Using BibTeX

Considering how many times we need to cite other works in our papers, BibTeX is the greatest thing since curved goalie sticks.

To use BibTeX, you need to do the following:

1. Prepare one or more .bib files, each containing BibTeX entries. Appendix B of Lamport tells you about .bib files.

2. Include a `\bibliographystyle` command in your .tex file to indicate how you want the bibliography to appear. It has to come somewhere after the `\begin{document}` command. The standard options are

   - `\bibliographystyle{plain}`: entries are sorted alphabetically and referenced by numbers.
   - `\bibliographystyle{alpha}`: entries are sorted alphabetically and referenced by labels formed from the author names and year of publication. For example, a paper that has just me as the sole author and was published in 1998 would have the label [Cor98], whereas CLRS has the label [CLRS01].
   - `\bibliographystyle{unsrt}`: entries are in order of first citation and referenced by numbers.
   - `\bibliographystyle{abbrv}`: like `plain` except that first names, months, and journal names are abbreviated.

   If you hunt around in the directories /usr/share/texmf/bibtex/bst, you can find other styles, like `siam`, `acm`, and `ieeetr`.

3. At the location where you want the bibliography to appear, put a `\bibliography` command. It takes as an argument a comma-separated list of the .bib files it should use. You can omit the .bib suffix. I find it best to not use the tilde form of a directory. Luckily, on our department systems, /u/thc is the same as ˜thc.

4. Use the `\cite` command for all citations.

The .bib files

You can have as many .bib files as you like. I have one for I/O-related papers and one for general papers. I also have one for FFT-related papers.

In retrospect, I should probably have had just one .bib file for everything. I've been migrating toward doing so, and my file ˜thc/mytex/papers.bib has almost everything I've put into BibTeX form.

We'll go over papers.bib. Things to note:

- emacs has a BibTeX mode. In BibTeX mode, a 3-key combination beginning C-c C-e will give you all the fields for a particular type of BIBTEX entry. For example, to get all the fields for an article, type C-c C-e C-a. Personally, I cannot remember all of the key bindings, so I either (a) hunt around until I get what I want, or (b) type C-h m when in BibTeX mode to get a list of key bindings. After you have filled in all the fields you want to fill in, type C-c C-c; this command "cleans" the entry by removing all unused fields and removing "OPT" from the optional fields that you have filled in.

  I recommend that you put the following line in your .emacs file:

  ```
  (set-variable 'bibtex-maintain-sorted-entries t)
  ```

  This line will ensure that BibTeX mode keeps your entries sorted by key.

- The way that BibTeX mode operates has changed over the years. I think that every entry in papers.bib was generated with BibTeX mode, but whereas now entries are enclosed in curly braces, they used to be enclosed in double-quotes.

- A hyphen in the pages field is considered to denote a range, and BIBTEX automatically converts it to an en-dash. If you have a page number with a hyphen in it (e.g., pages C-10–C-22), you're in trouble. If you put the en-dash in the pages field, it's left as a single en-dash.

- If you put curly braces around something, it will be typeset exactly as you have typed it.

  - For titles that are of papers (i.e., @Article or @InProceedings), BIBTEX normally makes only the first word and a word following a colon start with a capital letter. If you want "I/O" to appear in a title as "I/O" and not "i/o", put curly braces around it.
  - If you use the 3-letter month abbreviations, remove the curly braces from around them.
  - If you use abbreviations for journals or conferences (i.e., you take advantage of @string and the builtin @string commands), remove the curly braces from around their uses.

- Yes, you use @string to define your own strings, typically for conferences and journal. BIBTEX has a handful of builtin @string commands. You can see which ones BIBTEX has builtin by examining the .bst files in /usr/share/texmf/bibtex/bst, especially in the directory /usr/share/texmf/bibtex/bst/base. Search for `MACRO`; you'll find how months and journal names are predefined.

- When writing the first edition of CLR, we adopted the convention that the key we use is

  1. the full last name of the first author, followed by
  2. the first two letters of the last name of each of the other authors, followed by
  3. two digits for the year of publication, optionally followed by
  4. 'a', 'b', 'c', etc., as necessary to make the key unique.

  Now, you may choose a different scheme. For example, if you look at Dave Kotz's bibliography file ~dfk/bib/whole.bib, you'll see that Dave uses keys that are the full last name of the first author, followed by a colon, followed by a descriptive phrase. Dave also often includes the abstract and/or some comment text. I usually don't, since there is no way that's going to appear in any use of the entry.

- When BIBTeX was devised, there was no such thing as the Web, and hence no notion of publishing on the Web or of URLs. If something is published only on the Web, I use the @Misc entry type. If you just type in a URL into the howpublished field, you're likely to get some overfull lines. Since you won't have the option of rewriting, you need to allow line breaks. You can use either the `\allowbreak{}` command or the `\linebreak[1]` command to do so. For example, see the entries with keys Cormen03a, Cormen03b, and Cormen97a in papers.bib.

  Note also that if you want to typeset a tilde without resorting to the typewriter-like font you get from the `\verb` command, you can just enter `\~{ }`.

Citing bibliography entries

There are a few things about using the `\cite` command that you should know.

- Do not run the citation into the text preceding or following it.

