A university thesis class: Automation and its pitfalls

Peter Flynn

Abstract

Despite the number of thesis classes available, there are always features that can better be met by writing Yet Another Thesis Class. There are also variations in the documentation, assumptions about preloaded packages, and the extent to which the author can modify the layout.

In the case of UCC, the official requirements were very simple, avoiding the tendency to over-specify detail. The class was required to be usable in any discipline, so preloaded packages were kept to a minimum.

The class attempted to automate as much of the front matter as possible, based on the class options, to avoid unwanted variations in the metadata; and to ensure that the required 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 now 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* [3]. Recommendations on how to actually *write* a thesis are covered in the companion paper to this [5] and in an earlier summary [2].

1 Yet Another Thesis Class

A recent retrieval from CTAN[1] shows 42 thesis or thesis-related packages currently 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 elsewhere 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

---

[1] Scripted with the assistance of *dog*, *tidy*, and *lxprintf*, thanks to the robustness of the directory and link structure implemented by the CTAN team.

adfathesis Australian Defence Force Academy thesis format.
afthesis Air Force Institute of Technology thesis class.
beamer2thesis Thesis presentations using beamer.
classicthesis A 'classically styled' thesis package.
ebsthesis Typesetting theses for economics.
elteikthesis Thesis class for ELTE University Informatics wing.
fbithesis Computer Science thesis class for University of Dortmund.
gatech-thesis Georgia Institute of Technology thesis class.
hepthesis A class for academic reports, especially PhD theses.
jasthesis A 'standard' thesis class.
jkthesis Document class for formatting a thesis.
msu-thesis Class for Michigan State University Master's and PhD theses.
muthesis Classes for University of Manchester Dept of Computer Science.
pitthesis Document class for University of Pittsburgh theses.
pittetd* Electronic Theses and Dissertations at Pitt
psu-thesis Package for writing a thesis at Penn State University.
ryethesis Class for Ryerson Unversity Graduate School requirements.
sapthesis Typeset theses for Sapienza University, Rome.
seuthesis LaTeX template for theses at Southeastern University.
suthesis Typeset a Stanford University thesis.
thesis Typeset thesis.
thesis-titlepage-fhac Little style to create a standard titlepage for diploma thesis.

thuthesis Thesis template for Tsinghua University.
uafthesis Document class for theses at University of Alaska Fairbanks.
ucdavisthesis A thesis/dissertation class for University of California Davis.
ucthesis University of California thesis format.
ucthesis209 LaTeX 2.09 document style for UC theses.
uhthesis University of Houston thesis document style.
uiucthesis UIUC thesis class.
umich-thesis University of Michigan Thesis LaTeX class.
umthesis Dissertations at the University of Michigan.
unamthesis Style for Universidad Nacional Autónoma de México theses.
unswthesis UNSW theses.
uothesis Class for dissertations and theses at the University of Oregon.
uowthesis Document class for dissertations at the University of Wollongong.
uscthesis USC thesis style for LaTeX 2.09.
utorontothesis A thesis class definition for University of Toronto.
utthesis Thesis package for the University of Texas at Austin.
ut-thesis University of Toronto thesis style.
uwthesis University of Washington thesis class.
uwthesis209 LaTeX 2.09 style for University of Washington theses.
withesis University of Wisconsin-Madison Thesis LaTeX Class.
york-thesis A thesis class file for York University, Toronto.

* There are potentially many other non-CTAN classes in Electronic Thesis and Dissertation (ETD) sites worldwide.

