

Scholarship Skills

Andrew Black
Winter 2020

Week 1:
**Course Introduction, Background
Research, How to Read Papers**

All material © 1996–2020
David Maier, Tim Sheard,
Andrew Black,
Todd Leen.

Course Mechanics

Instructor

Andrew Black

apblack@pdx.edu, FAB 115-10. 503-725-2411

Time: Winter Quarter 2020

- Monday and Wednesday 08:15 – 09:30
Might be some variation in time
Final Exam slot: Monday 16th March 08:00–09:50
- Classroom: FMH B129
- Class Piazza
 - sign-up: <https://piazza.com/pdx/winter2020/cs669>
If you want a non-PSU address, email me
 - Home: <https://piazza.com/pdx/winter2020/cs669/home>
- Web Page: <http://web.cecs.pdx.edu/~black/ScholarshipSkills/>

Assessment Scheme

- **No Exams**
- **Exercises**
 - Usually due two classes after the one in which they are assigned.
 - Often used in-class the day they are due for discussion activities.
 - Graded on a “check off” basis
- **Projects**
 - Six projects:
 1. Annotated bibliography
 2. Revision of a 1-page summary
 3. 5-minute presentation of a paper
 4. 5-page paper on a technical topic
 5. Revision of 5-page paper
 6. 15-minute presentation based on paper above
 - Some presentations may be outside of regular class hours; all are encouraged to attend.
 - One week or more allowed for each project.
 - Graded with points

Grading

- **Grading Scheme**

- Exercises: 15%
- Projects: 85%
 - P1: Bibliography: 15%
 - P2: Revision of 1-page summary: 10%
 - P3: 5-min. presentation: 10%
 - P4: 5-page technical paper: 15%
 - P5: Revision of 5-page research paper: 20%
 - P6: 15-min. presentation: 15%

Texts

Required Texts:

- Lyn Dupré. Bugs in Writing: A Guide to Debugging Your Prose (2nd edition). Addison-Wesley, 1998. ISBN 020137921X
- Available used, but please get an edition more recent than the first.

Other Useful Books – see Class web page

Policies:

- **All deadlines are firm, unless explicitly extended.**
We will be as flexible as possible in accommodating special circumstances; but *advance notice* will make this a lot easier.
- **We follow the standard PSU guidelines for academic integrity.**
 - Discussion with your classmates is good!
 - Reading and critiquing each other's work is required.
 - Items turned in should be your own, individual work.
Be extremely careful to avoid plagiarism – anything copied verbatim from another source must be typographically distinguished as a quotation; the source must also be cited.
Paraphrasing and representing someone's ideas as your own are also plagiarism.

Academic Integrity

Students are expected to be honest in their academic dealings. Dishonesty is dealt with severely.

Homework. Make an effort to complete every piece of work. Pass in *only your own work.*

Writing assignments. Students are expected to do their own writing. Critique of your writing by others is encouraged, but you must *fix your mistakes on your own.*

PSU Code of Conduct

The Office of Student Affairs at Portland State University maintains a Student Code of Conduct: <https://www.pdx.edu/dos/psu-student-code-conduct>

The following constitutes conduct, proscribed by the university, for which a Student or Recognized Student Organization or group is subject to disciplinary action:

Academic Misconduct includes, but is not limited to: (a) cheating, (b) fraud, (c) plagiarism, such as word for word copying, using borrowed words or phrases from original text into new patterns without attribution, or paraphrasing another writer's ideas; (d) the buying or selling of all or any portion of course assignments and research papers; (e) performing academic assignments (including tests and examinations) in another person's stead; (f) unauthorized disclosure or receipt of academic information; (g) falsification of research data (h) unauthorized collaboration; (i) using the same paper or data for several assignments or courses without proper documentation; (j) unauthorized alteration of student records; and (k) academic sabotage, including destroying or obstructing another student's work

All expectations of the student conduct code will be enforced strictly in class.

Plagiarism

The Writing Center at Portland State University has prepared a guide to help cite references correctly and avoid plagiarism.

See <https://www.pdx.edu/writing-center/guided-tour-step-nine-checking-your-use-of-research>

Scholarly work resulting from plagiarism or cheating will receive no credit, and will result in additional disciplinary actions.

The Writing Center

Use it!

<https://www.pdx.edu/writing-center/>

Book an appointment in advance if you want help in a timely manner.

Projects

Projects involve a technical topic of your own choosing, reporting on prior work in that topic area.

1. Annotated Bibliography

- A bibliography of at least 10 references in your topic area. Citations and short written descriptions of each paper.

2. Revision of a 1 page summary

- 1-page (approx. 500 word) summary of one of the papers from your bibliography. (Creation of the original summary is an exercise)

3. 5-min. Presentation

- A short (five-minute) presentation that corresponds to your summary paper (Project 2).

Projects

4. First draft of 5-page Paper

- A background paper on a technical topic of 5 pages (~1500 words). It should be a complete paper with abstract, introduction and conclusions.

