Advanced Software Security

1. Introduction

Kihong Heo

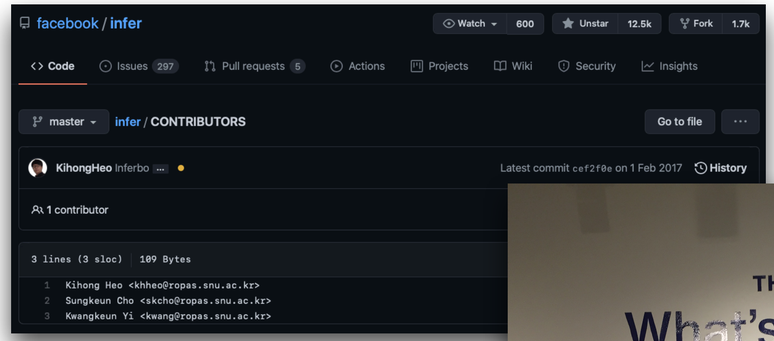
KAIST

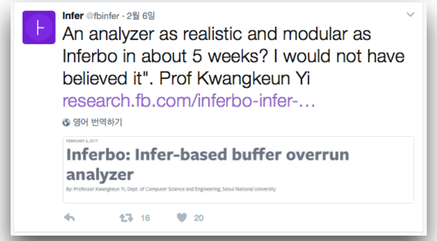
About Me

- Instructor: Kihong Heo (허기홍, kihong.heo@kaist.ac.kr)
- KAIST CS / GSIS / Programming Systems Lab.
- Homepage: https://kihongheo.kaist.ac.kr / https://prosys.kaist.ac.kr
- Office: N5 2321
- Office Hours: Tue 10:30 - 11:30

My Research

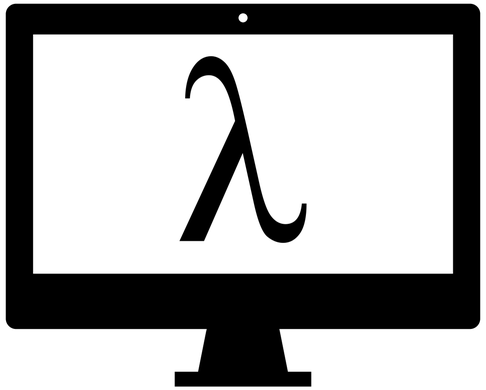
- Goal: solid PL theories ⟺ powerful programming systems
- Keywords: program analysis, programming language, SW security
- Good (also fierce) memories:





*https://research.fb.com/blog/2017/02/inferbo-infer-based-buffer-overrun-analyzer/

