How To Write A Dissertation

or

Bedtime Reading For People Who Do Not Have Time To Sleep

To The Candidate:

So, you are preparing to write a Ph.D. dissertation in an experimental area of Computer Science. Unless you have written many formal documents before, you are in for a surprise: it's difficult! There are two possible paths to success:

- Planning Ahead.
  Few take this path. The few who do leave the University so quickly that they are hardly noticed. If you want to make a lasting impression and have a long career as a graduate student, do not choose it.

- Perseverance.
  All you really have to do is outlast your doctoral committee. The good news is that they are much older than you, so you can guess who will eventually expire first. The bad news is that they are more practiced at this game (after all, they persevered in the face of their doctoral committee, didn't they?).

Here are a few guidelines that may help you when you finally get serious about writing. The list goes on forever; you probably won't want to read it all at once. But, please read it before you write anything.

The General Idea:

1. A thesis is a hypothesis or conjecture.
2. A PhD dissertation is a lengthy, formal document that argues in defense of a particular thesis. (So many people use the term "thesis" to refer to the document that a current dictionary now includes it as the third meaning of "thesis").
3. Two important adjectives used to describe a dissertation are "original" and "substantial." The research performed to support a thesis must be both, and the dissertation must show it to be so. In particular, a dissertation highlights original contributions.

4. The scientific method means starting with a hypothesis and then collecting evidence to support or deny it. Before one can write a dissertation defending a particular thesis, one must collect evidence that supports it. Thus, the most difficult aspect of writing a dissertation consists of organizing the evidence and associated discussions into a coherent form.
5. The essence of a dissertation is critical thinking, not experimental data. Analysis and concepts form the heart of the work.
6. A dissertation concentrates on principles: it states the lessons learned, and not merely the facts behind them.
7. In general, every statement in a dissertation must be supported either by a reference to published scientific literature or by original work. Moreover, a dissertation does not repeat the details of critical thinking and analysis found in published sources; it uses the results as fact and refers the reader to the source for further details.
8. Each sentence in a dissertation must be complete and correct in a grammatical sense. Moreover, a dissertation must satisfy the stringent rules of formal grammar (e.g., no contractions, no colloquialisms, no slurs, no undefined technical jargon, no hidden jokes, and no slang, even when such terms or phrases are in common use in the spoken language). Indeed, the writing in a dissertation must be crystal clear. Shades of meaning matter; the terminology and prose must make fine distinctions. The words must convey exactly the meaning intended, nothing more and nothing less.
9. Each statement in a dissertation must be correct and defensible in a logical and scientific sense. Moreover, the discussions in a dissertation must satisfy the most stringent rules of logic applied to mathematics and science.

What One Should Learn From the Exercise:

1. All scientists need to communicate discoveries; the PhD dissertation provides training for communication with other scientists.
2. Writing a dissertation requires a student to think deeply, to organize technical discussion, to muster arguments that will convince other scientists, and to follow rules for rigorous, formal presentation of the arguments and discussion.

A Rule of Thumb:

Good writing is essential in a dissertation. However, good writing cannot compensate for a paucity of ideas or concepts. Quite the contrary, a clear presentation always exposes weaknesses.

Definitions and Terminology:

1. Each technical term used in a dissertation must be defined either by a reference to a previously published definition (for standard terms with their usual meaning) or by a precise, unambiguous definition that appears before the term is used (for a new term or a standard term used in an unusual way).
2. Each term should be used in one and only one way throughout the dissertation.
3. The easiest way to avoid a long series of definitions is to include a statement: "the terminology used throughout this document follows that given in [CITATION]." Then, only define exceptions.

4. The introductory chapter can give the intuition (i.e., informal definitions) of terms provided they are defined more precisely later.

Terms and Phrases To Avoid:

- adverbs
  Mostly, they are very often overly used. Use strong words instead. For example, one could say, "Writers abuse adverbs."
- jokes or puns
  They have no place in a formal document.
- "bad", "good", "nice", "terrible", "stupid"
  A scientific dissertation does not make moral judgements. Use "incorrect/correct" to refer to factual correctness or errors. Use precise words or phrases to assess quality (e.g., "method A requires less computation than method B"). In general, one should avoid all qualitative judgements.
- "true", "pure",
  In the sense of "good" (it is judgmental).
- "perfect"
  Nothing is.
- "an ideal solution"
  You're judging again.
- "today", "modern times"
  Today is tomorrow's yesterday.
- "soon"
  How soon? Later tonight? Next decade?
- "we were surprised to learn..."
  Even if you were, so what?
- "seems", "seemingly",
  It doesn't matter how something appears;
- "would seem to show"
  all that matters are the facts.
- "in terms of"
  usually vague
- "based on", "X-based", "as the basis of"
  careful; can be vague
- "different"
  Does not mean "various"; different than what?
- "in light of"
  colloquial
- "lots of"
  vague & colloquial
- "kind of"
  vague & colloquial
- "type of"
  vague & colloquial

