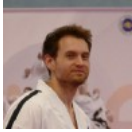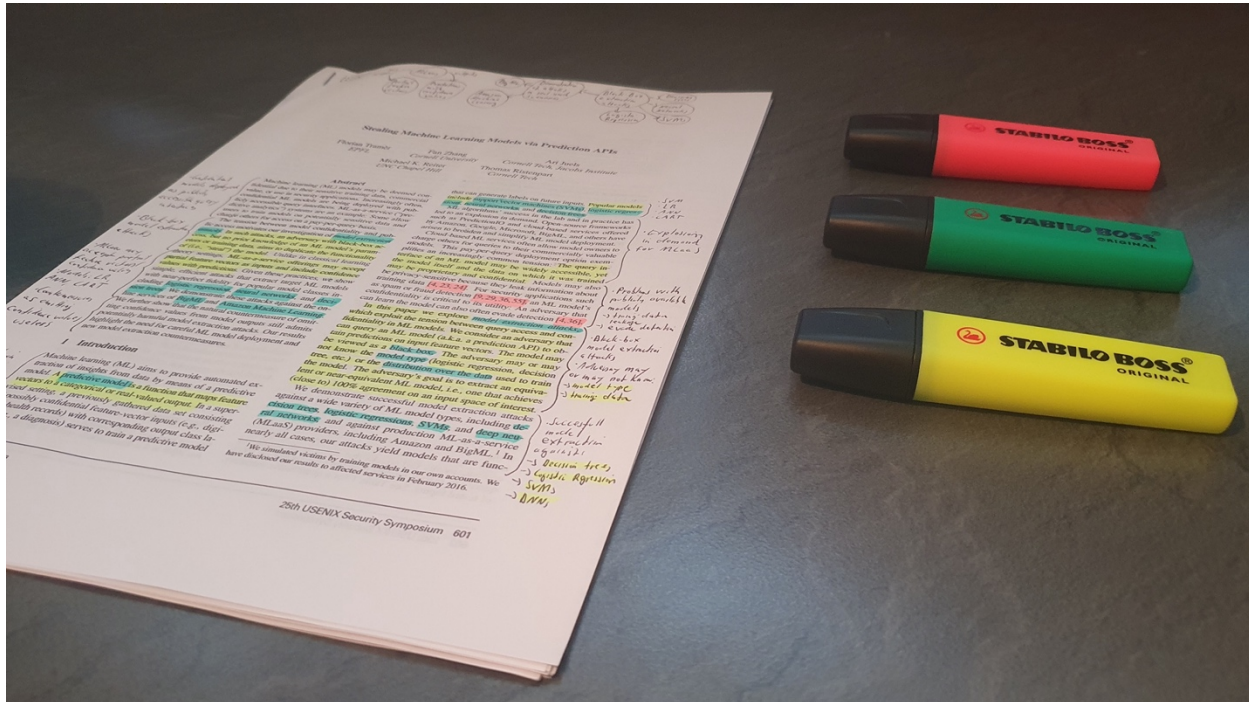# Medium

How To Read Scientific Papers

Increase your efficiency with the three-pass approach

[Christoph Schmidl](#) Apr 12 · 15 min read

Highlighted paper lying on my very own desk

Goal of this article

This article should serve as a rough guide on how to read a scientific paper because this skill is rarely taught at Universities and can lead to massive frustrations. Most of the time it is assumed that students already know some methods on how to read research papers but I have to admit that I knew none of them in the beginning.

When I had to read my first papers, I just started to read them from the beginning to the end. Like a book. I looked at every table, figure and math equation, and tried to make sense out of it. I wanted to understand it all and do not miss one single piece of information! It just so happens that there is a fitting term for that: [the](#)

[fear of missing out (FOMO)](#). But when I came to the end after several hours of frustration and background reading, I realized that the paper was not as helpful as I thought in the beginning. And I already forgot about the big picture or never had it in the first place. I got lost in details. Not a very effective way of reading a paper, especially when you are doing a literature survey or just have to read multiple papers in a day.