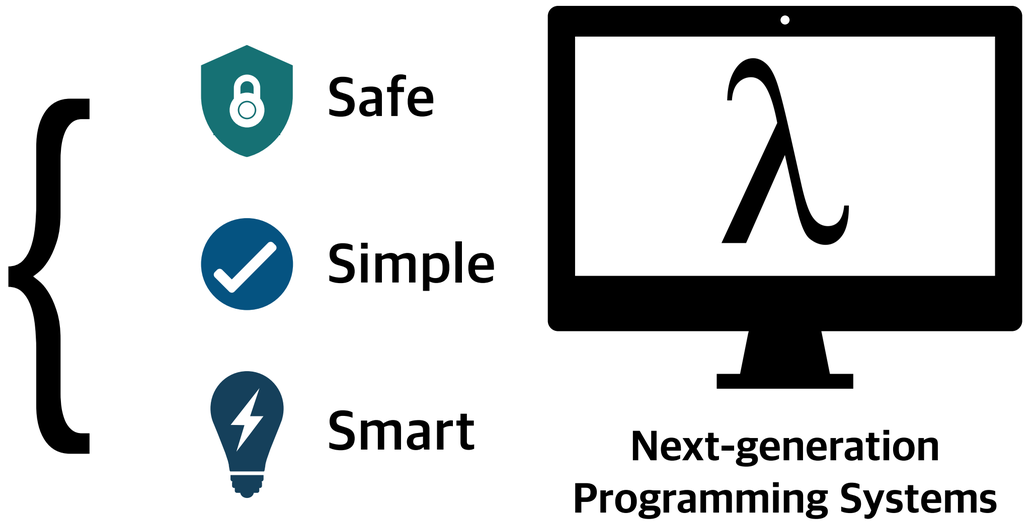
My Research



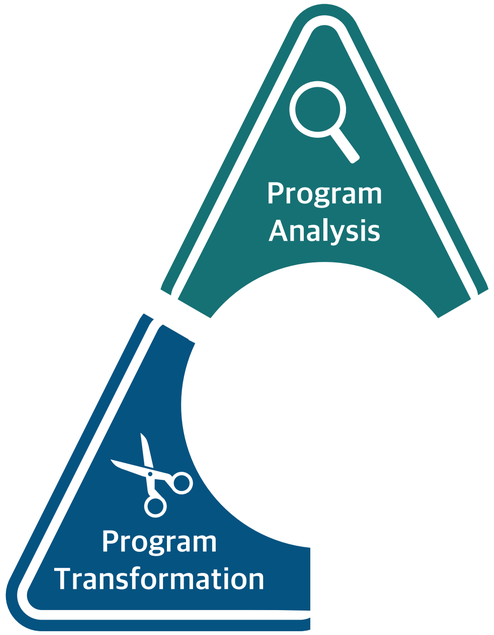
Next-generation
Programming Systems

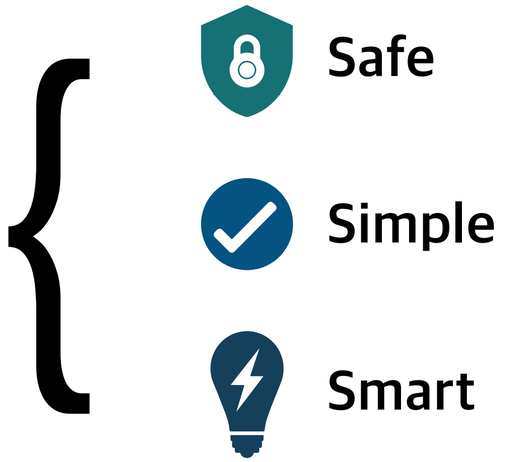
My Research

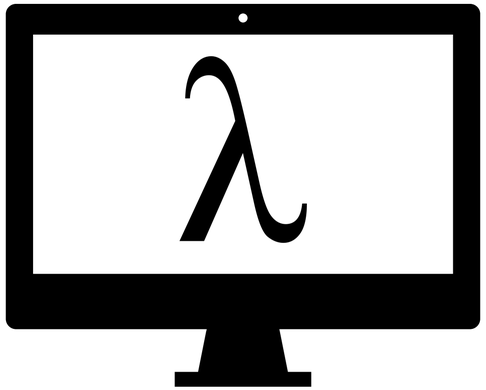

Program
Analysis


Safe
Simple
Smart
Next-generation
Programming Systems

My Research

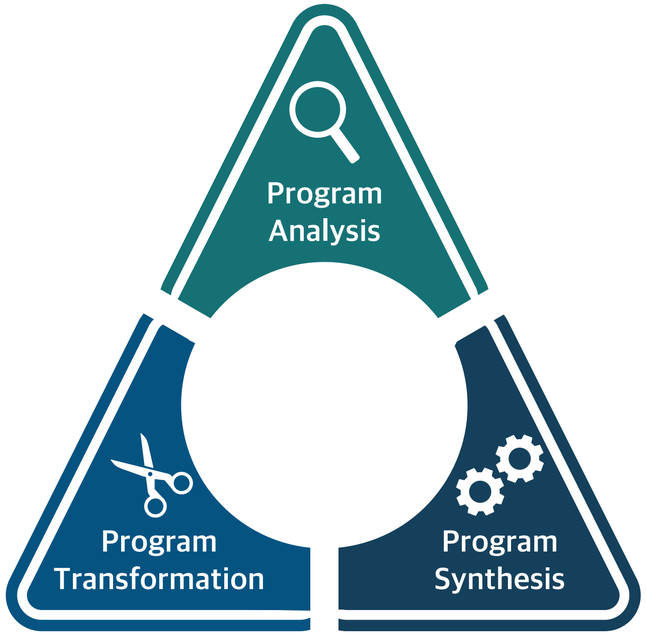






Next-generation Programming Systems

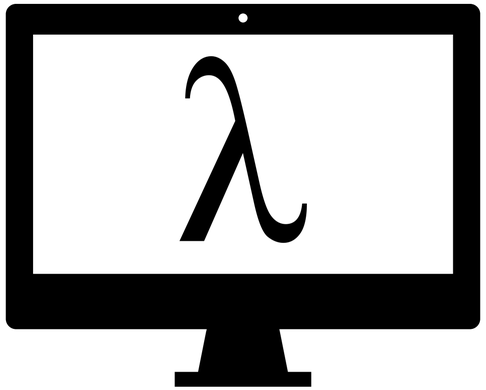
My Research



Safe

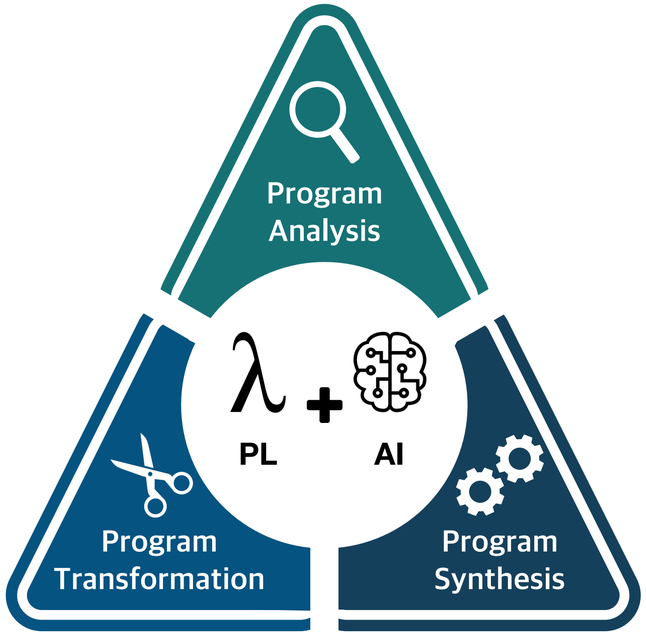
Simple

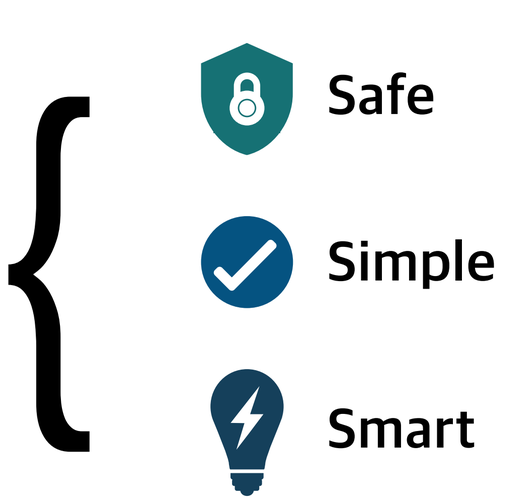
Smart

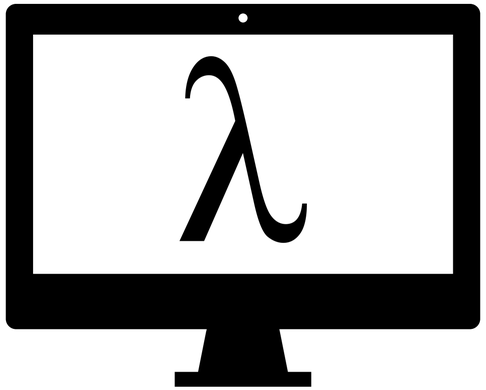


Next-generation Programming Systems

My Research







Next-generation Programming Systems

Course Information

- Course Website: https://github.com/prosyslab-classroom/is593-2022-spring
- Q&A Board: https://github.com/prosyslab-classroom/is593-2022-spring/issues
- TAs (mailing list: is593.ta@prosys.kr)
  - Hyunsu Kim (김현수, hyunsu.kim00@kaist.ac.kr)
  - Seungwan Kwon (권승완, seungwan.kwon@kaist.ac.kr)
- Textbook:
  - Lecture slides will be provided
  - See the course webpage

Important Notice (1): Academic Integrity

