

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
1 Описание предметной области	5
1.1 Логическое и реляционное программирование	5
1.1.1 miniKanren	7
1.2 Специализация программ	9
1.2.1 Специализация логических языков	10
1.2.2 Методы суперкомпиляции	11
1.3 Постановка задачи	12
2 Имеющиеся наработки	13
2.1 Алгоритмы суперкомпиляции	13
2.2 Язык μ Kanren	16
2.3 Библиотека для специализации miniKanren	17
2.4 Обобщённый алгоритм суперкомпиляции	20
3 Суперкомпиляция miniKanren	23
3.1 Реализация суперкомпилятора	23
3.1.1 Граф процессов	24
3.1.2 Раскрытие определений отношений	26
3.1.3 Построение остаточной программы	27
3.2 Модификации суперкомпилятора	28
3.2.1 Стратегии свёртки	28
3.2.2 Стратегии развёртки	28
3.2.3 Стратегии обобщения	31
3.2.4 Расширение μ Kanren оператором неэквивалентности . .	34
4 Экспериментальное исследование	37
4.1 Тестовое окружение	37
4.1.1 Набор тестов	38
4.2 Пример работы базового суперкомпилятора μ Kanren	39
4.3 Сравнение вариаций суперкомпилятора μ Kanren	42
4.3.1 Исследование модификации с ограничением неэквивалентности	45

4.4 Сравнение суперкомпилятора с существующими решениями . .	47
ЗАКЛЮЧЕНИЕ	50
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	51
ПРИЛОЖЕНИЕ А	56

ВВЕДЕНИЕ

Реляционное программирование — это чистая форма логического программирования, в которой программы представляются как наборы математических отношений [4]. Отношения не различают входные и выходные параметры, из-за чего одно и то же отношение может решать несколько связанных проблем. К примеру, отношение, задающее интерпретатор языка, можно использовать не только для вычисления программ по заданному входу, но и для генерации возможных входных значений по заданному результату или самих программ по спецификации входных и выходных значений.

miniKanren — это семейство встраиваемых предметно-ориентированных языков программирования [4]. miniKanren был специально сконструирован для поддержки реляционной парадигмы, опираясь на опыт логических языков, таких как языки семейства Prolog [48], Mercury [40] и Curry [11].

Реляционная парадигма довольно сложна, хотя потенциал её весьма велик. Часто наиболее естественный способ записи отношения не является эффективным. В частности, при задании функциональных отношений как сопоставления выходов входам, как это наблюдается в примере с интерпретатором, поиск входов по выходам практически всегда работает медленно.

Специализация — это техника автоматической оптимизации программ, при которой на основе программы и её частично известного входа порождается новая, более эффективная программа, которая сохраняет семантику исходной. Для специализации логических языков используются методы частичной дедукции [25], самый проработанный из которых — это конъюнктивная частичная дедукция [6]. ECCE, реализация конъюнктивной частичной дедукции для Prolog, показывает хорошие результаты [28], однако, в силу различий между реляционным и логическим программированием, можно предположить возможность разработать более подходящий метод специализации. Уже существует адаптация конъюнктивной частичной дедукции для miniKanren [32], однако её результаты нестабильны: несмотря на то, что в некоторых случаях производительность программ улучшается, в других — она может существенно ухудшиться.

Другой подход для специализации — это суперкомпиляция, техника автоматической трансформации и анализа программ, при которой программа символично исполняется с сохранением истории вычислений, на основе ко-

торой строится оптимизированная версия кода. Суперкомпиляция успешно применяется к функциональным и императивным языкам, однако для логических языков не сильно развита. Существуют работы, посвящённые демонстрации сходства процессов частичной дедукции и суперкомпиляции [10], а также предназначенный для Prolog суперкомпилятор APROPOS [7], который, однако, довольно ограничен в своих возможностях и требует ручного контроля.

В данной работе предлагается способ адаптации и реализации суперкомпилятора для реляционного языка `miniKanren`, а также рассматриваются его возможные вариации, приводящие к дальнейшему повышению производительности реляционных программ, и производится экспериментальное исследование результата.

1 Описание предметной области

1.1 Логическое и реляционное программирование

Логическое программирование — это вид декларативного программирования, основанный на формальной логике. Программы представляются в виде утверждений, представляющихся логическими формулами и описывающих определённую область проблем. “Вычисление” программы в контексте логического программирования производится в форме *поиска* доказательства утверждений на основе заданных фактов — аксиом — и правил вывода в соответствии с заданной *стратегией поиска* [48].

Стратегия поиска задаёт, каким образом происходит обход пространства поиска ответов, и, как следствие, определяет как, какие и в каком порядке будут находиться ответы. Стратегия поиска, при которой каждый возможный ответ будет со временем выдан, называется *полной*. Чаще применяются *неполные* стратегии, поскольку они менее требовательны к вычислительным ресурсам [12].

Самые известные языки логического программирования — языки семейства Prolog. Prolog применяется для доказательства теорем [14], проектирования баз знаний, создания экспертных систем [34] и искусственного интеллекта [3]. Prolog строится на логике предикатов первого порядка в форме дизъюнктов Хорна (то есть дизъюнктов только с одним положительным литералом) и использует *метод резолюций*, основанный на доказательстве от противного, для решения задач. Prolog вводит разнообразные синтаксические конструкции с *побочными эффектами*, то есть с действиями, приводящими к изменению *окружения* программы, к примеру, оператор отсечения (англ. *cut*), который влияет на способ вычисления программы, предотвращая нежелательные вычисления. Также он использует стратегию обхода в глубину, что приводит к тому, что поиск может “зациклиться” и никогда не выдать оставшиеся решения, но, тем не менее, благодаря своим расширениям Prolog остаётся хорошим решением для задач из своей области применения [48].

Реляционное программирование — это форма чистого логического программирования, в которой программы задаются как набор математических *отношений* [4].

Исторически, понятие реляционного программирования появилось

раньше [33] и задавало саму концепцию программы как отношения, когда логическое — появилось позже и предоставляло реализацию его идей [48]. Однако в данной работе под реляционным программированием понимается непосредственно то, что указано выше в определении.

В терминах реляционного программирования, к примеру, сложение $X + Y = Z$ может быть выражено отношением¹

$$\text{add}^o(X, Y, Z),$$

которое в зависимости от того, какие переменные заданы, порождает все возможные значения переменных, при которых отношение выполняется:

- $\text{add}^o(1, 2, 3)$ — проверка, что значения находятся в отношении;
- $\text{add}^o(1, 2, A)$ — поиск всех таких A , при которых $1 + 2 = A$;
- $\text{add}^o(A, B, 3)$ — поиск всех таких A и B , при которых $A + B = 3$;
- $\text{add}^o(A, B, C)$ — поиск всех троек A, B и C , при которых $A + B = C$.

Чистые отношения не предполагают функциональных зависимостей между переменными, поэтому поиск можно проводить в разных “направлениях”, в зависимости от того, какие переменные заданы, как показано выше в примере.

Для функциональных отношений можно ввести понятие направления: для запуска в “прямом” направлении задаются входные аргументы, а для запуска в “обратном” задаётся результат.

К примеру, отношение “меньше” для двух чисел X и Y можно задавать как $\text{less}^o(X, Y)$ и получить чистое реляционное отношение. Отношение же с функциональной зависимостью — $\text{less}_2^o(X, Y, R)$, где R сообщает, состоят ли X и Y в отношении, и тогда задание X и Y будет прямым направлением, а задание R — обратным.

Одно из применений реляционной парадигмы — *реляционные интерпретаторы*. Для языка L его интерпретатор — это функция eval_L , которая принимает на вход программу p_L на этом языке, её вход i и возвращает некоторый выход o :

$$\text{eval}_L(p_L, i) \equiv \llbracket p_L \rrbracket(i) = o$$

¹Символ o традиционно используется для обозначения отношения.

Реляционную версию интерпретатора можно представить как отношение:

$$\text{eval}_L^o(p_L, i, o).$$

При запуске такого отношения в разных направлениях можно не только вычислять результат, но и по программе p_L и выходу o искать возможные входы i или вовсе генерировать программу по указанным выходам и входам.

Можно отметить, что в языках, основанных на классическом Prolog, производить подобные вычисления для получения осмысленных результатов значительно труднее.

1.1.1 miniKanren

miniKanren – семейство встраиваемых предметно-ориентированных языков, специально спроектированное для реляционного программирования [4].

Основная реализация *miniKanren* написана на языке Scheme [8], однако существует множество реализаций в ряде других языков, в том числе Clojure, OCaml, Haskell и другие².

miniKanren предоставляет набор базовых конструкций: унификация (\equiv), конъюнкция (\wedge), дизъюнкция (\vee), введение свежей переменной (*fresh*), вызов реляционного отношения, — представляющий ядро языка, и разнообразные расширения, к примеру, оператор неэквивалентности (англ. *disequality constraint*) или операторы, не являющиеся чистыми, к примеру, предоставляющие функциональность отсечения из Prolog. Хотя в *miniKanren* введены операторы с эффектами, его использование как реляционного языка подразумевает работу только с чистыми операторами.

Классический пример — программа конкатенации двух списков — указан на рисунке 1. Здесь список R является конкатенацией списков X и Y в случае, когда список X пуст, а Y равен R , либо когда X и R декомпозируются на голову и хвост, а их хвосты состоят в отношении конкатенации с Y .

Для выполнения конкатенации над списками необходимо сформировать *запрос* (или *цель*). В запросе в аргументах указываются либо замкнутые термы, либо термы со свободными переменными. Результатом выполнения является список подстановок для свободных переменных, при которых отношение

²Официальный сайт языка *miniKanren*: minikanren.org, дата последнего посещения: 19.05.2020

```

1 appendo X Y R =
2   X ≡ [] ∧ Y ≡ R ∨
3   fresh (H X' R')
4     (X ≡ H :: X') ∧
5     (R ≡ H :: R') ∧
6     appendo X' Y R'

```

Рисунок 1 — Пример программы на miniKanren ($::$ — конструктор списка)

выполняется; когда свободных переменных нет, подстановка, соответственно, пустая. Полное отсутствие каких-либо подстановок говорит об отрицательном результате выполнения запроса.

На рисунке 2 приведён пример запроса, в котором мы хотим найти возможные значения переменных Y и R . Потенциально может быть бесконечное число ответов, к примеру, когда все аргументы в запросе — переменные, поэтому в системах miniKanren есть возможность запрашивать несколько первых ответов; в примере, это число 1. Ответы могут содержать в себе как конкретные замкнутые термы (к примеру, числа), так и свободные переменные, которые в примере обозначаются как $_n$. В примере одна и также свободная переменная $_0$ назначена и Y , и R . Это означает, что какое бы ни было значение Y , оно всегда будет являться хвостом хвоста R .

```

1 > run 1 (Y R) (appendo [1, 2] Y R)
2 Y = \_0
3 R = 1 :: 2 :: \_0

```

Рисунок 2 — Пример выполнения запроса к отношению конкатенации.

В определение miniKanren входит особый алгоритм поиска ответов — чередующийся поиск (англ. *interleaving search*), который рассматривает всё пространство поиска и является полным [20]. Это свойство чередующегося поиска определяет, вместе с отсутствием нечистых расширений, реляционность miniKanren.

Хотя miniKanren уже применяется в индустрии для поиска лечения редких генетических заболеваний в точной медицине [35], на данном этапе своего развития используется в основном в исследовательских целях:

- реляционные интерпретаторы на miniKanren для решения задач поиска [32];

- техника программирования по примерам [5];
- для доказательства теорем [36] ;
- в области вычислительной лингвистики [45].

Однако miniKanren обладает рядом существенных недостатков. Несмотря на то, что он ближе всего подошёл к реализации чистой реляционности, вычислительно он всё же зависим от сложности и ветвистости программ, из-за чего их запуск в разных направлениях может работать с разной скоростью, в особенности запуск в обратном направлении, который зачастую работает очень медленно. Также, описывать сложные задачи в качестве отношений — нетривиальная задача, наивно написанные реляционные программы вычисляются крайне неэффективно. Из-за чего, к примеру, существует транслятор из функционального языка в miniKanren [31], однако порождаемые им отношения — функциональные, а запуск их в обратном направлении, как указывалось выше, крайне непроизводителен [32].

Одно из возможных решения проблем производительности — специализация.

1.2 Специализация программ

Специализация — это метод автоматической оптимизации программ, при котором из программы удаляются избыточные вычисления, на основе информации о входных аргументах программы [17].

Специализатор spec_L языка L принимает на вход программу p_L и часть известного входа этой программы i_s (*статические данные*) и генерирует новую программу \hat{p}_L , которая ведёт себя на оставшемся входе i_d (*динамические данные*) также, как и оригинальная программа (формула 1).

$$\llbracket \text{spec}_L(p_L, i_s) \rrbracket(i_d) \equiv \hat{p}_L(i_d) \equiv \llbracket p_L \rrbracket(i_s, i_d) \quad (1)$$

Специализатор производит все вычисления, зависимые от статических данных, протягивание констант, инлайнинг и другие.

