

INFO 290T

Human-Centered Data Management



Announcements!

- Presentation slots open — see email!
 - For next week, freeze of slots is Saturday
- We will circulate the link to submit paper reviews
 - Only applies if you're not presenting
- Any questions about logistics?



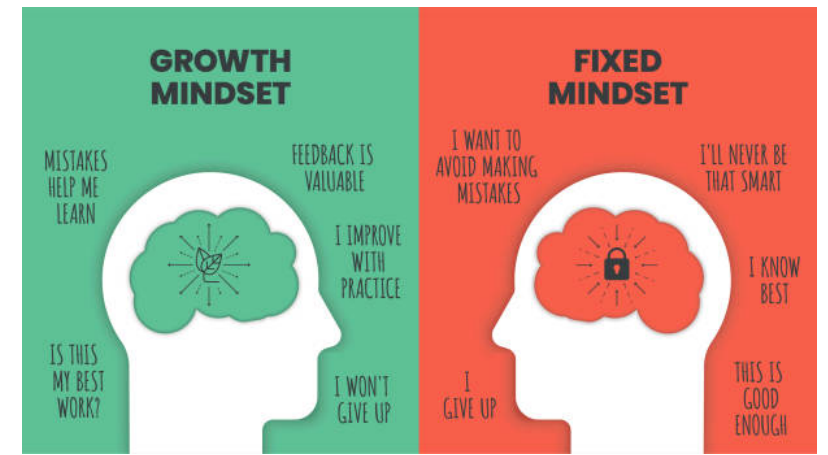
Today...

Primers!

- Reading papers primer
- Visualization primer
- HCI/Usability primer
- Systems primer (unlikely we'll get to it)



Reading Research Papers



- For those who are reading research papers for the first time...
- Like everything else, reading and understanding research papers is a practiced skill — you get better at it with more effort
 - Marginal effort for reading the n th research paper in an area is usually less than that for reading the n -kth research paper
 - Quickly pattern match, separate what's important from what's not, figure out where it fits in
- Important: don't believe everything you read!
 - Question assumptions as you go along : unlike how you may have read textbooks!
 - Often key threats to validity, reproducibility, correctness, adoption, usability, ... hidden in plain sight!
 - “Most papers are wrong, but some are useful”

What Should you Learn from a Research Paper

- **What are the motivations for this work?**
 - **People problem.** What is the end-user problem that hasn't been adequately addressed before?
 - **Technical problem.** Why doesn't it have a trivial solution? Why are previous solutions inadequate?
 - **Research question.** The technical and people problem lead to this.
- **What is the proposed solution?** Why is it believed that this solution will work, and be better than previous solutions? How is the solution achieved (designed and implemented)?
- **What is the evaluation of the proposed solution?** What argument, implementation, and/or experiment makes the case for the value of the ideas? What benefits or problems are identified?
- **What is your analysis of the identified problem, idea and evaluation?** Is this really going to work, who would want it, what it will take to give it to them, and when might it become a reality?
- **What are the contributions?** Beyond the insights on the research question, a few additional possibilities include: ideas, software, experimental techniques, benchmark, or an area survey.
- **What are future directions for this research?** What ideas did you come up with while reading the paper? May be identified as shortcomings or other critiques in the current work.
- **What questions are you left with? What is your takeaway?**



The Multiple Pass Approach

Adapted from “How to Read a Paper” by Keshav

The three pass approach

- Pass 1: Bird’s eye view
- Pass 2: Understand content, but not details
- Pass 3: Re-derive

How to Read a Paper

S. Keshav
David R. Cheriton School of Computer Science, University of Waterloo
Waterloo, ON, Canada
keshav@uwaterloo.ca

ABSTRACT
Researchers spend a great deal of time reading research papers. However, this skill is rarely taught, leading to much wasted effort. This article outlines a practical and efficient three-pass method for reading research papers. I also describe how to use this method to do a literature survey.
Categories and Subject Descriptors: A.1 [Introductory and Survey].
General Terms: Documentation.
Keywords: Paper, Reading, Hints.