- DO NOT share the course contents (e.g., assignments or exams) with others
  - Esp., Github public repository, chegg.com, etc
- DO NOT discuss the details of solutions with others
- Any integrity violation: at LEAST F
  - See the KAIST CS honor code
- If you have questions: QnA board > TAs > instructor

전산학부 명예규정 2022 봄학기 / School of Computing Honor Code 2022 Spring

카이스트 전산학부가 운영하는 모든 수업에 참여하는 학생은 개인의 명예와 타인의 권리를 함께 존중하며 성실성과 정직성을 지키기 위하여 최선을 다합니다. 모든 시험 및 과제물 작성에 있어 허가되지 않은 어떤 형태의 도움도 받지 않습니다. 다음의 행위들은 학업의 성실성과 정직성을 위반하는 것으로 간주됩니다:

- 본인 이외의 사람/기관이 작성한 답안지, 숙제, 프로그램 소스 코드, 보고서 등을 참고 및 이용하는 행위
- 시험 및 과제물과 관련해 chegg.com과 같이 정답을 공유하는 온라인 서비스를 이용하는 행위
- GitHub 등의 코드 저장소에 본인의 과제 답안을 공개하거나, 타인이 공개된 답안을 참고 및 이용하는 행위
- 다른 학생이 본인이 작성한 답안지, 숙제, 프로그램 소스 코드, 보고서 등을 참고하도록 용인하는 행위
- 다른 학생이 작성한 결과물을 자신의 것인 양 제출하는 행위
- 다른 학생을 대신해 시험을 치루는 행위
- 개인이 수행하도록 되어있는 take-home 시험이나 과제물 작성에 있어 허락 없이 공동 작업을 하거나 부적절한 도움을 받는 행위
- 표절: 길이와 무관하게 적절한 인용이나 언급 없이 타인의 창작물(참고서적, 문헌, 온라인상의 자료)을 무단으로 사용하는 행위

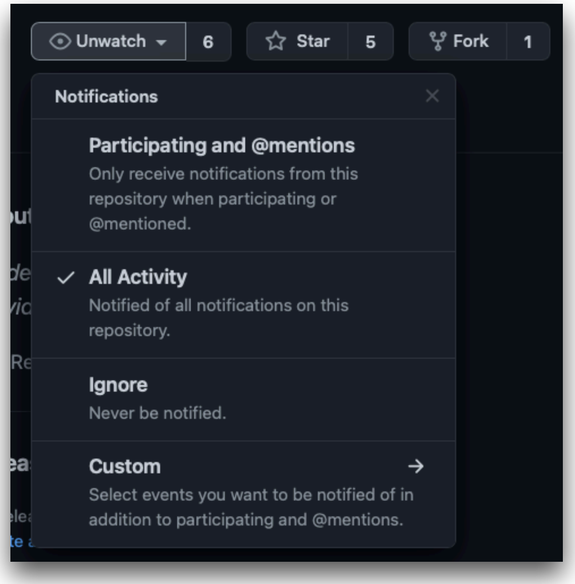
규정 위반 여부의 판단과 처벌 수위는 교과목 담당 교수에 의해 결정됩니다. 모든 부정행위는 전산학부 학사주임 교수 및 학부장님께 보고되며, 적발시 전산학부 내부적으로 아래와 같은 제약을 받습니다.