Одно из интересных теоретических применений специализации — это *проекция Футамуры* [9]. Процесс специализации интерпретатора на программу на языке L $\text{spec}_L(\text{eval}_L, p_L)$ порождает *скомпилированную* программу \hat{p}_L , а процесс специализации специализатора на интерпретатор языка L

$\text{spec}_{L''}(\text{spec}_{L'}, \text{eval}_L)$, в свою очередь, порождает *компилятор*. Это первая и вторая проекции Футамуры соответственно. Однако реализация специализаторов, которые бы не оставляли в порождаемой программе следы интерпретации, сложная и труднодостижимая задача [17].

Специализация разделяется на два больших класса: *online* и *offline* алгоритмы:

- *offline*-специализаторы — это двухфазные алгоритмы специализации, в первой фазе которого происходит разметка исходного кода, к примеру, с помощью анализа времени связывания [17], и во второй фазе — непосредственно во время специализации — *только* на основе полученной разметки принимаются решения об оптимизации;
- *online*-специализаторы, напротив, принимают решения о специализации на лету и могут произвести вычисления, для которых *offline*-специализатор сгенерировал бы код.

1.2.1 Специализация логических языков

Частичная дедукция — класс методов специализации логических языков, основанных на построении деревьев вывода, которые отражают процесс вывода методом резолюций, и анализе отдельно взятых атомов логических формул [25].

Реализации методов частичной дедукции успешно применяются для Prolog [18], в частности, система *offline* частичной дедукции LOGEN показывает хорошие результаты при специализации интерпретаторов и для некоторых интерпретаторов достигает для генерируемых программ отсутствие накладных расходов на интерпретацию, однако требует ручной модификации разметки [30].

Конъюнктивная частичная дедукция — одно из расширений метода частичной дедукции, отличительная особенность которой состоит в том, что конъюнкции рассматриваются как единая сущность наравне с атомами [6]. С помощью конъюнктивной частичной дедукции возможно добиться различных оптимизационных эффектов, среди которых выделяется дефорестация (англ. *deforestation*) [46] — оптимизация, при которой удаляются промежуточные структуры данных, — и таплинг (англ. *tupling*) [15] — оптимизация, при ко-

торой множество проходов по одной структуре данных заменяется на один проход. Это наиболее проработанный и мощный метод частичной дедукции.

Реализация методов частичной дедукции, включая конъюнктивную частичную дедукцию, для Prolog представлена в виде системы ЕССЕ [27].

В работе [32] представляется адаптация конъюнктивной частичной дедукции для miniKanren. Реализация добивается существенного роста производительности, однако, как будет показано в разделе 4, в силу особенностей метода и его направленности на Prolog, нестабильно даёт хорошие результаты и в некоторых случаях может затормозить исполнение программы.

1.2.2 Методы суперкомпиляции

Суперкомпиляция — метод анализа и преобразования программ, который отслеживает обобщённую возможную историю вычислений исходной программы и строит на её основе эквивалентную ему, более оптимальную программу. Впервые метод был предложен в работе [44].

Упрощение достигается путём удаления или преобразования некоторых избыточных действий: удаление лишнего кода, выполнение операций над уже известными данными, избавление от промежуточных структур данных, инлайнинг, превращение многопроходного алгоритма в однопроходный и другие [41]. Также суперкомпиляция может применяться для специализации программ.

Суперкомпиляция включает в себя частичные вычисления, однако не сводится к ним полностью и может привести в глубоким структурным изменениям оригинальной программы.

Суперкомпиляторы, которые используют только “положительную” информацию — то есть информацию о том, что свободные переменные чему-то равны, — называют позитивными (англ. *positive supercompilation*) [43]. К примеру, при достижении условного выражения **if** $x = a$ **then** t_1 **else** t_2 позитивный суперкомпилятор при вычислении t_1 будет учитывать то, что $x = a$, однако при вычислении t_2 он не будет знать, что $x \neq a$. Расширение позитивного компилятора с поддержкой такой “негативной” информации — идеальный суперкомпилятор (англ. *perfect supercompilation*) [39].

Техника суперкомпиляции в основном применяется для функциональных [2, 43] и императивных [21] языков. Для логических языков суперкомпи-

ляция слабо развита, однако существует посвящённые ей работы: в работе [10] демонстрируется схожесть подходов частичной дедукции и суперкомпиляции, в работе [7] представлен позитивный суперкомпилятор APROPOS для Prolog, однако эта версия довольно ограничена в своих возможностях и требует ручного вмешательства.

1.3 Постановка задачи

Таким образом, целью данной работы является улучшение качества специализации реляционных программ на miniKanren с использованием методов суперкомпиляции. Для этого были поставлены следующие задачи:

- реализовать базовый суперкомпилятор;
- рассмотреть и реализовать возможные модификации алгоритма суперкомпиляции, учитывающие свойства реляционных языков;
- произвести экспериментальное исследование результата.

2 Имеющиеся наработки

В этом разделе описывается более подробно метод суперкомпиляции, а также библиотека для специализации `miniKanren`, на основе которой велась разработка.

2.1 Алгоритмы суперкомпиляции

Общая схема суперкомпилятора представлена на рисунке 3.

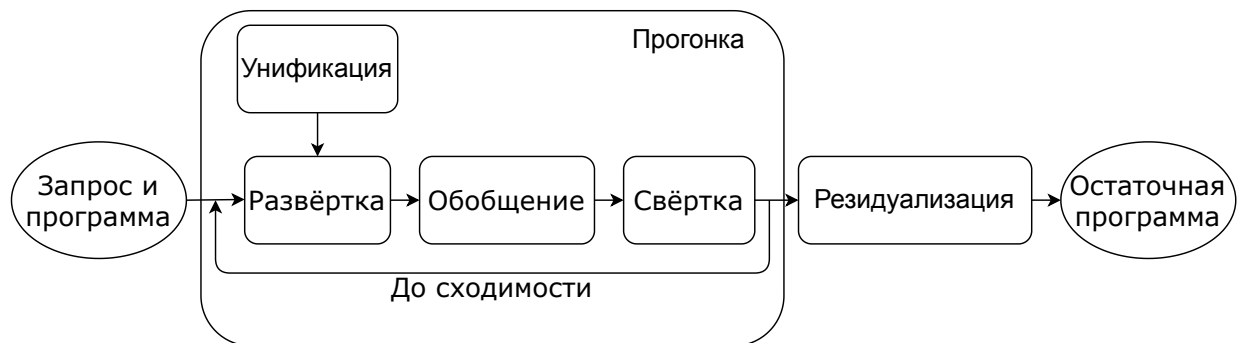


Рисунок 3 — Общая схема суперкомпилятора.

История вычислений при суперкомпиляции представляется в виде *графа процессов* — корневого ориентированного графа, в котором каждая ветвь — это отдельный путь вычислений, а каждый узел — состояние системы или *конфигурация*. Конфигурация обобщённо описывает множество состояний вычислительной системы или её подсистемы. К примеру, конфигурацией можно назвать выражение $1 + x$, в котором параметр x пробегает все возможные значения своего домена (допустим, множество натуральных чисел) и задаёт таким образом множество состояний программы [44].

Прогонкой (англ. *driving*) называется процесс построения графа процессов. Во время прогонки производится шаг символьных вычислений, после которого в граф процессов добавляются порождённые конфигурации; множество конфигураций появляется тогда, когда ветвления в программе зависят от свободных переменных. В данной работе шаг символьного вычисления именуется *развёрткой*.

В процессе прогонки в конфигурациях могут появляться новые свободные переменные, которые строятся из исходной конфигурации: если при вычислении выражения его переменная перешла в другую переменную (к примеру, из-за сопоставления с образцом), то в итоговую конфигурацию будет

подставлена новая переменная и связь старой и новой сохранится в некоторой *подстановке*. Подстановка — это отображение из множества переменных в множество возможно замкнутых термов. Применение подстановки к выражению заменит все вхождения переменных, принадлежащих её домену, на соответствующие термы.

Пример графа процессов представлен на рисунке 4. Здесь при исполне-

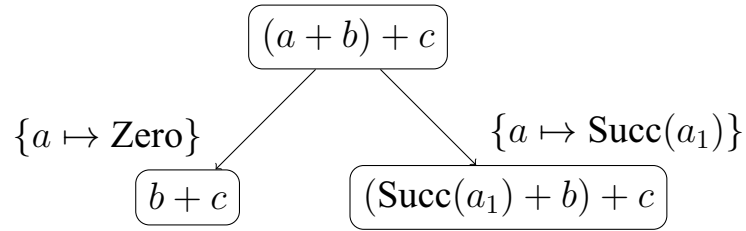


Рисунок 4 — Пример части графа процессов.

нии выражения $(a + b) + c$, где a, b, c — натуральные числа, были рассмотрены возможные значения a : либо оно равно нулю (конструктор *Zero*), либо это некоторое число a_1 , которому прибавили единицу (конструктор *Succ*). Эти два случая могут задавать различные пути исполнения и, соответственно, добавлены в дерево процессов как два различных состояния, в одно из которых войдёт программа при исполнении.

Процесс прогонки может быть бесконечным, к примеру, когда происходят рекурсивные вызовы. Для превращения бесконечного дерева вычисления в конечный объект, по которому можно восстановить исходное дерево, используется *свёртка*.

Свёртка (англ. *folding*) — это процесс преобразования дерева процессов в граф, при котором из вершины v_c добавляется ребро в родительскую вершину v_p , если выражение в конфигурации в v_c и в v_p равны с точностью до переименования. Пример ситуации для свёртки изображён на рисунке 5, на котором свёрточное ребро изображено пунктиром.

Однако существуют ситуации, при которых свёртка не приведёт к тому, что граф превратится в конечный объект. Такое может произойти, к примеру, когда два выражения структурно схожи, но не существует переименования, уравнивающих их.

Для решения этой проблемы используется *обобщение* [42]. Обобщение — это процесс замены одной конфигурации на другую, более абстрактную, описывающую больше состояний программы. Для обнаружения “похожей”

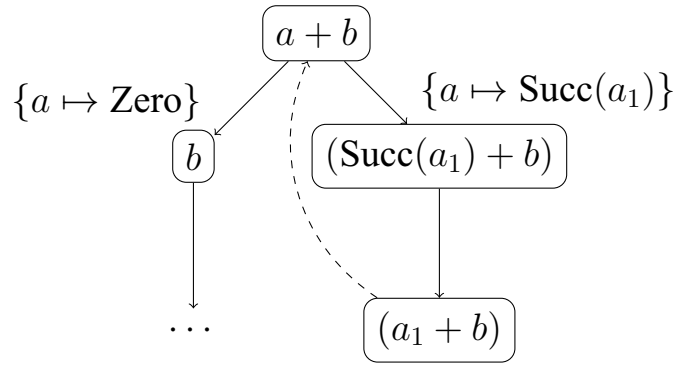


Рисунок 5 — Пример свёртки.

конфигурации используется предикат, традиционно называемый *свистком*: свисток определяет, похожа ли конфигурация на кого-то из родителей текущей конфигурации. В случае, когда свисток сигнализирует о найденной схожести, применяется обобщение. Сам шаг обобщения может произвести действия двух видов:

- *обобщение вниз* приводит к тому, что новая конфигурация заменяет текущую в графе процессов;
- *разделение* (англ. *split*) используется для декомпозиции выражения, элементы которого затем будут обработаны отдельно.

Пример обобщения представлен на рисунке 6

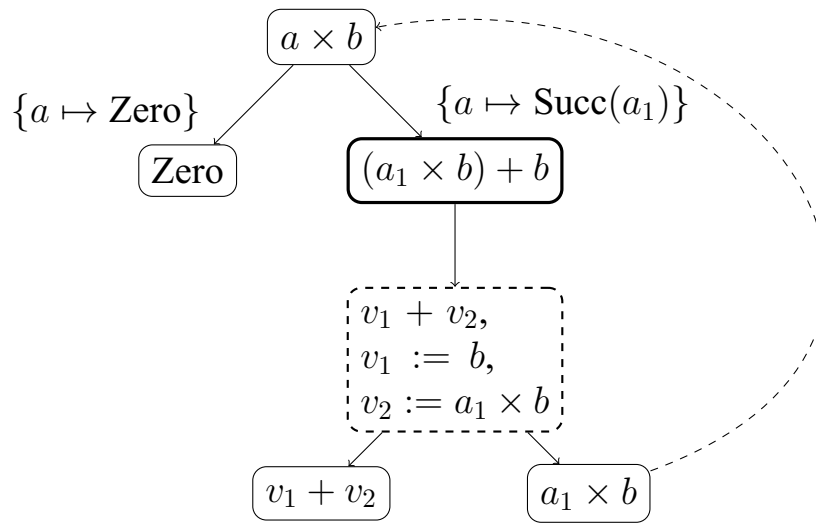


Рисунок 6 — Пример обобщения.

Построение программы по графу конфигураций называется *резидуализацией*, а построенная программа — *остаточной* (англ. *residual*). Алгоритм

выявления остаточной программы основан на обходе дерева, но в остальном полностью зависит от языка.

2.2 Язык μKanren

В данной работе для специализации был выбран μKanren — минималистичный диалект языка miniKanren [13]. μKanren содержит только чистые операторы, что значительно упрощает процесс специализации.

