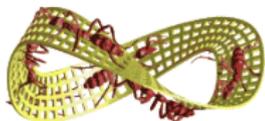


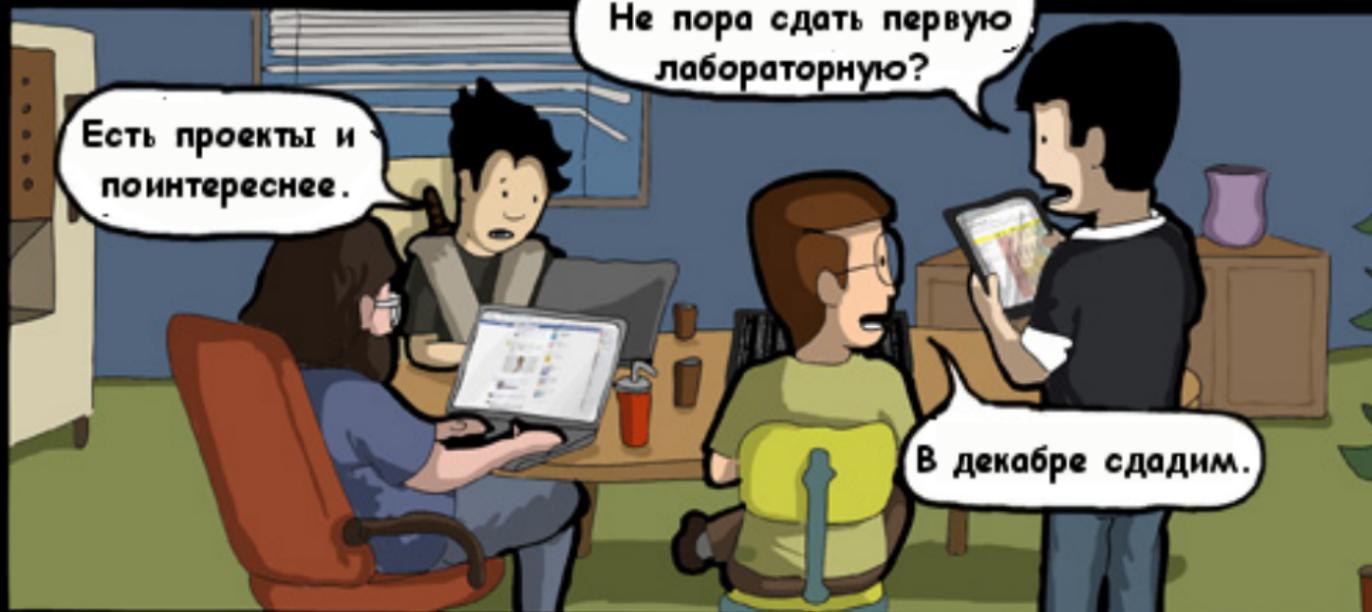
Эксперимент по созданию QUICK-and-DIRTY пружеров в рамках РК



А. Н. Непейвода
ИПС им. А.К. Айламазяна РАН

*Конференция «СПО в высшей школе»,
20 мая 2022, Переславль-Залесский*

ТИПичный ИУ9



Есть проекты и поинтереснее.

Полсеместра прошло.
Не пора сдать первую лабораторную?

В декабре сдадим.



Предусловие

- Специальность «Анализ, преобразование и порождение программного кода»
 - Курс «теория формальных языков» (5 семестр)
 - База: хорошая традиционно математическая (алгебра), хорошая практическая. По информатике — один семестр дискретной математики (алгоритмы).
-
- Много бывших участников олимпиад;
 - Много уже занятых на работе.



Особенности предметной области

- 1 Проблемы EXPTIME- и большей сложности;
- 2 Большинство практически интересных проблем (верификация, извлечение оптимального кода...) — неразрешимы;
- 3 Небольшие изменения в синтаксисе задачи (входных данных) \Rightarrow нелокальные изменения в решении.
- 4 Повсеместно распространённые техники построения аппроксимаций задач;
- 5 Вовлечение абстрактной алгебры для унификации описания преобразуемых структур.



