preparation of the cross-references and index is humanly impossible. Computers, however, get neither bored, nor tired, so they can accomplish the job, given suitable software. At the start of this project, such software did not exist.

3 Preparing to solve the problem

To understand how I solved the problem, it is useful to review how citations in the running text eventually produce a reference to a bibliographic entry. While the description here assumes $(\text{L}^{\text{A}})\text{T}_{\text{E}}\text{X}$ and $\text{BIB}\text{T}_{\text{E}}\text{X}$, any document formatting and bibliographic systems will have to do something similar.

First, the presence of a citation in the running text, such as `\cite{Lampport:1994:LDP}`, results in the entry on an auxiliary file of a line containing the citation label, such as

```
\citation{Lampport:1994:LDP}
```

Since the expansion of that citation in the running text may not be known until the bibliographic data has been retrieved, the amount of space that it requires is initially uncertain, and consequently, the typesetting system must make at least two passes over the document before line- and page-breaking decisions can be finalized.

Second, when the bibliographic software reads the auxiliary file, it must locate the requested citations in one or more bibliographic database files, and then format them for inclusion in the bibliography according to the speci-

fied style. For $\text{BIB}\text{T}_{\text{E}}\text{X}$, this might mean that the database entry

```
@Book{Lampport:1994:LDP,
  author = "Leslie Lampport",
  title = "{\LaTeX}: {A} Document
    Preparation System: User's
    Guide and Reference
    Manual",
  publisher = pub-AW,
  address = pub-AW:adr,
  edition = "Second",
  pages = "xvi + 272",
  year = "1994",
  ISBN = "0-201-52983-1",
  LCCN = "Z253.4.L38L35 1994",
  acknowledgement = ack-nhfb,
  bibdate = "Wed Aug 10 09:55:59 1994",
}
```

is retrieved and reformatted in an output bibliography file as

```
\bibitem{Lampport:LDP94}
Leslie Lampport.
\newblock {\em {\LaTeX}: {A} Document
  Preparation System: User's Guide and
  Reference Manual}.
\newblock Ad{\-d}i{\-s}on-Wes{\-l}ey,
  Reading, MA, USA, second edition, 1994.
\newblock \showISBN{0-201-52983-1}.
\newblock xvi + 272 pp.
\newblock \showLCCN{Z253.4.L38L35 1994}.
```

which is finally typeset by $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ in the form shown in the bibliography at the end of this document.

Several modifications of the citation process are required for our book indexing project:

- each citation recorded in the auxiliary file must also include its page number;
- the page numbers must not interfere with the bibliographic data extraction;
- the page number lists for each cited publication must be collected, sorted, and merged to remove duplicates, and then included in the corresponding bibliography entry;
- the authors and/or editors of each cited entry must be extracted and added to an auxiliary index file;
- the author/editor index file must be sorted, merged, and formatted for inclusion in an index.

For $(\text{L}^{\text{A}})\text{T}_{\text{E}}\text{X}$, access to page numbers is complicated by the asynchronous nature of $\text{T}_{\text{E}}\text{X}$'s

page breaking algorithm: the output routine can be called at any time, either implicitly or explicitly, and it may choose to delay some of the accumulated potential output material until the next page. Thus, T_EX does not have a single variable that reliably records the ‘current page number’. Fortunately, T_EX’s author saw the need for this, and provided a `\write` command whose argument is not evaluated until the output routine has made its page-breaking decision and the text is sent to the output file [5, pp. 215, 217]. Thus, the same mechanism that is used for writing page numbers in table-of-contents, list-of-figures, list-of-tables, and index files can be used for recording the citation page numbers.

4 The user interface

Once the package is properly installed in the local T_EX system, to use it, you first need to load it near the start of your document file:

```
LATEX 2e \usepackage{authidx}
LATEX 2.09 \documentstyle[authidx]{...}
AmSTEX
e-plain TEX \input authidx.sty
plain TEX
```

After loading the package, you need to request that data for the author/editor index be written on this run:

```
\makeauthoreditorindex
```

This command might be commented out during development to save a bit of time.

Although it is rarely necessary, or desirable, to do so, it *is* possible to write extra entries to the indexing file without citing them in the running text:

```
\authoreditorindex{BIBTEX-label}
```

This command is used internally by the modified `\cite` command to record the citation label in the indexing file.