Абстрактный синтаксис языка представлен на Рисунке 7.

\mathcal{C}	$= \{C_i\}$	конструктор с арностью i
\mathcal{X}	$= \{x, y, z, \dots\}$	переменные
\mathcal{T}_X	$= X \cup \{C_i(t_1, \dots, t_i) \mid t_j \in \mathcal{T}_X\}$	термы над множеством переменных
\mathcal{D}	$= \mathcal{T}_\emptyset$	замкнутое выражение
\mathcal{R}	$= \{R_i\}$	реляционный символ с арностью i
\mathcal{G}	$= \mathcal{T}_X \equiv \mathcal{T}_X$	унификация
	$\mathcal{G} \wedge \mathcal{G}$	конъюнкция
	$\mathcal{G} \vee \mathcal{G}$	дизъюнкция
	$\text{fresh } \mathcal{X} . \mathcal{G}$	введение свежей переменной
	$R_i(t_1, \dots, t_i), t_j \in \mathcal{T}_X$	вызов отношения
\mathcal{S}	$= \{R_i^j = \lambda x_1 \dots x_i . g_j; \}$	спецификация программы

Рисунок 7 — Синтаксис языка μKanren [38].

- Унификация двух термов $t_1 \equiv t_2$ порождает подстановку θ , называемую *унификатором*, такую что её применение к термам уравнивает их: $t_1\theta = t_2\theta$.

Алгоритм унификации языков семейства miniKanren использует проверку вхождения (англ. *occurs check*), что гарантирует корректность получаемых унификаторов, однако довольно сильно замедляет выполнение программ.

- Конъюнкция двух целей $g_1 \wedge g_2$ подразумевает одновременное успешное выполнение выражений g_1 и g_2 .
- Дизъюнкция двух целей $g_1 \vee g_2$ подразумевает, что достаточно, чтобы хотя бы одно из выражений g_1 или g_2 выполнялось успешно. Следует

отметить, что при выполнении g_1 выражение g_2 также будет вычисляться.

- Введение свежей переменной `fresh` в языках `miniKanren` нужно указывать явно, в отличие, к примеру, от `Prolog`, где это происходит неявно.
- Вызов реляционного отношения приводит к тому, что переданные в отношение термы унифицируются с аргументами отношения и подставляются в тело отношения.

В контексте вычислений важно различие между *синтаксическими* переменными, которые определяются в тексте программы и обычно представляются строковыми литералами, и *семантическими* переменными, которые непосредственно используются в процессе вычислений и представляются целыми числами, с которыми легче работать и генерировать свежие.

μ Kanren является ядром языка `miniKanren` и может быть без труда расширен необходимыми конструкциями.

2.3 Библиотека для специализации `miniKanren`

Реализация суперкомпилятора для μ Kanren строилась на основе проекта по специализации μ Kanren с помощью конъюнктивной частичной дедукции³ на функциональном языке программирования Haskell. Результаты специализации μ Kanren представлены в работе [32].

Библиотека вводит ряд структур данных и алгоритмов для реализации конъюнктивной частичной дедукции, однако существует возможность её переиспользования для суперкомпиляции в силу нескольких доводов:

- схожесть методов частичной дедукции и суперкомпиляции [10], из чего следует, что ряд вспомогательных функций и определений и для частичной дедукции, и для суперкомпиляции будут совпадать;
- алгоритм обобщения конъюнктивной частичной дедукции [6], о котором подробнее будет рассказано позже, имеет ряд общих черт с процессом обобщения в суперкомпиляции;

³Проект https://github.com/kajigor/uKanren_transformations/, дата последнего посещения: 10.05.2020

- библиотека предоставляет возможность преобразования сгенерированной программы на `miniKanren`, при котором удаляются излишние унификации и происходит удаление лишних аргументов (англ. *redundant argument filtering*) [29], что приводит к увеличению производительности программы, поскольку унификация — операция дорогая.

В библиотеке введены структуры данных, описывающие термы и выражения языка в соответствии с синтаксисом `μKanren` на рисунке 7. Над ними введён ряд важных структур и алгоритмов, речь о которых пойдёт ниже. Основные операции над выражениями в конъюнктивной частичной дедукции производятся над конъюнкциями вызовов реляционных отношений.

Во-первых, в библиотеке реализован алгоритм унификации двух термов. Операция унификации находит наиболее общий унификатор (англ. *most general unifier*), причём единственный, то есть такой унификатор θ , что для любого другого унификатора θ' существует подстановка σ , с которой композиция наиболее общего унификатора даёт θ' : $\theta' = \sigma \circ \theta$ [24]. К примеру, для двух термов $f(X, 2)$ и $f(1, Y)$ наиболее общим унификатором является подстановка $\{X \mapsto 1, Y \mapsto 2\}$, когда подстановки вроде $\{X \mapsto 1, Y \mapsto Z, Z \mapsto 2\}$ также унифицирует термы, однако содержат в себе лишние элементы. Поиск наиболее общего унификатора уменьшает размер итоговой подстановки и является предпочтительным.

Во-вторых, реализованы предикаты над термами, которые проясняют описанные ниже возможные связи термов.

- Выражение e_2 является *экземпляром* выражения e_1 ($e_1 \preceq e_2$), если существует такая подстановка θ , применение которой приравнивает два выражения $e_1\theta = e_2$; также говорят, что e_1 более общий, чем e_2 . К примеру, $f(X, Y) \preceq f(Y, X)$.
- Выражение e_2 является *строгим* экземпляром выражения e_1 ($e_1 \prec e_2$), если $e_1 \preceq e_2$ и $e_2 \not\preceq e_1$. К примеру, $f(X, X) \prec f(X, Y)$, но не наоборот.
- Выражения e_1 и e_2 *варианты* друг друга $e_1 \approx e_2$, если они являются экземплярами друг друга.

Предикат над вариантами определяет, являются ли два терма переименованием друг друга, и поэтому может быть использован для свёртки графа

процессов. Предикаты над экземплярами определяют схожесть термов и используется в обобщении и в алгоритмах суперкомпиляции [43].

В-третьих, в качестве свистка используется отношение *гомеоморфного вложения* [42]. Отношение гомеоморфного вложения \sqsubseteq определено индуктивно:

- переменные вложены в переменные: $x \sqsubseteq y$;
- терм X вложен в конструктор с именем C , если он вложен в один из аргументов конструктора:

$$X \sqsubseteq C_n(Y_1, \dots, Y_n) : \exists i, X \sqsubseteq Y_i;$$

- конструкторы с одинаковыми именами состоят в отношении вложения, если в этом отношении состоят их аргументы:

$$C_n(X_1, \dots, X_n) \sqsubseteq C_n(Y_1, \dots, Y_n) : \forall i, X_i \sqsubseteq Y_i.$$

К примеру, выражение $c(b) \sqsubseteq c(f(b))$, но $f(c(b)) \not\sqsubseteq c(f(b))$.

Преимущество использования гомеоморфного вложения, в первую очередь, состоит в том, что для этого отношения доказано, что на бесконечной последовательности выражений e_0, e_1, \dots, e_n обязательно найдутся такие два индекса $i < j$, что $e_i \sqsubseteq e_j$, вне зависимости от того, каким образом последовательность выражений была получена [43]. Это свойство позволяет доказать завершаемость алгоритма суперкомпиляции.

Однако отношение гомеоморфного вложения допускает, чтобы термы $f(X, X)$ и $f(X, Y)$ находились в отношении $f(X, X) \sqsubseteq f(X, Y)$ в силу того, что все переменные вкладываются друг в друга. Однако обобщение $f(X, X)$ и $f(X, Y)$ не привело бы к более общей конфигурации.

Отношение *строгого* гомеоморфного вложения \sqsubseteq^+ вводит дополнительное требование, чтобы терм X , состоящий в отношении с Y , не был *строгим* экземпляром Y [26]. В таком случае отношение $f(X, X) \not\sqsubseteq^+ f(X, Y)$, поскольку $f(X, Y)$ является строгим экземпляром $f(X, X)$ из-за того, что существует подстановка $\{X = X, Y = X\}$.

В рамках конъюнктивной частичной дедукции понятие гомеоморфного вложения было расширено на конъюнкции выражений. Пусть $Q = A_1 \wedge \dots \wedge A_n$ и Q' — конъюнкции термов, тогда $Q \sqsubseteq Q'$, тогда и только тогда, когда $Q' \not\sqsubseteq Q$ и

существует упорядоченные подконъюнкции $A'_1 \wedge \dots \wedge A'_n$ конъюнкции Q' , такие что $A_i \sqsubseteq A'_i$ [6]. Конъюнкция Q' может содержать в себе больше выражений за счёт того, что в этом случае при обобщении произойдёт шаг разделения. Это расширение было реализовано в рассматриваемой библиотеке.

В-четвёртых, реализован алгоритм обобщения для конъюнктивной частичной дедукции. В общем, алгоритмы обобщения основаны на понятии *наиболее тесного обобщения*.

- *Обобщение* выражения e_1 и e_2 — это выражение e_g , такое что $e_g \preceq e_1$ и $e_g \preceq e_2$. На пример, обобщением выражения $f(1, Y)$ и $f(X, 2)$ является $f(X, Y)$.
- *Наиболее тесное обобщение* (англ. *most specific generalization*) выражений e_1 и e_2 — это обобщение e_g , такое что для каждого обобщения $e'_g \preceq e_1$ и $e'_g \preceq e_2$ выполняется $e'_g \preceq e_g$ [43]. Функция обобщения принимает на себя два терма t_1 и t_2 и возвращает тройку $(t_g, \theta_1, \theta_2)$, такую что $t_1\theta_1 = t_g$ и $t_2\theta_2 = t_g$, при этом θ_1 и θ_2 назовём *обобщающими унификаторами* (англ. *generalizers*).

Алгоритм обобщения для конъюнктивной частичной дедукции согласован с определением гомеоморфного вложения и используется в рамках конъюнктивной частичной дедукции для построения более сложных алгоритмов обобщения, свойственных методам частичной дедукции. Однако его можно использовать как самостоятельный алгоритм для суперкомпиляции, который соединяет в себе возможность произвести шаги обобщения и разделения вместе, не выделяя отдельные шаги для это в процессе прогонки.

2.4 Обобщённый алгоритм суперкомпиляции

На основе вышесказанного можно построить обобщённый алгоритм суперкомпиляции, который в псевдокоде представлен на рисунке 8.

Алгоритм суперкомпиляции принимает на себя программу и запрос, на который необходимо специализировать программу, и после инициализации начальных значений, включающих в себя некоторое *окружения программы* (`env` на строке 8), в котором хранятся все вспомогательные структуры, запускает процесс прогонки. Прогонка производится до схождения и производит следующие действия в зависимости от состояния:

```

1  supercomp(program, query):
2    env ← createEnv program
3    configuration ← initialize query
4    graph ← emptyTree
5    drive(env, graph, configuration)
6    return residualize graph
7
8  drive(env, graph, configuration):
9    if configuration is empty
10   then add(env, graph, success node)
11   else if  $\exists$  parent: configuration  $\approx$  parent
12   then add(env, graph, renaming node)
13   else if  $\exists$  parent: parent  $\trianglelefteq^+$  configuration
14   then
15     add(env, graph, abstraction node)
16     children ← generalize(configuration, parent)
17      $\forall$  child  $\in$  children:
18       drive(env, graph, child)
19   else
20     add(env, graph, unfolding node)
21     children ← unfold(env, configuration)
22      $\forall$  child  $\in$  children:
23       drive(env, graph, child)

```

Рисунок 8 — Обобщённый алгоритм суперкомпиляции.

- если конфигурация пустая (строка 9), это означает, что вычисления успешно сошлись в конкретную подстановку. В таком случае происходит добавление в граф листового узла с этой подстановкой;
- если существует такая родительская конфигурация, что она является вариантом текущей (строка 11), то происходит свёртка и в граф добавляется листовой узел с ссылкой на родителя;
- если же среди родителей находится такой, на котором срабатывает свисток (строка 13), тогда производится обобщение, порождающее дочерние конфигурации (строка 16), на которых продолжается процесс прогонки;
- иначе происходит шаг символьного вычисления (строка 19), на котором порождаются конфигурации.

Окружение для суперкомпиляции должно сохранять следующие объекты:

- подстановку, в которой содержатся все накопленные непротиворечивые унификации, необходимую в процессе прогонки для проверки новых унификаций;
- первую свободную семантическую переменную, необходимую для генерации свежих переменных, к примеру, при абстракции;
- определение программы, необходимое для замены вызова на его тело при развёртывании.

После завершения этапа прогонки из графа процессов извлекается остаточная программа (строка 6).

3 Суперкомпиляция miniKanren

3.1 Реализация суперкомпилятора

В качестве конкретного языка для реализации суперкомпилятора был выбран Haskell. Самая существенная причина выбора в том, на этом языке написана разобранная в предыдущем разделе библиотека для специализации языка μ Kanren, которую можно эффективно переиспользовать для реализации суперкомпилятора. Несмотря на то, что некоторые техники суперкомпиляции сложно выразить в Haskell, в основном они с лёгкостью перекладываются на функциональную парадигму.

