

T_EX and medicine

Elizabeth Dearborn

Buffalo, New York

elizabeth (at) medicalese dot org

<http://www.MeDiCaLeSe.org>

1 Introduction

Way back in 1985, when I worked at a struggling graphic design shop, I learned to set type on a Mergenthaler phototypesetting machine which was old even then. Remember width cards and photographic chemicals? In 1987, I went into medical transcription, which is an extremely knowledge-intensive line of work with very tight deadlines. I wanted to be good at what I did, so I maintained my own word list, which grew exponentially with time.

I was interested in computer programming and the Internet. I studied these things on my own, and in 2001 I started my website at <http://www.MeDiCaLeSe.org>. Using a JavaScript site search program I found on the Internet, I put my word list online in searchable form. Although the website and the database were available to everyone, my main objective at the time was to be able to search my own word list quickly. Whenever a person searched the database, my entire word list was loaded into a temporary file on the client's machine. This also required JavaScript to be enabled on the client side.

After a couple of years, the word list had gotten huge, the JavaScript search engine was taking a long time to load, and I was beginning to feel that the products of my research were worth money. I wanted to find a way to serve *only* the requested results and serve them up faster, and, if possible, I wanted to accomplish this without making my code visible to the end user.

2 The php programming language

I looked, but I could not find a free or inexpensive package that would do what I wanted. Hiring a programmer was out of the question, as was going back to school. I'm not the type who takes courses; everything I've ever wanted to know about, I've learned on my own, with some degree of success. I checked out the various programming languages which are used on the Internet, such as C++, Pascal, php, and Perl. Of these, php seemed best for my purposes. It is free, runs entirely on the server side, and virtually every web host offers it to its customers.

My approach to learning enough php to get the

job done was very similar to the way I later went about learning T_EX. First, I defined exactly what I wanted to do. Next, I searched the official php website at <http://www.php.net> to see if a command existed that would accomplish this. If not, I would need to write my own. Sound familiar?

I used two books, both published by O'Reilly Media: *Programming PHP* by Rasmus Lerdorf and Kevin Tatroe, and *PHP Cookbook* by David Sklar and Adam Trachtenberg. Between these two wonderful books, the vast resources of the Internet, and my own determination, the improved site search was ready to go live in about two months' time, in the summer of 2003. The original version of my php site search had the Google-style highlighting as it does today, but did not then include the capability to exclude one string from the search results. I invite you to visit <http://www.MeDiCaLeSe.org> and try it out.

3 A database in plain text files

If you type the word "Jones" into the search box on the site, it will return 12 results, two of which are shown below:

Matches: jones: 12

Orthopedics: Jones fracture - fractured proximal fifth metatarsal.

Diagnostics: Jones silver stain.

If you view the page source, you won't see a listing of the php code that made the page, but you will be able to see the HTML code for the results. This is not very different from what the plain text entries in the database actually look like, which is:

```
<B>Orthopedics: </B>Jones fracture -  
fractured proximal fifth  
metatarsal.|06/04/04
```

```
...  
<B>Diagnostics: </B>Jones silver  
stain.|06/02/04
```

Short and sweet, and in no particular order. Anything following the vertical line does not show up on

the results page.

I do extensive research on medical terminology every single day, and I enter the new or changed data into the two text files which make up the database and upload them to the server every night. The php search program calculates the number of words, number of entries, and the last time the files were updated.

MeDiCaLeSe, the book and the website, concentrates on words that are new, ambiguous, or difficult to find, and includes words that are not found in the standard medical references, such as clinical trial acronyms, homeopathic and herbal remedies, products available only outside the U.S., discontinued drugs, and unapproved cancer therapies.

4 Show me the money!

I worked hard for little compensation as a transcriptionist, and in the case of the website for no compensation at all. In 2004 someone suggested I turn my word database into a book. At first I resisted, because the daily updates and corrections were a large part of the value of the website, and I thought it was impossible to publish even a reasonably current book on such an enormous, constantly changing, and complicated subject, if typographic quality were to be a consideration.

I had seen a few quickly-printed books which were published in a hurry to cash in on current events, but their quality was atrocious — full of typographical errors and crookedly printed on brittle, yellowing newsprint.

I had guest-edited several medical terminology books for a large medical publisher in the mid-1990s. These books were very nicely produced, but as a transcriptionist I felt the information they offered was not quite enough. And, by now, I enjoyed having total control over my own website and was not about to relinquish that just to get a book published. If I went to the trouble to put a manuscript together, I felt that the big medical publishers would turn it away. I was good at Internet research, and I started looking into print-on-demand publishing.

I would have to typeset my own manuscript, but this didn't intimidate me, since I had worked in typesetting before. I had heard of T_EX in the context of mathematical typesetting, and I started studying it on the Internet. I didn't have much money, and I appreciated the fact that T_EX, besides being free, requires fairly minimal computer hardware. I continued to study T_EX while organizing my database in manuscript form. I downloaded and experimented with T_EX, joined TUG, and lurked on `comp.text.tex` and the T_EXhax list.

