# Автоматический вывод индуктивных инвариантов программ с алгебраическими типами данных

#### Костюков Юрий Олегович

Научный руководитель: д. т. н., доцент Кознов Дмитрий Владимирович

2023



#### Содержание

Обзор предметной области

Постановка задачи

Результаты

Научная новизна

Публикации и выступления

Результаты

# Дизъюнкты Хорна над АТД

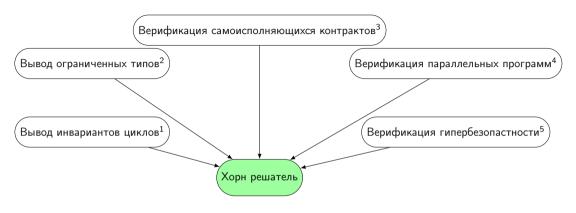
Пример программы на языке HASKELL:

```
data Nat = Z | S Nat
data List = nil | cons Nat List
drop Z xs = xs
drop _ nil = nil
drop (S n) (cons(_, xs)) = drop n xs
assert (¬∃ n xs . xs /= nil && drop n xs == drop (S n) xs)
```

Условия верификации в виде дизъюнктов Хорна над АТД:

$$op drop(Z,xs,xs)$$
 $op drop(S(n),nil,nil)$ 
 $drop(n,xs,rs) op drop(S(n),cons(x,xs),rs)$ 
 $op(xs=nil) \wedge drop(n,xs,ys) \wedge drop(S(n),xs,ys) o ot$ 

### Применения Хорн решателей



<sup>&</sup>lt;sup>1</sup> Gurfinkel и др. The SeaHorn Verification Framework. CAV'15

 $<sup>^{\</sup>rm 2}$  Tan и др. SolType: refinement types for arithmetic overflow in solidity. POPL'22

<sup>&</sup>lt;sup>3</sup> Alt и др. SolCMC: Solidity Compiler's Model Checker. CAV'22

<sup>&</sup>lt;sup>4</sup> Hoenicke и др. Thread Modularity at Many Levels. POPL'17

<sup>&</sup>lt;sup>5</sup> Shemer и др. Property Directed Self Composition. CAV'19

# Дизъюнкты Хорна формально

**Дизъюнкт Хорна** C — это формула первого порядка следующего вида:

$$\varphi \wedge P_1(\overline{x}_1) \wedge \ldots \wedge P_n(\overline{x}_n) \to H$$

- ightharpoonup ограничение arphi это формула теории алгебраических типов данных
- ightharpoonup голова H это либо ложь  $\perp$ , либо атом  $P(\overline{x})$
- ▶  $P_1, ..., P_n, P$  это неинтерпретированные символы
- ▶ все переменные (неявно) универсально квантифицированы

Система дизъюнктов Хорна — это конъюнкция дизъюнктов Хорна

Пусть  $\mathcal{H}$  — модель теории АТД,  $\mathcal{S}$  — система дизъюнктов Хорна. Индуктивный инвариант  $\mathcal{I}$  — расширение модели  $\mathcal{I} = \langle \mathcal{H}, \mathcal{R} \rangle$ , такое что  $\mathcal{I} \models \mathcal{S}$ .

Пусть  $\mathcal{H}$  — модель теории АТД,  $\mathcal{S}$  — система дизъюнктов Хорна. **Индуктивный инвариант**  $\mathcal{I}$  — расширение модели  $\mathcal{I} = \langle \mathcal{H}, \mathcal{R} \rangle$ , такое что  $\mathcal{I} \models \mathcal{S}$ .  $x = Z \land v = S(Z) \rightarrow inc(x, v)$  $x' = S(x) \land y' = S(y) \land inc(x, y) \rightarrow inc(x', y')$  $x = v \wedge inc(x, v) \rightarrow \bot$  $\mathcal{I}_1 = \mathcal{H}\Big\{\mathsf{inc} \mapsto \{(x,y) \mid y = S(x)\Big\}$  $\mathcal{I}_2 = \mathcal{H} \Big\{ inc \mapsto \{(x,y) \mid x \neq y \quad \Big\}$  $\mathcal{I}_3 = \dots$ 

