A university thesis class: automation and its pitfalls

Peter Flynn

Abstract

Despite the large collection of thesis classes available, there are always features that an institution needs which can be better met by writing Yet Another Thesis Class. There are also variations in the quality and availability of documentation, assumptions about preloaded packages, and the ease (or otherwise) with which the author can modify the layout.

In the case of UCC, the official requirements were very simple, avoiding the tendency to over-specify details found in some university formats. The class was also required to be generally applicable to any discipline, so only a minimum of packages was needed (although this turned out to be more than anticipated).

The major component was an attempt to automate as much of the front matter as possible, based on options to tokenize the discipline and class of degree. This was done to avoid accidental omissions, variations, and misspellings in the titling (or even deliberate rewording); and to ensure that the relevant compulsory components appeared in the right place without the author having to do anything.

The result has been piloted with 20–30 PhD candidates for a year, and needs only a few final changes before release. Two other institutions in the state have already expressed an interest in basing their own thesis classes on this one.

A summary of some of the points covered here was published in a recent *TUGboat* [2].

1 Yet Another Thesis Class

A recent retrieval from CTAN1 shows that there are currently 42 thesis or thesis-related packages available for \LaTeX\ (see Figure 1). These are almost all institution-specific, and implement a wide variety of rules and restrictions which fall into five broad groups:

- title page metadata;
- sequencing of preliminary pages;
- wording of formal declarations;
- formatting and layout;
- markup abbreviation and shortcuts.

The level of detail required by each institution varies so widely that using a thesis class from else-

1 Scripted with the assistance of dog, tidy, and \texttt{latexref}, thanks to the robustness of the directory and link structure implemented by the CTAN team.

\begin{itemize}
  \item adfhesis \hspace{1em} Australian Defence Force Academy thesis format.
  \item afthesis \hspace{1em} Air Force Institute of Technology thesis class.
  \item beamerthesis \hspace{1em} Thesis presentations using beamer.
  \item classicthesis \hspace{1em} A ‘classically styled’ thesis package.
  \item ebsthesis \hspace{1em} Typesetting theses for economics.
  \item elteikthesis \hspace{1em} Thesis class for ELTE University Informatics wing.
  \item fbi thesis \hspace{1em} Computer Science thesis class for University of Dortmund.
  \item gatechthesis \hspace{1em} Georgia Institute of Technology thesis class.
  \item hepthesis \hspace{1em} A class for academic reports, especially PhD theses.
  \item jasthesis \hspace{1em} A ‘standard’ thesis class.
  \item jkthesis \hspace{1em} Document class for formatting a thesis.
  \item msuthesis \hspace{1em} Class for Michigan State University Master’s and PhD theses.
  \item muthesis \hspace{1em} Classes for University of Manchester Dept of Computer Science.
  \item pitthesis \hspace{1em} Document class for University of Pittsburgh theses.
  \item psuthesis \hspace{1em} Package for writing a thesis at Penn State University.
  \item ryethesis \hspace{1em} Class for Ryerson University Graduate School requirements.
  \item sapthesis \hspace{1em} Typeset theses for Sapienza University, Rome.
  \item seuthesis \hspace{1em} \LaTeX\ template for theses at Southern Methodist University.
  \item suthesis \hspace{1em} Typeset a Stanford University thesis.
  \item thesis \hspace{1em} Typeset thesis.
  \item thesis-titlepage-fhac \hspace{1em} Little style to create a standard titlepage for diploma thesis.
  \item thuthesis \hspace{1em} Thesis template for Tsinghua University.
  \item uafthesis \hspace{1em} Document class for theses at University of Alaska Fairbanks.
  \item ucdavisthesis \hspace{1em} A thesis/dissertation class for University of California Davis.
  \item ucthesis \hspace{1em} University of California thesis format.
  \item ucthesis2009 \hspace{1em} \LaTeX\ 2.09 document style for UC Theses.
  \item uethesis \hspace{1em} University of Houston thesis document style.
  \item uiuc thesis \hspace{1em} UIUC thesis class.
  \item umichthesis \hspace{1em} University of Michigan Thesis \LaTeX\ class.
  \item umthesis \hspace{1em} Dissertations at the University of Michigan.
  \item umanthesis \hspace{1em} Style for Universidad Nacional Autónoma de México theses.
  \item unswthesis \hspace{1em} UNSW theses.
  \item unothesis \hspace{1em} Class for dissertations and theses at the University of Oregon.
  \item uowthesis \hspace{1em} Document class for dissertations at the University of Wollongong.
  \item uschthesis \hspace{1em} USC thesis style for \LaTeX\ 2.09.
  \item utorontothesis \hspace{1em} A thesis class definition for University of Toronto.
  \item utthesis \hspace{1em} Thesis package for the University of Texas at Austin.
  \item utthesis2009 \hspace{1em} \LaTeX\ 2.09 thesis class for University of Toronto.
  \item utthesis2009 \hspace{1em} \LaTeX\ 2.09 style for University of Toronto.
  \item witsthesis \hspace{1em} University of Wisconsin-Madison Thesis \LaTeX\ Class.
  \item yorkthesis \hspace{1em} A thesis class file for York University, Toronto.
\end{itemize}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{thesispackages.png}
\caption{Thesis and thesis-related packages available from CTAN as of May 2012.}
\end{figure}
where usually means some re-configuration and re-programming, which may be beyond the skills of the author. In some cases there is extensive documentation and an example thesis document; in others there is very light formatting and specification, and authors are left to modify the document as they see fit.

