**General Advice**

Obviously, having a close look at a number of good thesis reports is a good idea. There are no firm rules on how to write a thesis, and there is certainly a lot of advise available. I'll try to concentrate on a few main points (which tend to apply to a lot of technical writing, not only to theses but also to conference papers).

**Structure and Coherence**

Make sure your thesis is well structured, that each major section does what it is supposed to do, and that the whole thing hangs together. The basic structure is often as follows (but other structures are possible). In particular, don't think you need to have exactly as many major sections or chapters as there below list implies; sometimes it makes sense to merge things, sometimes it makes sense to move things (e.g., the literature review is in many papers deferred until after the results), sometimes it makes sense to split a logical part into several individual sections. Use common sense!

**Title**

Use a descriptive title for your work. Don't use a title that promises more than you'll deliver, don't use a title that implies something different from what you've done. (The focus of a thesis often shifts in the course of a year, don't be afraid to adjust the title, in consultation with your supervisor.)

**Abstract**

A short (1–3 paragraphs) summary of the work. Should state the problem, major assumptions, basic idea of solution, results. Avoid non-standard terms and acronyms. The abstract must be able to be read completely on its own, detached from any other work (e.g., in collections of paper abstracts). Don't use reference in an abstract.

**Introduction**

Introduce the problem (gently!) Try to give the reader an appreciation of the difficulty, and an idea of how you will go about it. It's like the overture of an opera: it plays on all the relevant themes.

Make sure you clearly state the vision/aims of your work, what problem you are trying to solve, and why it is important. While the introduction is the part that is read first (ignoring title and abstract) it is usually best written last (when you actually know what you have really achieved). Remember, it's the first thing that is being read, and will have a major influence on the how the reader approaches your work. If you bore them now, you've most likely lost them already. If you make outrageous claims pretend to solve the world's problems, etc, you're likely fighting an uphill battle later on. Also, make sure you pick up any threads spun in the introduction later on, to ensure that the reader things they get what they have been promised. Don't create an expectation that you'll deliver more than you actually do. Remember, the reader may be your marker (of a thesis) or referee (of a paper), and you don't want to piss them off.

**Exposition of problem**

The basic problem should have been stated in the intro, here is the place to go into detail. (Whether this is the same section or a different one is besides the point.) Make it clear you know what you are talking about (and this includes being complete, don't jump right into things, give the reader a chance to follow). Give a thorough and complete discussion of the problem, enough so an educated reader whose speciality is outside yours can appreciate that you're trying to attack an interesting problem, and also appreciates what's interesting about it. Btw, don't call this section“exposition of the problem”, or you'll be immediately exposed as someone who can only follow recipes. Same applies to the next bit.

**Literature review**

(often called “related work”)

This is really important. If you cannot demonstrate that you know, and understand, what others have done, you only demonstrate that you're clueless. For an undergraduate thesis this, together with a thorough understanding of the problem, should be the result of the first session's work. It is an unfortunate fact that many students do very little work during the first session of their thesis. It usually shows here (and is usually reflected in their mark). Don't think you can fool your thesis supervisor/assessor. And don't even dream about fooling the referee of a paper. If you haven't done your homework here, it's probably not worth going any further.

In this part you demonstrate that you are aware of what's going on in the field, and how it relates to your particular problem. In a thesis (unlike a conference paper) it may be ok to repeat work that has already been done elsewhere (usually in somewhat different circumstances). Be open, and explain why what you're doing is still worthwhile. In the more normal case that you're doing something that hasn't already been done, convince that reader that this is actually the case. One of the less convincing arguments goes along the line “a Google search on `frying giblets on StrongARM-driven toasters' didn't turn up anything”. You might as well pack up here. The way to convince the reader that your work hasn't been done before is to explain what has been done, what's different about what has been done, and, if you're good, why it hasn't been done already. There is always related work, and the more vague you are about it, the more obvious it is that you haven't done your homework. (And, no, looking at all the Google hits isn't enough.)

