Designing Well-Scoped Undergraduate Research Projects

Kelly Shaw
kas10@williams.edu
CRA UR2PhD
May 13, 2024
This presentation and associated resources were developed for the Computing Research Association's UR2PhD program by Kelly Shaw, Williams College, Department of Computer Science.

UR2PhD: Graduate Student Mentor Training Course © 2023 by Computing Research Association’s UR2PhD Program is licensed under Creative Commons Attribution-ShareAlike 4.0 International. To view a copy of this license, visit https://creativecommons.org/licenses/by-sa/4.0/
# Why Advise Undergraduate Researchers?

<table>
<thead>
<tr>
<th>Benefits</th>
<th>Concerns</th>
</tr>
</thead>
<tbody>
<tr>
<td>● Additional contributors can help advance your research project</td>
<td>● UGrads work more slowly than you</td>
</tr>
<tr>
<td>● Inspire and prepare the next generation of researchers</td>
<td>● UGrads require frequent guidance and oversight</td>
</tr>
<tr>
<td>● Instill skills and experience in students for their careers</td>
<td>● UGrads disappear quickly, abandoning projects before completion</td>
</tr>
<tr>
<td>● Improve your mentoring skills</td>
<td>● Students might have mismatched background/skills for a project</td>
</tr>
<tr>
<td>● Recruit future graduate students</td>
<td>● Bad research experiences turn UGrads away from research</td>
</tr>
</tbody>
</table>
Why Advise Undergraduate Researchers?

Planning Unlocks Benefits

Design projects to advance your research

Consider student skills in project design

Create well-scoped projects for given timeframe with frequent, concrete deliverables

Stitch multiple student research experiences together
Two Approaches for Beginner Project Design

**Approach 1**
Start a new exploratory project related to other research group efforts

**Approach 2**
Join an existing group project and work on an independent part
Suggestions for New Exploratory Projects

Projects with an initial hypothesis that can be evaluated empirically and whose results can lead to new questions to explore

Examples:

- Replicate an existing study and extend it using new data with different characteristics
- Study state of the art software tools on a new data set with different characteristics than the software was originally designed.
- Conduct a human subject study on a specific user interface or tool
- Modify and evaluate existing software to implement a new idea based on a hypothesis
- Analyze a new data set for specific characteristics
- Evaluate a system created for automating a task
Suggestions for Joining an Existing Project

Projects may involve starting in the middle of an existing project:

- Data collection or generation for existing system or study
  - Perform interviews or administer survey
  - Collect data using existing software on new data inputs
  - Create scripts to generate data
- Analysis of already collected data
  - Write scripts to retrieve and analyze existing data
- Visualization of already collected data
  - Write scripts to retrieve and visualize data
- Implementation of a new instrument
  - Add new functionality to an existing system
  - Create a software implementation of a theoretical algorithm
  - Create a new human subject survey
  - Install new data sets or applications for use with existing system
Suggestions for Joining an Existing Project

Projects may involve **starting in the middle** of an existing project

Examples:

- **Data collection or generation** for existing system or study
  - Perform interviews or administer survey
  - Collect data using existing software on new data inputs
  - Create scripts to generate data

- **Analysis** of already collected data
  - Write scripts to retrieve and analyze existing data

- **Visualization** of already collected data
  - Write scripts to retrieve and visualize data

- **Implementation** of a new instrument
  - Add new functionality to an existing system
  - Create a software implementation of a theoretical algorithm
  - Create a new human subject survey
  - Install new data sets or applications for use with existing system
Both Approaches Involve Subsets of Research Activities

- Search literature for papers
- Read technical papers
- Create instrument
  - Modify existing and/or create new
- Collect or generate data
- Analyze data
- Visualize data
Tips for Research Project **Design**

- Design authentic research projects that matter to your research group
- Design projects not on the critical path
- Start at a theoretical level students can understand
- Draw on skills students already possess or can learn quickly
- Design projects with clearly defined, frequent milestones
- Keep scope modest and flexible enough to be simplified or extended
- Create low-risk projects with a good chance of producing results within given time frame
Tips for Designing the Research Project's **Scope**

Turn the research problem into explicit, actionable steps

- Identify **deliverables or subgoals** that are contributions that move the project forward
- Identify steps that require **advanced knowledge** and plan to do them in advance of project start
- Identify **repetitive steps** that could be independent subgoals / deliverables
- Identify **skills or knowledge** to be learned
Tips for Designing the Research Project's **Timing**

- Create a **dependency graph** of tasks
  - Including concrete deliverables and skill learning
- Realistically **estimate time** required for each task
  - Estimate time for each individual repetitive step
- Map tasks to a **timeline** using a Gantt chart
  - For repetitive steps, indicate that each is its own subgoal
- **Adjust** deliverables to fit work into time frame
  - Identify and truncate decomposable or repetitive tasks
  - Employ team of student researchers to parallelize repetitive tasks
  - Turn subgoal into a final goal
Example Project

A paper that you just presented implemented a new caching policy on a multiprocessor system. For 2 applications studied, the approach did not perform as well due to network congestion.

**Goal:** Understand 2 applications' characteristics that caused network congestion.
Example Project

A paper that you just presented implemented a new caching policy on a multiprocessor system. For 2 applications studied, the approach did not perform as well due to network congestion.

Goal: Understand 2 applications' characteristics that caused network congestion.

Experimental framework:

- Large, complicated C code that simulates system and outputs results to files.
- Applications written in C.

