Shaping a Research Project

• A research project proceeds very differently from a software development project
• For a typical software development project, the task is to *find an efficient, workable solution* to a problem
• For a typical research project, the task is to *measure the quality of the solution* to a problem
• Set milestones to explicitly consider what is needed at the end, and then reason backwards
Shaping a Research Project

• The origin of a research investigation is a moment of insight
• You might choose to explore an idea that is intriguing, or has the potential to lead to something new, or might contradict conventional wisdom
• The final outcome is an objective scientific report, but curiosity and guesswork establishes the direction
Shaping a Research Project

• As early as possible, determine the answers to the following questions:
  – What is the broad problem to be investigated?
  – What are the specific initial activities to undertake and the outcomes to pursue?
  – What makes these questions interesting?
Shaping a Research Project

• Most research is incremental
  – Most research is an improvement on existing work, or a variation that improves, extends, or replaces work done by others

• What is the scope of the increment?
  – A trivial step that does no more than explore the obvious solution to a simple problem is not worth investigating
  – There must be a challenge and the possibility of unexpected discovery
Shaping a Research Project

- Planning
- Reading
- Implementing
- Testing
- Analyzing
- Thinking critically
- Writing
- Presenting
Planning a Research Project

• Example of a research hypothesis:
  – Reducing the number of memory accesses using a cache can make a program run faster, even if the number of instructions executed remains the same. Is it possible to make better use of CPU cache to reduce computational costs?

• Example of a research hypothesis:
  – Can we eliminate the copy to the auxiliary array in mergesort by switching the role of the input and the auxiliary array in each recursive call?
Planning a Research Project

• How do you come up with a research hypothesis?
  – Suppose P-lists are a well-known data structure used for a range of applications, in particular as an in-memory search structure that is fast and compact. You have developed a new data structure called a Q-list. Formal analysis has shown that the two structures have the same asymptotic complexity in both space and time, but you intuitively believe that the Q-list is superior in practice.
Planning a Research Project

- A non-testable hypothesis:
  - Q-lists are superior to P-lists.

- A testable hypothesis:
  - As an in-memory search structure for large data sets, Q-lists are faster and more compact than P-lists.

- Further qualification may be necessary:
  - We assume there is a skew access pattern, that is, the majority of the accesses will be to a small proportion of the data.
Planning a Research Project

• A testable hypothesis should be **concrete** and should be **capable of being disproved**

• Example of a vague hypothesis:
  – *Q-list performance is superior to P-list performance*

• Examples of non-disprovable hypotheses:
  – *Our proposed query language is relatively easy to learn*
  – *Our search engine can find interesting web pages in response to queries*
Planning a Research Project

• Ultimately, the hypothesis must be able to make successful predictions

• There is a vast difference between an observation and a tested hypothesis:
  – Observation: The algorithm worked on our data.
  – Tested hypothesis: The algorithm was predicted to work on any data of this class, and this prediction has been confirmed on our data.
Planning a Research Project

• The research goal is to test the hypothesis
  – Array-based data structures for sorting algorithms make better use of cache than tree-based data structures

• The phenomenon that should be observed if this hypothesis is correct is a trend:
  – As the number of items to be sorted increases, the tree-based method shows an increasingly high rate of cache misses as compared to the array-based method

• The data to be collected and analyzed is the number of cache misses for several sets of items to be sorted
Finding Relevant Literature

• A research project necessarily builds upon a body of prior work
• The “doing” and “describing” of research requires a thorough knowledge of the work of others
• Visit websites of research groups and researchers working in the area
• Look up references in research papers
• Browse recent issues of journals and conferences in the area
• Use obvious search terms to explore the web
Finding Relevant Literature

• Search digital libraries

• Explore conference websites and conference programs

• Use a citation index

• Go to the library

• Discuss your work with as many people as possible
Reading Literature

• Published papers are not textbooks and should not be read in the same way as a chapter in a textbook
• Begin with a brief browse through the paper to identify whether it is relevant to your particular project
• How do you know if you should believe what you read?
• Capture information about each paper that you expect to cite, or that may be peripherally relevant
• Classify the paper and cross-index it with other papers on the same topic
• Be organized from the start
Reading Literature