5. Revision of 5-page Paper

- Based on the feedback you have received, and other comments that you have collected, revise P4

6. 15-min. Presentation

- Prepare and deliver a 15-minute presentation corresponding to your 5-page paper. These will be scheduled during the final slot and at other times. You are encouraged to attend all the presentations.

Purpose of Course

Make you better scholars

- better researchers
- better writers
- better presenters
- better reviewers

Especially writing, and oral presentation

Learn to communicate

Written papers and articles

How to read

How to write

How to judge the writing of others

Oral Presentations

How to organize an oral presentation

How to present an oral presentation

How to listen to an oral presentation

Why?

***Why* are you taking this course?**

unprofessional, whereas in the 1990s many talks were unendurably awful. The growth in the use of good tools for presentations has been a key factor in this change.

Some cultural changes are less positive. A decade ago, I reported that many talks did not have a clear message and were merely a compilation of clever visuals, and that a well-described algorithm had become a welcome, rare exception; both these trends have persisted. Also, while the globalization of English has brought unquestionable benefits to international communication and collaboration, it means that today many papers are written, refereed, and published without passing through the hands of someone who is skilled in the language, so that even experienced researchers occasionally produce work that is far too hard to understand. The Web provides easy access to literature, but perhaps the necessity of using a library imposed a discipline that is now lacking, as past work appears to be increasingly neglected. Some issues concern the integrity of the scientific enterprise itself, such as the growing number of venues that publish work that doesn't meet even the most rudimentary standards.

The perspectives of all scientists are shaped by the research environments in which they work. My research has involved some theoretical studies, but the bulk of my work has been experimental. I appreciate theoretical work for its elegance, yet find it sterile when it is too detached from practical value. While experimental

Zobel Writing for Computer Science, p vi

Writing is not the end of the research process, but instead shapes it

ibid, p3

A scientist who can conceive of and explore interesting ideas in a rigorous way should be able to use much the same skills to solve the problem of how to explain and present those ideas to other people

ibid, p1

Background Research

Why Do Background Research?

- **Learn an area — not everything is in your textbooks**
 - Concepts, terminology
 - What are the important issues?
 - What are the main techniques?
 - e.g. modeling, simulation, proof techniques, statistics, algorithms
- **Who are the prominent people in the area?**
 - Where do they work?
 - What conferences do they attend?

Why Do Background Research?

- **To find a problem to work on**
 - What parts of a field are active
 - What do people consider to be today's significant questions?
- **To learn the customs of modern scientists**
 - What kinds of papers: theoretical, analysis, simulation, implementation, experience reports.
 - What forms do they take?
- **“Cultural” enrichment**

Why Do Background Research?

- **To situate your own work**
 - What are the trends in the field?
 - How does your work differ from the work of others?
 - Are you going against the grain?
- **Avoid repeating the work of others**
 - Most forums for research require novelty
 - Except: survey papers
- **Find out about the current “state of the art”**
 - Avoid getting scooped
- **When your research is complete**

you need to be confident that you have read and understood all of the ... literature that has a significant connection to your work [Zobel, p20]

Background Research

Finding relevant material

No single source that you can rely on

You have to use a combination of approaches

Not everything is on the web

- Not everything on the web is valuable – no quality control!
- Need to be able to track down print forms of books & papers.

Direct Search

- **Library (physical and digital)**
 - Books
 - Proceedings
 - Collections of papers (e.g., "Readings in X", Comm. ACM's first 25 years)
 - Journals
 - often have an annual index
 - on-line search of journal index very common
 - most journals have back issues on line (ACM Digital Library)
 - Online access to electronic collections
 - arXiv: preprints

Direct Search

- **Others' offices, group libraries**
 - Who subscribes to what?
 - Who goes to which conferences?
- **I usually try to find a recent article first**

Chaining backwards – can be easier than chaining forward

CiteSeerX: <http://citeseerx.ist.psu.edu/> is useful for chaining forward, as is Google Scholar.

Indexing and Abstract Services

- **ACM Guide to Computing Literature**
now part of the ACM Digital Library – includes non-ACM publications
- **On-line indexing services**
searchable via author, title, keywords
sometimes include complete abstracts
 - Inspec
 - PubMed
- **Area Indexes: e.g., DBLP**
- **Citation Indexes**
 - Web of Science, CiteSeerX
 - limited in what they index
 - Google Scholar

Organizing what you find

- **At some point, I try to identify the earliest articles in a field**
 - Sometimes these are the seminal works
 - other times, later works are more organized and make a better introduction – e.g., survey articles
- **What are the most-often cited works?**
 - Oft-cited works are probably important works
 - Google.com ranks papers found in a search by how often they are cited (on the Internet)
 - CiteSeer graphs citations by year
- **Start trying to determine seminal papers, definitive references**
 - How do the papers fit together?
 - Make a dependency graph of what papers cite others. This helps you view the development of a field

Leveraging What You've Already Found

- **Bibliographies of books and articles already obtained**
- **Survey articles**
 - ACM Computing Surveys
 - Foundations and Trends in Theoretical Computer Science
- **Bulletins and notices**
 - ACM SIGS
 - IEEE technical committees
 - bibliographies
 - research group archives
- **IEEE Computer, IEEE Software**
 - overview articles, for general audience

