

CSE 501: Implementation of Programming Languages
Autumn 2010

Logistics

Course website: <http://www.cs.washington.edu/501/>

Class meetings: TuTh 12:00–1:20, room CSE 403

Lecturer: Michael Ernst

<http://www.cs.washington.edu/homes/mernst/>

Office hours: by appointment, see <http://www.cs.washington.edu/homes/mernst/contact-calendar.html>

While there are no set office hours, I welcome discussions with students. Please don't be shy about contacting me if you wish to meet!

no TA

1 Structure of the course

Many compiler classes cover all parts of a compiler, from beginning (parsing) to end (code generation), and the assignments have you create a compiler for a toy language, again from start to finish.

CSE 501 is different. It explicitly ignores parts of a compiler that are very well-understood and that you can learn on your own, such as the front end (lexical analysis and parsing). It is not a broad course, but focuses on specific topics, particularly ones that are more recent.

For a schedule, including a list of specific topics that may be covered, see the course webpage.

2 Prerequisites

You do not have to have already taken a compiler class. However, undergraduate are recommended to do so.

3 Grades

Grades will be assigned based on the project (75%), class participation including the class presentation (20%), and instructor discretion (5%). You will submit 5 versions of your project writeup, counting the proposal, and these will be weighted 10%, 10%, 10%, 15%, 30%. As this is a graduate class, the class will be A-centered, but students are not guaranteed a grade of A.

4 Papers

4.1 Discussions

CSE 501 is a graduate paper-reading seminar. Each class session will begin with a brief discussion and presentation of material, after (or during) which the floor will be open to rebuttals, discussion of related work, criticism, brainstorming about follow-on research, etc. At this level in your career, you should no longer be a passive listener to lectures but an active participant in the discussion.

Here are two motivations for participating:

- Class participation is required, and is a component of your grade. It is also a way to help the instructor know that you are doing the required preparation. Unacceptable participation will result in a low grade.

- When you actively participate, you will learn the material more effectively than by just sitting and listening to lectures. This is true even if – nay, especially if – you do not fully understand the material. The class sessions are structured as discussions to challenge you to think, make connections, or just be clear about what you do and do not understand. The latter is perfectly acceptable in terms of participation, and will also help you to learn the material better and thus to get more out of the class.

Some people are reluctant to participate. Here are some reasons:

- You may believe that no one else has the same questions you do, or that your level of understanding is lower than that of other students, or that your preparation is inadequate, or that everyone else understands the material. Based on my extensive teaching experience, it is highly unlikely that any of these are true; rather, you are suffering from Imposter Syndrome, which is *extremely* common in academics. It is likely that there are one or two people in the class who already understand a given topic well, but the vast majority of your classmates are greatly relieved when you ask the same "stupid" question that they wish they had the courage to ask – I've seen this in their faces many times! Furthermore, the class does not exist for the benefit of those who already understand the material, but to explain it to those who do not yet understand it. So long as you have made a good-faith effort to understand the material, your question will not be stupid or out of place.
- You may be congenitally shy, or come from a culture or educational background that discourages questions, or be unsure of your English. While these may be true reasons (unlike the false ones noted immediately above), they are not excuses. To be a success in a technical field, it is a requirement to have the courage to ask questions and to speak up in groups. If you cannot do this, then no matter how great your accomplishments, you will never fulfill your potential. So, you should treat this class as teaching you that critical skill for your career, and let go of your inhibitions to class participation.

I am happy, at any time, to talk about these or any other issues; just make an appointment with me.

4.2 Summaries

To help you prepare, you will write a one-paragraph commentary on each paper, and submit it at least 24 hours *before* the class meets to discuss the paper. You will post your commentary to the course webpage for viewing by the instructor and by other students. The commentary should reflect your understanding and analysis of the issues raised by the paper, and should also help direct (both your and others') preparation for in-class discussion.

You may write the commentary in whatever style you prefer that meets the goals listed above. One good format for the commentary is to critique the paper, listing the following three points: its biggest contribution (and, briefly, why that result was not already obvious), its biggest mistake (in motivation, methodology, algorithm, data analysis, conclusions, or some other area), and the biggest question that it raises (or the most interesting and important follow-on work that it suggests). Another acceptable format is to summarize the paper, describing its thesis, approach, and conclusions, and stating why it is significant. The commentary should also list questions that you have about the paper, such as about technical points or connections to related work.

It's OK if you read the paper and there are issues you do not understand. Please ask questions about those issues — both in your summary and in class — and we will all gain by the discussion. It's best to explain *why* something makes no sense to you. For example, don't just say, "I didn't understand section 2", but state where there is a logical fallacy, a conclusion that does not follow, an undefined term, etc. The lecturer will use these questions to help shape the lectures.

I also encourage you to use the summaries to ask questions. If you have a question, it is likely that many other people have the same question but are too shy (or vain, or insecure) to ask it; they will appreciate your raising the point. However, do come to class prepared: carefully read the paper and get as much as you can out of it on your own. Doing so will make the class time that much more productive.

You will have access to all the other students' submissions. Please read them, because reading the other summaries is a good way for you to get perspective. You can see what you missed (or what other people missed), and whether you agree with their take on the key ideas. It will help to make the class sessions more productive.

Please submit your summary in PDF form, and include your name.

