

ACADEMIC SKILLS IN COMPUTER SCIENCE

Lecture 4: Reading Research Reports

Markus Kröttsch
Knowledge-Based Systems

TU Dresden, 23th April 2019

Finding Literature: (2) Effective search and retrieval

Goals for today

Learning goals of this lecture:

- (1) Review techniques to effectively search and access research publications
- (2) Learn how to approach research papers when gathering information
- (3) Understand the standard structure of research reports
- (4) Get to know general strategies for literature search

Searching for research

Searching literature is much easier today than it ever was

- Unified search of every possible publication
- Search by content, citations, and publication time
- Almost no delay between author and reader
- Many texts easily accessible online

... all thanks to the world wide web and modern search engines

Basic search process:

- (1) Find publications on the Web (coming next)
- (2) Judge if they are relevant/trustworthy/useful (previous lecture)

Finding literature (1)

For direct text-based search, major search engine providers are usually best:

Example: [Google Scholar](#) is a useful service for searching research texts (rather than the whole web). Features:

- Full text search (like usual web search)
- See authors, venue, publication time, and citations with results
- Links to PDF when available
- Search in citing documents
- Specify publication time

Like web search in general, the service's results are neither complete nor correct; rankings can be manipulated

Finding literature (2)

Specialised databases may offer more finely configurable search features and higher data quality.

Example: [DBLP](#) is the most important literature database in computer science.

- Freely available, high quality data (but no citation data)
- Simple, clean interface, centred on author or venue

Example: [Scopus](#) is a literature database and search service by Elsevier.

- Advanced search by many criteria (author, title, host institution, venue, funding body, ...)
- More careful selection of sources than Google scholar
- Display citing documents; possibly filtering self-citations

Unfortunately, it is not free and offers a less-than-ideal user experience.

Other search techniques

Text search is not the only option:

- **Citation-based:** given an interesting paper, look at its citations (backward search), and search for other papers that cite it (forward search) [Google, Scopus]
- **Wikipedia-based:** look at what Wikipedia cites (use to find classical key references to established fields and definitions; be careful with very recent and niche topics)
- **Person-based:** given an interesting paper, figure out what its authors have published recently [DBLP, Google, author's web page]
- **Venue-based:** browse recent editions of a relevant conference or journal [DBLP, Scopus]

→ combine techniques for best results; interleave reading/rating and further search

Accessing literature

Many research papers in computer science are readily accessible:

- Google Scholar links to known legal PDF versions on the Web
- Even many historic publications have been scanned and OCR'ed
- Many authors use their right to publish preprints on their sites
- Some articles are free for the general public (open access)
- Other articles are available only through university subscription: access them from the university network (eduroam or VPN suffice)
- If all fails, you can email the authors or ask friends in other places (maybe other universities have access)
- Nevertheless, some works might still remain inaccessible ...

Accessing literature

Legal note: The use of the online services [Library Genesis \(libgen\)](#) and [Sci-Hub](#) to gain immediate free access to research articles and textbooks can be illegal (whether or not the Tor network is used to avoid tracking). The sites' practice of granting access to tax-paid research without collecting millions in license fees jeopardises the business of major publishing houses, and undermines their efforts of offering some of their content to a wider public under specific circumstances. Therefore, you should not use these sites, even if studies suggest that many students and academics, including from top universities, are doing this as part of their regular work.^a

^aBohannon, John (28 April 2016). "Who's downloading pirated papers? Everyone". *Science*. 352 (6285): 508–512. doi:10.1126/science.aaf5664

Reading Research Reports

Reading

We have already learned some basic skills:

- How to discover potentially relevant work
- How to judge what is most relevant and promising

Next step: actually read the works

Why read?

- Learn about new results and gain new insights
- Learn about useful methods and research techniques
- Learn about hot topics and active fields of research
- Discover limitations of current works and important open problems
- Find out if a work (by certain authors, published at certain venues) is useful/interesting/of good quality
- Understand the typical scope/form/structure of works in this area/venue

Organising the search for literature

A **structured literature search** is more than just browsing the web for interesting papers!