In the present case, there were other requirements which led to the decision to write a local thesis class rather than modify an existing one:

- the class had to be usable by all disciplines, not just those in which \LaTeX{} has traditionally been used the most;
- it had to automate (where possible) those areas where the author would not in any case have a choice;
- it had to allow for the writing of a thesis in the Irish language.

Some guidance was available in the form of the Thesis Requirements published by the university. I am indebted to the staff of the Registrar’s Office, the Graduate Studies Office, and the Boole Library; to the individual colleges, departments, faculties, and schools; and to students and other users for all their comments and suggestions.

2 Building the thesis class

We took an early decision to base the class on the standard report class because it appeared to be the one most familiar to existing users for writing theses. This also meant we could adopt (or in some cases, prohibit) existing class options.

Among the suggestions we received in feedback from users (when we discussed developing a thesis class), was to keep it simple and make it obvious. We interpreted this to mean that we should as far as possible keep the existing meanings for existing commands and environments, and not introduce new ones which were not easily memorable.

Other suggestions included adding optional arguments to certain commands for features which were very frequently used, as repeatedly having to specify something manually was seen as part of the tedium of using \LaTeX{} by users who lacked the programming skills needed to write their own macros.

2.1 Title page metadata

This is often technically one of the simplest parts of a document class design, but it has an impact out of proportion to its position. It is, after all, just one page in a document that will run to hundreds of pages, but it is the first page people see, and the first place that critical eyes will look for errors. It is also hard to convince new postgraduate students that they do not have much choice in how this page is laid out.

The current layout (see Figure 2) is vertically- and horizontally-centred, and contains conventional metadata:

- title and subtitle;
- author and [optional] registration number and qualifications;
- institutional identity (crest\footnote{The crest is a character in a METAFONT font which was drawn for us by Jeremy Gibbons in 1989, and is only accessible for download locally.} and name);
- divisional identity (college, school, etc);
- class of degree;
- date;
- names of supervisors and head of discipline.

The only less conventional addition was the acknowledgement of a sponsor — this is commonplace in the Acknowledgements, but its appearance on the title

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Title page layout for the UCC thesis}
\end{figure}

Peter Flynn

\footnote{Users later reported the most compelling reason was that the report class supports the \texttt{abstract} environment, which the book class does not.}
Figure 3: Commands for the title page metadata

page is becoming a condition of some funding agencies. There is provision for the numbering of multi-volume theses with a \volume command, but I am informed by the Graduate Studies Office that multi-volume theses are a great rarity nowadays.

Because of complexities in the university structure, we had decided to enforce the discipline and degree options in the \documentclass command, rather than allow authors to name their affiliation in an uncoordinated manner. This enabled us to use the information before anything else, to set the string names for the faculty (college) and department (school) and the bibliographic reference format prevalent in the discipline. The remaining metadata, therefore, is given with conventional commands as shown in Figure 3.

The \title, \author, and \date are standard; the remainder were added with the commonly-used method of defining an internal command default which is then redefined when the author uses the equivalent external command, for example:

\def@subtitle{\relax}
\newcommand{\subtitle}[1]{% \
\gdef{\subtitle[#1]}{}
}

We can then test the internal commands for equality to \relax during the processing of \maketitle to see whether or not the metadata commands were used, so that their absence can be accommodated in the spacing or replaced by warning messages, for example:

\if\relax\@subtitle You need to give the name of your head of discipline \else\@subtitle\fi

All the additional commands are given defaults or warnings in this way, so that a beginner accustomed only to \title, \author, and \date (or a user who has not Read The Fine Manual) will not be faced with \LaTeX errors.

2.2 Sequencing of preliminary pages

The university rules require only that the Table of Contents comes immediately after the title page. Because the underlying report is invoked with the oneshide option, there are no blank pages. The \tableofcontents command is therefore contained in the \maketitle command, so that it cannot be omitted or moved.