Sometimes some relevant background work is quite old; the discipline goes in cycles and it isn't all that infrequent that people rediscover things that have been done 30 years ago. In such case please note that the language has changed a fair bit in the meantime. You're not doing your reader a favour of reporting an old paper's findings in that paper's language (and in the informed reader's mind you'll raise the suspicion that you don't understand what's going on). Talk about the work of the paper in contemporary language. This makes it easier to compare to other work, including yours.

**Design of your solution**

Having explained the problem, and what others have done in similar situations, now explain your approach. Again, give a general overview of your design first, and then go into detail. Make sure that the document (particularly a thesis) is self-contained: It should be possible for a reader familiar with the general area to understand your design. Again, be forthright about the limitations of your design. Also, make sure you justify any shortcuts/limitations convincingly.

**Implementation**

In many (not all cases) there is a clear difference between the general approach (design) and its implementation in your particular circumstances. The design may be more general than what you can do given time and resources. Or you have developed a general design, and are now implementing a prototype on particular hardware. Give all required details. It should be possible to understand all this without referring to the source code.

This will, in general, include extracts of actual algorithms and hardware components used. Don't list pages of C code, an electronic copy of the source will accompany the submission and should be available to the marker, so there's no point in killing extra trees to put it into the report.

Make sure you describe your implementation in enough detail. Someone who has nothing else but your thesis report to go by should be able to repeat your work, and arrive at essentially the same implementation. Reproducibility is an important component of scientific work. Also, clearly state the limitations of your implementation, and justify them.

**Experiments**

A thesis almost always has an experimental part, typically some comparison to other approaches. This is usually its weakest part. Many students debug their code less than a week prior to the submission deadline (typical indication of having started too late) which makes it hopeless to do any real benchmarking. Benchmarking takes time, for running the experiments, but also for thinking them up in the first place, and for analysing the results.

Probably the majority of theses I mark is really deficient in this part, typically for lack of attention (often resulting from a late start). Think about what makes sense to measure, what you want to learn from your measurements. Think about what is really the relevant contribution of your thesis, and how you can prove that you have achieved your goals. Think about what you can measure in order to get a good insight into the performance of various aspects of your design, how you can distinguish between systematic and accidental effects, how you can convince yourself that your results are right. If you get surprising results, don't just say “surprise, surprise, performance isn't as good as hoped”. Find out why. Surprises without explanation indicate either that you are clueless about what's going on, or that you have made a mistake (most likely both). Unconvincing results tend to imply unconvincing marks. (Of course, this could be avoided if the results were available more than a couple of days prior to the thesis deadline.)

It is amazing how few students have even the faintest clue of the most basic statistics and their use. Measurements always have statistical (sampling) errors. Owing to the deterministic nature of simulations these are sometimes very small, as disturbing factors can be designed. However, the reader should be given an indication of how statistically significant the results are. This is done by providing at least a standard deviation in addition to averages. Whenever the results of several runs are averaged, a standard deviation can (and must) be supplied. After all, you average to reduce statistical errors.

The reproducibility argument applies here just as much as for the implementation. Give enough detail on what you measure, and how you measure it, so that someone who has your implementation (but not your test code) or has re-done your implementation independently, should be able to repeat your measurements and arrive at essentially the same results. I read many theses which contain results which seem outright wrong. In most cases not enough detail is provided to allow me to pinpoint the likely source of the error. In many cases the cause is systematic errors resulting from an incorrect measurement technique. If it seems wrong, and the text doesn't convince me that it is not wrong, I will assume that it is wrong.

**Discussion**

Discuss and explain your results. Show how they support your thesis (or, if they don't, come up with a damned good reason real quick). It is important to separate objective facts clearly from their discussion (which is bound to contain subjective opinion). If the reader doesn't understand your results, you probably do neither. And this will be reflected in the assessment.

**Conclusion**

Don't leave it at the discussion: discuss what you/we can learn from the results. Draw some real conclusions. Separate discussion/interpretation of the results clearly from the conclusions you draw from them. (So-called “conclusion creep” tends to upset reviewers. It means surrendering your scientific objectivity.) Identify all shortcomings/limitations of your work, and discuss how they could be fixed (“future work”).

