Paper Reading

CS 2001: Research Topics in Computer Science
Fall 2013

Dietrich School of Arts of and Sciences
Department of Computer Science
Research papers are the lifeblood of science

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

—Isaac Newton, 1676
Today, we’ll cover

Typical paper structure

How to read

- Structuring your reading session
- What to look for
- Comprehension strategies

Filling in gaps in your knowledge
Papers in computer science often follow a somewhat predictable format

**Abstract**
- Introduction
- Related work*
- Proposed design/system/method
- Evaluation
- Discussion
- Related work*
- Conclusions & Future work

**References**

**Content:**
- Very short (~250 words)
- Brief description of purpose
- Highlight main results

**Purpose:**
- “Hook” the reader
  - Why is this paper interesting?
  - Why should I spend my time reading this?
- What do you claim to do?
- Set the stage for the paper
Papers in computer science often follow a somewhat predictable format

**Content:**
- Usually 1-1.5 pages
- Main elements
  - What is the problem?
  - Why is the state of the art insufficient?
  - Overview of the solution
  - Novel contributions of the work?
  - How is impact evaluated?
- Balance content and conciseness!

**Purpose:**
- Motivate the work
- Inform the reader of what is to come
- Many reviewers will make their initial decisions after reading only the intro!
Papers in computer science often follow a somewhat predictable format

Abstract
- Introduction
- Related work*
- Proposed design/system/method
- Evaluation
- Discussion
- Related work*
- Conclusions & Future work

References

Content:
- Overview of past research/results
- Comparison to claimed contributions
- Not a book report!

Purpose:
- Motivate the work (How?)
- Inform the reader that you are aware of prior results
- Clearly demonstrate the novelty in your approach

Note: Related work may occur at the beginning or end of a paper
- Beginning: Prior work is necessary for understanding this paper
- End: Prior work is only tangentially related
Papers in computer science often follow a somewhat predictable format

Abstract
- Introduction
- Related work*
- Proposed design/system/method
- Evaluation
- Discussion
- Related work*
- Conclusions & Future work
References

Content:
- Maybe more than one section...
  - Requirements, Design
  - Syntax, Semantics, Enforcement
  - Design, Implementation
  - ...
- This is the novel content of a paper

Purpose:
- Proposal of original idea(s)
- This is the authors’ contribution!
- Should be detailed enough for others to replicate the work (in theory)
Papers in computer science often follow a somewhat predictable format:

- **Abstract**
- **Introduction**
- **Related work**
- **Proposed design/system/method**
- **Evaluation**
- **Discussion**
- **Related work**
- **Conclusions & Future work**
- **References**

**Content:**
- Could be any number of things
  - Performance measurements
  - Simulation results
  - Analysis of user study data
  - Formal proofs
  - ...

**Purpose:** “Prove” that the stated contributions are meaningful.

**Note:** A incomplete/incorrect evaluation can kill an otherwise good paper!
Papers in computer science often follow a somewhat predictable format

Abstract
- Introduction
- Related work*
- Proposed design/system/method
- Evaluation
- Discussion
- Related work*
- Conclusions & Future work

References

**Content:**
- Not all papers have this
- If included, typically contains
  - An interpretation of results/evaluation
  - Discussion of open problems
  - Description of limitations

**Purpose:**
- Papers do not often “close” a topic
- This is where you reflect on what has been done, and what is still open
Papers in computer science often follow a somewhat predictable format

**Abstract**
- Introduction
- Related work*
- Proposed design/system/method
- Evaluation
- Discussion
- Related work*
- Conclusions & Future work

**References**

**Content:**
- **Far too often:** Rehash of the paper
- **Ideally:** Reflection on contributions

**Purpose:**
- One last summary of contributions given the whole context of the work
- Identification of promising future research directions
Preparing to read

Reading a research paper is different than other reading!
- 10 pages of news: < 10 minutes
- 10 pages of fiction: < 20 minutes
- 10 pages in a textbook: < 30 minutes
- 10 page research paper: 20 minutes - several hours!

Prior to reading, make sure you gather the appropriate supplies

Quiet Environment  Appropriate Time (How much?)  Note-Taking Supplies (?!)

Flatline
Why are you planning to read that paper?

There are many legitimate reasons for reading a paper:
- I heard someone talking about this result
- It’s related to a problem I am working on
- My advisor told me to
- This provides context for another problem
- The conference talk interested me
- I think that I might want to explore this area
- ...

**Take-away point:** Why you plan to read a paper will—to some degree—dictate how you should go about reading it.
A multi-pass approach to reading is generally good

Keshav* has a nice paper on a three-pass reading approach

Pass 1: Basic comprehension
- What is the main topic of the paper?
- What are the authors’ claimed contributions?
- What do they cite?