Some suggestions:

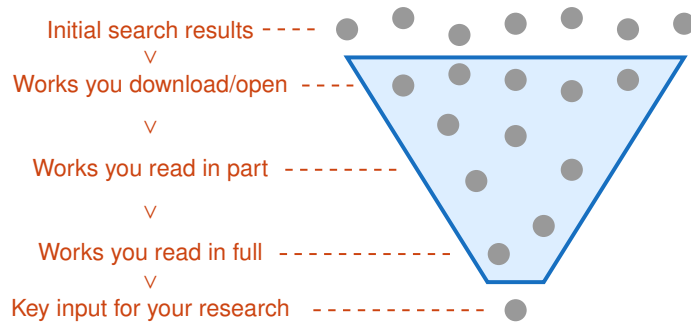
- (1) Start with a concrete goal/question in mind
- (2) Read in a goal-oriented, analytical way (coming up next)
- (3) Take notes (goal: safe yourself from having to repeat the whole effort later)
- (4) Store or bookmark the actual texts so you can find them later
- (5) Organise & relate papers/researchers/venues/topics/ideas as you read (this cannot be perfect but will change as you learn more)

Taking notes:

- Essential for memorising and building understanding
- Many forms exist: plain text, mind maps & other drawings, structured collections, . . .
~> use whatever format/language/medium works for you
- Prefer flexible, low-tech, "analogue" methods

Reading ≠ actually reading every word

The review of relevant literature is a process of continued focussing and selection:



How to filter

Most ways of searching for works (Google Scholar, crawling publication lists, looking through references from some paper, etc.) let you access **key information** immediately:

- **What:** title
- **Who:** authors
- **Where:** publication venue
- **When:** year
- an possibly other details, such as citation counts

~> can be used to judge relevance for your current search (easier with experience)

Works who pass this test need to be looked at in more detail . . .

The Structure of Research Reports

To read papers efficiently, it helps to understand the following:

General structure of research reports:

1. **Title, authors, affiliations**
2. **Abstract**
3. **Introduction**
4. The Research: details, methodology, results
5. **Conclusion**
6. **References**

This form is highly standardised:

- the parts in **bold** are almost always there, and have a largely standardised function
- the middle part is specific to the research that was done

1. Title, authors, affiliations

The standard header of any report

Read and remember the authors and affiliations!

Neither means much to you when you start, but you want to get beyond this level!

- You probably saw some of this before deciding to open the document
- **Author list and order** might help to classify paper
 - Non-alphabetic orders usually have a meaning:
first author is lead researcher, last author might be supervisor
(alphabetic order means: you don't know anything)
 - Rule of thumb in computer science:
many authors ~> system-oriented
few authors ~> foundational (exception: invited paper)
- **Author names and affiliations** can tell you much about what to expect
(if authors/groups are known for a particular research programme or topical focus or research style)
- The **title** is the authors' own preferred short description of their work
(information: what is the work about?; meta-information: how confident/original/witty are the authors?)

2. Abstract

The brief summary text at the start of papers

The paper in a nutshell

Abstracts are there to help readers decide if they should read the paper

- Self-contained (can be read without even opening paper)
- Should clarify: area of work, research question, main results & methods, possibly limitations/open issues
- Standardised abstract formats are not common in computer science yet (in contrast to, e.g., medicine)

4. The Research

The actual essence of the paper

The “main part” of the work

This will include several sections/chapters of technical material.

- Should be highly factual, comprehensive, and objective
- Usually follows a clear methodological structure
- Goal:

Provide all necessary details
to understand, validate, and appreciate
the approach and findings

- Non-goal: cater to non-expert readers

(exception: text books, tutorial papers, background chapters of well-written theses)

→ to be read in a goal-oriented fashion, possibly skipping pieces

3. Introduction

The introductory section or chapter

The teaser and overview

The introduction positions and motivates the work, and advertises the key results.