I repeat: don't stick slavishly to this structure. Also, remember that the thesis must be:

* honest, stating clearly all limitations;
* self-contained—don't write just for the locals, don't assume that the reader has read the same literature as you, don't let the reader work out the details for themselves.

**Things Students Frequently Get Wrong**

This is my list of things that people most frequently get wrong, listed in no particular order, except that the most annoying ones are at the top. From now on I expect you to consult this list, and fix up your prose before getting your draft to me. If you don't, you risk having it returned to you unread.

**Spelling**

There is **no** excuse for presenting a draft that hasn't gone through a spell checker. If you're too lazy to do this, then I'm too lazy to read your work. Period.

**Apostrophes**

* Apostrophes are used to mark possessions and attributions. Like the thread's priority. Note that there is no apostrophe in the case of the personal pronouns he, she and it: the thread used up its time slice. Bob's pretty clear about this one!
* Apostrophes are used for omissions. Like I can't, or it's time. Note that these are generally not used in formal prose (such as reports and papers) as they sound colloquial.

That's pretty much it. Almost any other use of apostrophes is wrong. But keep in mind that apostrophes are actually useful, so don't leave them off completely!

**Capitalization**

Don't Randomly capitalize Words. Looks Ridiculous, doesn't it? Capitals are used for:

* words beginning a sentence;
* names (proper nouns)
* acronyms
* certain types of words in high-level headings.

Capitals shouldn't be used for definitions (see below), and even less without any obvious reason.

**Commas**

This is probably what I get most often wrong myself (partially because of totally different rules in German and English).

[Commas] have a vital role to play in longer sentences, separating information into readable units, and guiding the reader as to the relationship between phrases and items in a series.

1. A single comma to ensure correct reading of sentences which start with a longish introductory element: Before the close of the last Ice Age, Tasmania was joined to the mainland of Australia.
2. Pairs of commas help in the middle of a sentence to set off any string of words which is either a parenthesis or in apposition to whatever went before: The desert trees, casuarinas and acacias, were sprouting new green needles. (Apposition) The dead canyons, all nature in them reduced to desiccation, came alive with the sound of rain slithering down the crevasses. (Parenthesis)

Note that a pair of [em-]dashes could have been used instead of commas with the parenthesis, in both formal and informal writing.

1. Sets of commas are a means of separating:
   1. strings of predicative adjectives, as in: It looks big, bold, enticing.
   2. items in a series, as in: The billabongs at sunset drew flocks of galahs, gang-gangs, budgerigars and cockatoos of all kinds.

**Colons (and lists)**

Colons are used to indicate that examples or specific details are to come:

* If a full sentence follows after the colon, the word after the colon is capitalized.
* If the sentence continues, the word after the colon isn't capitalized, unless it would be capitalized anyway, or it's a slogan or a motto.
* Alternatively, the examples or details may be given as several complete sentences, in which case they should start in a new paragraph.
* Bullet lists or enumerated lists set as paragraphs (so-called vertical lists) are introduced by a colon. Regarding their capitalization and punctuation, there are three cases to distinguish:
  1. If the list items are short (few words or simple phrase) and without internal punctuation, their first word is not capitalized and no punctuation is used (except possibly at the end of the last one).
  2. If the list items contain internal punctuation, but are not all complete sentences, then their first word is not capitalized and each item is terminated by a semicolon (except the last, which is terminated by a full stop).
  3. If the list items are each (one or more) complete sentences, they are written as such: first word capitalized, and each terminated with a full stop.

**Period (full stop)**

The period is used to end a sentence, as well to identify an abbreviation. The two are actually distinguished in type-setting: a period designating an abbreviation (and nothing else) is followed by a normal inter-word space, while a period at the end of a sentence is followed by a longer inter-sentence space. Many formatters (incl. web browsers) automatically produce an inter-sentence space after each period; this is wrong if it is not the actual end of the sentence, and must be overwritten by forcing an inter-word space. LaTeX does it right for abbreviations ending in capitals, but otherwise the period must be followed by a backslash.