Pass 2: First look at real details
- Focus on details: evaluation, figures, methods
- Ignore proofs

Pass 3: Depth!
- Fully understand all details

Your first pass over the paper should help you decide how much time you need to invest in it

Focus your attention on:
- Title and Abstract
- Full details of the Introduction
- Section and Sub-Section headings in the body
- Full details in the Conclusion
- Skim references, note what you’ve read

After this, you should know about the “5 Cs”
- **Category:** Experimental paper? Theory? Measurement?
- **Context:** What does this paper cite?
- **Correctness:** Do any assumptions seem reasonable?
- **Contributions:** What do the authors (claim) to contribute?
- **Clarity:** Can you follow the paper?

You can probably accomplish this for most papers in ~10 minutes
Audience Participation!


Let’s talk a little bit...

- **Category**: Experimental paper? Theory? Measurement?
- **Context**: What does this paper cite?
- **Correctness**: Do any assumptions seem reasonable?
- **Contributions**: What do the authors (claim) to contribute?
- **Clarity**: Can you follow the paper?
The second pass over a paper is all about breadth of knowledge

**General idea:** Read the whole paper, but skip super-intricate details like proofs.

Focus on:
- Understanding methodology, evaluation, figures, etc.
- Mark relevant references for later reading (**more breadth!**)  
- Being able to explain the main ideas of the paper to someone else

This process can take **up to an hour** for a standard 10-page paper

**Why so long?**
- Perhaps you’re new to the subject area
- Authors use methodologies or techniques that are unfamiliar
- Paper is just badly written...
After breadth reading, you should be able to answer many questions about a paper.

Important questions include:

- What are the motivations for this work?
- What is the proposed solution? Is it novel?
- How is this solution evaluated?
- What do you think about the problem, solution, and evaluation?
- What are the contributions of this work?
- What does this paper close an area of research? Open a new one? Lead to interesting future work?
- What questions do you still have?

Griswold has a nice template for answering these questions. I’ve linked to it on the course page.
Audience Participation!


Let’s talk a little bit...

- What are the motivations for this work?
- What is the proposed solution? Is it novel?
- How is this solution evaluated?
- What do you think about the problem, solution, and evaluation?
- What are the contributions of this work?
- Does this paper close an area of research? Open a new one? Lead to interesting future work?
- What questions do you still have?
Your third pass over a paper should focus on developing an intricate understanding of the subject matter.

**Main focus:** Everything you’ve glossed over so far
- Thorough scrutiny of assumptions
- What alternative solutions might have been possible?
- Does the evaluation cover enough meaningful cases?
- Detailed examination of proofs and proof techniques

After a thorough pass, you should (ideally) be able to replicate the results presented in the paper.

This is a **time-intensive** process
- 4-5 hours for beginners
- Around an hour for more experienced readers
Note-taking can help build your understanding of a paper and manage the many papers that you’ll eventually read.

Note taking **while you read** helps capture the context of your reading session for later reference.

Use a highlighter to mark major points, definitions, and theorems for quick reference later.

Make notes in the margin:
- Write down questions as they pop into your head.
- Answer previous questions as you find answers.
- Summarize tables, graphs, etc.
- Add details to incomplete/unclear examples.
Note-taking can help build your understanding of a paper and manage the *many* papers that you’ll eventually read.

Note taking **after you read** can help:
- Ensure complete understanding of relevant papers
- Manage large collections of papers as your progress in your studies

Consider making a document per research area.

For each paper, write up:
- A technical summary of the work
- A brief description of the paper’s relation to other works
- Relationships to your ongoing/planned research
- Any cool ideas for future work that come to mind

A few examples...
Initially, you will have an incomplete knowledge of a research area. How can you fix this problem?

**Step 1:** Read up on prior work!

**Step 2:** Understand how this paper fits into more recent research

There are research tools to help aid these processes

- **ACM portal:** [http://portal.acm.org](http://portal.acm.org)
- **IEEEXplore:** [http://ieeexplore.ieee.org/Xplore](http://ieeexplore.ieee.org/Xplore)
- **Google scholar:** [http://scholar.google.com](http://scholar.google.com)
- **Citeseer:** [http://citeseerx.ist.psu.edu/](http://citeseerx.ist.psu.edu/)
Let’s do a little tracing...
Conclusions

Paper reading is an essential skill for PhD students (and researchers in general!)

At first, this is a slow process, but gets easier with practice

Multi-pass reading can help aid comprehension

- **Pass 1**: Overview
- **Pass 2**: General understanding, expand breadth of knowledge
- **Pass 3**: Details, details, details

Next time: Writing paper reviews