Рассмотрим более внимательно схему суперкомпилятора на рисунке 9. На нём жёлтым цветом выделены модули, которые были взяты из библиотеки специализации для построения суперкомпилятора.

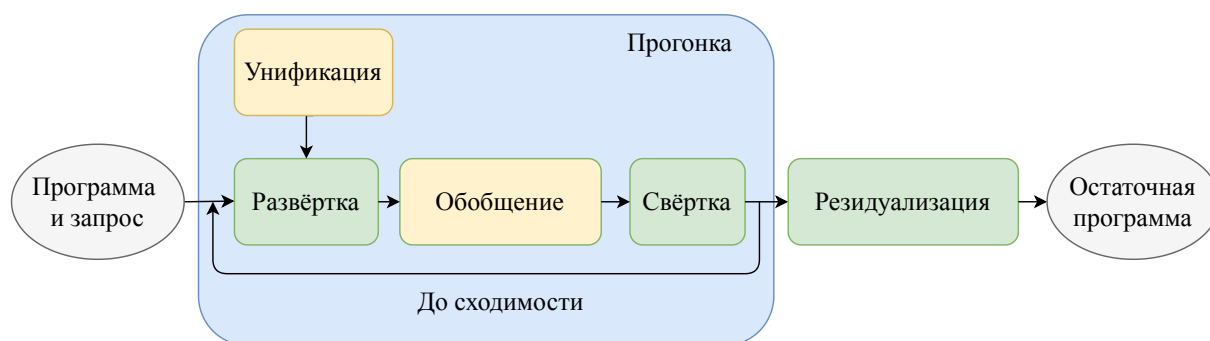


Рисунок 9 — Уточнённая схема суперкомпилятора.

Тогда необходимы модули свёртки, развёртки, резидуализации, а также модифицированный под задачу модуль обобщения, которые затем нужно собрать в единый суперкомпилятор:

- реализация модуля свёртки заключается всего лишь в стратегии выбора множества конфигураций, на которые можно применять непосредственно операцию свёртки;
- реализация модуля развёртки требует проработки стратегий символического вычисления применительно для μ Kanren;
- реализация всего суперкомпилятора требует проработку структур данных, самая важная из которых — граф процессов;
- реализация резидуализации состоит из обхода графа процессов, во время которого происходит построение остаточной программы.

3.1.1 Граф процессов

Представление графа процессов в Haskell затруднено тем, что графовые структуры данных обычно требуют ссылок на произвольные узлы, что приводит к появлению перекрёстных ссылок. Прямая реализация этой идеи сложна в разработке и поддержке и не является идиоматичной, хотя и используется в реализациях суперкомпиляторов для небольших функциональных языков [47]. Использование *IORef*⁴, хотя и предоставляет изменяемость, приводит к неоправданному усложнению кода всего проекта, лишая код функциональной чистоты. Также есть возможность использовать структуру данных отображения для представления графа процессов, как это сделано в реализации на Haskell работы [22], однако в данной работе используется метод, который обычно применяется в частичной дедукции.

Заметим, что история вычислений представляется в виде графа из наличия *обратных рёбер* — то есть рёбра от детей к их родителям, — которые появляются при свёртке, когда ребёнок является переименованием родителя. Тогда, если уметь сохранять или восстанавливать информацию об этой связи, то достаточно будет представить граф в качестве *дерева* процессов. Древоподобная структура однозначно отображается на процесс символьных вычислений, а также с ними легко и идиоматично работать в Haskell.

Структура дерева процессов представлена на рисунке 10.

```

1 type Conf = Conjunction (RelationalCall FreeVar)
2
3 type Subs = Variable  $\mapsto$  Term
4
5 data Tree where
6   Failure      :: Tree
7   Success      :: Subst  $\rightarrow$  Tree
8   Renaming     :: Conf  $\rightarrow$  Subst  $\rightarrow$  Tree
9   Abstraction  :: Conf  $\rightarrow$  Subst  $\rightarrow$  List Tree  $\rightarrow$  Tree
10  Generalizer   :: Subst  $\rightarrow$  Tree  $\rightarrow$  Tree
11  Unfolding     :: Conf  $\rightarrow$  Subst  $\rightarrow$  List Tree  $\rightarrow$  Tree

```

Рисунок 10 — Описание дерева процессов.

Конфигурация *Conf* определена как выражение со свободными пере-

⁴Документация модуля Data.IORef: <https://hackage.haskell.org/package/base-4.11.1.0/docs/Data-IORef.html>, дата последнего посещения: 14.05.2020

менными. В узле дерева процессов хранится конфигурация, приведённая к форме, содержащей только конъюнкцию вызовов реляционного отношения. Это сделано из тех соображений, что, во-первых, дизъюнкция представляет собой ветвление вычислений, посему, соответственно, представляется как ветвление в дереве процессов, во-вторых, унификации производятся во время символьных вычислений и добавляются в подстановку, в-третьих, так как введение свежей переменной оказывает влияние лишь на состояние, в котором производятся вычисления, неосмысленно сохранять его в конфигурации.

Подстановка `Subst` соответствует своему математическому определению как отображению из переменных в термы. Узлы дерева процессов представляют шаги суперкомпиляции и исходы вычисления выражений:

- `Failure` обозначает неудавшиеся вычисления. Такой исход случается при появлении противоречивых подстановок;
- `Success`, напротив, обозначает удавшееся вычисление, которое свелось к подстановке типа `Subst`;
- `Renaming` обозначает узел, конфигурация которой является переименованием какого-то родительского узла.
- `Abstraction` обозначает узел, который может быть обобщён на одного из родителей. После обобщения может появиться несколько конфигураций, которые являются результатом применения разделения. Эти конфигурации добавляются в качестве списка дочерних поддеревьев в текущий узел;
- `Generalizer` хранит себе обобщающий унификатор, который порождается во время обобщения двух термов, и поддерево с обобщённой конфигурацией;
- `Unfolding` обозначает шаг символьного вычисления, на котором произошёл шаг вычислений и по рассматриваемой на этом шаге конфигурации породились новые конфигурации.

Суперкомпилятор строит дерево процессов в глубину.

3.1.2 Раскрытие определений отношений

Стратегия символьного вычисления определяется на этапе развёртки: происходит “развёртывание” одного или более вызовов реляционного отношения, при котором происходит замена вызова на определение. Развёртка по данной конфигурации C порождает множество конфигурации $\{C_1, \dots, C_n\}$, описывающих состояния, в которые может перейти процесс реального исполнения программы. Классически, шаг символьного вычисления соответствует семантике языка, который суперкомпилируется, и для μKanren существует сертифицированная семантика [38], однако описание шага символьного вычисления μKanren для суперкомпиляции усложнено тем, что реляционные языки не исполняются привычным образом, как, к примеру, функциональные программы. Их “выполнение” происходит через *поиск*, который не отображается на классический процесс суперкомпиляции.

Тогда порождение конфигурации можно рассматривать не как непосредственный шаг символьного вычисления, но как определение набора состояний, которые будут рассмотрены. Такое состояние появляется путём раскрытия тела одного или нескольких конъюнктов конфигурации.

К примеру, рассмотрим часть программы на μKanren на рисунке 11, в котором определены отношения f и g .

1	$f(a) = f'(a) \vee f''(a)$
2	$g(a, b) = g'(a) \wedge g''(b)$

Рисунок 11 — Пример отношений для демонстрации шага символьных вычислений

Допустим, на шаге суперкомпиляции алгоритм обрабатывает конфигурацию $f(v_1) \wedge g(v_1, v_2)$ и нужно сделать шаг символьного вычисления. Рассмотрим несколько способов породить новые конфигурации.

- Если раскроется определение f , то будут получены новые конфигурации $f'(v_1) \wedge g(v_1, v_2)$ и $f''(v_1) \wedge g(v_1, v_2)$.
- Если раскроется определение g , то будет получена новая конфигурация $f(v_1) \wedge g'(v_1) \wedge g''(v_2)$.
- Если раскроются оба определения f и g , то будут получены новые конфигурации $f'(v_1) \wedge g(v_1) \wedge g''(v_2)$ и $f''(v_1) \wedge g(v_1) \wedge g''(v_2)$.

Последний набор конфигураций — это полный набор состояний, в которые процесс вычислений может прийти. В первых двух наборах, можно отметить, порождённые конфигурации не исключают возможные состояния процессов, отображённые в последнем наборе. Они могут появиться на последующих шагах вычисления, если перед этим ветвь исполнения не будет остановлена из-за противоречивой подстановки.

Таким образом, какой бы способ развёртывания определений ни был бы выбран, он не будет исключать состояния, в которые процесс вычисления теоретически может прийти, но выбор разных стратегий развёртывания может систематически приводить к разным деревьям процессов, а следовательно — к различным эффектам специализации.

Базовой стратегией порождения новых конфигураций выбрана *полная стратегия развёртывания*, пример которой был представлен выше, при которой заменяются определения всех реляционных вызовов конфигурации.

3.1.3 Построение остаточной программы

Выявление остаточной программы по дереву процессов — *резидуализация* — породит новые определения отношений. Больше одного отношения из дерева процессов может появиться в случае, когда узлы Renaming указывают на узлы, отличные от корня. Поэтому первой фазой происходит пометка узлов, задающих таким образом отношения, а также удаление поддеревьев, у которых все ветви вычисления пришли к неудаче.

Далее происходит обход дерева, во время которого генерируются узлы синтаксического дерева программы в зависимости от типа текущего узла дерева процессов:

- Узел Unfoldable приводит к появлению дизъюнкций подпрограмм, которые задают дети этого узла. Это обусловлено тем, что при прогонке в этом узле происходит ветвление вычислений;
- Узел Abstraction приводит к появлению конъюнкций подпрограмм, которые задают дети этого узла. Это обусловлено тем, что хотя операция обобщения выявляет подконъюнкции из конфигурации и рассматривает их отдельно, оба поддерева, задающиеся этими подконъюнкциями, должны выполняться в одно и то же время;

- Узел `Generalizer` задаёт обобщающий унификатор, который должен быть добавлен перед своим поддеревом;
- Узел `Renaming` формирует вызов реляционного отношения;
- Узел `Success` представляет собой успешное вычисление, предоставляющее непротиворечивую подстановку.

3.2 Модификации суперкомпилятора

Предполагается, что базовый алгоритм специализации должен быть хуже по скорости и, в некоторых случаях, по качеству специализации. Данная работа не затрагивает способы повышения эффективности процесса суперкомпиляции, однако демонстрирует некоторые продвижения в эту сторону из-за чисто прагматических соображений. Основное внимание уделяется подходам к суперкомпиляции и их влиянию на суперкомпиляционные эффекты.

3.2.1 Стратегии свёртки

В базовом алгоритме суперкомпиляции поиск узлов на которые происходят переименования происходит среди родителей. Это напрямую соотносится с понятием символьных вычислений: по достижении узла, который является переименованием уже встреченного, вычисление переходит на родительский узел. Однако довольно часто встречается такая ситуация, что в разных поддеревьях дерева процессов встречаются одинаковые конфигурации, поддеревья которых оказываются полностью идентичными. В таком случае, кажется очевидной и несложной оптимизация, при которой запоминаются вычисленные поддеревья и в случае, когда встречается схожая конфигурация, поддерево не вычисляется заново, а добавляется ссылка на него.

3.2.2 Стратегии развёртки

Как уже говорилось, разные стратегии развёртывания реляционных вызовов могут привести к разным эффектам специализации. К примеру, полная стратегия развёртывания, которая была принята за базовую, может приводить к дефорестации, описанной ранее.

Основной недостаток базового подхода в том, что для получения всех возможных состояний он производит декартово произведение нормализованных состояний тел вызовов в конъюнкциях, что приводит к сильному разрастанию и дерева процессов и, как следствие, сильно требователен к вычислительным ресурсам и приводит к большой ветвистости программ. Последнее оказывает негативный эффект на процесс вычисления и может ухудшить производительность. Вследствие чего реализация новых стратегий развёртывания производится в исследовательских и прикладных целях.

Для лёгкой подмены стратегий суперкомпиляции был разработан специальный интерфейс `Unfoldable` (рисунок 12).

```

1 class Unfoldable a where
2   initialize :: Conf → a
3   get       :: a → Conf
4   unfold    :: a → Env → List (Env, a)

```

Рисунок 12 — Интерфейс для различных стратегий развёртывания.

Предоставляемые интерфейсом функции используются в алгоритме суперкомпиляции следующим образом:

- `initialize` оборачивает конфигурацию в структуру, в которой может содержаться вспомогательная информация для процесса развёртывания;
- `get` позволяет получить конфигурацию для её применения к операциям, не зависящим от стратегий;
- `unfold` непосредственно проводит шаг вычисления на основе текущей конфигурации и её окружения, порождая новые конфигурации с соответствующими им состояниями.

В работе рассмотрен и реализован ряд стратегий, описанных ниже.

- **Модифицированная полная стратегия развёртки**, при которой сначала из цели раскрываются все нерекурсивные вызовы. Нерекурсивность определяется лишь тем, содержит ли определение реляционный вызов самого себя. Более сложный анализ структуры функций не мог бы быть использован в силу того, что тогда было бы необходимо реализовать класс алгоритмов анализа, что совершенно отдельная задача.