The `\bibliography` command has been modified to create `\bibdata` entries in both the normal auxiliary file, and the author/editor indexing file.

The index is printed inside a `theauthor editorindex` environment, which is generated automatically when the user issues the `\print authoreditorindex` command.

L^AT_EX users will recognize the similarity of these to `makeidx.sty`, `makeindex`, `\index`, the `index`, and `\printindex`.

For additional user control over the author/editor index, the user may provide alternate defi-

nitions of the following macros, with `\renewcommand` (in L^AT_EX) or `\def` (in other T_EX formats):

```
\def \AENAMEfontfirst #1{{\sc #1}}%
\def \AENAMEfontrest #1{{\rm #1}}%
\def \AEPAGEfontfirst #1{{\bf #1}}%
\def \AEPAGEfontrest #1{{\rm #1}}%
```

As shown, each takes a single argument, and allows typesetting the first author or page number differently from the remaining ones.

The reason for offering alternate typesetting of first and remaining authors is the tradition, in many fields, of referring to a multi-author work by the first author alone, or as a colleague once remarked to me:

If you see a reference to a paper by *Smith et al.*, then you know that Smith got the credit, but Al did the work!

Personally, I deplore this practice, and try very hard to avoid use of *et al.*, and the corresponding *and others* in BIB_TE_X files.

The default definitions of these font macros do not specify a font *size*, so in L^AT_EX, they will automatically inherit the prevailing font size. This allows you to use L^AT_EX input like this:

```
\begin{small}
  \printauthoreditorindex
  \printindex
\end{small}
```

to typeset the indexes in smaller type, as is commonly done.

The `theauthoreditorindex` environment is modeled very closely on L^AT_EX’s `theindex` environment. It switches to two-column mode, which starts a new page, then begins an unnumbered section or chapter with the title *Author/Editor Index*, sets the even- and odd-page running heads to *AUTHOR/EDITOR INDEX*, sets the page style to *plain* for the first page of the index, and defines `\item` to be the same as the internal L^AT_EX `\@idxitem` macro.

While this default behavior will often be adequate, many documents may require something different, such as for the three-column indexing used in our book [8].

Rather than trying to parametrize bits and pieces of this environment with user-customizable macros, it seems to make more sense to simply advise the user to redefine the `authoreditorindex` completely when needed, with `\renewenvironment` in L^AT_EX, or `\def` in other T_EX formats.

For example, for a journal article such as this one, starting a new page and changing page headers is unnecessary, so this simple redefinition suffices:

```
\renewenvironment
  {theauthoreditorindex}%
  {%
  \section*{Author/editor index}%
  \setlength{\parindent}{0pt}%
  \setlength{\parskip}{0pt plus 0.3pt}%
  \renewcommand{\item}%
  {%
  \par
  \setlength{\hangindent}{40pt}%
  }%
  }%
  {%
  \relax
  }
```

The corresponding redefinition of `theindex` differs from this only in the section heading.

In a large bibliography, it may be useful to include author/editor page references to the bibliography itself:

```
\bibliographyindextrue
```

You can turn off this feature with

```
\bibliographyindexfalse
```

The default setting is *true*.

Finally, the appearance of page-number cross-references in the bibliography can be controlled by redefining this macro:

```
\def \bibindex #1{\{\{\bf #1\}\}}
```

The argument to `\bibindex` is a comma-separated list of page numbers, in ascending order, with consecutive page numbers collapsed to page ranges.

In summary, here is what a typical input file for L^AT_EX 2_ε might look like:

```
\documentclass{book}
\usepackage{makeidx}
\usepackage{authidx}
\makeauthoreditorindex
\makeindex
\begin{document}
...
\printauthoreditorindex
\printindex
\end{document}
```

The AUTHIDX package must be included *after* the MAKEIDX package, because the latter fails to check for a prior definition of one macro (*\see*)

that AUTHIDX also must define in case it is used without the other package.

