Remaking ACM LaTeX styles

Boris Veytsman

Abstract

The Association for Computing Machinery is one of the largest publishers of computation texts in the world. It publishes more than fifty journals and many more conference proceedings every year. It was among the early adopters of TeX.

Unfortunately, over the years ACM styles accumulated many patches and haphazard changes. They diverged to the point when supporting became an impossible task. This warranted a complete refactoring.

This talk discusses the experience of rewriting ACM styles and the lessons learned.

1   Introduction

Five years ago I was asked to update BibTeX styles for the Association of Computing Machinery (ACM). I did not know at that time that this commission would start a very interesting line of work.

The ACM [2] is one of the largest publishers in the computing and information science in the world. It produces dozens and dozens of journals and conference proceedings. Thus I considered the work on this assignment to be a great honor and a large responsibility.

It befits the ACM mission and traditions that it is one of the early adopters of TeX. There are timestamps in the ACM style files going as far back as the middle of 1980s, i.e., even predating TeX3. As any computer specialist knows only too well, code this old requires much care and attention lest it become a crazy quilt of patches upon patches (the integrity of TeX itself over the years is an important exception rather than the general rule). This is especially true when the code is maintained by generations of programmers stressed by deadlines and production requirements.

In the case of the ACM files, both the LaTeX and BibTeX code and the output display the result of many temporary *ad hoc* decisions and show overlapping fingerprints of editors and coders, often with incompatible philosophies and approaches. As one frustrated TeXpert wrote me (name withheld by request),

> . . . 3 packages copied in with a comment (good!) that they are needed but without taking out `\endinput` that was in the code from the package copied in (bad :-) so after the first nothing else is ever used. . .

1. Class files:
   - (a) `acm_proc_article-sp.cls`
   - (b) `acmlarge.cls`
   - (c) `acmsiggraph.cls`
   - (d) `acmsmall-ec13.cls`
   - (e) `acmsmall.cls`
   - (f) `acmtog.cls`
   - (g) `acmtrans2m.cls`
   - (h) `sig-alternate-05-2015.cls`
   - (i) `sig-alternate.cls`
   - (j) `sigchi-ext.cls`
   - (k) `sigchi.cls`
   - (l) `sigplanconf.cls`
2. BibTeX styles:
   - (a) `ACM-Reference-Format-Journals.bst`
   - (b) `SIGCHI-Reference-Format.bst`
   - (c) `acmsiggraph.bst`
   - (d) `acm-abbrv.bst`
   - (e) `acm-alpha.bst`
   - (f) `acm-plain.bst`
   - (g) `acm-unsrt.bst`

Figure 1: Legacy code base (2015)

> . . . and it seems there is a redefinition of startsection inside that is broken—last night 30 min before my deadline I found 3 sections dangling at the bottom of columns. . .

> . . . and the footnotes are horror and the fonts too and. . .

> . . . looks worse than your average Word document . . .

These problems were exacerbated by the amount of copy and paste in the TeX code. Many times over the years whenever the need arose, the original code was cloned, changed in subtle (or not so subtle) ways, and a new class file was released. At the end of 2015 I found that I was dealing with as many as 12 class files and 7 BibTeX styles (Figure 1). Thus any update to the system required dozens of tantalizingly similar but slightly different changes in these files. This was not sustainable.

Another problem with the old styles was that interfaces to the elements like tables or figures were set long before the common standards were adopted. As the result, they looked quite strange for a LaTeX user. The unusual ways to do usual things were confusing to the authors and caused errors.

Thus, the decision of the senior staff of ACM to make a radical refactoring of the styles was excellent news. Both the typographic design and the coding were going to change. This was an opportunity to write the styles from scratch.

2 Organization of work

With many stakeholders, it took some effort to organize the writing of the styles and templates. The tasks were distributed as follows. The ACM editors updated the design and fonts selection. I wrote the LaTeX and BibTeX code. The company Aptara [1], which does typesetting for the ACM, developed word processor templates for the authors who do not use TeX, as well as tools for the extraction of metadata.