**Quotation marks**

* Quotation marks are for quotations. They are not to introduce new terms, they are for quoting someone/something, e.g., called “giblet” in [Bloe 99]. They are also used for irony (a small subset of what you'd use a smiley for), but this is rare in technical prose. E.g., Its “outstanding” performance made the system useless except for toy applications. Note that not all humour is irony!
* Quotation marks are normally double ticks. Single quotation marks are used only for quotations inside quotations.
* The begin and end quotation marks are different, as in the above examples.
* If the quotation extends over several paragraphs, it should start a new paragraph, and the begin-quotation marks must be repeated at the beginning of each paragraph. However, in such a case it is much better to set off the quotation by indentation (as with the LaTeX quotation environment) and use no quotation marks at all.
* There is some confusion about other punctuation. There are two basic cases:
  1. The quotation ends with an exclamation or question mark. In this case the mark goes inside the quotation, and no period follows, even if the quotation marks the end of the sentence.
  2. Otherwise, if the quotation is at the end of the sentence, put the period inside the quotation marks if the quotation would normally end in a period, otherwise put it outside.

**Definitions/introductions of new terms**

Use italics when introducing new terms. This makes it easy for the reader to find the definition again, particularly when not having the time to read the paper in one shot. Do not capitalize words when they are introduced (unless you'd normally capitalize them). Do not put them in quotations marks (see above).

Note: The LaTeX command \it is almost always the wrong way to use italics. Use the LaTeX \em command (or, better, the LaTeX2e \emph command) which will handle nested emphasis correctly.

**Acronyms and Initialisms**

Technically the difference between the two is that acronyms you pronounce as a word (NICTA) while initialisms are pronounced as individual letters (UNSW). The distinction is hardly ever made and both are generally lumped under the general term of “acronym”, as in the reminder of this document.

Properly define all acronyms on first use. Don't introduce too many acronyms, and use standardized ones whenever possible.

Don't introduce acronyms in headings! If a term for which you want to use an acronym appears first in a heading, define the acronym on the next appearance (the first one in paragraph mode). Also, don't introduce an acronym which is then not used for a long time. In such a case it is also better to defer the introduction of the acronym.

It sometimes happens that an acronym is introduced and used more-or-less heavily in an early part of a thesis or paper, is then not used for a long time, until it is used again towards the end. In such a case, remember that the reader may not read the whole thesis or paper in one go, and may have forgotten what the acronym stands for. In such a case (at least if it's an acronym that isn't widely used) it's better to re-state the definition when the term starts appearing again. A very gentle way to remind the reader of the meaning of an acronym is to use it just after its expanded form in a way that makes its meaning obvious. **Basic rule: Be nice to the reader!**

Acronyms are normally all upper case, however, in our discipline mixed case acronyms have become very common (e.g., QoS for quality of service). They should still start with a capital letter. Acronyms are almost never all lower case. The one exception is units of measurement (e.g. loc for lines of code, although journals would normally use LOC for this). If you find an all-capital acronym too imposing consider using SMALLCAPS. However, remember to be consistent: if you decide to use a special font for something like a specific acronym, make sure you always use the same font for the thing. Also, don't go overboard with fonts.

What's the plural of CPU? CPUs or CPU's? If you look at journals employing professional typesetters you'll find that the answer is clear: CPUs is a plural while CPU's indicates a possession or attribution. Example: Of the system's two CPUs, only one was operational. The second CPU's power supply had been disconnected.

A special case of this is acronyms ending in s, e.g. OS. I have found a (seemingly authoritative) reference which claims that in this case you need an apostrophe. I strongly recommend OSes over OS's for the plural, in order to clearly distinguish it from the possessive case. Note that UNIX is traditionally pluralized as UNIXen, like oxen, but I think that's tradition rather than a grammatical rule.

In rare cases using no apostrophe for the plural might create confusions with mixed-case acronyms. In that case use an apostrophe if you really think that it improves clarity.