5 Running the programs

Besides T_EX, B_IB_TE_X, and M_AK_EI_ND_EX, there is one additional program needed to complete the author/editor indexing job: AUTHIDX. Its task is to combine the data in the author/editor index file (*.aei*) created in the L^AT_EX runs with data extracted from the B_IB_TE_X database files to produce a modified index file (*.ida*) that is suitable for input to M_AK_EI_ND_EX.

Here is the sequence of job steps that needs to be run on a UNIX system:

```
latex mybook.ltx
bibtex mybook
latex mybook.ltx
bibtex mybook
authidx mybook.aei > mybook.ida
mv mybook.bbl mybook.tmp
authidx mybook.aei mybook.tmp \
  > mybook.bbl
makeindex -o mybook.ina mybook.ida
makeindex -o mybook.ind mybook.idx
latex mybook.ltx
```

The steps on most other operating systems would be very similar to these.

The second B_IB_TE_X run is needed to handle bibliographic data that contains `\cite` commands for cross-referencing; it can be omitted if this is not the case.

The second AUTHIDX run modifies the *.bbl* file to include page-number cross references.

Since the above steps will likely need to be done several times during document development, these commands should be encapsulated in a single command file, or better, a *Makefile*, since the M_AK_E program will check stated file dependencies to avoid running unnecessary steps.

6 Portability considerations

The AUTHIDX program is approximately 1000 lines of code in the AWK programming language [1, 2, 3, 11].

AWK is a simple, and clean, text processing language that is ideally suited to this task. I have written over a hundred thousand lines of code in that language, and found that equivalent programs in other languages are invariably significantly longer, and much harder to get right.

In addition, AWK is an IEEE POSIX [4, 9, 10, 7, 13] standard, and since POSIX conformance is

available (though not necessarily installed by default) for virtually all commercial operating systems, you should expect your system vendor to provide it.

To find out whether your system has POSIX compliance, visit <http://nemo.ncsl.nist.gov/posix/151-1reg> and <http://nemo.ncsl.nist.gov/posix/151-2reg> on the World-Wide Web, or send electronic mail to posix@nist.gov with this text:

```
send help
send 151-1reg
send 151-2reg
```

The most obvious omission among the many systems listed is the Apple Macintosh operating system; only the A/UX and MachTen operating systems on the Macintosh hardware platform have POSIX compliance. By searching on the Web, I located a reference to a port of AWK to the Macintosh Programmer's Workbench (MPW), but at the time of writing this, the link was unreachable.

There are also independent commercial implementations of the language for the IBM PC platform, and at least four freely-distributable implementations: BAWK, GAWK, MAWK, and NAWK, the last of which is from the original developers at AT&T [1], with continued improvements. The AWK development groups are in close contact, so the implementations retain fidelity to the original language definition [1].

The book about the Free Software Foundation's GNU implementation, GAWK [11], is also available in T_EXinfo form for use online in the GNU EMACS INFO system, or for separate typesetting and printing, but I use my bound copy a good deal too.

More details on sources of AWK implementations are included in the documentation in the AUTHIDX distribution.

7 Availability

The master archive location for the AUTHIDX package distribution is <ftp://ftp.math.utah.edu/pub/tex/authidx>. By the time you read this, it should also be mirrored to the Comprehensive T_EX Archive Network (CTAN); try [finger ctan@tug.org](mailto:finger.ctan@tug.org) to find a CTAN mirror site near you.

8 Conclusions

The bibliography and indexes which follow this section demonstrate, for this short document,

what the AUTHIDX package can do. For a much bigger example, see our mathematical biology book [8]. That book has separate bibliographies in each chapter, demonstrating that AUTHIDX can be used even for documents with multiple bibliographies.

The beauty of the package is that it makes much better bibliographies and indexes available for very little effort on the part of the author: three extra lines in the input file, and four extra commands to be executed during the typesetting.

As the utility of its features becomes more widely known and appreciated, I hope that this package will receive wide use in the T_EX community, and that publishers will adopt it for their routine technical book production.

Provided that B_IB_TE_X bibliographic databases are used, AUTHIDX and a very simple MAKEINDEX style file that produces SGML instead of T_EX would make the package available to those many publishers who use SGML markup for documents.

With an alternate bibliographic database format, the AUTHIDX program would have to be rewritten, but I expect that major portions of the existing AWK code could be reused, so that the revision might take only a day or two of work.

