ERSP: Independent Learning

S14 Week 4
Today’s objectives

• Teach your classmate something technical, to measure your own learning
• Describe important aspects of doing a literature search in computer science

Reminder: Don’t forget about your logs! (I’ll be sending you some feedback/gentle reminders)
Announcements: Possible Summer Research Opportunity

• We might have a limited number of research opportunities available this summer
• These opportunities would be:
  • Between 6-10 weeks
  • Full-time (40 hours/week), or possibly part time
  • For pay
• How to apply:
  • Send me an email with the subject “ERSP internship application” describing:
    1. Your interest in doing ERSP research over the summer
    2. Your other summer commitments and their dates
    3. The dates you are available to work
    4. Whether you can work full time or part time
The purpose of background reading at this two stages:

- Develop a sense of the research field
- Determine which specific objectives are still open problems
- Figure out what research problems others have addressed
- Learn about technical material you need to understand the research
Pre-session survey

• Fill out the pre-session survey about what you will be teaching your partner today
• Then write your name and your general topic on the board
Teaching session!

- Spend 30 minutes teaching your partner the technical skill you learned
- Students: you should really try to learn the material. Feel free to ask questions!
Post-session survey

• Both the teacher and the student fill out the post-session survey
Teaching session!

- Switch partners! Spend 30 minutes teaching your partner the technical skill you learned
- Students: you should really try to learn the material. Feel free to ask questions!
Post-session survey

- Both the teacher and the student fill out the post-session survey
Debrief

• Compare your pre- and post-session surveys with each other. Discuss:
  • What went well
  • What could have gone better
  • Any matches or mismatches between what the teacher was teaching and what the learner was learning
For next week: A Literature Search

- You just read a paper and you have a brilliant idea about a follow-up study/improvement. But has it already been done? Will people care? How do you know?
For next week: A Literature Search

• You just read a paper and you have a brilliant idea about a follow-up study/improvement. But has it already been done? Will people care? How do you know?
• Search the literature!
Where we are

The purpose of background reading at this two stages:

- **Develop a sense of the research field**
- **Determine which specific objectives are still open problems**
- Figure out what research problems others have addressed
- Learn about technical material you need to understand the research
For next week: A Literature Search

• Your goal for next week is to expand out from the paper you have read for last week to get a broader sense of the field. This is a fundamental skill in doing research, and usually the first thing that any researcher in a new area does.

• But how to do this? Google?

• Let’s try it!
• What problem does this research solve?

the set of most frequent queries in a given day. Most such computations are conceptually straightforward. However, the input data is usually large and the computations have to be distributed across hundreds or thousands of machines in order to finish in a reasonable amount of time. The issues of how to parallelize the computation, distribute the data, and handle failures conspire to obscure the original simple computation with large amounts of complex code to deal with these issues.

As a reaction to this complexity, we designed a new abstraction that allows us to express the simple computations we were trying to perform but hides the messy details of parallelization, fault tolerance, data distribution and load balancing in a library. Our abstraction is inspired by the
MapReduce: Simplified Data Processing on Large Clusters

by Jeffrey Dean and Sanjay Ghemawat

- What other solutions previously existed to solve this problem? Where could you look to find the answer?
• What other solutions previously existed to solve this problem? Where could you look to find the answer?

7 Related Work
Many systems have provided restricted programming models and used the restrictions to parallelize the computation automatically. For example, an associative function can be computed over all prefixes of an $N$ element array in $\log N$ time on $N$ processors using parallel prefix computations [6, 11, 14]. MapReduce can be considered a simplification and distillation of

MapReduce: Simplified Data Processing on Large Clusters

What is OSDI? Where can I find the papers?
What is OSDI? Where can I find the papers?
What other solutions have arisen since, that solve the same problems? Where could I look to figure this out?
What other solutions have arisen since, that solve the same problems? Where could I look to figure this out?
Main sources for finding related work

- Search backward: What was done before?
- Search forward: What has built on this work?
- Search the conference/journal: What’s being done and classified as similar by the community?
- Search the authors: What else are these same authors doing?
A guide to doing a literature search (Quality, not quantity)

- Quality can be measured by:
  - How many others cite the paper/who cites the paper
  - What conference/journal it was published in
  - Your own assessment of the work (harder, since you’re not an expert)

- Know the important journals/conferences:
  - ACM, IEEE
  - Most fields have a “flagship” conference or journal. Figure out what it is.

- Search backward references and forward

- Tools:
  - Google scholar, citeseer (http://citeseerx.ist.psu.edu), raw Google (with care!)