• Evaluate research papers by asking questions:
  – What is the main result?
  – How precise are the claims?
  – How could the outcomes be used?
  – What is the evidence?
  – How was the evidence gathered?
  – How were measurements taken?
  – How carefully are the algorithms and experiments described?
  – Why is the paper trustworthy?
  – Has the right background literature been discussed?
  – What would reproduction of the results involve?
Criteria for Judging Research

• Intellectual contribution is the main criteria for judging the quality of a research paper
  – A paper is a contribution if it has originality and validity

• The presence of a critical analysis is also important
  – Authors should correctly identify the strengths, weaknesses, and implications of their own work
Writing Up Research

• Remember, the paper is the *only part of the work that survives or is assessed*!

• Start writing early, typically before the project’s half-way mark
  – Delay increases the time between having ideas and having to write about them
  – The reading of relevant papers will never end
Writing Up Research

• Writing is the major stimulus for research
  – Fresh ideas will develop
  – Vague concepts and misunderstandings will come to the surface
  – The appropriate form for proofs and experiments will take shape
  – Gaps in the research will become apparent
Writing Up Research

- The first task for the write-up is to identify your aims
  - Write down everything that motivated you to start on the project
- Next, define the scope of the work that you plan to write up
  - Make choices about what should be included
Writing Up Research

• Who are the readers? Are you writing for specialists in your area, a thesis committee, or a general computer science audience?
• What is the key background work that has to be discussed?
• What assumptions or definitions need to be formalized before the main idea can be presented?
• Which results are the most surprising?
• What is the one result that other researchers might adopt?
• Are other outcomes independent enough to be published separately?
Telling a Story

• Once the material for a paper has been collected, it has to be organized into a coherent self-contained narrative

• What does the reader need to learn?
  – An effective paper educates its readers
  – Everything from the introduction to the conclusion should have the logical flow of a narrative
Telling a Story

• A narrative is a walk through the ideas and outcomes
  – “It is like a guided tour through a gallery, in which each room contains something new for the readers to comprehend”
  – There is also an expectation of logical closure
Telling a Story

• The logical order is usually in the form of a chain:
  – Start with a problem statement
  – Next, review the previous solutions and explain their drawbacks
  – Then present the new solution
  – Finally, demonstrate that the solution improves on its predecessors
Telling a Story

• When describing specific results, it is helpful to begin with a brief overview of whatever has been observed
  – The first sentence summarizes the specific results
  – The next few sentences provide the context
  – The remainder explains the implications
Organization

1. Title, author(s), and other front matter
2. Abstract
3. Introduction
4. Background (literature review)
5. Methodology
6. Results
7. Discussion
8. Conclusions
9. References
Title and Author(s)

• Include name(s), affiliations, and addresses
• Consider using your middle initial
• Use the same style on all of your papers
• Search terms, keywords, and key phrases are usually included
The Abstract

- An abstract is typically a single paragraph of about 50 to 100 words.
- It is a concise summary of the paper’s aims, scope, and conclusions.
- The function of the abstract is to allow readers to judge whether or not the paper is of relevance to them.
- Be as precise as possible:
  - Instead of writing “space requirements can be significantly reduced”, write “space requirements were reduced by 60%”.
The Introduction

• The introduction can be thought of as an expanded version of the abstract
• It should describe the paper’s topic, the problem being studied, references to key papers, the approach to the solution, the scope and limitations of the solution, and the outcomes
The Introduction

• General outline of the Introduction:
  1. A general statement introducing the broad research area of the particular topic being investigated
  2. An explanation of the specific problem (difficulty, obstacle, challenge) to be solved
  3. A brief review of existing or standard solutions to this problem and their limitations
  4. An outline of the proposed new solution
  5. A summary of how the solution was evaluated and what the outcomes of the evaluation were
Literature Review

• The literature review (background material) serves three main purposes:
  1. It demonstrates that your work is new
  2. It demonstrates your knowledge of the research area
  3. It serves as a pointer to background knowledge
Literature Review