Student info:

- Mostly experienced in Python programming.
- Some basic C programming skills.
- Taken introductory computer organization course.
Project Steps (Broad Strokes)

- Add functionality to simulator to collect data needed to examine network traffic and to output results to files
- Create code to analyze data in output files to identify network congestion
- Collect and analyze network traffic data for 2 applications, mapping network congestion back to the application source code

**Key insights**

- Student C programming skill not up to simulator modification
- Student will need to learn about networking and specific caching policy
- Student strength is Python programming skill
Project Steps: Pre-internship preparation

- Identification of papers / documentation for student to read
  - Research group orientation tutorials
  - Paper/video describing how to read a technical paper
  - Paper describing the caching policies and showing application results
  - Paper providing basic description of the 2 applications
  - Documentation describing the system and how to run applications on it
  - Background reading about network traffic in multiprocessor systems

- Implementation of data collection in system needed to explain network traffic resulting in congestion and output of data to files

- Creation of single microbenchmark application and generation of its output data

- Identification of Python libraries needed for analyzing output data
Project Steps for Undergraduate Researcher

- Read/discuss provided literature
- Create Python scripts for analysing network traffic in output files
- Write microbenchmark applications in C to test data collection and analysis code
- Provide feedback on correctness of data collection code (iterate)
- Run and analyze 2 applications' network traffic behavior
Project Steps for Undergraduate Researcher

- Learn to read a technical paper
- Read/discuss provided literature
- Learn to use git
- Learn to save artifacts / data
- Learn to use Python libraries for data analysis
- Create Python scripts for analysing network traffic in output files
- Learn to run system to generate data
- Write microbenchmark applications in C to test data collection and analysis code
- Provide feedback on correctness of data collection code (iterate)
- Run and analyze 2 applications' network traffic behavior
Create Dependency Graph With Timing

1. Learn to read papers  
   1 day

2. Read papers / background material  
   4 days

3. Learn to use Python library  
   1 day

4. Learn to save artifacts  
   <1 day

5. Learn to run system  
   <1 day

6. Create Python scripts for analyzing output  
   4 days

7. Create and debug microbenchmarks  
   2 days each

8. Debug system data collection code  
   1 day each error

9. Collect and analyze output data for app  
   2 week each
Create Timeline By Mapping to Gantt Chart

Shorter duration tasks early
Create Timeline By Mapping to Gantt Chart

Can add/remove microbenchmarks
Create Timeline By Mapping to Gantt Chart

Research progress still made even if these aren't completed
Additional Tips For Constructing a Feasible Timeline

- Identify tasks that contain rabbit holes and plan to proactively guide student through those tasks
- Actively engage with students and give feedback on each deliverable
- For large deliverables, create periodic check ins to give feedback
- Be willing to make a subgoal the final goal if time runs short
- Make sure all deliverables are documented and checked in for future student (at the time of delivery, not the end of the research experience)
Ensuring Project Completion: Stitching

Designing each project with many, small, concrete deliverables enables you to stitch multiple students' work together

- Same or new student could start at a subgoal that became a final goal for previous undergraduate research experience

Requires in progress sharing, documentation, and review of deliverables, including

- instruments that were created
- collected data and description of method collection
- analyses performed and data used in analysis
- figures generated and data used
- …
Guided Worksheet to Help With Process

- Ideas for brainstorming projects
- Questions guiding
  - Decomposition into deliverables
  - Identification of needed skills / knowledge acquisition
  - Identification of resources researcher will need
- Tips for mapping items to a timeline
- Advice for stitching multiple projects together
Guided Worksheet to Help With Process

- Ideas for brainstorming projects
- Questions guiding
  - Decomposition into deliverables
  - Identification of needed skills / knowledge acquisition
  - Identification of resources researcher will need
- Tips for mapping items to a timeline
- Advice for stitching multiple projects together

Also available: Completed worksheet using example from this presentation
Quick Tips for Mentoring Undergraduates

- Give students frequent and early wins
- Help students learn when to ask for help
- Help students develop resilience to failure and understand that research is open-ended
- Work with students to adapt timeline and goals as research unfolds
- Consider working with groups of students
- Consider who else in your lab can provide support to students
- Set expectations, including for meetings and communication
- Get to know students as multifaceted people
References

GRADUATE STUDENT MENTOR TRAINING COURSE

An 11-session course that provides mentors with the tools they need to cultivate strong, effective, and productive environments.

To be eligible for the course, students must:
- Be enrolled in a graduate program
- Be actively engaged in mentoring a group of undergraduate researchers

Priority will be given to mentors of participants in the undergraduate research methods course.

Available in the summer and the fall.

UNDERGRADUATE RESEARCH METHODS COURSE

An 11-session course that equips undergraduates with the skills they need to be successful as researchers.

To apply to the course, students need to:
- Identify a research team of 2-4 undergraduates and a mentor
- Identify a research project

No research experience is required. Applications are accepted on a rolling basis.

Available in the summer and the fall.

UNDERGRADUATE RESEARCH ENGAGEMENT & AWARENESS WORKSHOPS

One-hour interactive, virtual workshops designed for students to learn about research careers and pathways.

Available in the spring and the fall.

COMPUTING RESEARCH ENGAGEMENT & AWARENESS WORKSHOPS

One-hour interactive, virtual workshops designed for students to learn about research careers and pathways.

Available in the spring and the fall.

GRADUATE SCHOOL APPLICATION WORKSHOPS FOR COMPUTING PROGRAMS

One-hour interactive, virtual workshops designed for students to prepare to draft compelling graduate school applications.

Available in the fall.