

# Chapter 1

## Reading

*If I have seen further,  
it is only by standing on the shoulders of giants.*

Isaac Newton, letter to Robert Hooke,  
5 February 1675 os.

### 1.1 Introduction

As someone who plans to contribute to the sum of human knowledge, you are likely to spend more time reading than writing. Why? Because you have to be aware of the current state of your field before you can contribute to it, and because your own work must be firmly grounded on the work of others. Reading can be a powerful way of improving your own writing. Do you really like the organization, phraseology, typography or style of a paper that you are reading? Use this to improve your own writing. Is there something about an article that makes it hard to follow, difficult to comprehend, or just drives you crazy? Don't do that in your own work!

### 1.2 Why Read?

There are various reasons to read an article, and the way that you read will depend on your motivation. Are you looking to answer a particular question, trying to discover if your own work is novel, situating your work in a field, or simply educating yourself? You may also have been asked to review the article, write an abstract of it, or prepare a presentation on it; you may be looking for inspiration, or you may just be stuck on a bus and have nothing else at hand!

What might you get out of the reading process? My colleague James Hook shared the following list:

- The context of the paper.
- The thesis being investigated.
- The contribution.
- The method of investigation.
- The “power” of the results.
- The influence of the paper.
- The applicability of the results.
- Summary of the technical development.
- Details of any examples.

Your reason for reading the article will help you prioritize some of these benefits above others, and tell you where to focus your energies. If you don't know why you are reading an article, how can you know whether reading it will be worthwhile?

### 1.3 Deciding What to Read

When you select a topic, you may find yourself overwhelmed by materials from a wide variety of sources. How does one choose where to start? Some suggestions include:

- Ask your advisor or senior students about what conferences and journals are reputable in your area.
- Take a look at your advisor's web page to see where she is publishing.
- Pay attention to the reference sections of papers to see what conferences and journals are frequently mentioned and what papers are frequently cited.
- If possible, find surveys of your topic; when available, these make a great starting point.

Once you have found a good paper, take a look at the references to expand your reading list. Even a bad paper can sometimes have good references! In addition, the related work section is often a good source of background information on the context of the paper. Identify the research group that produced the paper and look at their publication list or annual reports to expand your reading list.

When you find a paper through a search engine, you will probably first approach it from its title. Read the title and ask yourself “what kind of paper is this?” Sometimes, the title will indicate the methodology used in the research, such as “Using bi-simulation to determine equivalence of DFAs”. Some titles suggest that the paper is a comparison work, such as “Four types of ...”. In general, titles provide a summary statement that either describes the research, or presents the main conclusion.

Consider the credibility of the paper. What is the affiliation of the author? What bias might that bring to the paper? What about the reputation of the conference or journal that published the paper? If the paper has been published for a while, has it been cited often? If the paper is too old, try to determine if it has been superseded.

## 1.4 The Mechanics of Reading

*Some people there are who, being grown,  
forget the horrible task of learning to read.*

John Steinbeck. *The Acts of King Arthur and his Noble Knights*. Pan Books, 1979.

Find a quiet place to read where you will not be disturbed or distracted. One of the reasons that I don't like reading at a computer is that it presents a constant source of distraction; find a situation that works for you. It is important to have some way of taking notes, electronic or on paper. Notes are important because they boost your understanding. If you highlight major points in your notes, you will not only have a handy reference if you need to summarize the article, but you will also have to decide what the major points *are*, and that will help you to maintain focus. Notes will also help you in the future. When you have just read an article, you may understand it well, but it may not be so clear in a week, or in a year. So take notes, and file them with the paper.

Reading is more than just passing your eyes over the words in consecutive order. Reading is understanding. You can maximize understanding by tackling a book or article in several passes, sometimes spread over time. How many passes? Over what time period? These depend on how difficult the paper is, how sharp you are feeling on a given day, how familiar you are with the topic, and on how thoroughly you need to understand the material. I'll describe one approach as an example.