- **Стратегия развёртывания самого первого элемента**, при которой ожидается, что конфигурации не будут часто разбиваться на подконъюнкции, что может привести к оптимизационным эффектам, схожим с полным развёртыванием.
- **Последовательная стратегия развёртывания**, при которой отслеживается, какой вызов был раскрыт на предыдущем шаге, чтобы на текущем раскрыть следующий за ним. Ожидается, что в этой стратегии при суперкомпиляции будут учтены, так или иначе, все или большинство конъюнтов конфигурации.
- **Нерекурсивная стратегия развёртывания**, являющаяся модификацией последовательной стратегии, при которой в первую очередь раскрывается нерекурсивный вызов в конфигурации.

Ожидается, что при нерекурсивной стратегии развёртывания из конфигураций будут как можно быстрее появляться выражения, которые могут быть сокращены или вовсе удалены из-за унификации (к примеру, отношения, кодирующие таблицы истинности, такие как `ando`) или привести к скорой свёртке.

- **Рекурсивная стратегия развёртывания**, являющаяся модификацией последовательной стратегии, при которой в первую очередь раскрывается рекурсивный вызов в конфигурации. Ожидается, что это может привести к скорому обобщению.
- **Стратегия развёртывания вызовов с минимальным количеством ветвлений**, при которой на каждом шаге вычисления будет появляться минимально возможное количество конфигураций, что приведёт к минимальной ветвистости дерева.
- **Стратегия развёртывания вызовов с максимальным количеством ветвлений**, при которой на каждом шаге вычисления будет появляться максимально возможное количество конфигураций, что, с одной стороны, увеличит количество возможных состояний, но потенциально может привести к скорому сворачиванию или обобщению.

3.2.3 Стратегии обобщения

Для модификации стратегии обобщения были рассмотрены два подхода.

Во-первых, увеличение множества рассматриваемых конфигураций, на которые производится обобщение, до всех уже рассмотренных конфигураций.

Покажем, что допустимо использовать обобщение на все вершины. Для этого рассмотрим конфигурацию C_w и некоторую конфигурацию C_p из множества конфигураций для обобщения, расширенное всеми уже обработанными конфигурациями. При обобщении C_p и C_w породится множество дочерних конфигураций, каждая из которых будет более общим подконъюнтом C_w и, в силу своей обобщённости, будет содержать меньше информации, чем исходная конфигурация, однако не будет ей противоречить. В итоге, обобщение на все обработанные вершины влияет только на то, каким образом разбивается рассматриваемая конфигурация, что может привести к скорой свёртке, и как много информации о переменных теряется.

Обобщение на вычисленные узлы приводит к тому, что деревья конфигураций быстрее сходятся, однако остаётся под вопросом, ухудшает ли этот подход качество специализации.

Во-вторых, допустимо ввести модификацию, которая запрещает производить операцию обобщения, если родительская конфигурация была получена с помощью обобщения. Такая эвристика может привести к тому, что программы будут чаще развёртываться, однако в купе с другими модификациями может привести к интересным результатам.

В-третьих, в суперкомпиляции, в отличие от методов частичной дедукции, для обобщения дополнительно может использоваться техника обобщения вверх, при которой происходит не подвешивание обобщённой конфигурации в качестве потомка конфигурации, которая обобщалась, но замена самого родителя на новую конфигурацию [43]. Старое же поддерево родителя уничтожается.

Для определения необходимости обобщать вверх введём предикат $e_1 < e_2$, который определяет, что $e_1 \prec e_2$ и $e_2 \not\prec e_1$. Такое ограничение необходимо из-за того, что суперкомпилятор оперирует конъюнкциями выражений и делает операции разделения и обобщения вниз за один шаг с конъюнкциями возможно разной длины, однако для обобщения вверх необходимо удостовериться, что замена родительского дерева, во-первых, не добавит конфигураций, кото-

рых там не может быть, во-вторых, предоставит только более общую информацию.

```

1  else if  $\exists$  parent: parent  $\leq$  configuration
2  then
3      node  $\leftarrow$  generalize(configuration, parent)
4      addUp(env, tree, parent, node)

```

Рисунок 13 — Расширение алгоритма суперкомпиляции.

Наличие операции обобщения вверх предполагает, что необходимо умение передвигаться по дереву вверх и изменять его. Реализация в Haskell этой идеи — задача крайне нетривиальная. Возможно представлять деревья в мутабельных массивах, однако при обобщении необходимо удалять целые поддеревья, что при таком подходе сложная операция.

Классическим способом решения этой проблемы являются *зипперы* (англ. *zipper*) [16], на основе которых, к примеру, создан система для суперкомпиляции, представленная в работе [37].

Идиома зипперов предлагает рассматривать структуру данных как пару из элемента, на котором установлен фокус, и контекста, который представляется как структура данных с “дыркой”, в котором сфокусированный элемент должен находиться. К примеру, зиппер для списка `[1, 2, 3, 4, 5]` при фокусе на 3 представляется таким образом: `(3, ([2, 1], [4, 5]))`. Тогда перефокусировка вправо или влево на один элемент происходит за константное время, как и замена элемента, для которой достаточно заменить первую компоненту пары. В то время как, в силу того, что операция взятия элемента в связном списке по индексу происходит за линейное время от длины списка, взятие элемента слева от 3 также будет происходить за линейное время, как и, соответственно, модификация списка. Для деревьев с произвольным количеством детей зиппер может выглядеть как пара из текущего узла и списка родителей, отсортированного в порядке близости к узлу (рисунок 14).

```

1 data Parent = Parent { children :: ListZipper Node }
2 type TreeZipper = (Node, List Parent)

```

Рисунок 14 — Пример структуры зиппера для деревьев

Родительский (структура `Parent`) список детей представлен в виде зип-

пера (поле `children`) для списка, в котором происходит фокус: у непосредственного родителя — на элемент в фокусе, а у остальных родителей — на предыдущего в порядке сортировки. Тогда передвижение вверх до первого подходящего узла происходит путём заполнения дыры при переходе на родителя, а передвижение в левого брата — простой сдвиг в зиппере списка `children`.

При представлении дерева процессов в идиоме зипперов основа алгоритма суперкомпиляции принимает форму описания действий при смене состояния зиппера. Дерево всё ещё строится в глубину и происходит это следующим образом: если мы в узле, порождающем другие узлы (то есть *Unfolding* или *Abstraction*), то порождённые узлы добавляются в дерево с пометкой о том, что они не достроены, а алгоритм спускается в первого ребёнка. Когда же алгоритм приходит в листовой узел, то ему нужно подняться до родителя, у которого существует помеченный потомок, и спуститься в этого потомка. Алгоритм завершается, когда не осталось помеченных детей.

Обобщение вверх приводит к тому, что происходит замена целого поддерева процессов предка, на которого обобщается конфигурация. Иногда это может приводить к потере связи между аргументами, из-за чего исчезает потенциал для возможных положительных эффектов, к примеру, протягивания констант.

К примеру, на рисунке 15 представлено дерево процессов, при котором происходит обобщение вверх.

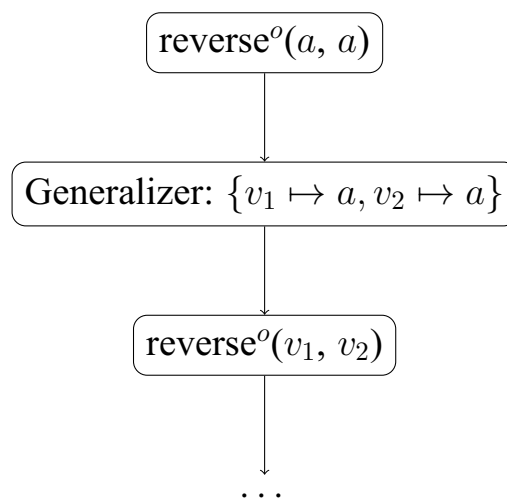


Рисунок 15 — Демонстрация потери информации при обобщении вверх.

С одной стороны, теряется потенциал для генерации более оптимальной

для цели $\text{reverse}^o(a, a)$ программы, но с другой стороны рассмотрим следующие соображения.

В данном примере процесс прогонки происходил следующим образом: некоторое время строился граф для конфигурации $\text{reverse}^o(a, a)$, затем вывелась конфигурация $\text{reverse}^o(a, b)$. Если бы обобщения вверх не происходило бы, то поиск ответов в результирующей программе малое время провёл бы в поддереве, оптимизированном под $\text{reverse}^o(a, a)$, а остальное — в обобщённом $\text{reverse}^o(a, b)$.

В общем случае это может происходить не только с корнем дерева, но и в каких-то его поддеревьях. Однако запрет на обобщение вверх в поддеревьях может сгенерировать слишком много частных случаев и привести к более неэффективным программам.

Из того, что, во-первых, есть потенциал оптимизации при сохранении информации в корне дерева и, во-вторых, необходимо сдерживать разрастание дерева конфигураций, допускается рассмотрение алгоритма с обобщением вверх с запретом на обобщение к корню дерева.

3.2.4 Расширение μKanren оператором неэквивалентности

Множество операций в оригинальном μKanren покрывает все нужды реляционного программирования, однако на ряде программ оно вычислительно допускает пути исполнения, которые не приводят к успеху, однако сообщить об этом не представляется возможным.

К примеру, на рисунке 16 изображена операция поиска значения по ключу в списке пар ключ-значение lookup^o .

1	$\text{lookup}^o \ K \ L \ R =$
2	$(K', V) :: L' \equiv L \wedge$
3	$(K' \equiv K \wedge V \equiv R \vee \text{lookup}^o \ K \ L' \ R)$

Рисунок 16 — Отношения поиска значения по ключу.

В соответствии с программой список L должен иметь в голове пару из ключа и значения (K', V) и либо этот ключ K' унифицируется с искомым ключом K и значение V — с результатом R , либо поиск происходит в хвосте списка L' . Проблема этой программы в том, что если унификация $(K', V) :: L' \equiv L$ прошла успешно и был найден результат, то поиск всё

равно продёт во вторую ветку с рекурсивным вызовом и будет искать значение дальше, хотя по прагматике поиска ключа в списке должен вернуться лишь одно значение. Более того, суперкомпилятору тоже придётся учитывать и, возможно, проводить вычисления, которые не принесут никакой пользы.

В miniKanren существует операция неэквивалентности $t_1 \not\equiv t_1$, вводящая ограничение неэквивалентности (англ. *disequality constraints*)[1]. Операция неэквивалентности определяет, что два терма t_1 и t_2 никогда не должны быть равны, накладывая ограничения на возможные значения свободных переменных терма.

Расширение синтаксиса μ Kanren представлено на рисунке 17.

$$\mathcal{G} = \dots \quad \mathcal{T}_x \not\equiv \mathcal{T}_x \quad \text{неэквивалентность}$$

Рисунок 17 — Расширение синтаксиса μ Kanren относительно указанного на рисунке 7.

Исправленная версия отношения lookup^o представлена на рисунке 18.

```

1 lookupo K L R =
2   (K', V) :: L' ≡ L ∧
3   (K' ≡ K ∧ V ≡ R ∨
4   K' ≢ K ∧ lookupo K L' R)
```

Рисунок 18 — Исправленное отношение поиска значения по ключу.

В такой реализации две по сути исключают друг друга ветви исполнения будут исключать друг друга и при вычислении запросов, и при суперкомпиляции.

Для реализации ограничения неэквивалентности вводится новая сущность под названием “хранилище ограничений” Ω (англ. *constraints store*), которое используется для проверки нарушений неэквивалентности. Окружение расширяется хранилищем ограничений, которое затем используется при унификации и при добавлении новых ограничений.

Тогда нужно ввести следующие модификации в алгоритм унификации конфигурации, который собирает все операции унификации в конъюнкции перед тем, как добавить её в множество допустимых конфигураций.

- При встрече операции неэквивалентности $t_1 \not\equiv t_2$ необходимо произвести следующие действия. Применить накопленную подстановку к термам $t_1\theta = t'_1$ и $t_2\theta = t'_2$ и унифицировать термы t'_1 и t'_2 . Если получился пустой унификатор, значит, эти термы равны и ограничение нарушено. В таком случае суперкомпилятор покинет эту ветвь вычислений. Если же термы не унифицируются, значит, никакая подстановка в дальнейшем не нарушит ограничение. Иначе необходимо запомнить унификатор в хранилище.
- При встрече операции унификации $t_1 \equiv t_2$ необходимо получить их унификатор. Если его не существует или он пуст, то дополнительных действий производить не нужно. Иначе нужно проверить, не нарушает ли унификатор ограничения неэквивалентности.

Указанное расширение было добавлено в библиотеку с реализацией сопутствующих алгоритмов.

4 Экспериментальное исследование

4.1 Тестовое окружение

В качестве основной конкретной реализации μ Kanren для тестирования использовался OCanren⁵[23], реализованный на OCaml[23].

Тесты запускались на ноутбуке со следующими характеристиками: Intel Core i5-6200U CPU, 2.30GHz, DDR4, 12GiB.

Для тестирования суперкомпилятора и его модификаций использовался следующий алгоритм.