But it seems like there is a better way on how to approach this problem: **the three-pass approach**.

The remainder of this article is structured like this and explains each topic in greater detail:

1. **The three-pass approach (**tl;dr**)**
2. **The first pass: The bird's-eye view**
3. **The second pass: Grasp the content**
4. **The third pass: Virtually re-implement the paper**
5. **Doing a Literature Survey**
6. **Optional extensions**

- Little boxes
- Highlighters
- Mindmaps
- Pomodoro sessions
- The Feynman technique and rubber duck debugging
- Parkinson's law and the Pareto principle

The three-pass approach (tl;dr)

In "How to Read a Paper"[1] by Srinivasan Keshav, he describes the three-pass approach which acts as a filtering system. It is an iterative and incremental way of reading a paper. This deductive method goes from a general overview to the specific details while each step takes more time than the previous one and gives you deeper insights in each iteration.

1. **The first pass:** Here you get the bird's-eye view or "the big picture" of the paper. This step usually takes 5 to 10 minutes. You skim through the structure of the paper and ignore any details like math equations but you should read the abstract, title, introduction and conclusions entirely. This step serves as a first check if the paper is worth reading in general. By following this approach you can already discard papers which are not helpful, e.g., in a literature review.

2. **The second pass:** Here you try to understand the content of the paper by reading it as a whole. This step can take up to 1 hour. You can still ignore details like math equations but try to make some notes at the margins and write down key points. Try to rephrase the key points in your own words.

3. **The third pass:** You have to be very certain that this paper is worth your time before continuing with this step because it can take up to 5 hours as a beginner. More experienced readers may be able to finish this step in 1 hour. Now is the time to read the complete paper with all its math equations and details. Try to virtually re-implement the paper or use any tools you like to recreate the results. If you are a reviewer then you probably have to take this step to give detailed feedback.

I would like to point out that this article is not about reviewing papers. However, if you are searching for any reviewing guidelines then take a look at the references at the end of this article [2],[3],[4].

The following sections describe each step of the three-pass approach in greater detail.

The first pass: The bird's-eye view

"The first pass is a quick scan to get a bird's-eye view of the paper." — Srinivasan Keshav

The goal of the first pass is to get the big picture of the paper and should not take longer than 10 minutes. You don't have to get into the details or even read the paper in its entirety.

Glance over the paper and see how it is structured. Look at the sections and sub-sections but ignore their content. While you are reading the sections and sub-sections, you are already priming your brain for the upcoming content and you may already come up with some vague questions in your head. This will make it easier for you to spot important or interesting passages later on if you decide to go further.

While you just glance over the structure you should read the following sections completely:

1. **Abstract**
2. **Title**
3. **Introduction**
4. **Conclusions**

These sections will give you enough information so that you know what the paper is about and if it's worth reading any further. While reading these sections you could also take a look at the references and see if something seems familiar to you or if something has already been mentioned in other papers you have read before.

At the end the first pass you should be able to answer the so-called "five C's" as Keshav[1] puts it:

1. **Category:** The category describes the type of the paper. Is this paper about a prototype? About a new optimization method? Is it a literature survey?

2. **Context:** The context puts the paper into perspective to other papers. What other papers are related to this one? Can you connect it to something else? You could also see the context as a semantic tree where you assign specific importance to the paper. Is it an important branch or an uninteresting leaf? Maybe you do not have any prior knowledge in this field and therefore you still have to build your semantic

tree from the ground up. This can be demotivating in the beginning but it is normal.

3. **Correctness:** Correctness is, just as the name suggests, a validity measurement. Are the assumptions valid? Most of the time the first pass won't give you enough information to answer this question with certainty but you probably have a hunch which is enough in the beginning.

4. **Contributions:** Most papers have a list of their contributions right in the introduction section. Are these contributions meaningful? Are they useful? Which problems do they solve? Are these contributions novel?