I write a little bit of short mystery fiction, and I knew several mystery writers who had either published with a subsidy house or gone out on a limb and published their own work. Subsidy publishing didn't interest me, as I needed to have the books available through the regular book-buying channels. Finally, a friend who started her own publishing company convinced me that I could do the same, and that I could have the books printed on demand by Lightning Source and distributed worldwide by Ingram. This was all I needed to hear! In April 2005 I obtained a business license and business checking account in the name of Blowtorch Press, and ordered a block of ISBN numbers. I named the book *MeDiCaLeSe 2005*, so that identifying subsequent revisions would be a no-brainer.

5 Why T_EX for medicine?

The large medical book publishers, such as Elsevier, usually have their own T_EX style files available. Information on manuscript preparation for medical journals is available from the International Committee of Medical Journal Editors, <http://www.icmje.org>. Over 400 journals use the Vancouver style files, available at <http://www.ctan.org/tex-archive/biblio/bibtex/contrib/vancouver>.

For those who need the old apothecary symbols for minim, dram, scruple, etc., they are available in the PIXymbols font for Windows and Mac, sold at http://www.vershen.com/psg_txtc.html.

The Computer Modern font has almost all the diacritical marks and special characters needed in medicine. The `textcomp` package is needed only for the R symbol. I wanted the μ symbol, the dot and umlaut in Åström, and the beta symbol in Dia β eta to be properly typeset, even though these characters are not used in medical transcription; in that context, these words are written as micro, Astrom, and DiaBeta respectively. Medical transcription is usually done in a word processing program, and turnaround time is crucial.

By mid-2004 I had found the T_EX editor I like best, which is L^AT_EX Editor by Shu Shen, a graduate student in Singapore. This software is free and very lightweight, which was an advantage since I didn't know much about T_EX beyond what was needed to get the job done.

The php code does all the heavy lifting for the website. For a book, I would need to organize the data somehow. I put the entries into 39 chapters called Drugs, Abbreviations, Vocabulary, Cardiology, Neurology, and so on, alphabetizing the entries within the chapters, and then duplicating the entries into the different chapters as needed. I aimed for as

much redundancy as possible, because my primary intended audience of medical transcriptionists would not be willing to purchase the book unless they knew they could find what they were looking for quickly, and with enough information to know whether the word was the correct one for the situation.

Also because of the special requirements of the book's intended audience, I included this statement in the preface:

NO ADDED HYPHENS: We did not introduce new breaks into any of the words in this book. If you see a hyphenated word at the end of a line, it means the word should always be written with the hyphen. For the medical transcription community, we felt it was important to be clear on this, even at the expense of aesthetically undesirable line breaks. We have tried to make the book attractive and easily readable in spite of our self-imposed constraints on hyphenation.

To the best of my knowledge, *MeDiCaLeSe 2005* is the only medical transcription reference book with no added hyphenation.

I immediately saw that I would have no end of trouble without a bulletproof method of producing dictionary-style pages in double columns. I also needed the first and last words to appear at the top of the page on which they were defined. I began an intensive study of the `fancyhdr`, `geometry`, and `multicol` packages, and I tweaked the code until I was able to produce dictionary-style pages.

For double columns, all point size commands active at the end of each line must be the same. Otherwise, the lines of type do not match up from one column to the next.

Some of the main L^AT_EX file for the book appears below. I included my personal commands, which all begin with “en” plus one letter.

```
% c:\m05\medicalese.tex
\documentclass{book}
\def\enl{\filbreak\small\textbf}
\def\ene{\filbreak\normalsize\textbf}
\def\enn{\normalsize}
\def\eno{\enspace\scriptsize}
\def\enc{\filbreak\small\texttt}
\def\enb{\protect\raisebox{-2pt}}
\def\enu{\protect\raisebox{2pt}}
\usepackage{fancyhdr}
\usepackage[papersize={6.14in,9.21in},
margin={0.5in},top={1in},bottom={0.75in},
headheight={34.545pt},centering,
```

```
verbose=true]{geometry}
\usepackage[none]{hyphenat}
\usepackage{textcomp}
\usepackage{multicol}
\newcommand{\enk}[1]{#1\markboth
{#1}{#1}}
\begin{document}
% \frontmatter
\mainmatter
% COPY THIS CODE WHEN CHANGING HEADERS
% \end{multicols}
% \eject
\input drugs
\input notes
...
\input discont
\input notes
% COPY THIS CODE WHEN CHANGING HEADERS
% \backmatter
\ejct
\end{document}
\bye
```