The List of Figures and List of Tables are also in the \maketitle command, but as the class must also be usable in disciplines where there may be no tables or figures at all, the need for these lists is determined by two Boolean switches:

\if\lof\listoffigures\fi
\if\lot\listoftables\fi

These are set to false at the end of the document if there were no figures or tables, using global counters defined in the table and figure code, and written to the .aux file, where they will take effect on the subsequent run:

\ifnum\c@totfigure=0 \write\@mainaux{\string\global\string\loffalse}\fi
\ifnum\c@tottable=0 \write\@mainaux{\string\global\string\lotfalse}\fi

However, there are also class options no\lot and no\lof which will prevent the LoT or LoF being used even when tables and figures are present — when there are only one or two tables or figures in use, a formal list may not be wanted.

The compulsory formal Declaration that this work is the student’s own is also produced automatically as part of the \maketitle command, after the ToC (and LoF and LoT, if present), set centred on a page to itself.

No other declaration is required by the university, and it is not usual for copies of any of the forms signed by the supervisor or Extern to be included, but these could clearly be implemented at the same point by the same method if needed.

A considerable number of students require preliminary (unnumbered) sections, before the thesis proper starts with the first part or chapter. These are needed to hold explanatory material such as an Introduction, a list of materials, or tutorial matter

A university thesis class: automation and its pitfalls
on a special topic. While this could be done with a \section* command, a new \prelim command was created to ensure a page-break beforehand, and to create an entry in the ToC which would otherwise be absent.

The abstract environment was also changed to use this \prelim so that it too would occupy a page to itself (the Abstract is limited by the rules to 300 words).

Two new environments were created, dedication and acknowledgements; the first sets the content centred on a page by itself; the second just uses the \prelim command to title the Acknowledgements. The rules do not specify an order for the Abstract, Dedication, Acknowledgements, or any other pre-lims.

No decision has been taken about the position of glossaries: these are not mentioned in the rules, and while the glossaries package is recommended, there is no compelling evidence one way or another for its placement either here or at the end of the document.

2.3 Formatting and layout
The current rules [5] are very undemanding in this regard:

- The text must be either, printed, typewritten or otherwise reproduced on good quality size A4 paper, with a left-hand margin 4 cm. Double or one and a half spacing is recommended. Copies must be bound or otherwise securely fastened and numbered consecutively, page numbers to be located centrally at the bottom of the page.
- No mention is made of the other margins, or of the typeface or size, or of the format of bibliographic references, so some unilateral decisions were made (the handling of bibliographic formats is dealt with in § 3).

• The top and bottom margins are set to 3cm and the right-hand (foreedge) to 2.5cm;
• The typeface defaults to Computer Modern;
• The body size defaults to 11pt;
• The \setspace package is used to set the default to 11/2 line-spacing — some leading would in any case have been needed for 11pt type on lines this length;
• Page numbering uses roman numerals from the title page to the beginning of the first part or chapter, at which point it restarts in arabic numbers.

After some discussion locally and on \texttt{comp.text}, we decided to make the default document setting \raggedright. This was partly because it helps avoid H&J problems, especially in the natural sciences where very long words are more frequent; partly because a thesis is not a professionally-typset publication like a book, and does not appear to benefit greatly from justification; and partly because ragged-right setting improves readability on a page width with relatively long lines.

However, \raggedright also turns off paragraph indentation, so the \parskip package is used to add space between paragraphs. This layout is in fact expected by students whose experience to date has been restricted to wordprocessors, where it is conventional to use an empty paragraph between paragraphs to simulate paragraph-spacing. Two extra options, \texttt{justified} and \texttt{indented} can be used to restore the state to book-style.

Running headers and footers were implemented with the \fancyhdr package, to provide navigational detail from \texttt{leftmark} and \texttt{rightmark} without the capitalization used in the default classes (this proved remarkably difficult to defeat), and with the use of a \texttt{hbox} in the definitions to allow long titles to be line-wrapped. In draft mode, the footer also provides the \jobname and a timestamp.

Some minor changes included more space above and below captions; the enforcement of the rule for caption positions (Tables: above, Figures: below) by restyling floats with the \float package;\footnote{This has the (possibly) useful side-effect of allowing the \texttt{[b]} positional specifier.} the allocation of more space to the page number in the ToC, LoF, and LoT; and the reassignment of wider values to the page-fractions for floats, along the lines suggested in the \TeX FAQ [1].

Two minor changes are made to the default layout of block quotations and description lists (below), but all the remaining parts, chapters, sectioning, lists, and other structural elements behave and appear as usual in a \LaTeX document. The \texttt{usepackage} command can be used in the normal way: a list of the packages already in use is contained in the class documentation.