5. **Clarity:** Based on the sections you just read, do you think that the paper is well written? Did you spot any grammar mistakes? Any typos?

This pass should serve as a quick, first filter. When you are done with the first pass you can decide to read further and continue with the second pass or you decide not to read further because:

- You are lacking background information

- You don't know enough about this topic

- The paper does not interest you or is not beneficial to you

- The paper is poorly written

- The authors make false assumptions

If that paper lies not in your area of expertise but may become relevant to you at a later point then this first pass is sufficient and you probably do not have to continue reading. If that's not the case then you can continue with the second pass.

Medium

The second pass: Grasp the content

"Sometimes you won't understand a paper even at the end of the second pass." — Srinivasan Keshav

The second pass can last up to 1 hour and here you should read the complete paper. Ignore details such as proofs or equations because most of the time you won't need that specific knowledge anyway and it costs you valuable time. Take some notes at the margins of the paper and write down the key points. Writing down little summaries or key points at the margins in your own words is a great way to see if you really understand what you've just read; and you will remember it way longer.

Look at any type of illustration in the paper like tables and figures and see if you can spot any mistakes or discrepancies. Do the illustrations make sense? What kind of information do they convey? Are the axes properly labeled? Do the figures and tables have proper captions? Sloppy work like this can already be a strong indicator of an overall badly written paper.

You can already mark relevant unread references for further reading which is a good way to learn more about the background. Build your semantic tree and see which papers are important branches and which ones are unimportant leaves.

At the end of the second pass it can happen that you still don't understand what you've just read. This could be due to many reasons. Maybe this is not your field of expertise or you are lacking background information. Do not feel discouraged because this happens all the time; even to Professors… so I was told.

Keep in mind that research groups often spent several months or even years to conduct their research. And now they had to compress their results and knowledge into a paper which may be had to meet certain requirements to get accepted by a conference, e.g., a certain page

count. If you think about it that way, then it does feel way less demotivating when you did not understand everything in 1 hour.

It sometimes helps when you write down what you did not understand. Then you have a great starting point to fill in the knowledge gaps later on through some background reading.

You now have different options available to you:

1. You can stop reading any further because the paper is not beneficial to you out of several reasons
2. Put the paper aside and continue reading after you read some background material
3. Continue with the third pass

The third pass: Virtually re-implement the paper

"This pass requires great attention to detail. You should identify and challenge every assumption in every statement." — Srinivasan Keshav

If you are a beginner then this pass probably takes 4 to 5 hours. This is a lot of work and you should carefully consider if this step is worth your time. On the other hand, if you are already an experienced reader then this step may only take you 1 hour. This step is mandatory if you are a designated reviewer or you already know for sure that you have to understand the paper with all its details.

Read the paper in its entirety and question every detail. Now it's time to get into the nitty-gritty math equations and trying to comprehend what is going on. Make the same assumptions as the authors and re-create the work from scratch. You can virtually re-implement the steps in your head or use any tool that you may deem fit. Use a piece of paper and draw a flowchart of the different steps or use pseudo-

code. It's really up to you. Most of the time I'm reading papers related to Artificial Intelligence and Computer Science and therefore it makes sense to re-implement things in raw Python or use Jupyter Notebooks. It really depends on your field.

At the end of this pass you should be an expert and know the paper's strong and weak points. You can make statements about missing citations and potential issues. You can reconstruct the structure and explain to someone in simple language what the paper is all about.

The concept of learning by teaching others is called the "Feynman technique" and is a great way to discover any gaps in your understanding.

Learning From the Feynman Technique

They called Feynman the "Great Explainer."

medium.com

Doing a Literature Survey

Doing a literature survey is a bit different than reading a single paper but you can still apply the three-pass approach.