Other Internet Resources

- **Newsgroups**
 - DBWorld (databases)
- **SIG Publications**
 - Software Engineering notes
- **Web pages**
 - Research groups
 - General sources — on-line bibliographies
- **Directories and search engines**
 - National tech report library (NCSTRL)
<http://www.ncstrl.org/>
 - Computing Resource Repository (CoRR)
<https://arxiv.org/corr/home>

- **Direct solicitation**

In mailing lists

make sure that your questions are appropriate

Email

might ask if there is more recent work,
ask specific question

If you have a technical report that is more than a few months old, it's highly likely that it has been published someplace, and probably revised

Miscellaneous

- **Conference announcements**
 - Advance copies of papers
- **Manuals — some systems have no general publications**
- **Tutorial notes**
 - usually from conferences, or summer schools
 - sometimes have videos
- **Articles in encyclopedias, other references**
 - tend to become outdated quickly

Miscellaneous

- **Notification services**
 - you subscribe to “keywords”, and are notified when new works mentioning those keywords are published
- **Video-recorded lectures**
 - Increasing numbers now on the web
- **Laboratory annual reports**
- **Recent theses**
 - especially the related work sections
 - ProQuest: theses and dissertations
 - Available through PSU library
 - Abstracts to 1962, full documents to 1997

Organizing Source Materials

- **Make your own copy**
 - I find it useful to make a copy of sources I'm using, even if I have the journal or proceedings.
 - You can write on it
 - Its portable
 - Easy to group similar papers
- **Or: capture a digital copy.**
- **Make sure you capture citation info when you copy an article.**
- **“Fair-use”**
 - it is okay to make a single copy of copyright material for personal use.

Filing Materials

Don't over-organize your articles

- Rule of thumb: If you can find it in 3 minutes its ok
- **Maier files by broad category, and then by author. e.g.**
 - Object oriented databases
 - Type systems
 - Operating system kernels
- **Black files alphabetically by first author, but creates a searchable index**

If you categorize too finely, you start agonizing about where to put an article.

Maybe build a web page for yourself?

Prioritizing Sources to Read

In a new area, I usually try to find a recent paper with a good related-work section to start with, or an overview article

to get notation, (terminology and lay of the land)

journal paper, book chapter, magazine article usually better than a conference paper (for this purpose). Exception: invited papers, usually more of a survey, maybe longer than non-invited papers

Try to figure out which are the main groups working on an area, read a paper from each

get recent papers

authors are good at citing their own work (backchain)

may need to go to earlier papers for details

Next, look for older, highly cited papers.

Can also be enlightening to read a critique paper.

less common in CS than in social sciences or humanities

How to read a technical article

Reading is more than just passing your eyes over the words in consecutive order.

Reading is understanding

Why Are You Reading the Paper?

To answer a particular question

Looking for help solving a problem

Educating yourself

To learn if your own work or ideas are novel

It has been assigned by someone

Reviewing or abstracting it

Preparing a presentation on it

Looking for inspiration

Waiting for the dentist, got nothing better to do

What should I get out of reading a paper?

- 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 development**
- Details of any examples**

Thanks to Jim Hook for this list.

How to Read

**Approach a book or paper in several passes,
sometimes spread over time.**

How many passes?

Over what time period?

Pass 1

Abstract

- to determine relevance
- to determine kind of paper

Pictures

tables, graphs, and diagrams
concepts

and References

do I recognize them or know what they're about?

Might never go into more detail than this.

Pass 2

Introduction, Chapters, Section beginnings, Examples, Summary

to determine organization and content
might decide on this basis only to read parts
of the book or paper

figure out if authors are good writers

if you have a choice, why not read the well written
papers?

help set priority/payoff

Pass 3

Full reading

Often **take notes** during this phase

Try to capture main contributions

What distinguishes it from other work?

What it is similar to?

“like Knuth-Morris-Pratt, but matching from the right of the pattern”

What are assumptions, or line of development?

Questions on what I don't understand

write these right on the front page or prominently in your notes (for when you read it again)

Pass 4

Detailed study

Go back over hard or unclear parts

Frame my own questions about what is left un-answered.

Cracking a Tough Paper

Have someone who understands it explain it

Find an article by same authors to a more general audience

Find a later article by same author(s)

Make up (additional) examples. Good way to

master definitions, understand algorithms, check out theorems.

Working through an example is a good way to explain or understand something quickly.

Reading groups

Draw a picture

Write code

Cracking a Tough Paper 2

Recast portions in alternative notation, such as logic, or a functional program

“Guess” what is going on.

See if your hypothesis is consistent with the rest of the paper

Try to find a counterexample

Often, this is the way proofs are first found

Send questions to the author

Try to be specific—

Cast your question so there is a 2-3 line answer

Do not force the author to go back and read their original paper

Learn to Do Late Binding

- **Useful skill: to be able to hold content without understanding it —**
 - Plow ahead if you don't understand
- **Later insights or definitions make it click**
- **Think back over papers periodically—**
 - One more reason to write questions on the front of the paper