Figure 1: Thesis and thesis-related packages available from CTAN as of May 2012

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.[2] 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

Human Interfaces
to Structured Documents

The usability of software
for authoring and editing

Peter Flynn
MA
102229277

NATIONAL UNIVERSITY OF IRELAND, CORK

SCHOOL OF APPLIED PSYCHOLOGY
DEPARTMENT OF APPLIED PSYCHOLOGY

**Thesis submitted for the degree of
Doctor of Philosophy**

31 December 2012

Supervisor: Dr Jurek Kirakowski
Head of Department/School: Prof John Groeger

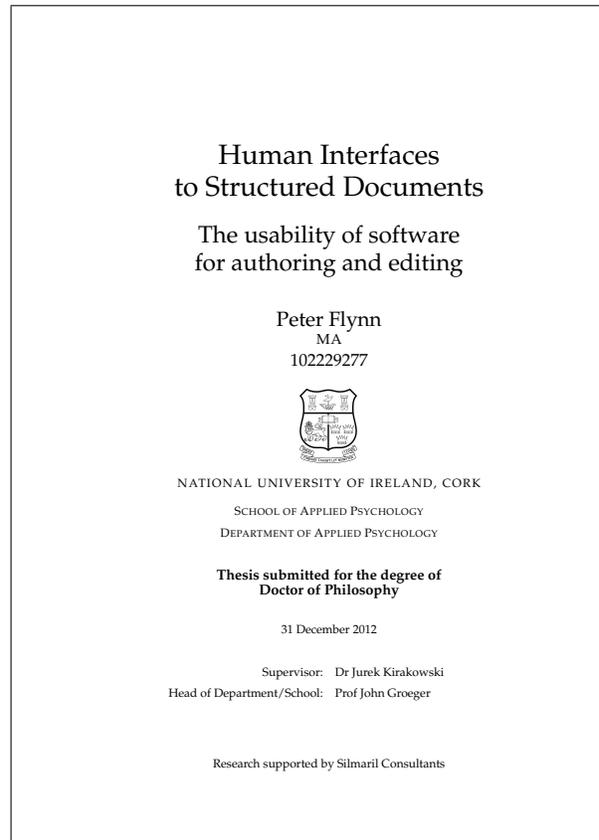Research supported by Silmaril Consultants

Figure 2: Title page layout for the UCC thesis

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[3] 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 page is becoming a condition of some funding agencies. There is provision for the numbering of multi-volume theses with a `\volume` command, but

---

[2] Users later reported the most compelling reason was that the report class supports the `abstract` environment, which the book class does not.

[3] 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.

```
\documentclass[apsych,phd,12pt]{uccthesis}
\usepackage{palatino}
\begin{document}
\title{Human Interfaces
     \\to Structured Documents}
\subtitle{The usability of software
        \\for authoring and editing}
\author[102229277]{Peter Flynn}
\qualifications{MA}
\department{Applied Psychology}
\professor{Prof John Groeger}
\supervisor{Dr Jurek Kirakowski}
\sponsor{Silmaril Consultants}
\date{31 December 2012}
\maketitle
...
\end{document}
```