Here is the beginning of one of the chapter files:

```
% c:\m05\labpath.tex
\pagestyle{fancy}\fancyhead{}
\fancyhead[LE,RO]{\textsf{\rightmark
\\leftmark}}
\fancyhead[C]{\textbf{Diagnostics}}
\fancyhead[RE,LO]{\textsl{\small includes
laboratory,\\pathology, and radiology}}
\renewcommand{\headrulewidth}{0.2pt}
\fancyfoot[LE,RO]{\textit{MeDiCaLeSe~2005}}
\fancyfoot[RE,LO]{\textsf
{www.medicalese.org}}
\renewcommand{\footrulewidth}{0.2pt}
\raggedright
\begin{multicols}{2}\\
% END OF CODE TO COPY WHEN CHANGING
% HEADERS ... diagnostics
\textbf{\enk{3-androstenedione}}\enn:
endocrine test ... \ene
{\enk{3D color Doppler}}\enn: developed ...
be measured by standard 2D imaging.\ene
```

And the end of the same file:

```
{\enk{Xplorer}}\enn~filmless radiographic
imaging system.\ene
{\enk{ZstatFlu}}\enn~throat swab, a
quick test for influenza A and B.\ene
\\
% COPY THIS CODE WHEN CHANGING HEADERS
\end{multicols}
\ejct
```

which produces:

Xplorer filmless radiographic imaging system.
ZStatFlu throat swab, a quick test for
influenza A and B.

6 Ugliness and badness

Because of the rule about not adding hyphens, the book contains some ugly paragraphs, like this:

Avalide: irbesartan/
hydrochlorothiazide, combination
angiotensin II receptor blocker/thiazide
diuretic.

Also, the warning

```
Underfull \vbox (badness 10000) has  
occurred while \output is active
```

happened quite often. I learned to disregard these.

Here's one that doesn't look that bad, but \TeX complains that the line is just a tad too long (for my actual page width), and suggests an incorrect hyphenation:

Markham-Meyerding hemilaminectomy
retractor.

```
Overfull \hbox (4.52963pt too wide) in  
paragraph at lines 689--689  
\OT1/cmr/bx/n/10 Markham-Meyerding  
\OT1/cmr/m/n/10 hemil-aminec-tomy
```

In a project such as this, the most insidious kind of badness is not seen until the `.pdf` file has been made and one is examining it page by page, since there is no way to know in advance where the page breaks will occur. I'm embarrassed to admit that I didn't catch this point size error in the guide words at the top of page 592 in time.

SaphLITE
Songer

(Songer is incorrectly printed in `\small`, like `\LITE`.) When something like this happens, I go back to the input file and find the entry corresponding to the *first* guide word on the offending page. Just before the two closing brackets that define the first guide word, add the command to revert to normal size type.

7 Finished!

I spent a couple of hours a day for about six months typesetting the book. I gave up my transcription job in August to devote more time to the book, and finished the typesetting on September 14, 2005.

The book contained essentially the same information as the website of four days prior. Now I had to tackle the cover! At this point I started studying the `pstricks` package in earnest, and I purchased *The L^AT_EX Companion*, which is still the only \TeX -related book I've ever bought. Because of memory limitations, I wasn't able to use \TeX to make the front cover, but I did use it for the publishing company logo and for the text overlays on the spine and the back cover.

Altogether, the cover took me three weeks to make using Paint Shop Pro 7.04. This is not the printer's recommended software, but it was all I had. I made the cover in three pieces—front, spine, back—and pasted it together as a giant 300 dpi `.tif` file. Lightning Source had provided the bar code and I pasted this on the back cover. Then I burned the cover `.tif` file and the `.pdf` files which made up the interior to a CD and mailed it off to Lightning. Eleven days later, I received my proof copy by overnight delivery. I knew the `.pdf` files of the text wouldn't cause problems, but I wasn't at all sure how the cover would turn out. To my delight, it was absolutely beautiful.

I signed off electronically on the proof copy, and within days my book appeared for sale online at Barnes & Noble, Powell's, the university distributor eFollett.com, and Amazon, as well as other book-sellers all over the world. The book is selling within its niche market.

As publisher, I decide on suggested retail price, discount rate, and when/if a book goes out of print. *MeDiCaLeSe 2005* lists for \$41. I have no control over the actual selling price, and have seen the book advertised for sale at prices ranging from \$25 to \$80.

8 *MeDiCaLeSe 2006* and beyond

Today is October 22, 2006, and I'm in the final stages of preparing *MeDiCaLeSe 2006* for publication. The page count is 744 this time around. I discovered along the way that the upper limit of what one \TeX file can process on my machine is about 713 pages of size 6.14" \times 9.21". As soon as I finish proof-reading the final printed `.pdf` output, I'll make the cover—this time with `pstricks`, specifying CMYK colors for high quality color reproduction. I now have 512 mb of RAM, which I hope will be enough.

When the total number of pages exceeds 828, I will either have to increase the page size or look for an alternative to print-on-demand publishing. The average age of the book's target audience of medical transcriptionists is late 50s or older, so it would not be practical to decrease the point size.