Индуктивные инварианты составляют решётку

Пусть  $\mathcal{H}$  — модель теории АТД,  $\mathcal{S}$  — система дизъюнктов Хорна. **Индуктивный инвариант**  $\mathcal{I}$  — расширение модели  $\mathcal{I} = \langle \mathcal{H}, \mathcal{R} \rangle$ , такое что  $\mathcal{I} \models \mathcal{S}$ .  $x = Z \land v = S(Z) \rightarrow inc(x, v)$  $x' = S(x) \land v' = S(v) \land inc(x, v) \rightarrow inc(x', v')$  $x = v \wedge inc(x, v) \rightarrow \bot$  $\mathcal{I}_1 = \mathcal{H}\Big\{\mathsf{inc} \mapsto \{(x,y) \mid y = S(x)\Big\}$  $\mathcal{I}_2 = \mathcal{H}\Big\{\mathsf{inc} \mapsto \{(x,y) \mid x \neq y \qquad \Big\}$  $\mathcal{I}_3 = \dots$ 

Как представлять эти бесконечные множества?

Пусть  $\mathcal{H}$  — модель теории АТД,  $\mathcal{S}$  — система дизъюнктов Хорна. Индуктивный инвариант  $\mathcal{I}$  — расширение модели  $\mathcal{I} = \langle \mathcal{H}, \mathcal{R} \rangle$ , такое что  $\mathcal{I} \models \mathcal{S}$ .  $x = Z \land y = S(Z) \to inc(x,y)$   $x' = S(x) \land y' = S(y) \land inc(x,y) \to inc(x',y')$   $x = y \land inc(x,y) \to \bot$   $\mathcal{I}_1 = \mathcal{H} \Big\{ inc \mapsto \qquad \qquad y = S(x) \Big\}$ 

 $\mathcal{I}_2 = \mathcal{H}\Big\{\mathsf{inc} \mapsto \neg(\mathsf{x}=\mathsf{y})\Big\}$ 

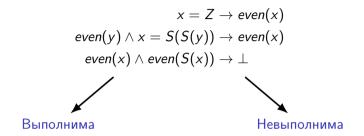
Как представлять эти бесконечные множества?

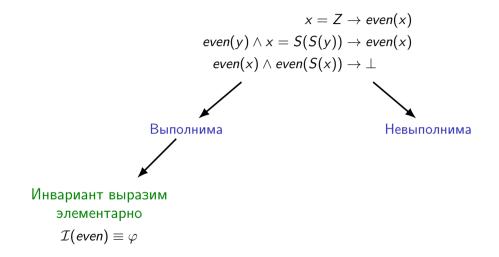
 $\mathcal{I}_3 = \dots$ 

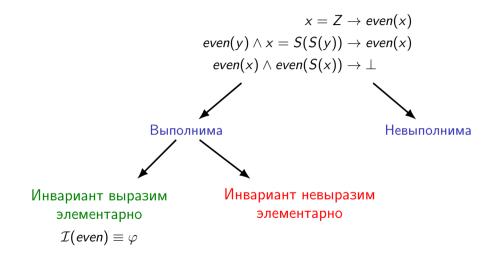
Инварианты обычно представляются в логике первого порядка (ЛПП) ЛПП задаёт т.н. класс элементарных инвариантов

$$egin{aligned} x &= Z 
ightarrow ext{even}(x) \ ext{even}(y) \land x &= S(S(y)) 
ightarrow ext{even}(x) \ ext{even}(x) \land ext{even}(S(x)) 
ightarrow ota \end{aligned}$$

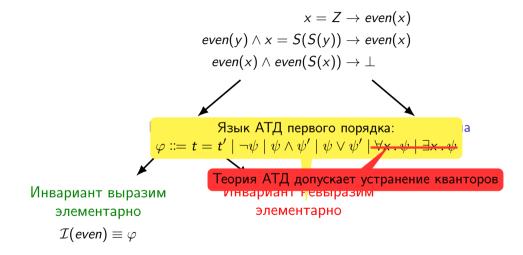
$$x=Z o even(x)$$
  $even(y) \wedge x=S(S(y)) o even(x)$   $even(x) \wedge even(S(x)) o ot$ 

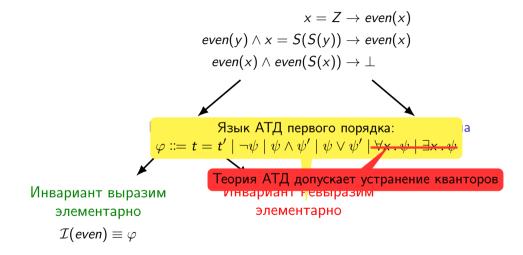
















#### Постановка задачи

**Цель работы** — предложение новых классов индуктивных инвариантов для программ с АТД и создание для них методов автоматического вывода. Задачи:

- 1. Предложить новые классы индуктивных инвариантов программ с АТД, позволяющие выражать рекурсивные и синхронные отношения
- 2. Создать методы автоматического вывода инвариантов в новых классах
- 3. Выполнить пилотную программную реализацию предложенных методов
- 4. Провести экспериментальное сопоставление реализованного инструмента с существующими на представительном тестовом наборе

#### Результаты

- 1. Предложен метод вывода регулярных инвариантов при помощи поиска конечных моделей
- 2. Предложен метод вывода синхронных регулярных инвариантов при помощи поиска конечных моделей
- 3. Предложен класс инвариантов, основанный на булевой комбинации элементарных и регулярных инвариантов Также предложен метод совместного вывода инвариантов в этом классе посредством вывода инвариантов в подклассах
- 4. Проведено теоретическое сравнение рассмотренных классов инвариантов Доказаны леммы о «накачке» для элементарных инвариантов
- 5. Выполнена пилотная реализация предложенных методов на языке F# в рамках инструмента RInGen Разработанный инструмент решил из бенчмарка «Tons of Inductive Problems» в 3.74 раза больше задач, чем наилучший из существующих инструментов

#### Соответствие результатов паспорту специальности 2.3.5

Результаты соответствуют направлению исследования № 1

 Модели, методы и алгоритмы проектирования, анализа, трансформации, верификации и тестирования программ и программных систем
 из паспорта специальности.

#### Научная новизна

- 1. Впервые предложен класс инвариантов, основанный на булевой комбинации элементарных и регулярных инвариантов
- 2. Впервые предложен алгоритм вывода инвариантов для программ с АТД, основанный на поиске конечных моделей
- 3. Предложен новый алгоритм совместного вывода инвариантов в комбинации классов инвариантов на базе методов вывода инвариантов в подклассах
- 4. Впервые введены и доказаны леммы о «накачке» для языков первого порядка в сигнатуре теории АТД

### Публикации по теме диссертации

- [1] Kostyukov Yurii, Mordvinov Dmitry, Fedyukovich Grigory. Beyond the Elementary Representations of Program Invariants over Algebraic Data Types // Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation. PLDI 2021. New York, NY, USA: Association for Computing Machinery, 2021. P. 451–465.
- [2] Kostyukov Yurii, Mordvinov Dmitry, Fedyukovich Grigory. Collaborative Inference of Combined Invariants // Proceedings of 24th International Conference on Logic for Programming, Artificial Intelligence and Reasoning / Ed. by Ruzica Piskac, Andrei Voronkov. — Vol. 94 of EPiC Series in Computing. — EasyChair, 2023. — P. 288–305.
- [3] Автоматическое доказательство корректности программ с динамической памятью / Юрий Олегович Костюков, Константин Аланович Батоев, Дмитрий Александрович Мордвинов и др. // Труды Института системного программирования РАН. 2019. Т. 31, № 5. С. 37–62.
- [4] Генерация слабейших предусловий программ с динамической памятью в символьном исполнении / Александр Владимирович Мисонижник, Юрий Олегович Костюков, Михаил Павлович Костицын и др. // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 5. С. 982–991.

#### Выступления по теме диссертации

- ► Международный семинар HCVS 2021 (28 марта 2021, Люксембург)
- ► Семинар компании Huawei (18-19 ноября 2021, Санкт-Петербург)
- Ежегодный внутренней семинар JetBrains Research (18 декабря 2021, Санкт-Петербург)
- Конференция PLDI 2021 (23-25 июня 2021, Канада)
- Внутренний семинар Венского технического университета (3 июня 2022, Австрия)
- Конференция LPAR 2023 (4-9 июня 2023, Колумбия)

Разработанный инструмент в 2021 и 2022 годах занял, соответственно, 2 и 1 место на АТД секции международных соревнований СНС-СОМР.