- "something like"
  vague & colloquial
- "just about"
  vague & colloquial
- "number of"
  vague; do you mean "some", "many", or "most"? A quantitative statement is preferable.
- "due to"
  colloquial
- "probably"
  only if you know the statistical probability (if you do, state it quantitatively
- "obviously, clearly"
  be careful: obvious/clear to everyone?
- "simple"
  Can have a negative connotation, as in "simpleton"
- "along with"
  Just use "with"
- "actually, really"
  define terms precisely to eliminate the need to clarify
- "the fact that"
  makes it a meta-sentence; rephrase
- "this", "that"
  As in "This causes concern." Reason: "this" can refer to the subject of the previous sentence, the entire previous sentence, the entire previous paragraph, the entire previous section, etc. More important, it can be interpreted in the concrete sense or in the meta-sense. For example, in: *"X does Y. This means ..."* the reader can assume "this" refers to Y or to the fact that X does it. Even when restricted (e.g., "this computation..."), the phrase is weak and often ambiguous.
- "You will read about..."
  The second person has no place in a formal dissertation.
- "I will describe..."
  The first person has no place in a formal dissertation. If self-reference is essential, phrase it as "Section 10 describes..."
- "we" as in "we see that"
  A trap to avoid. Reason: almost any sentence can be written to begin with 'we" because "we" can refer to: the reader and author, the author and advisor, the author and research team, experimental computer scientists, the entire computer science community, the science community, or some other unspecified group.
- "Hopefully, the program..."
  Computer programs don't hope, not unless they implement AI systems. By the way, if you are writing an AI thesis, talk to someone else: AI people have their own system of rules.
- "...a famous researcher..."

It doesn't matter who said it or who did it. In fact, such statements prejudice the reader.

- Be Careful When Using "few, most, all, any, every".
  A dissertation is precise. If a sentence says "Most computer systems contain X", you must be able to defend it. Are you sure you really know the facts? How many computers were built and sold yesterday?
- "must", "always"
  Absolutely?
- "should"
  Who says so?
- "proof", "prove"
  Would a mathematician agree that it's a proof?
- "show"
  Used in the sense of "prove". To "show" something, you need to provide a formal proof.
- "can/may"
  Your mother probably told you the difference.

Voice:

Use active constructions. For example, say "the operating system starts the device" instead of "the device is started by the operating system."

Tense:

Write in the present tense. For example, say "The system writes a page to the disk and then uses the frame..." instead of "The system will use the frame after it wrote the page to disk..."

Define Negation Early:

Example: say "no data block waits on the output queue" instead of "a data block awaiting output is not on the queue."

Grammar and Logic:

Be careful that the subject of each sentence really does what the verb says it does. Saying "Programs must make procedure calls using the X instruction" is not the same as saying "Programs must use the X instruction when they call a procedure." In fact, the first is patently false! Another example: "RPC requires programs to transmit large packets" is not the same as "RPC requires a mechanism that allows programs to transmit large packets."
All computer scientists should know the rules of logic. Unfortunately the rules are more difficult to follow when the language of discourse is English instead of mathematical symbols. For example, the sentence "There is a compiler that translates the N languages by..." means a single compiler exists that handles all the languages, while the sentence "For each of the N languages, there is a compiler that translates..." means that there may be 1 compiler, 2 compilers, or N compilers. When written using mathematical symbols, the difference are obvious because "for all" and "there exists" are reversed.

Focus On Results and Not The People/Circumstances In Which They Were Obtained:

"After working eight hours in the lab that night, we realized..." has no place in the dissertation. It doesn't matter when you realized it or how long you worked to obtain the answer. Another example: "Jim and I arrived at the numbers shown in Table 3 by measuring..." Put an acknowledgement to Jim in the dissertation, but do not include names (even your own) in the main body. You may be tempted to document a long series of experiments that produced nothing or a coincidence that resulted in success. Avoid it completely. In particular, do not document seemingly mystical influences (e.g., "if that cat had not crawled through the hole in the floor, we might not have discovered the power supply error indicator on the network bridge"). Never attribute such events to mystical causes or imply that strange forces may have affected your results. Summary: stick to the plain facts. Describe the results without dwelling on your reactions or events that helped you achieve them.

Avoid Self-Assessment (both praise and criticism):

Both of the following examples are incorrect: "The method outlined in Section 2 represents a major breakthrough in the design of distributed systems because..." "Although the technique in the next section is not earthshaking,..."

References To Extant Work:

One always cites papers, not authors. Thus, one uses a singular verb to refer to a paper even though it has multiple authors. For example "Johnson and Smith [J&S90] reports that..." Avoid the phrase "the authors claim that X". The use of "claim" casts doubt on "X" because it references the authors' thoughts instead of the facts. If you agree "X" is correct, simply state "X" followed by a reference. If one absolutely must reference a paper instead of a result, say "the paper states that..." or "Johnson and Smith [J&S 90] presents evidence that...".

Concept vs. Instance:

A reader can become confused when a concept and an instance of it are blurred. Common examples include: an algorithm and a particular program that implements it, a programming language and a compiler, a general abstraction and its particular implementation in a computer system, a data structure and a particular instance of it in memory.