  > Cormen, Davidson, and Chatterjee first described FG[5] in June 2003.

  is wrong.

  > Cormen, Davidson, and Chatterjee first described FG [5] in June 2003.

  is correct.

- To cite more than one reference at the same point in the text, use just one `\cite` command, with keys separated by commas.

  > Cormen and his coauthors produced several papers on out-of-core sorting [1], [2], [4], [3].

  is wrong.

  > Cormen and his coauthors produced several papers on out-of-core sorting [1, 2, 4, 3].

  is correct.

- The above example also illustrates one of my pet peeves. When citing multiple references in the same place, the citations should appear *exactly* in the order in which they appear in the bibliography. What I should have done above was

  > Cormen and his coauthors produced several papers on out-of-core sorting [1, 2, 3, 4].

  You may have to reorder the citations, depending on the bibliography style you're using. For example, `plain` and `alpha` may put the references in different orders in the bibliography.

- As a matter of style, try to avoid using a citation as a noun phrase. For example,

  > [3] showed that out-of-core columnsort is competitive with NOW-Sort.

  is not as good as

  > Chaudhry, Cormen, and Wisniewski [3] showed that out-of-core columnsort is competitive with NOW-Sort.

This advice is more of a guideline than a rule. There are times that using a citation as a noun phrase is almost unavoidable, but do what you can to avoid these situations.

- If you are referencing a book, it is a nice courtesy to the reader to give the appropriate pages or section. That's because it's usually not so hard to find material in a paper, but books are a lot bigger.

  To give a location within the citation, you use the optional argument to `\cite`:

  > Cormen, Leiserson, Rivest, and Stein [7, pp. 509–517] analyze the disjoint-set-union algorithm with path compression and union by rank via the potential method.

  Or,

  > Cormen, Leiserson, Rivest, and Stein [7, Section 21.4] analyze the disjoint-set-union algorithm with path compression and union by rank via the potential method.

  If you want to cite a location within a citation and you have more than one citation, you'll need to use multiple citations. In this case, you'll want to rewrite your text so that it flows smoothly.

  > Both editions of *Introduction to Algorithms* [6, pp. 450–458] and [7, pp. 509–517] analyze the disjoint-set-union algorithm with path compression and union by rank, but they use different methods for the analyses.

Running the commands

When your paper uses BIBTeX, you have to run four commands to create the .dvi file:

1. Run latex on your .tex file. That gets the bibliography style and citation keys into your .aux file.
2. Run bibtex on your .tex file. The bibtex program will read the .aux file to determine the bibliography style it should use and which keys it needs to find in the .bib files. The bibtex program produces two files: a .bbl file, which will turn into the bibliography in your paper, and a .blg file, which is just a log file.
3. Run latex on your .tex file again. That gets the citations into the output of latex, but you'll still get complaints about undefined citations.
4. Run latex on your .tex file one more time. Everything should be hunky-dory.

If my file is bozo.tex, then in tcsh, I type `latex bozo`, then `bibtex !$`, then I hit the uparrow key twice, and then one more uparrow.

Two final pieces of advice

Read over your bibliography. Do not assume that just because you think you typed the entries into the .bib file correctly, they'll print OK in your paper. Read the bibliography and verify. You would not believe some of the bibliography errors I see in papers submitted to conferences and even journals: names spelled incorrectly, missing page numbers, and even citations with just a URL and nothing else!

Do not copy bibliography entries from others. Do not steal entries in .bib files. Make your own. Do not copy a bibliography entry from a published paper. *Make sure that you have the cited work in your own hands, or at the very least, that you have verified that the entry is 100% correct!*

Believe it or not, I have seen—many times—authors cite *their own work* incorrectly. So you should not even copy a bibliography entry for a paper by author X from something else by author X!

Blind copying of bibliography entries is how bibliography errors propagate.

References

[1] Geeta Chaudhry and Thomas H. Cormen. Getting more from out-of-core columnsort. In *4th Workshop on Algorithm Engineering and Experiments (ALENEX 02)*, pages 143–154, January 2002.

[2] Geeta Chaudhry and Thomas H. Cormen. Stupid columnsort tricks. Technical Report TR2003-444, Dartmouth College Department of Computer Science, April 2003. Submitted to *Journal of Algorithms*.

[3] Geeta Chaudhry, Thomas H. Cormen, and Leonard F. Wisniewski. Columnsort lives! An efficient out-of-core sorting program. In *Proceedings of the Thirteenth Annual ACM Symposium on Parallel Algorithms and Architectures*, pages 169–178, July 2001.

[4] Geeta Chaudhry, Elizabeth A. Hamon, and Thomas H. Cormen. Relaxing the problem-size bound for out-of-core columnsort. In *Proceedings of the Fifteenth Annual ACM Symposium on Parallel Algorithms and Architectures*, June 2003. SPAA Revue paper.

[5] Thomas H. Cormen, Elena Riccio Davidson, and Siddhartha Chatterjee. Asynchronous buffered computation design and engineering framework generator (ABCDEFG). In *High-End Computing Revitalization Task Force Workshop*, June 2003. White paper.

[6] Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*. The MIT Press and McGraw-Hill, 1990.

[7] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. *Introduction to Algorithms*. The MIT Press and McGraw-Hill, second edition, 2001.