**Units of measurement and their prefixes**

I regularly see “KB”, “kb”, “Kb” all (intending to) refer to the same thing (1024 bytes), all wrong. Specifically: KB would be kelvin bytes, presumably a unit of information temperature, I don't think anyone has found a use for that unit yet; kb would be kilo bits, which these days is probably only used as part of a unit of bandwidth for really slow links; Kb would be the useless kelvin bits. So, bit is “b”, byte is “B”, kilo is “k”, not “K”. Furthermore, the unit prefixes “k”, “M”, “G”, etc. strictly refer to powers of ten, i.e., 103, 106, 109. In IT contexts they usually (but not always) stand for powers of two, i.e., 210, 220, 230. This is of course confusing. If you think it is not, can you tell me whether a Gigabit Ethernet is supposed to have a bandwidth of 109b/s or 230b/s?

There are in fact proper SI prefixes for power-of-two multiples: “Ki”, “Mi”, “Gi”, etc. These are, unfortunately, not widely used yet, but are becoming more popular. Use them systematically!

**Headings**

Capitalize or not? Generally speaking, only top-level or, for larger documents, second-level section headings should be capitalized. For other headings capitalize the first word (of course), but otherwise nothing you would not capitalize in normal text.

**Footnotes**

First rule: use them sparingly. Humanities people love them, scientists and engineers use them rarely. You are the latter. If you have more than an average of one footnote per page consider changing your degree.

Second rule: Footnotes should be fair-dinkum sentences, able to be read by themselves. A footnote like 5kB is a definitive no-no. Something like #define'd to 5kB. is very bad. Good is The buffer size is defined to be 5KiB.

**Hyphens, en-dashes and em-dashes**

These are three kinds of dashes used in text:

* The hyphen (LaTeX “-”, Unicode “-”, HTML “-”, plain ASCII “-”) is used for hyphenation (breaking words at the end of line), as well as for compound words. The former you never need to do explicitly, LaTeX does it for you. (You may help LaTeX in difficult cases, as in hy\-phe\-nate.) The hyphen is generally to be used to overwrite the default binding of English. Attributes preceding a noun are by default bound right to left in English, which can produce an incorrect meaning. For example, single address space is right, as address qualifies space, and single qualifies address space. However, if this is itself used to qualify another noun, it needs hyphens: single-address-space operating system. Without the hyphens, operating system would be qualified by space, and a space operating system is something different from what we are concerned with. Hyphens may not only be required by adjectives qualifying a noun: The syscall requires the invoking process to be root-owned. Finally there are compound terms which use a hyphen, such as know-how. Use them sparingly!
* The en-dash “–” (LaTeX “–”, Unicode “<Compose>–.”, HTML entity reference “&ndash;”, plain ASCII “-” as for the hyphen) is used for ranges, e.g. RAM sizes of 0.5–64GiB are supported. The en-dash is used between single words or numbers without surrounding space, but has surrounding space if it is between items that have internal space. Example: during the time of 12 March – 15 May. The HTML entity reference is &ndash;, in plain ASCII use - as for the hyphen.
* the em-dash “—” (LaTeX “—”, Unicode “<Compose>—”, HTML entity reference “&mdash;”, plain ASCII “–”) is used as a separator, somewhat similar to a semicolon. Note that LaTeX runs it right into the adjacent words, apparently that's what the rules say. If you don't like it, you can use \,—\, to force some space.

**Passive voice**

Avoid the excessive use of the passive voice, it is considered poor style (partially because it creates the impression that you are not really taking responsibility for what you've written). If 1/3 or more of your sentences use passive voice, your prose is poor.

Even worse is what I frequently see in undergraduate theses: people using passive voice as a way to avoid the first person, e.g., “a suitable protocol was designed to cope with that situation”, when the student means to say that they designed the protocol. This might be a case of shyness, but it comes across as trying to avoid responsibility for one's actions. At best it leaves the reader puzzling who had actually done the work. Show through your writing that you assume ownership and responsibility for what you have done, and make it always perfectly clear what you have contributed and what came from others!

**Like vs. such as**

When you are referring to a set, the members of which have in common a given characteristic, and you wish to give an example that is a member of that set, you should use such as. When you are referring to a set that does not include your example, but that contains members that resemble your example, you should use like. Examples: Students, such as those at TUM, sometimes are having fun. Sometimes they behave like children with a new toy.

**Spaces**

**Before the colons in definition lists**

Does not belong.