1. Подготавливается программа, реализованная на внутреннем представлении μ Kanren.
2. Программа и запрос, на который будет происходить специализация, подаются на вход суперкомпилятору.
3. По дереву процессов, порождённому суперкомпилятором, строится остаточная программа.
4. Остаточная программа транслируется в OCanren и запускается в заранее подготовленном окружении с тестовыми запросами.

Реализованный суперкомпилятор сравнивался с реализацией конъюнктивной частичной дедукции для μ Kanren⁶, а также с реализацией конъюнктивной частичной дедукции для Prolog — системой ECCE⁷. Другие специализаторы не рассматриваются, так как согласно работе [28], специализация с помощью конъюнктивной частичной дедукции в ECCE показывает лучшие результаты.

Для использования последнего требовалось оттранслировать программу на μ Kanren в Prolog, специализировать её на запрос, далее оттранслировать результирующую программу в OCanren. Это допустимо сделать в силу того, что между μ Kanren и чистым подмножеством Prolog есть взаимно однозначное синтаксическое соответствие. Все необходимые средства для этого также предоставлялись указанной библиотекой специализации.

⁵Проект OCanren: <https://github.com/JetBrains-Research/OCanren>, дата последнего посещения: 15.05.2020

⁶Проект: https://github.com/kajigor/uKanren_transformations, дата последнего посещения: 15.05.2020

⁷Проект ECCE: <https://github.com/leuschel/ecce>, дата последнего посещения: 15.05.2020

4.1.1 Набор тестов

Был выбран следующий набор тестов для тестирования и анализа суперкомпилятора и его модификаций.

- Программа `doubleAppend(xs, ys, zs, rs)`, которая производит конкатенацию трёх списков. Она классически используется для проверки эффекта дефорестации в специализированной программе. Специализация происходила на запрос `doubleAppend` со свободными переменными.
- Программа `maxLength(xs, max, len)`, которая находит в списке максимальный элемент и длину списка. Она классически используется для проверки эффекта таплинга в специализированной программе. Специализация происходила на запрос `maxLength` со свободными переменными.
- Программа сортировки `sort(list, result)`. Выбрана в силу показательности результатов. Специализация происходила на запрос `sort` со свободными переменными.
- Отношение, проверяющее принадлежность пути графу `isPath(path, graph, result)`. Специализация происходила на запрос `isPath(path, graph, True)`. Запросы к программе для тестирования:
 - поиск пути заданного размера в случайном графе:
`isPath(path, graph, True) ^ length(path, N);`
 - поиск фиксированного количества путей, принадлежащих данному графу.

В приложении А на рисунке 22 представлен код, который был использован для генерации графов. Функции на вход подаётся количество вершин, количество рёбер и зерно для генератора случайных чисел. В тестах приводятся три графа на 20 вершинах с 50 рёбрами с зёрнами: граф 1 — 42, граф 2 — 34, граф 3 — 106.

- Интерпретатор формул логики высказываний `logint(formula, subst, result)`. Интерпретатор специали-

зируется на то, чтобы всегда генерировать выполнимые формулы `logint(formula, subst, True)`.

- Интерпретатор лямбда-исчисления `lam(expr, result)`. Интерпретатор специализируется на выражение `lam(expr, result)` и используется для генерации n выражений в нормальной форме `lam(expr, expr)`;
- Проверка типов в просто типизированном лямбда-исчислении `infer(typ, expr)`. Отношение использовалось для генерации выражений, соответствующих заданной спецификации типа и выражения. Специализируется выражение:

$$\text{infer}(\text{typ}, \text{expr}) \wedge \text{typ} \equiv \text{TYPE_SPEC} \wedge \text{expr} \equiv \text{EXPR_SPEC}$$

Ко всем вышеперечисленным программам были применены суперкомпиляторы, и результаты выполнения остаточных программ были проверены на адекватность и соответствие задачам исходных программ. Проводилось 10 измерений, пока которым выводились средние значения. Следует отметить, что задачи доказательства сохранения семантики суперкомпилятором на стояло.

4.2 Пример работы базового суперкомпилятора μ Kanren

Разберём на примере результат работы базового суперкомпилятора.

Классическая программа для тестирования эффектов специализации — `doubleAppend`, представленная на рисунке 19, в которой происходит конкатенация списков трёх списков.

```

1 doubleAppend a b c d =
2   fresh (t)
3   (append a b t ^ append t c d)
4 append y4 y5 y6 =
5   (y4 ≡ [] ^ y6 ≡ y5) ∨
6   fresh (ty t h)
7   (y4 ≡ h :: t ^
8    y6 ≡ h :: ty ^
9    append t y5 ty)
```

Рисунок 19 — Программа для тестирования `doubleAppend`

Во многих бенчмарках [19, 28] это программа используется для проверки эффекта дефорестации.

Рассмотрим дерево процессов на рисунке 20 (для компактности `append` сокращён до `app`), которое порождается применением базового суперкомпилятора к программе `doubleAppend`, причём в качестве аргументов — простые свободные переменные, из-за чего просто оптимизируется сама структура программы.

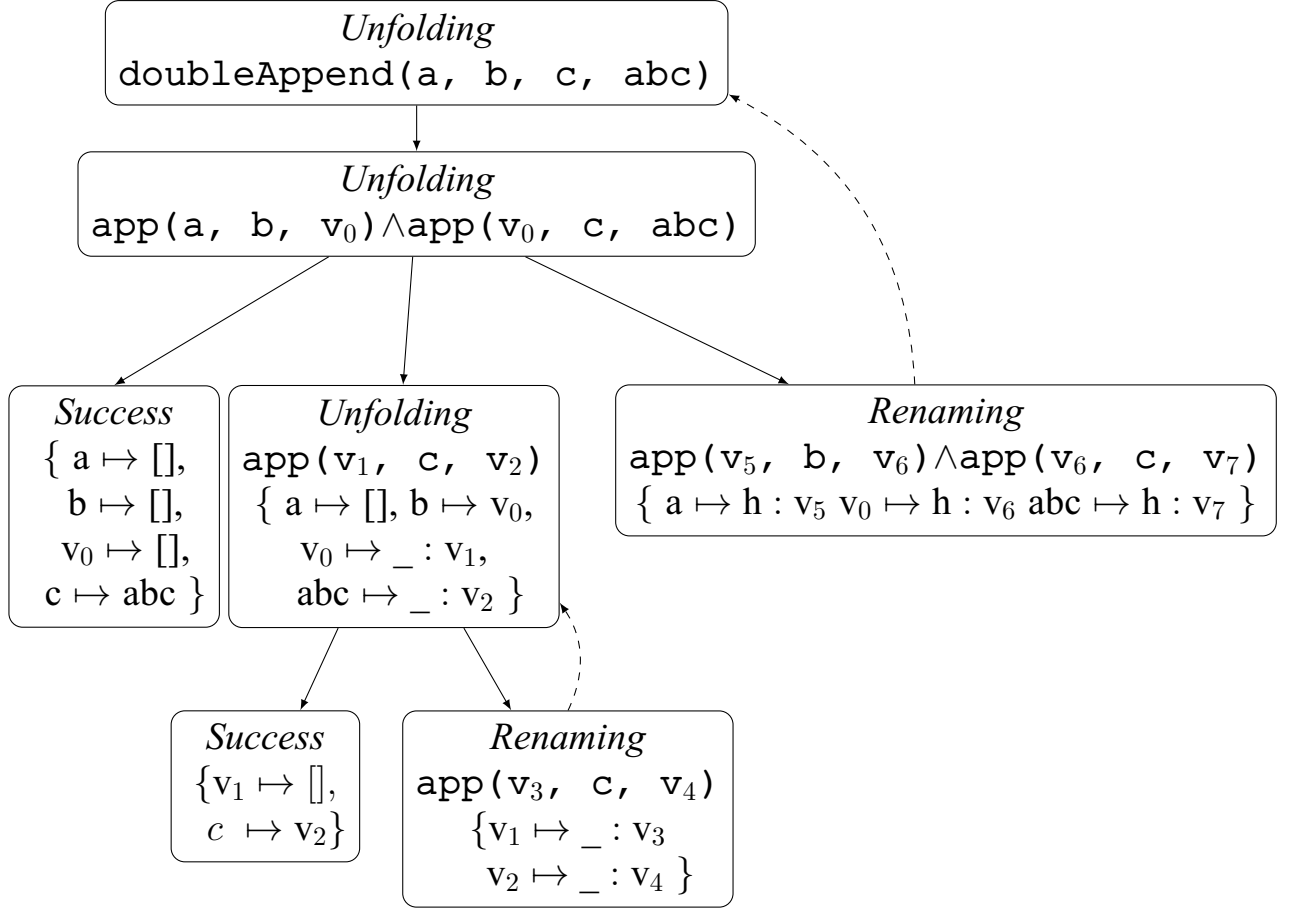


Рисунок 20 — Дерево процессов для программы `doubleAppend`

В приведённом дереве первым этапом происходит замена определения `doubleAppend` на его тело. Так как в определении нет дизъюнкций, получаем всего одну конфигурацию, в которой операцией `fresh` добавляется новая семантическая переменная `v0`. Далее происходит обработка конфигурации с более чем одним конъюнктом. Так как используется полная стратегия развёртывания, каждая из конъюнкций раскрывается по определению и рассматривается их дизъюнктивная нормальная форма, в которой всего четыре дизъюнкта, один из которых не унифицируется, поэтому всего появляется только три кон-

фигурации. Эти конфигурации рассматривают несколько возможных значений переменных:

1. когда a и b пустые списки, то результатом конкатенации является c ;
2. если же b не пуст, тогда результат — это конкатенация списков b и c . Отношение конкатенации двух списков также специализируется под задачу. В результате этой специализации порождается исходное отношение, поскольку никакой новой информации не было выявлено и исходная программа уже была оптимальна;
3. в третьем же случае выводится, что результат конкатенации трёх списков a , b и c — это конкатенация списков v_5 , b и c , где v_5 является хвостом списка, а в голову которой добавили голову a .

По данному дереву порождается остаточная программа, изображённая на рисунке 21.

```

1 doubleAppendo a b c d =
2   fresh (x4) (app3 a b x4 c d)
3 app3 a b t c d =
4   fresh (x7 x6 x5 x10 x9 x8)
5     (a ≡ [] ∧ b ≡ t ∧
6       (t ≡ [] ∧ c ≡ d ∨
7         (t ≡ x8 :: x9) ∧
8           d ≡ x8 :: x10 ∧
9             app2 x9 c x10
10        ) ∨
11        (a ≡ x5 :: x6 ∧
12          t ≡ x5 :: x7 ∧
13            d ≡ x5 :: x10 ∧
14              app3 x6 b x7 c x10) )
15 app2 a b c =
16   fresh (x13 x12 x11)
17   (a ≡ [] ∧ b ≡ c ∨
18     (a ≡ x11 :: x12 ∧
19       c ≡ x11 :: x13 ∧
20       app x12 b x13))

```

Рисунок 21 — Суперкомпилированная программа doubleAppend

В итоге, наблюдается, как двупроходный алгоритм становится однопроходным.

4.3 Сравнение вариаций суперкомпилятора μ Kanren

В таблицах используются условные обозначения для стратегий развёртывания:

- *Full* и *Full-non-rec* обозначают полную стратегию и полную стратегию развёртывания с приоритетом на нерекурсивные вызовы соответственно;
- *Seq* обозначает последовательную стратегию развёртывания;
- *Non-rec* и *Rec* обозначают нерекурсивную и рекурсивную стратегии соответственно;
- *Min* и *Max* обозначают минимальную и максимальную стратегии соответственно;
- *First* обозначает стратегию, при которой всегда развёртывается первый конъюнкт.

А также для суперкомпиляторов:

- *Б.С.* обозначает базовый суперкомпилятор с обобщением вниз на предков;
- *М.1* обозначает модификацию, при которой происходит запрет на обобщение сразу после обобщения;
- *М.2* обозначает модификацию, при которой обобщение происходит на все ранее вычисленные вершины;
- *М.3* обозначает модификацию, при которой происходит обобщение вверх на родительские вершины;
- *М.4* обозначает модификацию, при которой происходит обобщение вверх на родительские вершины, кроме корневой.
- *М.5* обозначает модификацию, при которой происходит обобщение вверх на родительские вершины, а также запрет обобщения после обобщения.

В таблице 1 представлены результаты сравнения модификаций с базовым суперкомпилятором при разных стратегиях развёртывания. При полных стратегиях для `doubleAppend` возникает эффект дефорестации. В остальных стратегиях этого не происходит, из-за чего исполнения по крайней мере

в два раза хуже. Из-за того, что программа довольно небольшая, модификации алгоритма суперкомпиляции хотя не и оказывают влияния, результат не ухудшают.