#### Результаты

- 1. Предложен метод вывода регулярных инвариантов при помощи поиска конечных моделей
- 2. Предложен метод вывода синхронных регулярных инвариантов при помощи поиска конечных моделей
- 3. Предложен класс инвариантов, основанный на булевой комбинации элементарных и регулярных инвариантов Также предложен метод совместного вывода инвариантов в этом классе посредством вывода инвариантов в подклассах
- 4. Проведено теоретическое сравнение рассмотренных классов инвариантов Доказаны леммы о «накачке» для элементарных инвариантов
- 5. Выполнена пилотная реализация предложенных методов на языке F# в рамках инструмента RInGen Разработанный инструмент решил из бенчмарка «Tons of Inductive Problems» в 3.74 раза больше задач, чем наилучший из существующих инструментов

#### Система дизъюнктов

$$op = even(Z)$$
 $even(y) op even(S(S(y)))$ 
 $even(x) \wedge even(S(x)) o ot$ 

АТД ограничения устранены















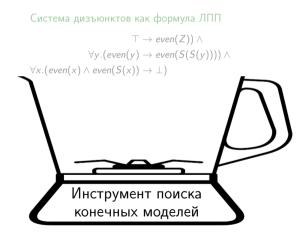


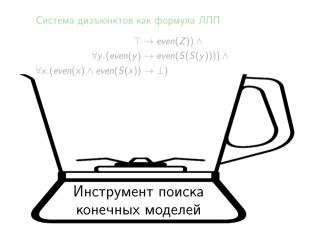


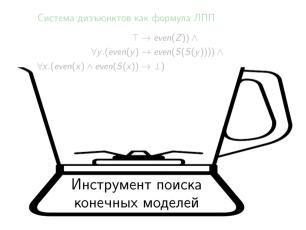


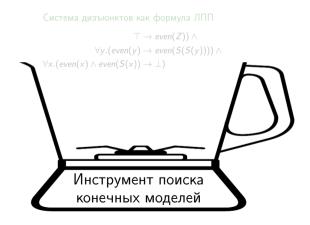


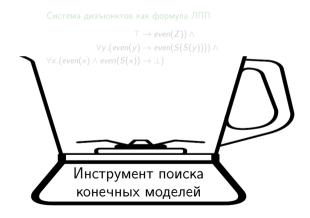


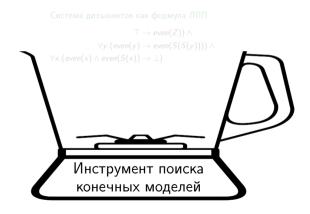




































$$|\mathcal{M}|_{Nat} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(even) = \{0\}$ 



$$\begin{aligned} |\mathcal{M}|_{Nat} &= \{0, 1\} \\ \mathcal{M}(Z) &= 0 \\ \mathcal{M}(S)(x) &= 1 - x \\ \mathcal{M}(\textit{even}) &= \{0\} \end{aligned}$$



$$|\mathcal{M}|_{\mathit{Nat}} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(\mathit{even}) = \{0\}$ 



$$|\mathcal{M}|_{Nat} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(even) = \{0\}$ 



$$|\mathcal{M}|_{\mathit{Nat}} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(\mathit{even}) = \{0\}$ 



$$|\mathcal{M}|_{Nat} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(\textit{even}) = \{0\}$ 



$$|\mathcal{M}|_{\mathit{Nat}} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(\mathit{even}) = \{0\}$ 



$$|\mathcal{M}|_{Nat} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(even) = \{0\}$ 



$$|\mathcal{M}|_{Nat} = \{0, 1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(\textit{even}) = \{0\}$ 



$$|\mathcal{M}|_{\mathit{Nat}} = \{0,1\}$$
  
 $\mathcal{M}(Z) = 0$   
 $\mathcal{M}(S)(x) = 1 - x$   
 $\mathcal{M}(\mathit{even}) = \{0\}$ 



$$egin{aligned} \left|\mathcal{M}
ight|_{\mathit{Nat}} &= \{0,1\} \ \mathcal{M}(Z) &= 0 \ \mathcal{M}(S)(x) &= 1-x \ \mathcal{M}(\mathit{even}) &= \{0\} \end{aligned}$$



$$egin{aligned} \left|\mathcal{M}
ight|_{ extit{Nat}} &= \{0,1\} \ \mathcal{M}(Z) &= 0 \ \mathcal{M}(S)(x) &= 1-x \ \mathcal{M}( extit{even}) &= \{0\} \end{aligned}$$



$$|\mathcal{M}|_{Nat} = \{0, 1\}$$

$$\mathcal{M}(Z) = 0$$

$$\mathcal{M}(S)(x) = 1 - x$$

$$\mathcal{M}(even) = \{0\}$$



$$|\mathcal{M}|_{Nat} = \{0, 1\}$$
 $\mathcal{M}(Z) = 0$ 
 $\mathcal{M}(S)(x) = 1 - x$ 
 $\mathcal{M}(even) = \{0\}$ 