**Before parantheses**

Why should an opening parenthesis be glued to the preceding word? No matter whether this introduces an acronym or a non-essential remark, the outside of the parentheses like air to breath.

**Citations**

Whether to use numeric or alphabetic references isn't all that important (unless prescribed by a conference or journal), but alphabetic tends to be more readable. Independent of citation style, the following rules should be followed:

* Use the LaTeX cite package. It doesn't give you additional commands, but it fixes a few quirks in LaTeX. Among others it automatically sorts multiple citations, and it correctly spaces the angular brackets (if you use the \cite command without leading white space).
* Citing several papers at one point should be done with a single \cite command. For example, use gives good results\cite{Bloe\_99, Jay\_87}, resulting in gives good results [3,5]. Do not use gives good results\cite{Bloe\_99}\cite{Jay\_87} which produces the ugly gives good results [3][5]. Also, note that there is no space between the \cite command and the preceding word, LaTeX (with the cite package) does the spacing correctly.
* Avoid citations of the kind [1] thinks that threads are cool, but [2] argues that they suck. This works a bit better if using alphanumeric citation labels. Better, though, use the author's names: Joe and Bloe [1] think that threads are cool, but O'Neill et al. [2] argue that they suck. Except that (of course) you'll never use such colloquialisms in formal prose. :-)
* BibTeX is a great tool, but you need to know how to use it. A regular trap is to forget that TeX knows more about typesetting than you do. So, for example, it changes the case of words in the title. If your title contains acronyms and proper names (most do), they tend to get down-cased. Any such words which should not have their case changed should be put into braces, e.g., {The {Mungi} {OS} and its Use in Merry-Go-Round Seat Allocation}.
* In citations don't abuse the category technical report. I see this happen a lot: people cite just about anything that hasn't been published in a journal or conferences as a TR. This is wrong! The concept of a TR is actually fairly well defined:
  + A TR is published in some sort. This is generally as part of a formal TR series of some institution, in hardcopy or on the web or both. (They aren't always called “technical report”, other common names are “research report”, “technical memorandum”, <institution> report” etc.) The publication (i.e. availability outside) is essential, otherwise it's at best an internal report.
  + A TR has a number (absolutely!), an institution (publisher), a date (month and year at least) and a publisher's address (besides all the other stuff bibentries have).

If your document doesn't have these features, it's not a TR. It's probably better categorized as a working paper. Even then it has a date and an institution address.

* Citing web pages is often unavoidable (but also often a sign of laziness). When citing web pages be aware that they may only be short-lived. Consider whether the reference will be of any use to the reader at all if the link is broken. Or whether your whole document only has a use-by date a few months past writing.
* Any cited document, whatever it may be, as a few features:
  + Date. Absolutely. If you don't have a date you're lazy.
  + Author/organization/creator/person responsible for contents. If you don't have it, see above.
  + Whatever information the reader needs to find that document. In most BibTeX entry types these are clearly identified as mandatory fields. Mandatory means that they aren't optional. Don't pretend they are. For a working paper these might be the contact details of the author.

**LaTeX**

* To represent URLs, don't just use \texttt{url} (which causes problems with the tilde character) or \verb|url| (which tends to produce vastly overfull lines). Instead use the command \url{url}, available with the url package. This will, by default, typeset the string in TTY font, but that can be changed to the more readable \urlstyle{sf}.
* Don't use bitmap formats for figures (nor bitmaps converted to EPS). They almost always lead to poor results.

**Miscellaneous**

* i.e. (id est, Latin for that is) is written with two full stops and no comma (which implies that in LaTeX you normally need to follow the second full stop with a backslash).

**Formalities**

This should go without saying, but, apparently, doesn't:

* every document (even an early draft) has a title
* every document (even an early draft) has an author (or several)
* every document (even an early draft, except a manuscript submitted for publication) has a date
* every document (even an early draft) has page numbers. Only exception is that camera-ready conference papers often are required to be submitted without page numbers. This shouldn't stop you from using page numbers in drafts, as well as in submissions for reviewing (reduces the chance of a reviewer messing up your paper while reading).

Additional information can be found at <http://cgi.cse.unsw.edu.au/~kevine/thesisguide.html>.