1. INTRODUCTION
Researchers must read papers for several reasons: to review them for a conference or a class, to keep current in their field, or for a literature survey of a new field. A typical researcher will likely spend hundreds of hours every year reading papers.
Learning to efficiently read a paper is a critical but rarely taught skill. Beginning graduate students, therefore, must learn on their own using trial and error. Students waste much effort in the process and are frequently driven to frustration.
For many years I have used a simple approach to efficiently read papers. This paper describes the ‘three-pass’ approach and its use in doing a literature survey.

2. THE THREE-PASS APPROACH
The key idea is that you should read the paper in up to three passes, instead of starting at the beginning and plowing your way to the end. Each pass accomplishes specific goals and builds upon the previous pass: The *first* pass gives you a general idea about the paper. The *second* pass lets you grasp the paper’s content, but not its details. The *third* pass helps you understand the paper in depth.

2.1 The first pass
The first pass is a quick scan to get a bird’s-eye view of the paper. You can also decide whether you need to do any more passes. This pass should take about five to ten minutes and consists of the following steps:

1. Carefully read the title, abstract, and introduction
2. Read the section and sub-section headings, but ignore everything else
3. Read the conclusions
4. Glance over the references, mentally ticking off the ones you’ve already read

At the end of the first pass, you should be able to answer the *five Cs*:

1. *Category*: What type of paper is this? A measurement paper? An analysis of an existing system? A description of a research prototype?
2. *Context*: Which other papers is it related to? Which theoretical bases were used to analyze the problem?
3. *Correctness*: Do the assumptions appear to be valid?
4. *Contributions*: What are the paper’s main contributions?
5. *Clarity*: Is the paper well written?

Using this information, you may choose not to read further. This could be because the paper doesn’t interest you, or you don’t know enough about the area to understand the paper, or that the authors make invalid assumptions. The first pass is adequate for papers that aren’t in your research area, but may someday prove relevant.
Incidentally, when you write a paper, you can expect most reviewers (and readers) to make only one pass over it. Take care to choose coherent section and sub-section titles and to write concise and comprehensive abstracts. If a reviewer cannot understand the gist after one pass, the paper will likely be rejected; if a reader cannot understand the highlights of the paper after five minutes, the paper will likely never be read.

2.2 The second pass
In the second pass, read the paper with greater care, but ignore details such as proofs. It helps to jot down the key points, or to make comments in the margins, as you read.

1. Look carefully at the figures, diagrams and other illustrations in the paper. Pay special attention to graphs. Are the axes properly labeled? Are results shown with error bars, so that conclusions are statistically significant? Common mistakes like these will separate rushed, shoddy work from the truly excellent.
2. Remember to mark relevant unread references for further reading (this is a good way to learn more about the background of the paper).

ACM SIGCOMM Computer Communication Review 83 Volume 37, Number 3, July 2007



Pass 1: Bird's Eye View

- Goal: to determine if you need to do more passes
- Read
 - Title, abstract, introduction
 - All the section/subsection headers
 - Conclusion
 - Skim the references
- Time allocated: 5 minutes
- Should be able to answer the 5 Cs
 - Category: What type of paper is it? (e.g., benchmarking, user study, survey, research paper)
 - Context: Where does it sit in in the literature?
 - Correctness: Do the assumptions appear to be valid on first glance?
 - Contributions: What are the main contributions?
 - Clarity: Is the paper easy to read/clear?
- Each of these could be deal-breakers for reading further, e.g., incorrect assumptions, irrelevant topic, not easy to read



Pass 2: Understand Content, Not Details

- Goal: to understand the paper's contributions in more detail
- Time: 1 hour
- Read through all sections, but:
 - Skip over all non-essential content: proofs, implementation details, methodological details
- Pay special attention to
 - figures & captions
 - where the paper fits in (may identify other papers to follow up)
- Should be able to summarize the main content of the paper, with supporting evidence, to someone else
- This level of reading is enough for a paper that you are interested in, but is not in your specific area of focus



Pass 3: Re-derive

- Time: 4-5 hours, with practice, can come down to 1-2 hours in papers in your area
- Goal: Virtually re-implement the paper:
 - Question everything
 - How would *you* tackle each step?
 - Does each assumption pass muster?
 - Can you prove each theorem (if any)?
- This is where you identify flaws in the paper, and ideas for follow up work
- You wouldn't want to do this for a lot of papers – just for papers that are in your specific niche