#### Вывод регулярных инвариантов при помощи поиска конечных моделей

$$|\mathcal{M}|_{Nat} = \{0,1\}$$
 $\mathcal{M}(Z) = 0$ 
 $\mathcal{M}(S)(x) = 1 - x$ 
 $\mathcal{M}(even) = \{0\}$ 



#### Вывод регулярных инвариантов при помощи поиска конечных моделей

$$|\mathcal{M}|_{Nat} = \{0,1\}$$
 $\mathcal{M}(Z) = 0$ 
 $\mathcal{M}(S)(x) = 1 - x$ 
 $\mathcal{M}(even) = \{0\}$ 



#### Вывод регулярных инвариантов при помощи поиска конечных моделей

$$|\mathcal{M}|_{Nat} = \{0,1\}$$
 $\mathcal{M}(Z) = 0$ 
 $\mathcal{M}(S)(x) = 1 - x$ 
 $\mathcal{M}(even) = \{0\}$ 
 $Z \rightarrow 0$ 
 $S$ 
 $1$ 
 $S$ 
 $I(even) = \mathcal{L}(A)$ 



# Вывод синхронных регулярных инвариантов при помощи поиска конечных моделей

Регулярные языки не позволяют представлять синхронные отношения:

$$T \to lt(Z, S(x))$$

$$lt(x, y) \to lt(S(x), S(y))$$

$$lt(x, y) \land lt(y, x) \to \bot$$

Идея: порождать декларативное описание синхронного автомата

$$R(q) 
ightharpoonup R(p(d(f,g,q),d(f,g,q)))$$
  
 $R(p(q_1,q_2)) 
ightharpoonup \left(F(q_1) 
ightharpoonup F(d(S,S,q_2))\right)$ 

Из модели можно извлечь определение синхронного автомата

$$A = \left\langle \{0, 1\}, \Sigma_F^{\leq 2}, \{1\}, \Delta \right
angle$$
 
$$\langle Z, Z \rangle \mapsto 0 \qquad \qquad Z \mapsto 0 \qquad \qquad S(q) \mapsto 0$$
 
$$\langle Z, S \rangle (q) \mapsto 1 \qquad \qquad \langle S, Z \rangle (q) \mapsto 0 \qquad \qquad \langle S, S \rangle (q) \mapsto q$$
 
$$\mathcal{L}(A) = \left\{ \langle S^n(Z), S^m(Z) \rangle \mid n < m \right\}$$

ELDARICA GOLEM

FreqHorn

DUALITY QARMC

Z3/SPACER

HOICE

Хорн решатели

PROVER.