- 향후 2학기 동안 모든 포상 및 장학금 수여/추천 대상에서 제외함
- (주전공이 전산학부가 아닌 경우) 향후 2학기 동안 전산학부로의 전과 금지

위반의 심각성에 따라 학교 전체의 상벌위원회에 회부될 수 있습니다. 카이스트 학사규정이 허용하는 징계의 범위는 아래 첨부된 학생 징계 양형 기준을 참고하세요 (학생 핸드북 한글판 67페이지 , 다운로드는 https://portal.kaist.ac.kr/ennotice/student_notice/11614934649338)

Important Notice (2): Out-of-class

- All Q&A and public notices: Github issue board
  - “Watch” all notifications
- Private notices (grading, etc): KLMS
- Questions are always welcome except for
  - Too detailed ones (TAs are not debuggers!)
  - Directly related to the solutions



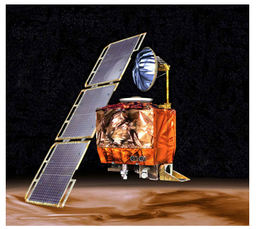
SOFTWARE BUGS

Software Bugs: A Persistent Problem

- A long time ago, far far away

The Patriot Missile (1991)
Floating-point roundoff
28 soldiers died

The Ariane-5 Rocket (1996)
Integer Overflow
$100M



NASA's Mars Climate Orbiter (1999)
Meters-Inches Miscalculation
$125M

- Unfortunately, it becomes your own problem now

Software Bugs: A Persistent Problem

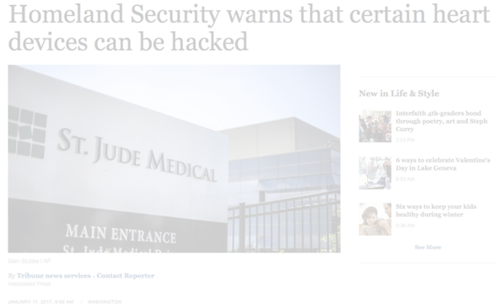
- A long time ago, for far away

COST OF A SOFTWARE BUG

$100	$1,500	$10,000
If found in Gathering Requirements phase	If found in QA testing phase	If found in Production

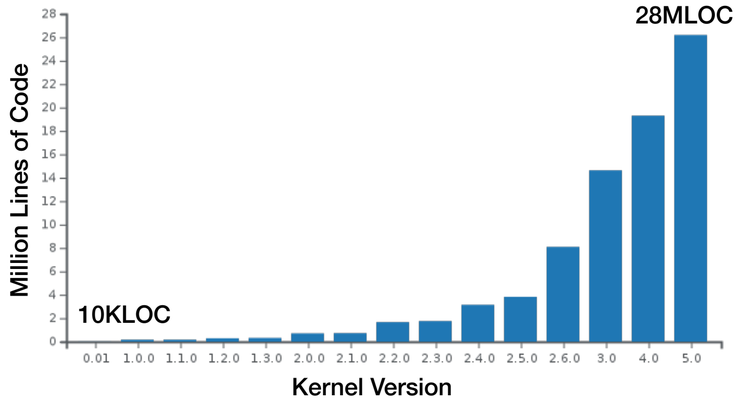
- U

- IBM Systems Sciences Institute, 2015

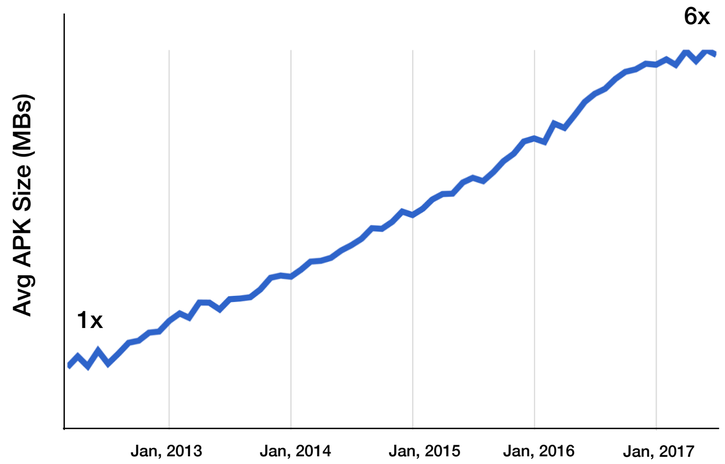


Why Software Still Fails?