### 1.4.1 Establishing context

Even before you start reading a paper, it is important to build a little context. Take a look at the authors' name, the group they belong to, and the year in which the paper was published. Knowing a little about the authors will help you understand their motivation a little better. A fairly recent paper gives you a technical summary of the research in that field. Conversely, if you find the background overwhelming, it may be better to pick an older paper cited in the references (perhaps from the same authors) to build up some expertise.

### 1.4.2 Goals of the first pass

On the first pass you will be *skimming*; skimming gives you an overview of the authors' work and helps you establish a conceptual map of the various parts of the paper. This map helps you make sense of the sections of the paper and how they relate to each other, and helps prepare you for a more complete reading. It's also possible that after skimming a paper, you decide to read no further. That's OK; you are learning to be selective in what you read.

On the first pass, it's probably sufficient to read the abstract, look at the pictures, and skim the list of references. The abstract will help you to determine if the paper is relevant to your goals, and to find out what kind of paper it is: a review, a minor incremental result, a landmark paper in a field, or a paper that started a new field. The "pictures" — the tables, graphs and diagrams — should give you a clue as to the major concepts and results. You should also see if you recognize the cited references or know what they are about — if not, you will probably have to do some background reading if you need to understand everything in the article in front of you.

### 1.4.3 Goals of the second pass

On the second pass, read the introduction, look at the start of each section or chapter, look at the examples, and read the summary or conclusion. The introduction should help you to understand the organization and the context of the article, and to decide if the authors are good writers. It took me a long time to figure this out, but I finally realized that there are more interesting papers out there than I will ever be able to read. Since I have to be selective, why not select the papers that are well written: the ones where the authors have made my task as a reader easy? The second pass should help you to decide how much work a more thorough reading will be, and whether the payoff is likely to outweigh the effort. On this basis, you might stop here, or you might decide to read only parts of the book or article. If the paper was published some time ago, checking to see how many times it has been cited

can help to establish the importance of the paper, and help you determine if it is worth another pass.

### 1.4.4 Goals of the third pass

The third pass is a full reading, and you will probably find it useful to take notes as you read. You rarely read an article in isolation: you are more often building up your own knowledge of a particular research area. So try to capture in your notes the main contributions of the article, and how it is similar to and distinguished from other work.

During this pass, mark definitions of unfamiliar terms in the margin. Note for yourself the assumptions that the authors make, and try to challenge these assumptions. Ask yourself whether or not the authors are making reasonable assumptions, and how these assumptions affect the soundness of the results. The answers to these questions can help you understand applicability of the proposed solutions, and how effectively the authors have evaluated them.

Critically examine the experimental methods used in the evaluation. Do these methods measure what the authors claim? For example, if the paper is from the proceedings of the Very Large Database (VLDB) Conference, does it test against a petabyte-scale database? If the paper deals with distributed systems, are the experimental nodes located on the same virtual machine, or are they located in Australia, Asia, Europe, and the Americas? Can the authors explain their observations? Did they have adequate controls? Were tests carried out in a standard way?

If statistical methods are applied to numerical results, ask yourself if they are appropriate for the subject, and if they were properly applied. Are the results statistically significant? Is there an error analysis? A sensitivity analysis? Should there be?

Approach the conclusions with scientific skepticism. Make sure that the data support the conclusions. Are there other plausible explanations for the observed results? Check the data to see if there are “inconvenient” correlations that the authors have ignored. By challenging the conclusion in these ways, you can better understand how the authors reached their conclusions, and learn to apply similar reasoning to your own research.

Throughout the full reading, make notes on what you *don't* understand, and put them on the front page, or in your notebook in a prominent place. Don't stop just because something is unclear or difficult; you will often find that if you push ahead, the authors will give you a clue that will explain the difficulty. This process of “late-binding” — holding content in your head that you don't yet understand until, sometime later, “the penny drops” — is a useful skill to acquire. “Sometime later” may be weeks or months later, when

some other paper gives you the insight that makes an earlier one “click”. Once you understand, go back and add a note explaining the difficulty.

### 1.4.5 Goals of the fourth pass

The fourth pass is when you get down to detailed study; you go back over the hard parts or the parts that were unclear, and frame your own questions about what the authors leave unanswered. You may find that you do this pass on a minority of the articles that you read.