Since many conference committees (SIGs) wanted to be involved in the process, LaTeX and BibTeX code was put in a Github repository (`https://github.com/borisveytsman/acmart`). Github-based development turned out to be quite efficient for our purposes: the testers and SIG representatives could quickly assess the changes, submit bug reports and even contribute the code. Github seems to be a mature environment for free software development.

Sometimes it was difficult to accommodate the wishes of all the stakeholders, but we tried to keep in the spirit of compromise and consensus.

3 Design features

Instead of many class files (Figure 1) we use one class, *acmart*, with options corresponding to the output version. I sincerely hope this decision (one document class with options rather than several classes) will prevent the proliferation of copy-and-paste that plagued the old styles.

As suggested by the name, *acmart* is based on the famous *amsart* class [4], so all $\mathcal{AMS}$-LaTeX advanced math typesetting features are available by default. You can use environments like `cases`, `gather` or `multline`, commands like `\dfrac` and `\tfrac` or `\text` in math mode, as well as AMS-style theorem definitions (the class itself defines several theorem-like constructs and theorem styles).

There are three journal options: *acmsmall* for small trim size journals, *acmlarge* for large trim size journals and *acmtog* for *Transactions on Graphics*, which traditionally uses two-column format. There are five proceedings options: *sigconf* for most conference proceedings, *siggraph*, *sigplan* and *sigchi* for specific proceedings with distinct formatting, and *sigchi-a* for the special SIGCHI Extended Abstract. The latter is quite unusual: it has wide margins with marginal figures and tables. Another option, *manuscript*, is for a generic manuscript.

In Figures 2, 3, 4 and 5 some examples of the output are shown. Additional samples can be found in the documentation on CTAN (`http://ctan.org/pkg/acmart`) or in your TeX distribution.

Another important decision was to eliminate use of proprietary fonts. The Libertine fonts [6] with

Figure 2: Journal output, small trim size: *acmsmall*

newtxmath [5] give the pages a clean and crisp look. The footnotes are no longer cramped. In general, we tried to add a little air to the pages, while keeping in mind that the authors are constrained by page count limits.

One of the main principles of the design is the integrity of the interface. While the typesetting of the journals and proceedings is quite different, the interface is the same. The author should be able merely to change *acmsmall* to *sigconf* option in the `\documentclass` command in order to typeset the manuscript in a different category. The only exception are the marginal figures and tables for the *sigchi-a* option, which have no corresponding material in the other formats.

Another principle is the logical markup with most visual decisions made by LaTeX. This can be demonstrated by the way authors' information is encoded. In the old design the authors should manually set the number of authors and align their addresses on the page using tabular-like commands. The new design does this automatically.

Since the TeX file is used both for typesetting and for automatic extraction of metadata by Aptara

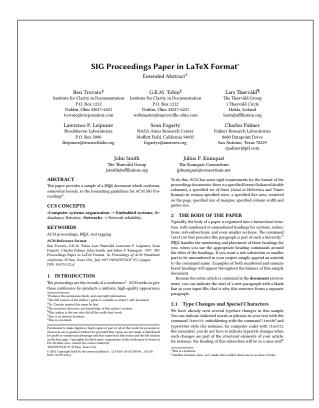**Figure 3**: Journal output, two columns: *acmtog*

Figure 4: Proceedings output: *sigconf*

Figure 5: SIGCHI Extended abstract: *sigchi-a*

```
\author{Ben Trovato}
\authornote{Dr.~Trovato insisted his
  name be first.}
\orcid{1234-5678-9012}
\email{trovato@example.edu}
\author{A. U. Thor}
\email{author@example.edu}
\affiliation{%
  \institution{Institute for Clarity
    in Documentation}
  \streetaddress{P.O. Box 1212}
  \city{Dublin}
  \state{Ohio}
  \postcode{43017-6221}
  \country{USA}}
```

Figure 6: Example of author information commands

tools, the commands are highly structured. For example, the authors' information is typed using the commands like \streetaddress or \city (Figure 6). There are special commands for grant sponsors and grant numbers, etc.