4.3 Presentations

Each student will present one paper during the term. The presentations will be done by pairs of students if there are enough students registered. Presenters will meet with the lecturer a week before the class meeting to receive feedback and improve their presentation. You should come to that meeting prepared with questions, not just expecting to run through all of your slides. The goal of your presentation should be that the class understands the main ideas afterward. (The goal of your presentation is *not* to prove that you know the material!)

5 Projects

The centerpiece of the course is a term-long project, done in groups of 2–4 students. The project should make a research contribution in the area of programming language implementation. As a research contribution, it should answer a question whose answer no one previously knew.

A non-exhaustive list of types of acceptable projects are:

- proposing and evaluating a fundamental new technique
- developing and assessing new algorithms to replace currently-used ones
- translating a methodology to a new problem domain
- applying known techniques to new problem domains, such as operating systems, networks, embedded systems, security, biology, aerospace, etc.
- evaluation of existing techniques or tools, via case studies or controlled experiments
- port an existing tool to a new domain (e.g., a new programming language or a new version of one)
- implementation of a proposed but never implemented technique
- address an item from the “future work” or “limitations” section of a published paper
- reimplement or reproduce a research result in a new context
- perform a proper experimental evaluation, for a promising paper that lacks experimental evaluation, or whose methodology is flawed

You should also consider building on your strengths and finding ways to leverage other work that you are doing. For instance, if you do work in AI, you might consider how machine learning could be applied to problems of language implementation. If you work in graphics, there are specific challenges for that domain. And so forth. You also might consider using a project that you are working on, or that you worked on in the past, as a test case for a tool. Synergies with your other work are welcomed!

Your final project report should be written in the style of a conference paper, and you will present it in a CSE 501 “conference” at the end of the term. (You are not required to submit your work to a conference or workshop, but you might choose to do so, and this is the standard of work expected.) Please use the ACM proceedings style (<http://www.acm.org/sigs/publications/proceedings-templates>), but *not* the tighter alternate style since the class does not impose a page limit.

You need not close the book on the area of research you choose, but you should do a good job of discovering and clearly communicating new knowledge in that area. Do not feel overwhelmed by the final project or final report. The instructor will meet with you repeatedly for discussions, feedback, and assistance on all aspects of your project. For tips about writing a technical paper, you should read (among other resources) <http://www.cs.washington.edu/homes/mernst/advice>.

5.1 Timetable

October 7: Proposal

Select a group, select a project, and write a short (1–3 pages) project proposal that includes its thesis and technical approach.

Be sure to motivate your project. Your paper should start with the problem. You also must give a use case — essentially, a story about how the tool is used and why it is important. For example, a particular program does not run correctly or efficiently. Be specific about the problem and about why current techniques do not address it. Your tool solves that problem. Include an architectural diagram of how the pieces of your system fit together.

Include evaluation criteria. The philosopher of science Karl Popper states that a theory is scientific only if it is falsifiable — if some experiment exists that would disprove it. You need to clearly state an experimentally refutable thesis or hypothesis. You should also state your evaluation metric, which should be, to the greatest extent possible, quantifiable. Another way to think of this is: how will you convince a skeptical reader (or your instructor) that you succeeded? They aren't going to just take your word that your project was a success.

Clearly state the scientific question or contribution. For example, how is your project an (incremental) advance over previous work? But you can also extend a technique to a new domain, or expand a problem statement. And ideally, your results will be transferable to other domains: they will change the way that others think or act in the future, or provide guidance to programmers or to tool developers or both. Oftentimes, much of your project is not novel. That's OK — engineering is necessary in order to do experiments or to build on previous work — so long as you can point out the part that is novel or what knowledge is produced.

October 12: Approval of proposal

Meet with the instructor for feedback on and approval of your project proposal.

November 2: Related work and methodology complete

Submit a version of your report that includes the introduction, description of the technical approach (e.g., algorithmic details), related work (in final form), and experimental methodology (i.e., evaluation plan). Each version of the paper should also address all previous comments from the course staff.

The evaluation plan should indicate what subject programs you will use and what metrics you will measure. It should explain why those are the right metrics, and what they reveal about your work. (For example, do they address the novel part of your technique, or are they end-to-end measures? Can they be directly compared to previous work, or not?) It should also indicate your criteria for success. (You should lay those out now, so that you have an objective measure when the research is complete.) This version of the paper should include all the tables or graphs (if any) that will appear in the final paper, and the structure of any textual sections of the evaluation. The tables or graphs will be empty for now, but you should know what they are.

You should have already divided the work among members of your group.

Do not stint on this component of the project: a prototype implementation and draft report is due in just three weeks.

November 23: Initial results

Submit a version of the report that includes at least partial results. If you are building a tool, you should have a prototype implementation that can run correctly end-to-end on a small example program. (For example, a compiler might be able to compile the factorial program.) Having a working implementation now will enable you to enhance it and to run experiments during the next several weeks; if you do not even yet have working code yet, you are unlikely to be able to complete a quality project in time.

As a rule of thumb, by this point you should have one complete row or column in *each* table filled in.

December 10 (last day of classes): Final report

Due by 4pm.

Thursday, December 16, 9:30-12:30: Project presentations

(This date is subject to change — for instance, to be moved earlier in the week — based on mutual agreement from the class.)

You should plan to talk for 20–25 minutes, followed by 5–10 minutes for questions (though the questions may come during rather than after the talk). As with any talk, be sure to practice it before you present in class. Be sure to include plenty of examples! (Even after receiving this advice, most groups fail to include enough to make their ideas comprehensible.) For more tips, see <http://www.cs.washington.edu/homes/mernst/advice/giving-talk.html>.