Size of Linux Kernel



Avg. Size of Android Apps



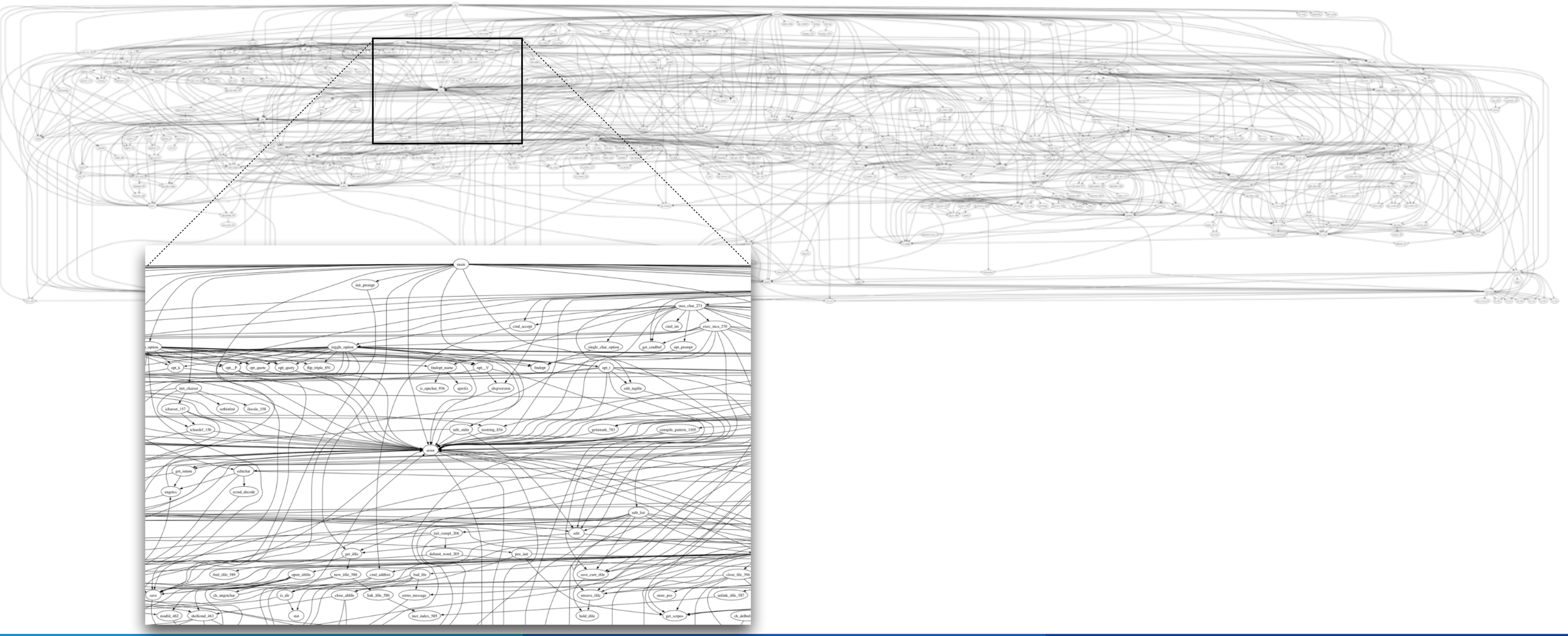
X

GitHub

10M+ New Developers
44M+ New Repositories
87M+ New Pull Requests
in 2019

Software Complexity

less-382 (23,822 LOC)



What is the Future of Programming?

How to automatically **prove** if a given program satisfies the specification?

How to automatically **generate** a program that satisfies the specification?

Course Objectives: Principles

- Program synthesis
  - How to specify specification?
  - How to generate a correct program?
- Program verification
  - How to specify specification?
  - How to generate a correctness proof?

Course Objectives: Opportunities

- What is the state-of-the art? What are the applications?
- Program synthesis: cryptography, de-obfuscation, programming-by-example, …
- Program verification: API protocol, compiler, web browser, OS, DNN, …

수업의 목표: 소통

- 과학은 소통이다 (Science is communication)
- 카이스트 연구자들의 두 가지 역할
  - 번뜩이는 연구 성과로 세계를 선도: 유창한 영어와 전문용어
  - 세계를 선도하는 연구성과를 국내에 전달: 쉬운 우리말
    - 자라나는 새싹
    - 우리 전공에 관심이 많은 일반인
    - 관련 부처 공무원

글쓰기

- 과학 글쓰기의 기본 원칙: 두괄식
- 제대로 된 우리말 번역과 전문 용어 원문 병기
  - 예: 프로그램 검증 (program verification)
- 바보들의 국문 글쓰기 알고리즘:

  1. 영문 용어를 단어별로 쪼갠다
  2. 각 용어를 사전에서 찾아 제일 처음 나오는 우리말 단어로 바꾼다
  3. 바뀐 단어들을 순서대로 나열한다
- 일류 과학자의 국문 글쓰기 알고리즘: 개념을 제대로 이해하고 핵심을 설명하는 쉬운 우리말로