### 1.4.6 Remembering what you’ve read

When you are finished, write a summary of the article in your own words. If you have doubts about the authors’ thesis, note your objections. If you find a phrase that you think you may wish to quote, note that too.

When you have finished reading a group of papers, write a summary of all of them. The summary should include the *relationships* between the papers, for example, how one paper’s key idea is originated from another, or how the later article one improved something introduced earlier. Not only will these notes be useful when you come to write the “related work” sections of your own papers or thesis, they will also give you practice writing.

## 1.5 Cracking a tough paper

Some papers are particularly tough. Here are some strategies that can help.

- Have someone who understands the paper explain it to you.
- Try to explain what you find difficult to someone else.
- Find an article by the same authors targeting a more general audience.
- Find a later paper by the same authors — maybe by then they understood their own work better, and could do a better job of explaining it.
- Make up (additional) examples. This can be a good way to master definitions, understand algorithms, or check out theorems. Working through an example is a good way to explain — or understand — something quickly.
- Draw a picture.
- Write code.

- Recast portions in alternative notation, *e.g.*, logic, functional programming language.
- “Guess” what is going on. See if your hypothesis is consistent with the rest of the paper.
- Try to find a counterexample to a theorem. Often, this is the way that proofs are first found.
- Send questions to the authors. You may be surprised to find that authors are often happy to answer *specific* questions about their work. Phrase your question so that there is a 2–3 line answer; quote enough of the context so that the authors do not have to go back and read their original paper.

## 1.6 Organizing Source Materials

Over the course of your career, you are likely to read many hundreds, probably thousands, of technical articles. You will accumulate them steadily, and it’s easy to become overwhelmed if you don’t have a system for keeping track of them.

Any system that you use should be simple enough that you will keep on using it. It doesn’t have to be perfect: if you can find the article that you are looking for within 2 or 3 minutes, your system is working fine. As an undergraduate, I simply filed research papers with the notes for the course to which they pertained. When I became a research student, there were no courses, so I made photocopies of the articles and filed them alphabetically *by first author* in a cabinet. Importantly, I also kept an index—on 3×5 inch cards—that categorized each article by keyword, sometimes several cards per article. I made it a rule that nothing went into the file cabinet without my first creating at least one index card. Other people group their papers by broad category, and then by author. Another commonly adopted rule is that whenever you make a copy of an article, *always* copy enough information to re-create a full citation. An easy way to do this is by copying the title page of the journal or proceedings.

I used to make paper copies of an article that I was working from, even if I had the full set of proceeding or the journal issue. That was because I could write freely on a copy, and because the copy was more portable. If you file your papers by topic, then filing is easier too. Today, a personal pdf file serves the same purpose as those old paper copies: you can annotate it, and you can carry it around.

Nowadays, I have replaced the file cabinet by a set of folders on my hard disk, one per author, but still organized alphabetically. These folders

currently contain about four thousand articles. The card index has been replaced by a *bibtex* database that contains the full citation information for each paper, as well as keywords of my choice that categorize it. Using *bibtex* for the index means that I can create the *References* section of an article with minimal effort; this used to be a major chore, and one that once caused me some embarrassment.

## 1.7 Chapter Summary

This chapter discussed techniques for effectively reading articles, and how these techniques might change depending on your motivation for reading. Here are the main ideas.

- Have your goal in mind when you read.
- Take notes as you read, and file them with the paper.
- Read in multiple passes; don't try to read from the front to the back in one session.
- Challenge the authors' assumptions and methods.
- Don't give up when the going gets tough.
- Don't be afraid to ask for help.
- Organize your technical articles so that you can find them in the future, and so that you can cite them correctly.

A handy one-page reference on reading is available at <http://www.cs.pdx.edu/~black/ScholarshipSkills/LectureNotes/efficientReading.pdf>. It summarizes the material in this chapter, and also contains some additional suggestions that you may find useful. It was prepared by Michael J. Hanson, and has been updated by Dylan McNamee and others.

## Acknowledgments

This chapter is based on lecture notes originally prepared by David Maier, and revised over a period of many years by Todd Leen, Tim Sheard and myself. James Hook contributed the list on page 2, and Dylan McNamee the notes on reading.