• Don’t use citations to support common knowledge
  – For example, the use of a binary tree in an algorithm does not require a reference to a data structures textbook
Literature Review

• Results from cited papers should be described fairly and accurately
  – Sometimes cited papers are not accessible to the reader
  – Any criticism should be based on sound, logical argument
  – Only rarely is it acceptable to offer opinions, and it is never acceptable to use flattery or scorn
Literature Review

• Do not use statements that might be interpreted as pejorative
  – Instead of
    • Robinson’s theory suggests that a cycle of handshaking can be eliminated, but he did not perform experiments to confirm his results (Robinson, 2011).
  – Use
    • Robinson’s theory suggests that a cycle of handshaking can be eliminated (Robinson, 2011), but as yet there is no experimental confirmation.
Literature Review

• Be sure to show the relationship to your own work
  – For example, a reference might show a general case, but you use a special case

• If you claim that concepts are equivalent, ensure that the equivalence is clear to the reader
Methodology

• Describe your approach in sufficient detail so that the experimental results can be replicated and verified

• A flow-chart or similar type of diagram helps to convey the steps that were involved

• Each step must be described in detail

• Include the source of all data (even if you generated it yourself) along with a description of the data and any other special equipment (hardware, software, etc.) that is necessary to run the experiment
Results

• Evidence needs to be convincing
• In general, there are four kinds of evidence that can be used to support an hypothesis:
  – Formal proof
  – Modeling
  – Simulation
  – Experimentation
Formal Proof

• A formal proof is a mathematical argument that the hypothesis is correct

• Not all hypotheses are amenable to formal proof, particularly hypotheses that involve the real world in some way
  – Human behavior is intrinsic to questions about interface design
  – System properties can be intractably complex
Modeling

• A model is an abstraction of the hypothesis, or some component of the hypothesis such as an algorithm whose properties are being considered
• A model is usually a mathematical description of the properties under investigation
• There should be a demonstration that the model and the hypothesis do indeed correspond
• The model should be valid and verifiable
Simulation

- A simulation is an implementation of a simplified form of the hypothesis
  - The difficulties of a full implementation are approximated
  - For example, a parallel algorithm could be tested on a sequential machine by using an interpreter that counts machine cycles and communication costs between simulated processors
Experimentation

- An experiment is a full test of the hypothesis, based on an implementation of the proposal, and on real (or at least realistic) data
  - An experiment should be conducted based on predictions made by a model, so that it confirms some expected behavior
  - An experiment should be severe – look for tests that are likely to fail if the hypothesis is false
Defending the Results

• Anticipate objections and doubts

• Example hypothesis:
  – This new string hashing algorithm is fast because it doesn’t use multiplication or division

• Possible objections:
  – But multiplication and division are not a problem on pipelined machines with floating-point accelerators
  – Is there also an array look-up? That can be slow.
  – What happens if the hash table size is not a power of 2?
Discussion

• Issues that need to be addressed:
  – Will the reader believe that the results are new?
    • Only if the authors have done a careful literature review, and fully explored and explained previous relevant work
  – Will the reader believe that the results are sensible?
    • The author should identify potential problems, and either concede them (with an explanation) or dismiss them through some cogent argument
Conclusion

- The conclusion draws together all of the topics discussed in the paper
- The conclusion should include a concise statement of the paper’s important results and an explanation of their significance
- Restate the limitation of the work
  - Shortcomings in the experiments, problems that the theory does not address, etc.
Conclusion

• The conclusion is an opportunity for the researcher to look beyond the current context to other problems that were not addressed, questions that were not answered, variations that could be explored

• Future work or possible extensions to this work should be included in the conclusion
References and Citations

• References should not be anonymous
  – Instead of
    • Other work [16] shows that . . .
  – Use
    • Marsden [16] has used an approach in which . . .
    • Other work (Marsden 1991) has used an approach in which . . .
References and Citations

• When discussing a reference with more than three authors, all but the first author’s name can be replaced by *et al.*
  – *Howers, Mann, Thompson, and Wills [9] provide another example.*
  – *Howers et al. [9] provide another example.*

• Do not use *et al.* in the Bibliography