Course Activities and Assignments

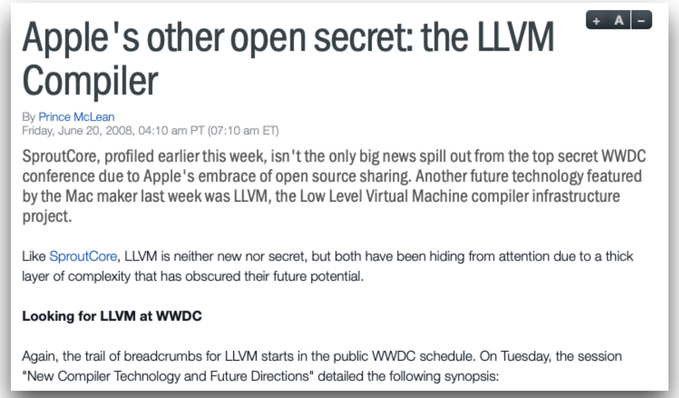
- 4 reading critiques: papers and articles
- Mid-term project 1: implement your own program synthesizer and present your idea
- Mid-term project 2: implement a program verifier
- Final project:
  - write a proposal
  - write reviews and a rebuttal
  - implement your idea and give a presentation

Participation

- Mandatory 1: attendance
  - If you have a problem (e.g., Covid), email me in advance and join via Zoom (see KLMS)
- Mandatory 2: questions
  - At least one question for each presentation of other students
- Highly recommended: further discussion online/offline

The LLVM Compiler Infrastructure

- The de-facto standard & well-structured compiler toolchain
  - parser, code optimizer, linker, loader, debugger, etc
- A wide variety of frontends: C/C++, Obj-C, Swift, Fortran, etc
  - translated to the LLVM IR (intermediate representation)



Apple's other open secret: the LLVM Compiler

By Prince McLean
Friday, June 20, 2008, 04:10 am PT (07:10 am ET)

SproutCore, profiled earlier this week, isn't the only big news spill out from the top secret WWDC conference due to Apple's embrace of open source sharing. Another future technology featured by the Mac maker last week was LLVM, the Low Level Virtual Machine compiler infrastructure project.

Like SproutCore, LLVM is neither new nor secret, but both have been hiding from attention due to a thick layer of complexity that has obscured their future potential.

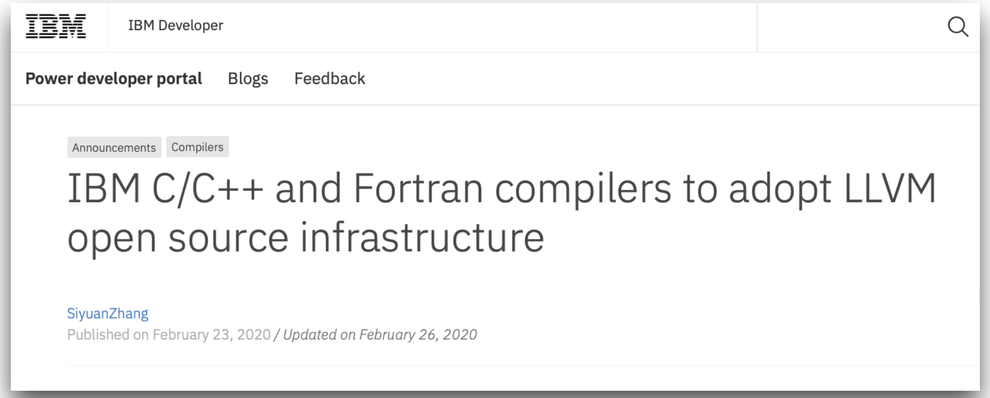
Looking for LLVM at WWDC

Again, the trail of breadcrumbs for LLVM starts in the public WWDC schedule. On Tuesday, the session "New Compiler Technology and Future Directions" detailed the following synopsis:

Google Chrome is replacing Microsoft's C++ compiler with Clang

By Muhammad Jarir Kanji · Mar 6, 2018 14:06 EST · HOT!

Alongside bringing better touch support and automatic ad-blocking for 'intrusive' ads to the desktop version of Chrome, Google is also making some changes to its browser under the hood. The company is now starting to build Chrome for Windows using the Clang compiler which it already uses for other platforms like macOS and Linux.



IBM Developer

Power developer portal Blogs Feedback

Announcements Compilers

IBM C/C++ and Fortran compilers to adopt LLVM open source infrastructure

SiyuanZhang
Published on February 23, 2020 / *Updated on February 26, 2020*

The OCaml Language

- Simple, safe, realistic and high-level programming language
- Official OCaml bindings to LLVM and Z3 API supported
- A lot of growing demands from academia and industry
- Why OCaml?
  - Simple
  - Safe
  - High-level

The Z3 Theorem Prover

- State-of-the-art automated theorem prover by Microsoft Research
- Solving satisfiability modulo theory (SMT) problems
  - first-order logic with background theories
    (e.g., arithmetic, bit-vectors, arrays, datatypes, uninterpreted functions, etc)