Figure 3: Commands for the title page metadata

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
    \else{\large\@subtitle\par}\fi
\if\relax\@professor You need to give the
    name of your head of discipline
    \else\@professor\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 oneside 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:

```
\iflof\listoffigures\fi
\iflot\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 nolot and nolof 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 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).

In addition, 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 prelims.

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 [7] 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 (foredge) 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 $1^1/_2$ 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 `comp.text.tex`, 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-typeset publication like a book, and does not appear to benefit greatly from justification; and partly because raggedright setting improves readability on a page width with relatively long lines.

However, as `\raggedright` also turns off paragraph indentation, 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, `justified` and `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 `\leftmark` and `\rightmark` *without* the capitalization used in the default classes (this proved remarkably difficult to defeat), and *with* the use of an `\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;[4] 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 `\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

The LaTeX default for quotations is unusual, in that it does not make the type size smaller, nor does it

---

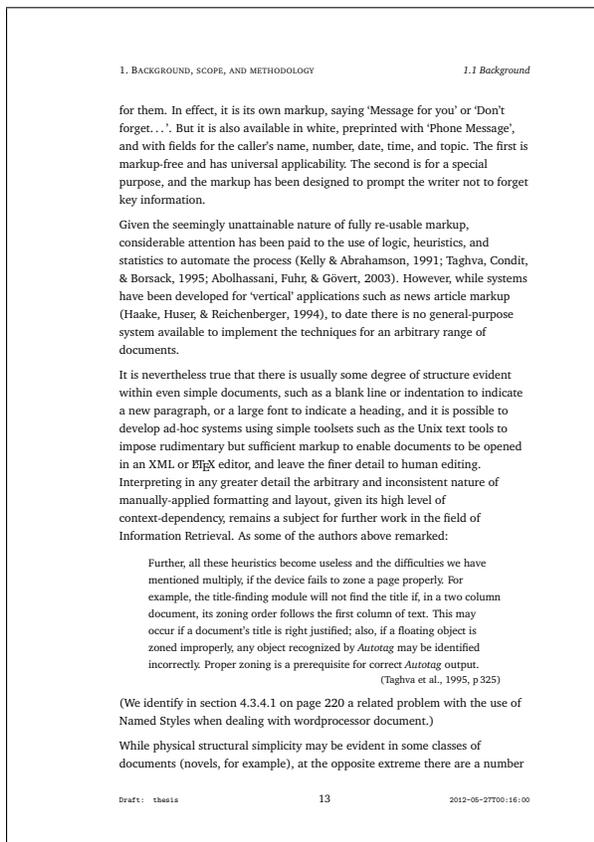[4] This has the (possibly) useful side-effect of allowing the `[H]` positional specifier.

Figure 4: Default page layout for the UCC thesis

defeat indentation on the first line, both of which are established conventions.[5] In academic work, a block quotation also usually requires a citation.

To implement this, the `quotation` environment was modified to change the size and start with a `\noindent`; and to take an argument, the BibTeX key of the cited passage. This enables the quotation to be set with a right-aligned citation immediately below. Although such a citation would be compulsory, it is currently defined as optional in order for its presence to be tested, and to allow for uncited use where the context already makes the origin obvious. However, if an optional argument such as this is to contain page or chapter references, which are themselves optional to the `\cite` command, additional armour is required:

`\begin{quotation}[{[p.36]{smith92}}]`

In response to user requests for an Epigraph at the start of chapters, an `epigraph` environment was created, in the same way as the modified `quotation` above. This has two arguments, however: a compul-

---

[5] I have never been able to find a use for the `quote` environment, so the use of `quotation` is recommended to authors.

sory one for the BibTeX key, but if the citation is to be informal and non-rigorous, that argument can be empty, and the optional argument can be used, for example:

`\begin{epigraph}[Popular saying]{}`

In both cases this is cumbersome and needs regularising, so a future version will probably use the `xargs` package to handle the additional metadata.

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 [6] 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 we have a complex and overlapping transitional 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 command. This also avoids misspellings, and the unfortunate 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 important.

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 option now both sets the correct naming *and* presets 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 `.ins` 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 [4], 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 IEEETR 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

[1] Robin Fairbairns, editor. *TeX Frequently Asked Questions.* UK TeX Users Group, Cambridge, UK, Mar 2012.

[2] Peter Flynn. Sorry, Professor, the dog ate my thesis: How to expect the unexpected when using LaTeX. In *Living and Working with LaTeX.* London Mathematical Society, Oct 2006.

[3] Peter Flynn. Typographers' Inn. *TUGboat*, 33(1):8–10, 2012.

[4] Peter Flynn. UCC Thesis Document Class. http://research.ucc.ie/latex/#uccthesis, Apr 2012.

[5] David Latchman. Preparing your thesis in LaTeX. In *TUG 2012*, Boston, MA, Jul 2012. TeX Users Group.

[6] Frank Mittelbach, Michel Goossens, Johannes Braams, David Carlisle, and Chris Rowley. *The LaTeX Companion.* Addison-Wesley/Pearson Education, Boston, MA, 2nd edition, 2004.

[7] University College Cork. Thesis Submission: Layout of Thesis. http://www.ucc.ie/en/graduatestudies/current/thesis/Layout/, Nov 2009.

⋄ 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