First pass

In the first pass you have to collect potentially useful papers. You can use a search engine like Google Scholar and type in keywords to find 3 to 5 *recent* papers. What I usually do is to create a simple list of papers clustered by their topic together with the publishing year and the count of citations. The citation count is usually a good indicator of a paper is important. Just typing in your keywords into Google can also lead to surprisingly well results.

5.2 Classifier Stealing

These papers have been accessed on the 2nd December 2019.

Paper	Year	Citations	Reference
GDALR: An Efficient Model Duplication Attack on Black Box Machine Learning Models	2019	0	Joshi and Tammana (2019)
Efficiently Stealing your Machine Learning Models	2019	0	Reith et al. (2019)
PRADA: protecting against DNN model stealing attacks	2019	23	Juuti et al. (2019)
Knockoff nets: Stealing functionality of black-box models	2019	13	Orekondy et al. (2019)
Model Extraction and Active Learning	2018	3	Chandrasekaran et al. (2018)
Stealing hyperparameters in machine learning	2018	72	Wang and Gong (2018)
How to steal a machine learning classifier with deep learning	2017	28	Shi et al. (2017)

When you have your little collection of initial papers ready, you can continue with the usual first-pass on each of them to get the big picture. You can also skim through the references to see

if the papers have any citations in common. Common citations are good candidates to include in your survey.

Second pass

When you identified common citations and repeated authors, you can visit their websites and see if you can spot any recent work. Also download the commonly cited papers and apply the three-pass approach for single papers again.

Third pass

Here you can try to visit the websites of the top conferences or journals and look through the recent proceedings. Try to identify related high-quality work and apply the three-pass approach for single papers again.

Optional extensions

Keep in mind that these optional steps might add some time to the three-pass approach. They might not be beneficial to you if you are just skimming through papers to see if they are helpful or not. On the other hand, if you already know that you have to read and understand the paper in its entirety and there is no way around it, then these steps might help you too. These are my personal practices and I'm constantly trying to improve them.

Little boxes

When you take a look at the following picture, you can see that I surrounded math equations, figures and tables with boxes. I like to look at papers in terms of clearly separated boxes and separate the text from the rest. I usually do this during the first-pass while I'm skimming through the paper. This helps me to quantify how many details in terms of math equations I can expect later on and it just seems more pleasant to my eyes.
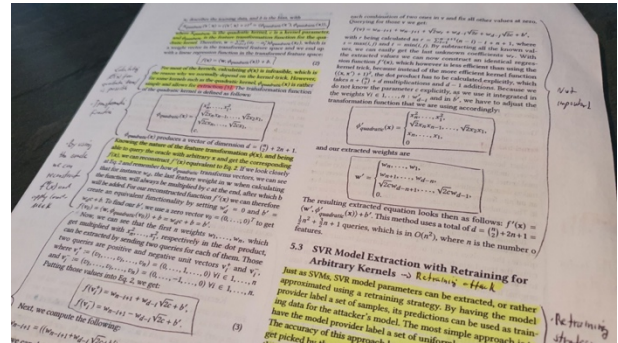
*Little boxes around math equations*

Highlighters

Highlighters are a great tool to mark sections in your paper and give them distinctive meanings. You can try to come up with your own highlighting system or use an existing one. Try to give every color a distinctive meaning and stick to it.


A nice set of highlighters

During the second-pass, I usually use yellow for interesting or important sentences. Orange is for citations and green for definitions or catchphrases. However, feel free to use whatever system you may please. Keep in mind that highlighting does not replace note-taking! During the second-pass you can take notes at the margins, draw little diagrams for better understanding and use highlighters in combination.
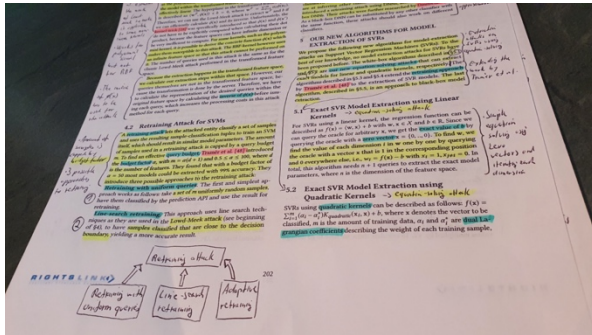
Each color has a distinctive meaning