Boolean Satisfiability Problem (SAT)

$$(\neg A \vee B) \wedge (\neg B \vee C) \wedge (A \vee \neg C \vee B)$$

Satisfiable when
A = false
B = true
C = true

Satisfiability modulo theory (SMT)

$$x + 2 = y \implies f(read(write(a, x, 3), y - 2)) = f(y - x + 1)$$

Arithmetic | Array | Uninterpreted Functions

Homework

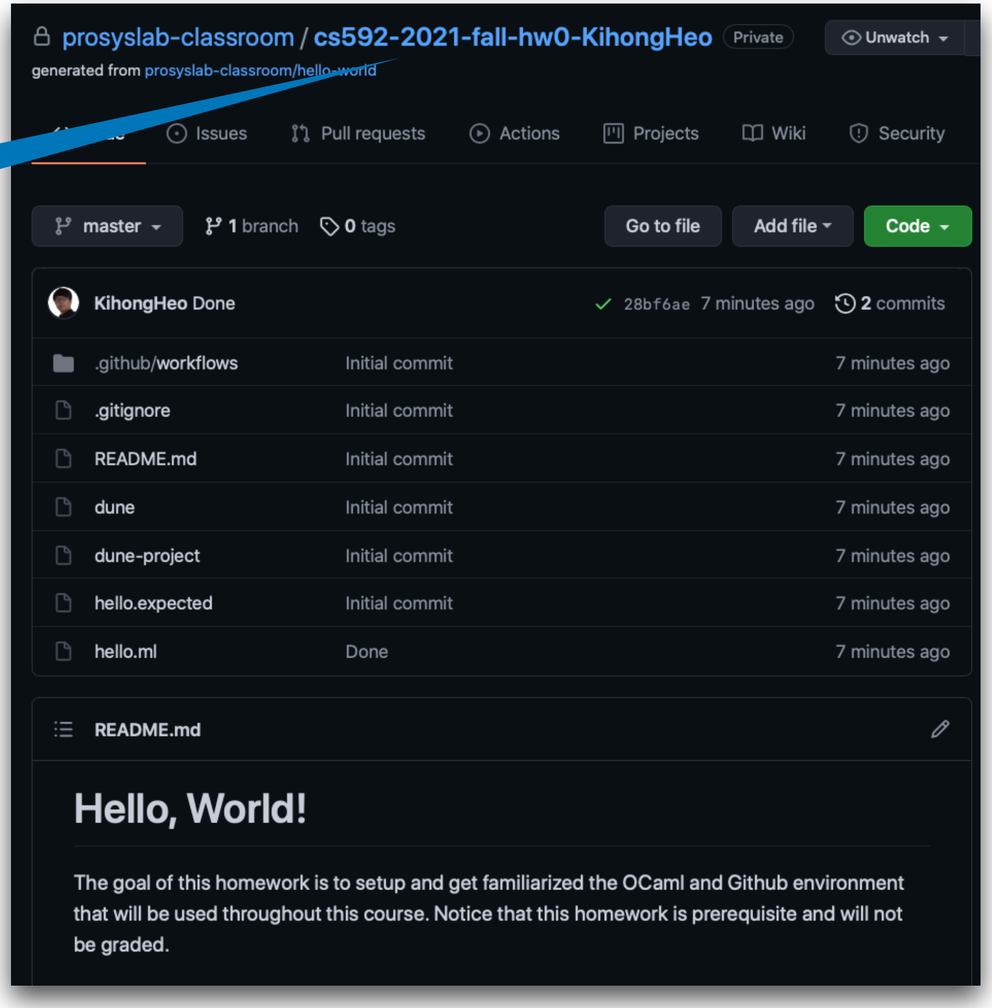
- All submissions will be managed using Github / Github Classroom
  1. For each HW, a unique invitation URL will be posted at KLMS
  2. Once you accept, a private repo for your assignment will be created
  3. You can push as many commits as you want before the deadline
  4. The final commit of your master branch will be graded
- 80% credit for 1-day late, 50% credit for 2-days late, NO credit otherwise

Homework 0: OCaml Programming

- Goal: setting up and getting familiarized with OCaml and Git environments
  - Implement your program in OCaml
  - Test on your machine
  - Push to your Github repository
  - See the result in Github Action
- The invitation URL will be posted at KLMS