	<i>Б.С.</i>	<i>М.1</i>	<i>М.2</i>	<i>М.3</i>	<i>М.4</i>	<i>М.5</i>
<i>Full</i>	0.0049	0.0055	0.0056	0.0050	0.0048	0.0053
<i>Full-non-rec</i>	0.0053	0.0048	0.0049	0.0050	0.0037	0.0039
<i>Seq</i>	0.0098	0.0119	0.0099	0.0127	0.0097	0.0102
<i>Non-rec</i>	0.0100	0.0122	0.0099	0.0099	0.0097	0.0098
<i>Rec</i>	0.0094	0.0133	0.0103	0.0098	0.0097	0.0100
<i>Min</i>	0.0096	0.0122	0.0100	0.0093	0.0094	0.0099
<i>Max</i>	0.0096	0.0096	0.0092	0.0101	0.0119	0.0091
<i>First</i>	0.0120	0.0094	0.0094	0.0099	0.0093	0.0093

Таблица 1 — Результат для `doubleAppend` с конкатенацией трёх списков длины 120, секунды

В таблице 2 указаны результаты выполнения для программы `maxLength` на списке `[1..200]`.

	<i>Б.С.</i>	<i>М.1</i>	<i>М.2</i>	<i>М.3</i>	<i>М.4</i>	<i>М.5</i>
<i>Full</i>	0.623	0.638	0.250	0.250	0.251	0.293
<i>Full-non-rec</i>	0.633	0.611	0.602	0.256	0.258	0.298
<i>Seq</i>	0.285	0.290	0.289	0.289	0.287	0.286
<i>Non-rec</i>	0.284	0.289	0.285	0.285	0.287	0.287
<i>Rec</i>	0.856	0.893	0.560	0.577	0.569	0.888
<i>Min</i>	0.317	0.280	0.303	0.284	0.279	0.280
<i>Max</i>	0.279	0.287	0.279	0.278	0.284	0.281
<i>First</i>	0.864	0.858	0.564	0.963	0.949	0.565

Таблица 2 — Запуск для `maxLength` на списке `[1..200]`, секунды

Модификация алгоритма *М.2-М.4* на стратегии *Full* породили программы с эффектом таплинга, при котором максимальный элемент и длина списка высчитывались за один проход. Интересное наблюдение состоит в том, что в других случаях, к примеру, при применении стратегии *Seq*, таплинга не происходит, но сама программа выполняется не существенно медленнее.

В таблице 3 указаны результаты тестирования для программы `sort` при сортировке списков случайного содержания (числа от 0 до 50) длины 50.

Примечательно, что на всех списках указанного свойства алгоритмы отработали практически одинаково. Более примечательно, что какие бы модифи-

	<i>Б.С.</i>	<i>М.1</i>	<i>М.2</i>	<i>М.3</i>	<i>М.4</i>	<i>М.5</i>
<i>Full</i>	0.239	0.252	0.238	0.232	0.239	0.244
<i>Full-non-rec</i>	0.241	0.240	0.241	0.241	0.245	0.242
<i>Seq</i>	0.242	0.240	0.239	0.238	0.240	0.254
<i>Non-rec</i>	0.245	0.242	0.247	0.236	0.244	0.242
<i>Rec</i>	0.239	0.242	0.252	0.240	0.250	0.243
<i>Min</i>	0.245	0.252	0.242	0.242	0.292	0.279
<i>Max</i>	0.246	0.242	0.249	0.246	0.246	0.239
<i>First</i>	0.250	0.248	0.263	0.245	0.239	0.247

Таблица 3 — Запуск для `sort`, секунды

кации ни взяли, порождаются довольно схожие программы. Это обосновывается тем, что само отношение сортировки имеет весьма короткое рекурсивное определение.

В таблицах 4, 5 и 6 указаны результаты тестирования для программы `logint`, которая использовалась для генерации всех выполнимых формул без свободных переменных, с одной свободной переменной и с двумя свободными переменными соответственно.

	<i>Б.С.</i>	<i>М.1</i>	<i>М.2</i>	<i>М.3</i>	<i>М.4</i>	<i>М.5</i>
<i>Full</i>	-	-	0.132	0.105	-	-
<i>Full-non-rec</i>	0.076	0.072	0.210	0.111	0.126	0.081
<i>Seq</i>	0.168	0.181	0.149	0.090	0.091	0.113
<i>Non-rec</i>	0.078	0.092	0.136	0.088	0.114	0.099
<i>Rec</i>	0.081	0.074	0.095	0.064	0.126	0.090
<i>Min</i>	0.080	0.064	0.110	0.092	0.117	0.082
<i>Max</i>	0.164	0.193	0.144	0.077	0.111	0.146
<i>First</i>	0.181	0.164	0.176	0.070	0.189	0.136

Таблица 4 — Запуск `logint` для генерации формул без переменных, секунды.

Прочерки в таблицах обозначают то, что из-за большой требовательности к ресурсам, на тех модификациях, которые не приводят к быстрому сворачиванию программ, включая и базовый суперкомпилятор, не удалось получить оптимизированные программы. Однако ожидается, что при должном количестве ресурсов порождённые программы окажутся слишком большими, что приведёт к значительным проблемам производительности, из-за чего добиваться результатов для данных случаев в данной работе считается нецелесообразным.

	<i>Б.С.</i>	<i>М.1</i>	<i>М.2</i>	<i>М.3</i>	<i>М.4</i>	<i>М.5</i>
<i>Full</i>	-	-	0.078	0.068	-	-
<i>Full-non-rec</i>	0.056	0.045	0.125	0.084	0.082	0.051
<i>Seq</i>	0.109	0.110	0.086	0.063	0.074	0.069
<i>Non-rec</i>	0.046	0.038	0.081	0.072	0.067	0.045
<i>Rec</i>	0.055	0.050	0.074	0.055	0.079	0.055
<i>Min</i>	0.053	0.041	0.066	0.055	0.079	0.059
<i>Max</i>	0.100	0.117	0.108	0.057	0.074	0.077
<i>First</i>	0.118	0.103	0.091	0.068	0.101	0.084

Таблица 5 — Запуск `logint` для генерации формул с одной переменной, секунды.

	<i>Б.С.</i>	<i>М.1</i>	<i>М.2</i>	<i>М.3</i>	<i>М.4</i>	<i>М.5</i>
<i>Full</i>	-	-	0.078	0.062	-	-
<i>Full-non-rec</i>	0.137	0.040	0.093	0.042	0.069	0.040
<i>Seq</i>	0.086	0.082	0.066	0.049	0.050	0.041
<i>Non-rec</i>	0.043	0.031	0.063	0.044	0.055	0.046
<i>Rec</i>	0.037	0.034	0.045	0.040	0.051	0.049
<i>Min</i>	0.037	0.039	0.049	0.041	0.054	0.045
<i>Max</i>	0.068	0.070	0.067	0.036	0.062	0.071
<i>First</i>	0.104	0.100	0.110	0.095	0.137	0.073

Таблица 6 — Запуск `logint` для генерации формул с двумя переменными, секунды.

В представленных таблицах наблюдается, что при увеличении количества свободных переменных, в среднем, уменьшается время генерации формул. Это обосновывается тем, что в деревьях исполнения порожаемых программам больше ветвей оканчиваются успешным поиском значения в подстановке.

4.3.1 Исследование модификации с ограничением неэквивалентности

Для исследования влияния ограничения неэквивалентности рассматривается отношение интерпретатора лямбда-исчисления `lam`, с помощью которого происходил поиск 50-ти термов в нормальной форме.

В таблице 7 приведено время выполнения программы без введения ограничения неэквивалентности. Здесь наблюдается интересная закономерность:

на очередном интерпретаторе модификации *M.1* и *M.5* показывают лучшие результаты.

	<i>Б.С.</i>	<i>M.1</i>	<i>M.2</i>	<i>M.3</i>	<i>M.4</i>	<i>M.5</i>
<i>Full</i>	0.106	0.088	0.102	0.106	0.095	0.084
<i>Full-non-rec</i>	0.106	0.086	0.111	0.105	0.083	0.082
<i>Seq</i>	0.107	0.085	0.105	0.103	0.083	0.080
<i>Non-rec</i>	0.112	0.087	0.100	0.110	0.086	0.084
<i>Rec</i>	0.111	0.083	0.112	0.111	0.090	0.083
<i>Min</i>	0.109	0.095	0.110	0.114	0.093	0.085
<i>Max</i>	0.107	0.098	0.105	0.105	0.086	0.086
<i>First</i>	0.106	0.089	0.111	0.118	0.093	0.092

Таблица 7 — Запуск *lam* для поиска термов в нормальной форме без оператора неэквивалентности, секунды.

В таблице 8 указано время исполнения на программе с использованием оператора неэквивалентности только в отношении подстановки, которое используется интерпретатором.

	<i>Б.С.</i>	<i>M.1</i>	<i>M.2</i>	<i>M.3</i>	<i>M.4</i>	<i>M.5</i>
<i>Full</i>	0.040	0.025	0.052	0.046	0.024	0.024
<i>Full-non-rec</i>	0.042	0.025	0.042	0.044	0.024	0.025
<i>Seq</i>	0.044	0.026	0.046	0.045	0.022	0.027
<i>Non-rec</i>	0.040	0.026	0.046	0.042	0.025	0.022
<i>Rec</i>	0.045	0.023	0.041	0.044	0.027	0.022
<i>Min</i>	0.042	0.034	0.045	0.045	0.023	0.027
<i>Max</i>	0.043	0.026	0.043	0.040	0.027	0.028
<i>First</i>	0.044	0.027	0.045	0.042	0.025	0.023

Таблица 8 — Запуск *lam* для поиска термов в нормальной форме с оператором неэквивалентности, секунды.

Из таблицы виден прирост производительности для методов *M.1*, *M.4* и *M.5* практически в четыре раза, а для остальных – в два.

Разные подходы показывают себя хорошо на разных классах задач. При этом стабильно самой неудачной стратегией развёртки себя показывает стратегия *First*. Самыми неудачными модификациями оказались *M.1* и *M.2*. Модификация *M.2* с обобщением на все рассмотренные вершины стабильно приводит к слишком раннему сворачиванию дерева, из-за чего теряется точность

специализации. Модификация *M.1*, наоборот, делает слишком много развёртываний, что также приводит к потере производительности. В целом, смена стратегии базового суперкомпилятора позволила увеличить производительность большинства порождаемых программ. В среднем, лучше всего работают стратегии *Seq* и её модификация *Non-rec*. Реализация обобщения вверх *M.3* значительно улучшила производительность программ. Её модификация *M.4* выигрывала в больших случаях, однако периодически значительно проигрывает, как и модификация *M.5*, которая хорошо показала себя только на интерпретаторах.

В итоге, оптимальной модификацией считается суперкомпилятор с обобщением вверх с нерекурсивной стратегией развёртывания.

4.4 Сравнение суперкомпилятора с существующими решениями

В таблице 9 представлены результаты сравнения базового алгоритма суперкомпиляции (*B.C.*) и выбранной наиболее успешной модификации суперкомпилятора (*M.C.*) с оригинальными программами (*Оригинал*), а также с оптимизированными программами с помощью системы ECCE для Prolog с использованием трансляции (*ECCE*) и адаптации конъюнктивной частичной дедукции для miniKanren (*CPD*).

В случае с `doubleAppend`, которому на вход подали три одинаковых списка длины 120, базовому суперкомпилятору удалось добиться эффекта дефорестации и сравниться с *ECCE*, модифицированный суперкомпилятор уступает *ECCE*, однако всё ещё быстрее как оригинальной программы, так и *CPD*.

Однако в тесте `maxLength`, которому на вход подали список [1..200], базовый суперкомпилятор пусть и делает таплинг, но слишком сильно ветвит программу, из-за чего она работает хуже оригинальной почти в два раза, однако модифицированный алгоритм пусть не ускоряет исполнение программы, ухудшает её незначительно, в отличие от *CPD*. Стоит заметить, что *ECCE* пусть и улучшил производительность, отличие от оригинальной программы оказалось несущественным.

Исполнение программы сортировки `sort` на случайном списке длины 50 и элементами не более 50, в свою очередь, заметно ускорилось относитель-

Параметр	Оригинал	ECCE	CPD	Б.С	М.С.
doubleAppend	списки фиксированной длины				
120	0.0135	0.0051	0.0133	0.0049	0.0098
maxLength	фиксированный список				
[1..200]	0.257	0.230	0.727	0.623	0.287
sort	случайный список фиксированной длины				
50	8.42	12.28	13.2	0.239	0.242
isPath	10 путей				
граф 3	> 300	1.03	1.19	2.43	1.81
isPath	произвольный путь длины 7				
граф 3	62.12	1.08	1.15	1.34	0.85
isPath	произвольный путь длины 10				
граф 1	12.51	1.01	1.20	1.28	0.48
граф 2	> 300s	1.73	2.09	0.85	0.48
граф 3		9.90	12.73	3.29	1.23
logint	размер подстановки				
0	> 300	0.17	2.7	-	0.11
1		0.09	1.7	-	0.07
2		0.08	0.9	-	0.05
lam	термы в нормальной форме				
50 термов	> 300	2.98	0.08	0.08	0.04
genBySpec	генерация по спецификации				
2000	> 300	0.53	0.62	-	0.36

Таблица 9 — Результаты сравнения алгоритмов специализации, секунды

но и оригинала, и специализированных программ с помощью конъюнктивной частичной дедукции. Это может объясняться тем, что в некоторых случаях методы конъюнктивной частичной дедукции делают лишние развёртывания, что в конечном счёте приводит к ухудшению программы.