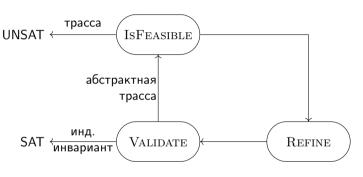
Vampire

ZIPPERPOSITION

Инструменты <sub>вывода теорем</sub>

CVC5

 $\mathbf{E}$ 



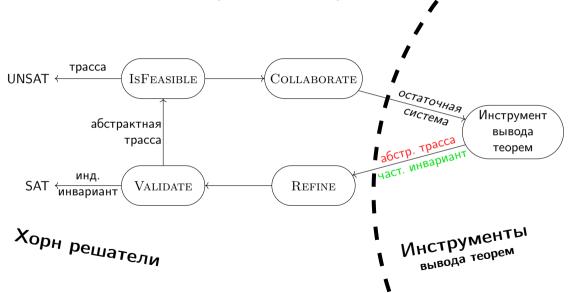
Хорн решатели

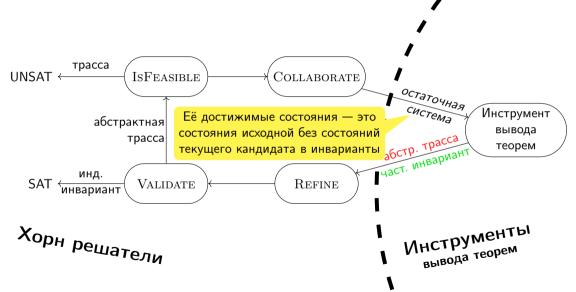
IPROVER CVC5

VAMPIRE E

ZIPPERPOSITION

Инструменты <sub>вывода теорем</sub>





## Теоретическое сравнение классов инвариантов

Класс	ELEM	SizeElem	Reg	Reg_	$\mathrm{Reg}_{\times}$	ELEMREG
Свойство				'		
Замкнут по ∩	Да	Да	Да¹	Дa <sup>2</sup>	Да <sup>2</sup>	Да
Замкнут по ∪	Да	Да	Да¹	Да <sup>2</sup>	Дa <sup>2</sup>	Да
Замкнут по \	Да	Да	Да¹	Дa <sup>2</sup>	Дa <sup>2</sup>	Да
Разрешимо $\overline{t} \in \mathit{I}$	Дa <sup>3</sup>	Да⁴	Да <sup>5</sup>	Да <sup>7</sup>	Да <sup>9</sup>	Да <sup>10</sup> Да <sup>10</sup>
Разрешимо $I=\varnothing$	Дa <sup>3</sup>	Да⁴	Да <sup>6</sup>	Да <sup>8</sup>	Да <sup>9</sup>	Да <sup>10</sup>
Выразимы рекурсив-	Нет	Частично	Да	Да	Да	Да
ные отношения		146171 1116				
Выразимы синхрон-	Да	Да	Нет	Частично	Да	Да
ные отношения				lacin mo		

Класс	Elem	SizeElem	Reg	Reg+	$\mathrm{Reg}_{\times}$	ElemReg
Elem	Ø	Ø	lr <sup>1,4,5</sup>	<i>lr</i> <sup>1,5</sup>	$lr^1$	Ø
SizeElem	$\infty$	Ø	<i>lr</i> <sup>1,4,5</sup>	<i>lr</i> <sup>1,5</sup>	$lr^1$	lt <sup>3</sup>
Reg	even <sup>2</sup>	even <sup>2</sup>	Ø	$\varnothing^4$	$\varnothing^{4,5}$	Ø
Reg+	even <sup>2,7</sup>	even <sup>2,4</sup>	$\infty^4$	Ø	$\varnothing^{5}$	/t <sup>3</sup>
$Reg_{\times}$	even <sup>2,4,5</sup>	even <sup>2,4,5</sup>	$\infty^{4,5}$	$\infty^{5}$	Ø	/t <sup>3,5</sup>
ElemReg	$\infty$	even <sup>2</sup>	$\infty$	<i>lr</i> <sup>1,5</sup>	lr <sup>1</sup>	Ø

#### Реализация

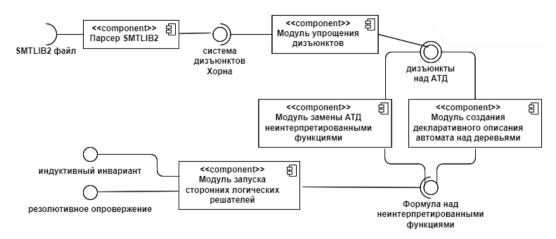


Рис.: Хорн-решатель RINGEN: https://github.com/Columpio/RInGen

### Сравнение Хорн-решателей с поддержкой АТД

Инструмент	Класс	Метод	Возвращает	Полностью
	инвариантов		инвариант	автоматический
SPACER	Elem	IC3/PDR	Да	Да
Racer	Catelem	IC3/PDR	Нет	Нет
Eldarica	SizeElem	CEGAR	Да	Да
VeriCaT	_	Трансф.	Нет	Да
HoIce	ELEM	ICE	Да	Да
RCHC	Reg+	ICE	Да	Да
RInGen(cvc5)	Reg	Трансф. +	Да	Да
		FMF		
RInGen(Vampire)	_	Трансф. +	Нет	Да
		Насыщение		
RInGen-Sync	$\mathrm{Reg}_{ imes}$	Трансф. +	Да	Да
		FMF		
RInGen-CICI(cvc5)	ElemReg	$CEGAR(\mathcal{O})$	Да	Да
RInGen-CICI(VAMPIRE)	_	$CEGAR(\mathcal{O})$	Нет	Да

#### Эксперименты

Инструмент	SAT	UNSAT
RACER	26	22
Eldarica	46	12
VeriCaT	16	10
cvc5-Ind	0	13
RInGen(cvc5)	25	21
RInGen(Vampire)	135	46
RInGen-Sync	43	21
RInGen-CICI(cvc5)	117	19
RInGen-CICI(VAMPIRE)	189	28