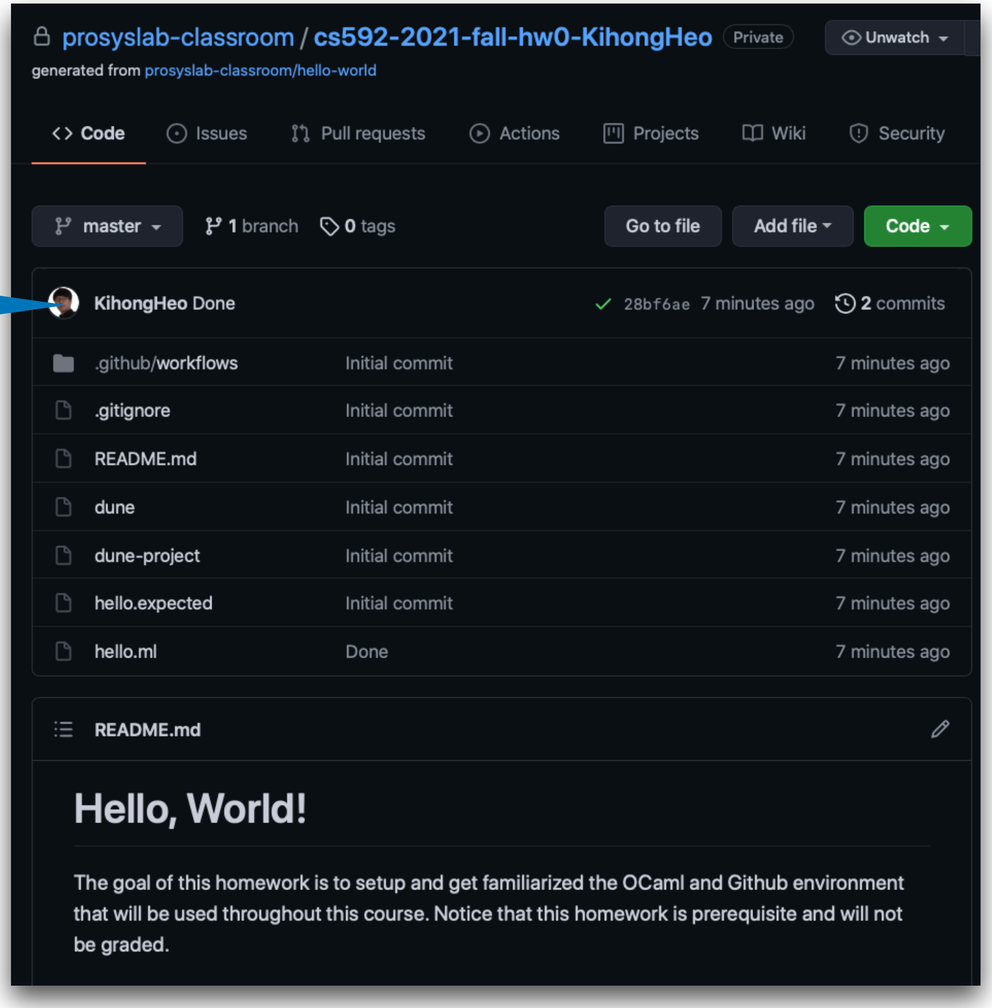
Homework 0: OCaml Programming

1. Accept the invitation and have your repository



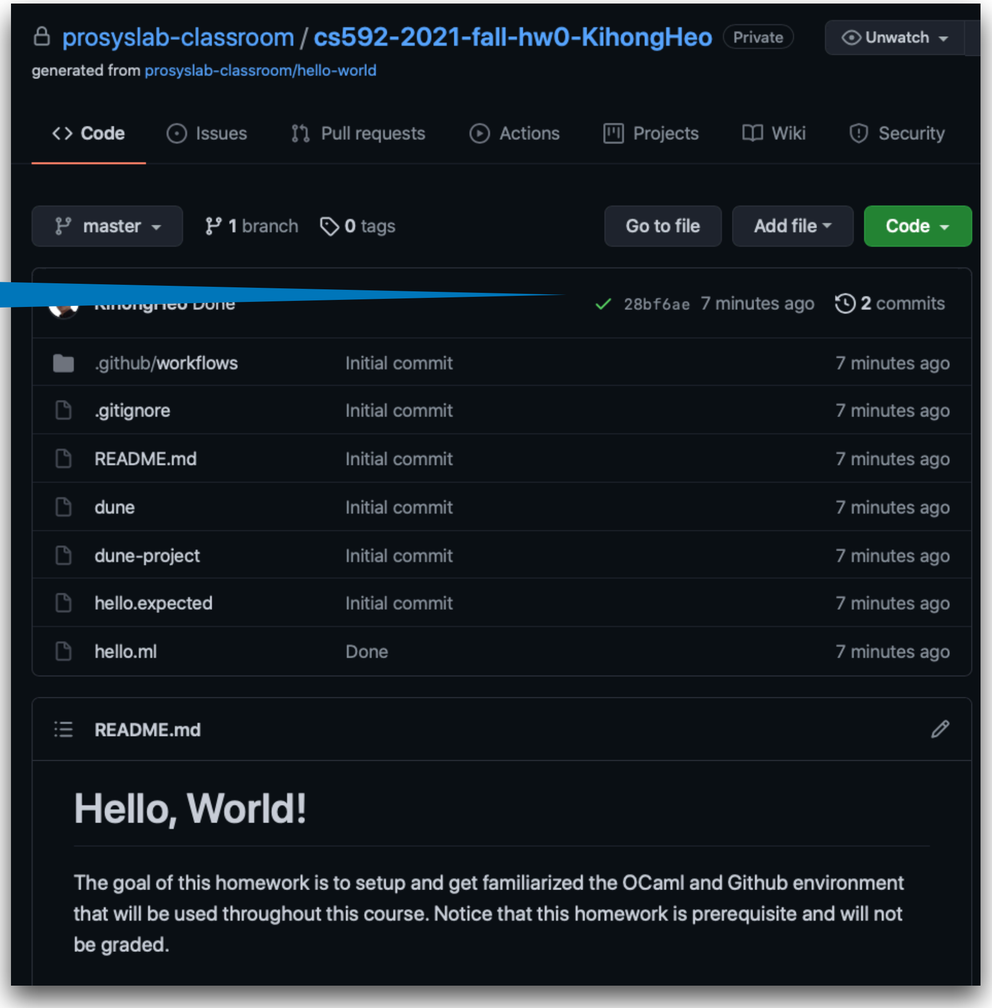
Homework 0: OCaml Programming

2. Commit your code



Homework 0: OCaml Programming

3. See your result



Misc

- Submit your Github account via the google form (see the Github issue board)
- Rules for programming assignments
  - Preserve the structures (directories, files, types, etc)
  - Don’t install further Github App