Interesting or important references at the end of the paper get the same color as before.
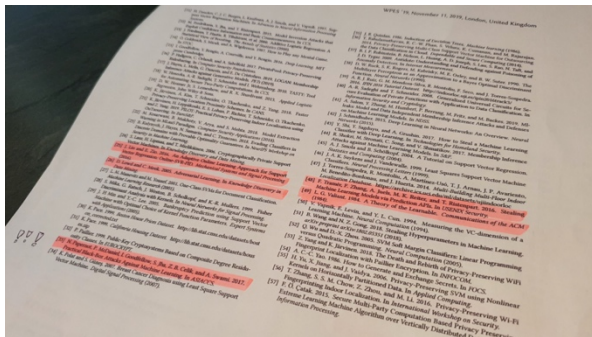
Marking references with orange

Mindmaps

If you are more visual and want to get a better overview of the paper, mind maps may be a suitable fit. There are no strict rules in creating mind maps and I just started with the title of the paper in the center. Big arrows are pointing to the main section titles and are highlighted with orange. These are the big branches. First-level subsections are highlighted with green. Anything else gets no highlighting. Feel free to come up with your own system.

This step usually adds 25 minutes to the first-pass and I'm still not sure if it is worth the time. On the other hand, if you continue to the second-pass and want to write down an important note, you can put that directly into the corresponding node of the mind map. This may help you to get the big picture more visually. This may also be a faster way to

refresh your memory about a paper after some time has passed.
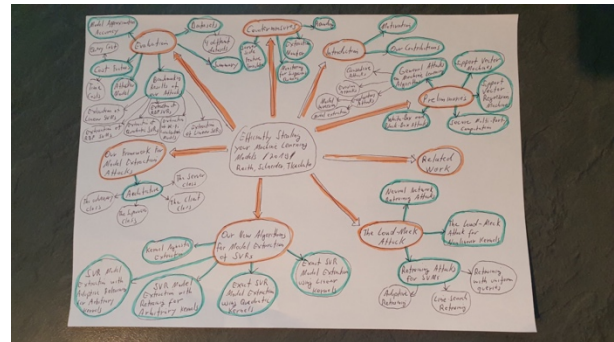
Pomodoro sessions

The Pomodoro technique[5] is a great tool if you are lacking motivation. Sometimes it's not the case that you do not know how to read a paper but more that you feel intimidated by it and lack the motivation to even get started. Procrastination kicks in and you are missing an important deadline for a review.

Get a timer and set it to 25 minutes. Do not expect any results. Just set it to 25 minutes and start. Eliminate any distractions and follow the three-pass approach until the 25 minutes are up. You may not finish the whole three-pass approach but at the end of the 25 minutes you will likely be surprised what you achieved. You now know what the paper is about and you probably feel less intimidated. You probably feel like you could set the timer for another 25 minutes.

By using this timeboxing approach you gain momentum and can follow the three-pass approach more easily. The nice thing is: you can apply the Pomodoro technique to any task.

The Feynman technique and rubber duck debugging

As mentioned earlier, the Feynman learning technique is a great tool to spot gaps in your understanding. The general steps are:

1. **Choose a concept** you want to learn and write its name at the top of a piece of paper.

2. **Pretend you are teaching the concept** to someone who has no prior knowledge about it. Try to use simple language and do not simply recite. Use your own words!

3. **Review your explanation**. Was is accurate despite the usage of simple language? Identify weak points in your explanation and write it on the piece of paper. Go back to your learning material and see if you can clarify these points.

4. **Simplify your explanation** if you used lots of technical terms or complex language in areas of your explanation.