References

- [1] Alfred V. Aho, Brian W. Kernighan, and Peter J. Weinberger. *The AWK Programming Language*. Addison-Wesley, Reading, MA, USA, 1988. ISBN 0-201-07981-X. x + 210 pp. LCCN QA76.73.A95 A35 1988. {**1004**, **1005**}
- [2] Dale Dougherty. *sed and awk*. O'Reilly & Associates, Inc., 981 Chestnut Street, Newton, MA 02164, USA, November 1990. ISBN 0-937175-59-5. 414 pp. LCCN QA76.76.U84 D69 1991. US\$27.95. {**1004**}
- [3] Dale Dougherty and Arnold Robbins. *sed and awk*. O'Reilly & Associates, Inc., 981 Chestnut Street, Newton, MA 02164, USA, second edition, February 1997. ISBN 1-56592-225-5. 420 pp. US\$29.95. {**1004**}
- [4] Bill Gallmeister. *POSIX.4: Programming for the Real World*. O'Reilly & Associates, Inc., 981 Chestnut Street, Newton, MA 02164, USA, January 1995. ISBN 1-56592-074-0. xviii + 548 pp. LCCN QA76.76.O63 G34 1995. US\$29.95. {**1004**}

- [5] Donald E. Knuth. *The T_EXbook*. Addison-Wesley, Reading, MA, USA, 1984. ISBN 0-201-13448-9. ix + 483 pp. LCCN Z253.4.T47 K58 1984. {**1002**}
- [6] Leslie Lamport. *L^AT_EX: A Document Preparation System: User's Guide and Reference Manual*. Addison-Wesley, Reading, MA, USA, second edition, 1994. ISBN 0-201-52983-1. xvi + 272 pp. LCCN Z253.4.L38L35 1994. {**1001**}
- [7] Donald A. Lewine. *POSIX programmer's guide: writing portable UNIX programs with the POSIX.1 standard*. O'Reilly & Associates, Inc., 981 Chestnut Street, Newton, MA 02164, USA, 1991. ISBN 0-937175-73-0. xxvii + 607 pp. LCCN QA76.76.O63 L487 1991b. US\$34.95. March 1994 printing with corrections, updates, and December 1991 Appendix G. {**1004**}
- [8] H. G. Othmer, F. R. Adler, M. A. Lewis, and J. C. Dallon, editors. *Case Studies in Mathematical Modeling—Ecology, Physiology, and Cell Biology*. Prentice-Hall, Englewood Cliffs, NJ 07632, USA, 1997. ISBN 0-13-574039-8. viii + 411 pp. LCCN QH541.15.M3C37 1997. The book has its own World-Wide Web page at <http://www.math.utah.edu/books/csmmepcb/index.html>. {**1001**, **1003**, **1005**}
- [9] *Portable Operating System Interface (POSIX)—Part 2: Shell and Utilities (Volume 1)*. Information technology—Portable Operating System Interface (POSIX). IEEE Computer Society, 345 E. 47th St, New York, NY 10017, USA, 1993. ISBN 1-55937-255-9. 1–750 pp. {**1004**}
- [10] *Portable Operating System Interface (POSIX)—Part 2: Shell and Utilities (Volume 2)*. Information technology—Portable Operating System Interface (POSIX). IEEE Computer Society, 345 E. 47th St, New York, NY 10017, USA, 1993. ISBN 1-55937-255-9. 751–1195 pp. {**1004**}
- [11] Arnold Robbins. *Effective AWK Programming*. Specialized Systems Consultants, P.O. Box 55549, Seattle, WA 98155, 1996. ISBN 0-916151-88-3. viii + 322 pp. LCCN QA76.73.A95 R63 1996. US\$27.00. URL <http://www.ssc.com/ssc/eap/>. {**1004**, **1005**}
- [12] Ian H. Witten, Alistair Moffat, and Timothy C. Bell. *Managing Gigabytes: Compressing and Indexing Documents and Images*. Van Nostrand Reinhold, New York, NY, USA, 1994. ISBN 0-442-01863-0. xiv + 429 pp. LCCN TA1637 .W58 1994. US\$54.95. The software for full-text indexing described in this book, and errata for the book, are available from <ftp://munnari.oz.au/pub/mg>. The BIBSEARCH package and its required extended version of MG are available at <ftp://ftp.math.utah.edu/pub/mg/mg-1.3x>. {**1001**}
- [13] Fred Zlotnick. *The POSIX.1 standard: a programmer's guide*. Benjamin/Cummings Pub. Co., Redwood City, CA, USA, 1991. ISBN 0-8053-9605-5. xi + 379 pp. LCCN QA76.76.063 Z57 1991. {**1004**}