- Here, the authors can tell the [story of their research](#)
- Goal 1: [Introduce reader to the paper's research question\(s\)](#):
 - Define and (very) briefly introduce research (sub)field
(could grow into an overview of the [recent] history of a field in long reports, e.g., theses or books)
 - From recent developments & state of the art to open problems
- Goal 2: [Reveal all relevant results of the work](#)
- Goal 3: [Clarify structure of remaining work \(if needed\)](#)

→ highly relevant, but requires very critical reception!

5. Conclusion

Final discussion, usually (much) shorter than introduction

Final words for the initiated reader

Concluding remarks that comment on the (technical) findings and outline limitations, directions & open questions.

- Should not just be a short reprise of the introduction
- Goal 1: [Discuss what was achieved](#)
 - To what extent are the results satisfactory answers to the questions?
 - Sometimes: another summary of the main results (e.g., using notation/terminology that would not have been understood in the introduction)
- Goal 2: [Provide an outlook to future research](#)
 - What remained open?
 - Which new directions of research do the findings enable?
- Research theses have an extended concluding chapter
- Books usually have conclusions per chapter (or not at all)

→ often informative even without reading whole work

6. References

The literature list of works that are referred to in the main text

Tell me who you cite, and I tell you who you are

A valuable resource to learn about a paper's scope and viewpoint.

- The references can be skimmed to understand the paper's topic and context
- Which approaches do the authors consider related?
- Are they aware of key works in the field?
- Do they have a wide or narrow view of the context of their research?
- Is their perception overly self-centric?

↪ extremely informative once you have a basic coordinate system of relevant researchers/topics/venues

Reading strategy (1)

Reading strategies differ, but reading start-to-end is rarely effective.

"I read every paper from the beginning until I hit the first Greek letter; then I jump to the conclusion." – An anonymous colleague

Which reading order makes most sense?

Possible general tactics:

- (1) Title, authors, affiliations – interesting?
- (2) Abstract – what's in this work (high level)?
- (3) Introduction (skip motivation if you are already motivated) – what did they do?
- (4) Conclusion – what was achieved?

+ scan references at any point for additional context

↪ then decide if to read further parts of the main research content

X. Related work

Dedicated section of some papers to discuss work in relation to other research

Positioning the work within the area

This is where authors distinguish their work from other approaches.

- Many papers just have minimal discussions of related work in introduction or conclusion
- If included, related work sections are typically right after the introduction or right before conclusion
- Good related work sections can be extremely valuable to new readers
- Bad related work sections are mindless name-dropping to please reviewers

↪ potentially useful to understand this work and also wider research area

Reading strategy (2)

If the paper is interesting, then you will likely need to read it several times

- You may have to revisit single words, sentences, paragraphs, and sections to fully understand them
- Speed reading techniques? Probably not – research papers are often highly compressed and have little redundancy that one could skip; small details do matter
- It can help to do two passes through a paper:
 - First pass: look for overarching ideas; accept that some details remain unclear
 - Second pass: try to completely follow every step
- Always read with a goal/question in mind and check if you still make progress towards that goal

Reading results

For every work you read, answer at least the following questions:

- (A) Where and when was the work published?
- (B) Who are the authors, and what are their affiliations?
- (C) Which problem did they address? What was their research question?
- (D) What did they propose as a solution?
- (E) Which methods did they use to validate their research?
- (F) What was the outcome? What did they learn?
- (G) Strong points of the work?
- (H) Weak points of the work?
- (I) Where would you place this work in comparison to its area/your own research?
- (J) What did you find hard to understand? Which further information do you need?

Beginners often do the following:

- Find a PDF online, without taking note about who, when and where.
- Spend all time & effort on understanding the proposed solution (D).

Summary

Finding computer science literature works best by combining several strategies using Google Scholar, DBLP, and possibly Scopus

It is easy to access almost any publication on the web

Effective literature (re)search is focussed and keeps track of its findings

The standard structure of research works supports efficient filtering & reading

What's next?

- Writing structure
- Writing style
- Typesetting