The class offers a number of useful features like

canned copyright statements (vetted by the ACM lawyers), review mode with line numbers printed, anonymous mode with the information about the authors, affiliations, grants and acknowledgments suppressed (for a blind review), etc. This anonymous mode is just one of the options for conditional typesetting: the authors could also have different versions for the online and hard-copy; for example, the online version could include supplementary materials. There are provisions to include CCS "concepts": hierarchical keywords generated by the ACM website.

The class uses standard LaTeX 2ε interfaces to common elements, such as figures and tables, as much as possible. The only area with ACM-specific commands is the front matter: unfortunately all publishers use their own systems to indicate the authors and their affiliations, and ACM is no exception here.

4   Bibliography

Historically some ACM publications used author-year citations, while other used numbered cites. Even the author-year ones were not uniform: some used *natbib*, while some used their own interface. There were pervasive differences in bibliography formatting. This led to a large number of "official" ACM BibTeX styles (see Figure 1).

The new *acmart* package uses only one BibTeX style, which is *natbib*-compatible and defaults to numeric citations. Fortunately, the *natbib* package [3] allows the user to choose either author-year or numbered citations, thus allowing SIGs to customize their bibliographies. Even when the citation style is numeric, commands like `\citeyear` and `\citeauthor` are allowed.

Another interesting feature of the citation style is that the bibliographic output is highly structured for use by the cross-referencing software. This is done transparently to the user, creating the entries like the one shown on Figure 7.

5   Conclusions and acknowledgments

This large work of creating the new ACM styles would not be possible without the help of many people. I would like to express my gratitude to:

- ACM editors: Craig Rodkin, Bernard Rous.
- Aptara: Neeraj Saxena, Sehar Tahir.
- Testers, users and SIG representatives: Chris Guccio, Wayne Graves, Matthew Fluet, Jofish Kaye, Frank Mittelbach, John Owens, Tobias Pape, David A. Shamma, Stephen Spencer.
- Authors of the early versions of ACM TeX and BibTeX styles.

Boris Veytsman

```
\bibitem[\protect\citeauthoryear{Akyildiz,
    Melodia, and Chowdhury}{Akyildiz
    et~al\mbox{.}}{2007}]%
  {Akyildiz-02}
\bibfield{author}{\bibinfo{person}{I.~F.
    Akyildiz},
  \bibinfo{person}{T. Melodia}, {and}
  \bibinfo{person}{K.~R. Chowdhury}.}
\bibinfo{year}{2007}).
\newblock \showarticletitle{A Survey on
  Wireless  Multimedia Sensor Networks}.
\newblock \bibinfo{journal}{{\em Computer
    Netw.\/}}
\bibinfo{volume}{{51}, 4},
\bibinfo{pages}{921--960}.
```

Figure 7: Bibliography entry made by the new ACM `bst` file

The new ACM styles are available on CTAN and the ACM web site, as well as in the major TeX distributions like TeX Live and MikTeX.

As mentioned above, development is hosted at Github, `https://github.com/borisveytsman/acmart`. The Github interface is the best way to send me bug reports or feature suggestions.

References

[1] Aptara. `http://www.aptaracorp.com`.

[2] Association for Computing Machinery. `http://www.acm.org`.

[3] Patrick W. Daly. *Natural Sciences Citations and References (Author-Year and Numerical Schemes)*, 2010. `http://ctan.org/pkg/natbib`.

[4] Michael Downes and Barbara Beeton. *The amsart, amsproc, and amsbook document classes*. American Mathematical Society, 2015. `http://ctan.org/pkg/amsart`.

[5] Michael Sharpe. *New TX font package*, 2016. `http://ctan.org/pkg/newtx`.

[6] Bob Tennent. *LaTeX Support for Linux Libertine and Biolinum Fonts*, 2014. `http://ctan.org/pkg/libertine`.

⋄ Boris Veytsman
  Systems Biology School and
     Computational Materials
     Science Center
  MS 6A2
  George Mason University
  Fairfax, VA 22030 USA
  `borisv (at) lk dot net`
  `http://borisv.lk.net`