Author/editor index

Adler, F. R., *see* OTHMER, H. G., 1001, 1003, 1005
 AHO, ALFRED V., **1004**, **1005**

Bell, Timothy C., *see* WITTEN, IAN H., 1001

Dallon, J. C., *see* OTHMER, H. G., 1001, 1003, 1005
 DOUGHERTY, DALE, **1004**

GALLMEISTER, BILL, **1004**

Kernighan, Brian W., *see* AHO, ALFRED V., 1004,
 1005

KNUTH, DONALD E., **1003**

LAMPORT, LESLIE, **1001**

LEWINE, DONALD A., **1004**

Lewis, M. A., *see* OTHMER, H. G., 1001, 1003, 1005

Moffat, Alistair, *see* WITTEN, IAN H., 1001

OTHMER, H. G., **1001**, **1003**, **1005**

ROBBINS, ARNOLD, **1004**, **1005**

Robbins, Arnold, *see* DOUGHERTY, DALE, 1004

Weinberger, Peter J., *see* AHO, ALFRED V., 1004,
 1005

WITTEN, IAN H., **1001**

ZLOTNICK, FRED, **1004**

Subject index

.aei, 1004
.bbl, 1004
.ida, 1004

A/UX, 1005
Apple Macintosh, 1005
AUTHIDX, 1001, 1004, 1005
authidx.sty, 1003
\authoreditorindex{BIBTEX-label}, 1003
AWK, 1004, 1005

BAWK, 1005
\bibdata, 1003
\bibindex, 1004
\bibliography, 1001, 1003
\bibliographystyle, 1001
bibliography archive
 Karlsruhe, 1001
 T_EX User Group, 1001
BIBSEARCH, 1001, 1006
BIB_TE_X, 1001
BIBTEX, 1004
BIB_TE_X sample entry, 1002

\cite, 1003, 1004
Comprehensive T_EX Archive Network (CTAN), 1005

\def, 1003
\documentstyle[authidx]{...}, 1003

EMACS, 1005

Free Software Foundation, 1005

GAWK, 1005
GNU, 1005

IEEE, 1004
\index, 1003
INFO, 1005
\input, 1003
International Standard Book Number (ISBN), 1002
\item, 1003

Karlsruhe bibliography archive, 1001

L^AT_EX sample file, 1004

MachTen, 1005
Macintosh
 Apple, 1005
Macintosh Programmer's Workbench (MPW), 1005
MAKE, 1004
Makefile, 1004
MAKEIDX, 1004
MAKEINDEX, 1004, 1005
\makeindex, 1003, 1004
\makeauthoreditorindex, 1003, 1004
makeidx.sty, 1003

mathematical biology, 1001
MAWK, 1005
MG, 1001, 1006

National Science Foundation, 1001
NAWK, 1005

output routine, 1003

POSIX, 1004, 1005
\printauthoreditorindex, 1003, 1004
\printindex, 1003, 1004

\renewcommand, 1003
\renewenvironment, 1003

\see, 1004

T_EX User Group bibliography archive, 1001
theauthoreditorindex, 1003
theindex, 1003, 1004

UNIX, 1004
\usepackage{authidx}, 1003

World-Wide Web, 1005
\write, 1003

 ◇ **Nelson H. F. Beebe**
 Center for Scientific Computing
 University of Utah
 Department of Mathematics, 105 JWB
 155 S 1400 E RM 233
 Salt Lake City, UT 84112-0090
 USA
 Tel: +1 801 581 5254
 FAX: +1 801 581 4148
 WWW URL:
 http://www.math.utah.edu/~beebe/
 Internet: beebe@math.utah.edu,
 beebe@acm.org, beebe@ieee.org