2.4 Markup abbreviation
No attempt was made to abbreviate any of the commands: although students do this frequently for their own pattern of usage, I have not been able to see any common methods. There are a few conventions in some disciplines for short-name macros for commonly-occurring constructs, and authors are free to use them. However, as mentioned earlier, some frequently-occurring constructs were found to benefit from a small amount of automation, the most important of which is the block quotation.
2.5 Quotations

It does not make the type size smaller, nor does it
place a margin of 1cm on each side of a page. However, if an optional argument such as this is to
be set with a right-aligned citation immediately
within even simple documents, such as a blank line or indentation to indicate
a new paragraph, or a large font to indicate a heading, and it is possible to
implement rudimentary but sufficient markup to enable documents to be opened
in an XML or LaTEX editor, and leave the finer detail to human editing.

A university thesis class: automation and its pitfalls

2.6 Description lists

A frequent annoyance for users is the inability of
the label argument to an \item in the description
list environment to be broken at line-end when it is
very long. The default formatting is also unusual
compared with modern practice of setting the label
value on a line by itself (as with the default formatting
of HTML’s <dt> element type). A number of
alternative formats from the \LaTeX\ Companion [4]
have been tried, but no final decision has been taken
on this yet.

3 Adherence to university structure

The university is currently in transition from the
traditional hierarchical Faculties and Departments
to a more fluid structure of Colleges and Schools.
As a result there a complex and overlapping transi-
tional organisation in which disciplines are being
merged, split, and renamed. In order to ensure that
the correct names are used, the author’s affiliation
and class of degree must be given as class options,
rather than as free text in the argument of a com-
mand. This also avoids misspellings, and the unfor-
tunate tendency of some students to represent the
discipline in terms other than the official ones. With
the imminent arrival of electronic submission, where
the name of the discipline or school will be part of
the PDF metadata, regularity is becoming more im-
portant.

In creating these options, it became clear that
other data could also be keyed to it, in particular
the bibliographic reference format required for each
discipline. As a result, selecting an affiliation op-
tion now both sets the correct naming and presents
the .bst file (and any associated .sty file) for the
discipline.

As there are currently 87 options for affiliation
and 92 for class of degree, this method would have
been unworkable in terms of manual maintenance.
Fortunately, the class was developed using an XML-
based methodology which generates the .dtx and

Figure 4: Default page layout for the UCC thesis
files, so the relevant string names and tokens could simply be transformed from an annual XML extract from the databases maintained by the university administration. In any case, if and when the transitional phase of restructuring is completed, the class interface will be updated to use a key/value syntax rather than 179 separate options!

4 Testing, feedback, and adoption
Informal testing was initially carried out by use in my own thesis, but was extended to drafts provided by students who came looking for help with LaTeX formatting. Discussions were held with the Registrar’s Office and the Graduate Studies Office to ensure that the layout implemented conformed to the rules.

In January 2010, an early version of the XML-generated class was made available locally for download [3], and over the course of 18 months about 45 students used it for their theses, reporting bugs as they were discovered.

The feedback was largely positive, and the automation of the title page and prelims was seen as a major benefit. The most useful feedback came as bug reports, and led to a spate of updates over the next six months as various solutions to errors were tested and implemented.

Some unresolved issues remain:
- separate options for the disciplines and classes of degree need to be replaced by key/value pairs;
- glossaries appear to be much more common than was previously envisaged;
- the formatting problems of description lists remain unresolved;
- the debate continues over the default unindented, paragraph-spaced, ragged-right setting vs justified and indented setting;
- the identification of the ‘right’ bibliographic reference format for each discipline is problematic. An enquiry among colleagues representing each discipline provided only a 20% response, so there is a lot more data to collect (the default has been set to Harvard). A few disciplines have two common formats — Physics, for example, allows either IEEE-ETR or AIP — so a mechanism is required to allow that to be specified.

At the moment the assumption is that students will be using BibTeX, not biblatex, as not all the formats required are yet available in the latter.

The class has by no means been tested to destruction: it appears to work with all the common packages, including hyperref, but the more use it gets and the more bugs are reported, the more likely it is to work.

A number of academics who use LaTeX themselves have started recommending the class to their students, and some informal changes have been made to create variant formats within the class for essays, term papers, and minor dissertations. An approach has also been received from a group of students in another institution in the state (which has no resident LaTeX expertise or support), asking for help in writing a thesis class, and I am aware of at least three students in a third institution who have adopted the class and simply changed the identity by editing the .cls file.

Writing a class file is a non-trivial activity, and I was fortunate to have access to a development and maintenance methodology which made editing and creation very much easier than writing a .cls file by hand. Having a clearly-defined goal makes development easier, as does having a set of patterns to work to, and I am indebted to the many authors of the classes listed in Figure 1 for their work. The intention is that from version 1.00 the package will be available on CTAN (minus the crest, which is reserved to the institution).

References

Peter Flynn
Electronic Publishing Unit, University College Cork Cork, Ireland Phone: +353 21 490 2609 pflynn (at) ucc dot ie http://research.ucc.ie/profiles/H505/pflynn