Terminology For Concepts And Abstractions

When defining the terminology for a concept, be careful to decide precisely how the idea translates to an implementation. Consider the following discussion:
VM systems include a concept known as an address space. The system dynamically creates an address space when a program needs one, and destroys an address space when the program that created the space has finished using it. A VM system uses a small, finite number to identify each address space. Conceptually, one understands that each new address space should have a new identifier. However, if a VM system executes so long that it exhausts all possible address space identifiers, it must reuse a number.
The important point is that the discussion only makes sense because it defines "address space" independently from "address space identifier". If one expects to discuss the

differences between a concept and its implementation, the definitions must allow such a distinction.

Knowledge vs. Data

The facts that result from an experiment are called "data". The term "knowledge" implies that the facts have been analyzed, condensed, or combined with facts from other experiments to produce useful information.

Cause and Effect:

A dissertation must carefully separate cause-effect relationships from simple statistical correlations. For example, even if all computer programs written in Professor X's lab require more memory than the computer programs written in Professor Y's lab, it may not have anything to do with the professors or the lab or the programmers (e.g., maybe the people working in professor X's lab are working on applications that require more memory than the applications in professor Y's lab).

Drawing Only Warranted Conclusions:

One must be careful to only draw conclusions that the evidence supports. For example, if programs run much slower on computer A than on computer B, one cannot conclude that the processor in A is slower than the processor in B unless one has ruled out all differences in the computers' operating systems, input or output devices, memory size, memory cache, or internal bus bandwidth. In fact, one must still refrain from judgment unless one has the results from a controlled experiment (e.g., running a set of several programs many times, each when the computer is otherwise idle). Even if the cause of some phenomenon seems obvious, one cannot draw a conclusion without solid, supporting evidence.

Commerce and Science:

In a scientific dissertation, one never draws conclusions about the economic viability or commercial success of an idea/method, nor does one speculate about the history of development or origins of an idea. A scientist must remain objective about the merits of an idea independent of its commercial popularity. In particular, a scientist never assumes that commercial success is a valid measure of merit (many popular products are neither well-designed nor well-engineered). Thus, statements such as ``over four hundred vendors make products using technique Y'' are irrelevant in a dissertation.

Politics and Science:

A scientist avoids all political influence when assessing ideas. Obviously, it should not matter whether government bodies, political parties, religious groups, or other organizations endorse an idea. More important and often overlooked, it does not matter whether an idea originated with a scientist who has already won a Nobel prize or a first-year graduate student. One must assess the idea independent of the source.

Canonical Organization:

In general, every dissertation must define the problem that motivated the research, tell why that problem is important, tell what others have done, describe the new contribution, document the experiments that validate the contribution, and draw conclusions. There is no canonical organization for a dissertation; each is unique. However, novices writing a dissertation in the experimental areas of CS may find the following example a good starting point:

- **Chapter 1: Introduction**

  An overview of the problem; why it is important; a summary of extant work and a statement of your hypothesis or specific question to be explored. Make it readable by anyone.

- **Chapter 2: Definitions**

  New terms only. Make the definitions precise, concise, and unambiguous.

- **Chapter 3: Conceptual Model**

  Describe the central concept underlying your work. Make it a "theme" that ties together all your arguments. It should provide an answer to the question posed in the introduction at a conceptual level. If necessary, add another chapter to give additional reasoning about the problem or its solution.

- **Chapter 4: Experimental Measurements**

  Describe the results of experiments that provide evidence in support of your thesis. Usually experiments either emphasize proof-of-concept (demonstrating the viability of a method/technique) or efficiency (demonstrating that a method/technique provides better performance than those that exist).

- **Chapter 5: Corollaries And Consequences**

  Describe variations, extensions, or other applications of the central idea.

- **Chapter 6: Conclusions**

  Summarize what was learned and how it can be applied. Mention the possibilities for future research.

- **Abstract:**

  A short (few paragraphs) summary of the dissertation. Describe the problem and the research approach. Emphasize the original contributions.

Suggested Order For Writing:

The easiest way to build a dissertation is inside-out. Begin by writing the chapters that describe your research (3, 4, and 5 in the above outline). Collect terms as they arise and keep a definition for each. Define each technical term, even if you use it in a conventional manner. Organize the definitions into a separate chapter. Make the definitions precise and formal. Review later chapters to verify that each use of a technical term adheres to its definition. After reading the middle chapters to verify terminology, write the conclusions. Write the introduction next. Finally, complete an abstract.

Key To Success:

By the way, there is a key to success: practice. No one ever learned to write by reading essays like this. Instead, you need to practice, practice, practice. Every day.

Parting thoughts:

We leave you with the following ideas to mull over. If they don't mean anything to you now, revisit them after you finish writing a dissertation.

> After great pain, a formal feeling comes.
> -- Emily Dickinson

> A man may write at any time, if he will set himself doggedly to it.
> -- Samuel Johnson

> Keep right on to the end of the road.
> -- Harry Lauder

> The average Ph.D. thesis is nothing but the transference of bones from one graveyard to another.
> -- Frank J. Dobie