Отношение проверки принадлежности пути графу `isPath` использовалось для поиска 10 путей в указанном графе. Здесь результат работы суперкомпиляции проиграл примерно в два раза алгоритмам конъюнктивной частичной дедукции, но решил задачу значительно быстрее оригинала. Однако уже на более осмысленной задаче по поиску путей заданной длины суперкомпилятор значительно выиграл у методов конъюнктивной частичной дедукции.

Также был рассмотрен интерпретатор логических формул `logint`, который запускался для поиска формул и их подстановок, в которой они выполняются, для формул с нулём, одним и двумя свободными переменными. Это

пример запуска в “обратную” сторону, который, как можно наблюдать из таблицы, действительно крайней неэффективен. Специализированные же версии дают значительный прирост производительности, однако модифицированная версия суперкомпилятора выигрывает у всех методов. Прочерки стоят, как объяснялось в предыдущем подразделе, из-за того, что в данном случае стратегия полного развёртывания требовала слишком много ресурсов.

В примере с отношением *lam* базовый суперкомпилятор отработал идентично *CPD*, но модифицированный в два раза быстрее нашёл 50 термов в нормальной форме.

Довольно сложным примером для суперкомпилятора оказалось отношение по генерации термов по спецификации:

- спецификация терма: `app (app _ _) _`;
- спецификация типа: `_ → (_ → not-arrow _)`.

Суперкомпиляция с полной стратегией развертки вновь не уложилась в имеющиеся ограничения на ресурсы, однако модифицированная версия значительно повышает эффективность программы, даже относительно *ECCE* и *CPD*.

ЗАКЛЮЧЕНИЕ

В результате проделанной работы был разработан суперкомпилятор для miniKanren, реализованы его модификации, а также произведено их экспериментальное исследование.

Реализованный суперкомпилятор показал улучшение производительности на подавляющем большинстве рассмотренных программ относительно исходной программы, а также относительно реализаций конъюнктивной частичной дедукции; в иных случаях проблемы с производительностью оказывались несущественными.

Исходный код проекта можно найти на сайте <https://github.com/RehMaar/uKanren-спес>, автор принимал участие под учётной записью RehMaar.

Результаты работы были представлены на конференции TEASE-LP'20.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Claire E. Alvis и др. “cKanren miniKanren with Constraints”. в: *In Proceedings of the 2011 Workshop on Scheme and Functional Programming (Scheme '11), Portland, OR*. 2011.
- [2] Maximilian Bolingbroke и Simon Peyton Jones. “Supercompilation by Evaluation”. в: *SIGPLAN Not.* 45.11 (сент. 2010), с. 135—146. ISSN: 0362-1340. DOI: 10.1145/2088456.1863540. URL: <https://doi.org/10.1145/2088456.1863540>.
- [3] Ivan Bratko. *Prolog programming for artificial intelligence (4th ed.)* Harlow, England ; New York: Addison Wesley.
- [4] William Byrd. “Relational Programming in miniKanren: Techniques, Applications, and Implementations”. в: (сент. 2009), с. 297.
- [5] William E. Byrd и др. “A Unified Approach to Solving Seven Programming Problems (Functional Pearl)”. в: *Proc. ACM Program. Lang.* 1.ICFP (авг. 2017). DOI: 10.1145/3110252. URL: <https://doi.org/10.1145/3110252>.
- [6] Danny De Schreye и др. “Conjunctive partial deduction: Foundations, control, algorithms, and experiments”. в: *The Journal of Logic Programming* 41.2-3 (1999), с. 231—277.
- [7] Stephan Diehl. *A Prolog Positive Supercompiler*. 1997.
- [8] Daniel P. Friedman, William E. Byrd и Oleg Kiselyov. *The Reasoned Schemer*. The MIT Press, 2005. ISBN: 0262562146.
- [9] Yoshihiko Futamura. “Partial Evaluation of Computation Process - An Approach to a Compiler-Compiler”. в: *Systems, Computers, Controls* 2 (1999), с. 45—50.
- [10] Robert Glück и Morten Heine Sørensen. “Partial Deduction and Driving are Equivalent”. в: *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 1994.
- [11] M. Hanus (ed.) *Curry: An Integrated Functional Logic Language (Vers. 0.9.0)*. Available at <http://www.curry-language.org>. 2016.

- [12] M. Hanus, B. Peemöller и F. Reck. “Search Strategies for Functional Logic Programming”. в: *Proc. of the 5th Working Conference on Programming Languages (ATPS'12)*. Springer LNI 199, 2012, с. 61–74.
- [13] Jason Hemann и Daniel P. Friedman. “ μ Kanren: A Minimal Functional Core for Relational Programming”. в: *Proceedings of the 2013 Annual Workshop on Scheme and Functional Programming*. 2013.
- [14] Jieh Hsiang и Mandayam Srivas. “Automatic inductive theorem proving using prolog”. в: *Theoretical Computer Science* 54.1 (1987), с. 3–28. ISSN: 0304-3975. DOI: [https://doi.org/10.1016/0304-3975\(87\)90016-8](https://doi.org/10.1016/0304-3975(87)90016-8). URL: <http://www.sciencedirect.com/science/article/pii/0304397587900168>.
- [15] Zhenjiang Hu и др. “Tupling Calculation Eliminates Multiple Data Traversals”. в: *ACM SIGPLAN Notices* 32 (дек. 1997). DOI: 10.1145/258948.258964.
- [16] Gérard Huet. “The Zipper”. в: *J. Funct. Program.* 7.5 (сент. 1997), с. 549–554. ISSN: 0956-7968. DOI: 10.1017/S0956796897002864. URL: <https://doi.org/10.1017/S0956796897002864>.
- [17] Neil D Jones, Carsten K Gomard и Peter Sestoft. *Partial evaluation and automatic program generation*. Peter Sestoft, 1993.
- [18] Jesper Jørgensen и Michael Leuschel. “Efficiently generating efficient generating extensions in Prolog”. в: *Partial Evaluation, International Seminar, LNCS 1110, pages 238–262, Schloß Dagstuhl*. Springer-Verlag, 1996, с. 238–262.
- [19] Jesper Jørgensen, Michael Leuschel и Bern Martens. “Conjunctive partial deduction in practice”. в: *Proceedings of the International Workshop on Logic Program Synthesis and Transformation (LOPSTR'96), LNCS 1207*. Springer-Verlag, 1996, с. 59–82.
- [20] Oleg Kiselyov и др. “Backtracking, Interleaving, and Terminating Monad Transformers: (Functional Pearl)”. в: *SIGPLAN Not.* 40.9 (сент. 2005), с. 192–203. ISSN: 0362-1340. DOI: 10.1145/1090189.1086390. URL: <https://doi.org/10.1145/1090189.1086390>.

- [21] Andrei V. Klimov. *An Approach to Supercompilation for Object-oriented Languages: the Java Supercompiler Case Study*. 2008.
- [22] Ilya Klyuchnikov и Sergei Romanenko. *SPSC: a Simple Supercompiler in Scala*. 2009.
- [23] Dmitrii Kosarev и Dmitri Boulytchev. “Typed Embedding of a Relational Language in OCaml”. в: *Electronic Proceedings in Theoretical Computer Science* 285 (дек. 2018), с. 1—22. DOI: 10.4204/EPTCS.285.1.
- [24] Jean-Louis Lassez, Michael Maher и Kim Marriott. “Unification Revisited”. в: т. 306. янв. 2006, с. 67—113. DOI: 10.1007/3-540-19129-1_4.
- [25] Michael Leuschel. *Advanced Techniques for Logic Program Specialisation*. тех. отч. 1997.
- [26] Michael Leuschel. “Homeomorphic embedding for online termination of symbolic methods”. в: *In The essence of computation, volume 2566 of LNCS*. Springer, 2002, с. 379—403.
- [27] Michael Leuschel. *The ecce Partial Deduction System*. тех. отч. In Proc. of the ILPS’97 Workshop on Tools и Environments for (Constraint) Logic Programming, U.P, 1997.
- [28] Michael Leuschel, Bern Martens и Danny De Schreye. “Controlling Generalization and Polyvariance in Partial Deduction of Normal Logic Programs”. в: *ACM Trans. Program. Lang. Syst.* 20.1 (янв. 1998), с. 208—258. ISSN: 0164-0925. DOI: 10.1145/271510.271525. URL: <https://doi.org/10.1145/271510.271525>.
- [29] Michael Leuschel и Morten Heine Soerensen. “Redundant Argument Filtering of Logic Programs”. в: Springer-Verlag, 1996, с. 83—103.
- [30] Michael Leuschel и др. “Specialising Interpreters Using Offline Partial Deduction”. в: т. 3049. янв. 2004, с. 340—375. DOI: 10.1007/978-3-540-25951-0_11.
- [31] Boulytchev D. Lozov P. Vyatkin A. “Typed Relational Conversion”. в: Wang M., Owens S. (eds) *Trends in Functional Programming. TFP 2017. Lecture Notes in Computer Science, vol 10788*. Springer, Cham (2018).

- [32] Petr Lozov, Ekaterina Verbitskaia и Dmitry Boulytchev. “Relational Interpreters for Search Problems”. в: *Relational Programming Workshop*. 2019, с. 43.
- [33] Bruce J. MacLennan. *Introduction to relational programming*. тех. отч. 1981.
- [34] Dennis Merritt. *Building Expert Systems in Prolog*. Spring-Verlag, 1989.
- [35] Matthew Might. “The Algorithm for Precision Medicine (Invited Talk)”. в: *Proceedings Companion of the 2019 ACM SIGPLAN International Conference on Systems, Programming, Languages, and Applications: Software for Humanity*. SPLASH Companion 2019. Athens, Greece: Association for Computing Machinery, 2019, с. 2. ISBN: 9781450369923. DOI: 10.1145/3359061.3365208. URL: <https://doi.org/10.1145/3359061.3365208>.
- [36] Joseph Near, William Byrd и Daniel Friedman. “aleanTAP: A Declarative Theorem Prover for First-Order Classical Logic”. в: т. 5366. дек. 2008, с. 238–252. DOI: 10.1007/978-3-540-89982-2_26.
- [37] Jason S. Reich. *Optimus Prime: A new tool for interactive transformation and supercompilation of functional programs*. тех. отч. The University of York, Department of Computer Science, 2009.
- [38] Dmitry Rozplochas и Dmitry Boulytchev. “Certified Semantics for miniKanren”. в: *miniKanren and Relational Programming Workshop*. 2019.
- [39] Jens Peter Secher и Morten Heine Sørensen. “On perfect supercompilation”. в: *Journal of Functional Programming* 6 (1996), с. 465–479.
- [40] Zoltan Somogyi, Fergus James Henderson и Thomas Charles Conway. “The implementation of Mercury, an efficient purely declarative logic programming language”. в: *IN PROCEEDINGS OF THE AUSTRALIAN COMPUTER SCIENCE CONFERENCE*. 1995, с. 499–512.
- [41] Morten Heine Sørensen. *Turchin’s Supercompiler Revisited - An operational theory of positive information propagation*. 1996.
- [42] Morten Heine Sørensen и Robert Glück. “An Algorithm of Generalization in Positive Supercompilation”. в: *ILPS*. 1995.

- [43] Morten Heine Sørensen и Robert Glück. “Introduction to Supercompilation”. в: *Partial Evaluation - Practice and Theory, DIKU 1998 International Summer School*. Berlin, Heidelberg: Springer-Verlag, 1998, с. 246—270. ISBN: 3540667105.
- [44] Valentin F. Turchin. “The Concept of a Supercompiler”. в: *ACM Transactions on Programming Languages and Systems* 8 (1986), с. 292—325.
- [45] Zoltan Varju, Richard Littauer и Peteris Erins. “Using Clojure in Linguistic Computing”. в: февр. 2012.
- [46] Philip Wadler. “Deforestation: Transforming Programs to Eliminate Trees”. в: *Proceedings of the 2nd European Symposium on Programming*. ESOP '88. Berlin, Heidelberg: Springer-Verlag, 1988, с. 344—358. ISBN: 3540190279.
- [47] Илья Ключников. “Суперкомпиляция: идеи и методы”. в: *Практика функционального программирования*, №7 (2011).
- [48] Дж. Малпас. *Реляционный язык Пролог и его применение*. М.: Наука, 1990.

ПРИЛОЖЕНИЕ А

```

1 generateGraph :: Int -> Int -> Int -> [(Int, Int)]
2 generateGraph n m =
3     evalRand (do tree <- forM [1..n-1] $ \i ->
4         getEdge i <$>
5         getRandomR (0, i-1) <*>
6         getRandom
7         rest <- replicateM (m - n + 1) $
8             getRandomR (0, n-2) >>=
9             \i -> getEdge i <$>
10                getRandomR (i+1, n-1) <*>
11                getRandom
12                pure $ tree ++ rest) . mkStdGen
13 where getEdge :: Int -> Int -> Bool -> (Int, Int)
14         getEdge i j True = (i, j)
15         getEdge i j False = (j, i)

```

Рисунок 22 — Генератор случайного ориентированного графа с заданными параметрами.