If you want to apply the Feynman technique but don't have a little brother at hand for step 2, then the rubber duck may be for you.

The idea behind rubber duck debugging has its roots in Software Engineering and first occurred in the book The Pragmatic Programmer by Andrew Hunt and David Thomas. In the book, a programmer carries around a rubber duck and explains the code, line-by-line, to the duck to spot any mistakes. You can also use any other object for this. Do you have a cat? I'm sure she always wanted to know how Hamiltonian Monte Carlo sampling [6] works.
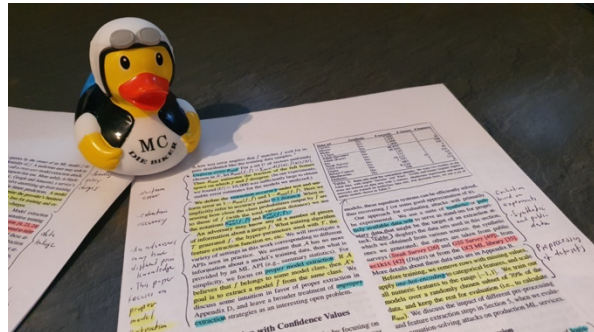
Explain it to someone who knows nothing about your topic

Parkinson's law and the Pareto principle

The following two approaches are not just limited to the task of reading papers but to any other task in general. If you combine these two then you come up with a capped timebox approach, e.g., plan 10 Pomodoro sessions for the whole paper and then stop. You can also try to give yourself a totally unrealistic timeframe to read a paper and then check your progress.

Parkinson's law states the following:

"Work expands so as to fill the time available for its completion" — **Cyril Northcote Parkinson**

If you plan 10 hours to read a paper, taking notes, writing summaries and so forth, then it will probably take you 10 hours.

The **Pareto principle** (also called the 80/20 rule) on the other hand states:

"For many events, roughly 80% of the effects come from 20% of the causes." — Vilfredo Pareto

This means that it takes you probably 20% of your overall effort and time to understand 80% of the paper. This 80/20 split is not fixed but is rather a rough estimate. It could also be something like 70/30.

Did you ever approach a deadline where you were left with 30 minutes to do a task that you thought would take you a couple of hours? And then you realized that you actually did quite well? Parkinson's law forced you into a 30-minute timeframe and the Pareto principle ensured that you only did the tasks which contributed the most to your final result. Try to simulate this situation by giving yourself unrealistic, tight deadlines.

Medium

I hope you enjoyed this article and I could help you during your academic journey.

Are you interested in upcoming updates about my articles or projects? My [newsletter](#) keeps you up-to-date once per month!

References

[1] Keshav, S. (Srinivasan)(2007). How to read a paper. *ACM SIGCOMM Computer Communication Review*, *37*(3), 83–84. (URL: https://dl.acm.org/doi/pdf/10.1145/1273445.1273458)

[2] Cormode, G. (2009). How not to review a paper: The tools and techniques of the adversarial reviewer. *ACM SIGMOD Record*, *37*(4), 100–104. (URL: https://dl.acm.org/doi/pdf/10.1145/1519103.1519122)

[3] Meier, A. (1992). How to review a technical paper. *Energy and Buildings*, *19*(1), 75–78. (URL: https://eta-intranet.lbl.gov/sites/default/files/how-to-review-a-technical-paper_0.pdf)

[4] Roscoe, T. (2007). Writing reviews for systems conferences. (URL: https://www.cl.cam.ac.uk/teaching/1011/R01/review-writing.pdf)

[5] Cirillo, F. (2018). *The Pomodoro Technique: The Life-Changing Time-Management System*. Random House.

[6] Betancourt, M. (2017). A conceptual introduction to Hamiltonian Monte Carlo. *arXiv preprint arXiv:1701.02434*. (URL: https://arxiv.org/pdf/1701.02434.pdf)

[Towards Data Science](#)

A Medium publication sharing concepts, ideas, and codes.