Termination Competition

- 2003, 3 пружера
- 2004, 5 пружеров
- 2005, 6 пружеров
- 2006, 8 пружеров
- ...
- 2018, 5 пружеров
- 2019, 6 пружеров
- 2020, 5 пружеров
- Все пружеры открытые;
- Список проблем тоже открытый;
- Секретные тесты;
- Отказ от ответа не карается;
- Высокий таймаут (не ниже 30 сек);
- Доказательство правильности — в свободной форме.



TermComp — PK Edition

- Наследование главных традиций TermComp;
- Лояльность к возможным ошибкам;
- Заранее известная структура тестов (но сами тесты выкладываются уже во время состязания вместе со скриптом тестирования);
- Доказательства не требуются;
- Скрипт тестирования открытый (можно взломать базу ответов).



Открытый декодер

```
Decode {  
  e.1'SYNTAX'e.2 = 'Syntax error';  
  e.x  
  , e.x : e.x1 s.1 e.z1 s.2 e.x2  
  , s.1 e.z1 s.2 s.1 e.z1 s.2  
  : e.z22 s.3 s.4 s.1 e.z1 s.2 e.z23  
  , s.1 e.z1 s.2 : e.z22 s.3 s.4 s.5 e.z24  
  = 'False';  
  e.x, e.x : e.x4 e.x3 e.x1 s.x1 e.x4 e.x2 s.x1 e.x3  
  = 'True';  
  e.x = 'Monster'; }
```

ГИДРА
БУХОЛЬЦА

ГИДРА
ТРДВ

ГИДРА
ИУ9





Структура тестов

- 60 TRS-завершаек; 30 TRS-незавершаек; 5 «монстров»; 5 на корректность синтаксиса.
- Ошибка карается только в половину стоимости теста (но есть дополнительные штрафы за монстров).
- 20 тестов — аналогичны тестам из базы TPDB; 10 тестов построили сами студенты; 60 тестов порождены автоматически.
- Каждый из тестов (кроме «монстров») с гарантией может быть успешно проанализирован одним из разобранных в курсе методов (анализ линейных функций в Фума, отношение Кнута–Бендикса, частичные вычисления, убывание на мономиальных порядках).



Монстры

$$f(x, g(x)) \rightarrow f(g(x), g(x))$$

$$f(D, x) \rightarrow D$$

$$g(E) \rightarrow f(E, E)$$

$$f(x, f(y, z)) \rightarrow f(f(x, y), z)$$

$$g(x) \rightarrow f(f(f(x, H), f(D, E)), E)$$

$$a(a(a(x))) \rightarrow b(a(b(x)))$$

$$b(b(b(x))) \rightarrow a(a(a(x)))$$

- Не разрешались актуальными лидерами TermComp;
- Правильный ответ \neq корректный ответ (предъявление конструктивного обоснования сомнительно);
- Без конструктивного обоснования за любой ответ кроме «неизвестно» или таймаута — штраф.



Система оценивания

- «Незавершайки» в среднем дороже «завершаек» (почти нет известных полиномиальных алгоритмов приближённого анализа);
- Баланс всё равно положительный (75/45);
- С учётом штрафа за «монстров» — рандомизация бесполезна. Недетерминизм допустим;
- Дополнительные поощрения за быстроедействие (на срезах по времени).



Неустойчивость

- Инвариантность относительно переименовок и перестановок.
- (Условная) инвариантность «завершаек» относительно специализации и убывания.
- (Условная) инвариантность «незавершаек» относительно обобщения и разрастания.



Результаты

- 10 заявок, 6 участников дали ссылки на репозитории (и одни первыми взломали базу тестов): 3× Python, 1×Java, 1×Typescript, 1×Go;
- Объём кода — от 400 строк на Python и TS до 750 на Java;
- Только две работы достоверно пересеклись по покрытиям;
- Работа с изгнанием «монстров» — продуктивный дебаг;
- Большинство участников скомбинировали несколько разных техник доказательства.



Выводы

- Практика в самостоятельном выборе алгоритмов и средств реализации;
- Практика в решении асимптотически трудных задач;
- Практика в дебаге аппроксимационных стратегий;
- Очень высокая мотивация.

Минусы:

- не для всех;
- занимает времени больше, чем стандартный РК.

Спасибо за внимание!

