



Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИНФОРМАТИКА И СИСТЕМЫ УПРАВЛЕНИЯ

КАФЕДРА КОМПЬЮТЕРНЫЕ СИСТЕМЫ И СЕТИ (ИУ6)

НАПРАВЛЕНИЕ ПОДГОТОВКИ 09.03.01 Информатика и вычислительная техника

ОТЧЕТ

О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

***Анализ синтаксиса и семантики стековых
языков программирования***

Студент

ИУ6-73Б

(Группа)

(Подпись, дата)

В.К. Залыгин

(И.О. Фамилия)

Руководитель

Б.И. Бычков

(Подпись, дата)

(И.О. Фамилия)

Оценка _____

2025 г.

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
**(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)**

З А Д А Н И Е
на выполнение научно-исследовательской работы

по теме Анализ синтаксиса и семантики стековых языков программирования

Студент группы ИУ6-73Б

Залыгин Вячеслав Константинович

(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

исследовательская

Источник тематики (кафедра, предприятие, НИР) кафедра

График выполнения НИР: 25% 4 нед., 50% 7 нед., 75% 11 нед., 100% 14 нед.

Техническое задание Выполнить аналитический обзор стековых языков программирования. Выделить особенности стековых языков программирования и провести их сравнительный анализ. Сделать выводы об области применения технологии и тенденциях в её развитии.

Оформление научно-исследовательской работы:

- 1) Отчет на 25-30 листах формата А4.
- 2) Перечень графического (илюстративного) материала (чертежи, плакаты, слайды и т.п.)
Необходимый иллюстративный графический материал включить в качестве рисунков в расчетно-пояснительную записку
- 3) Приложение А. Техническое задание на ВКРБ на 5-8 листах формата А4.

Дата выдачи задания « 1 » сентября 2025 г.

Руководитель

Б.И. Бычков
(И.О. Фамилия)

(Подпись, дата)

Студент

В.К. Залыгин
(И.О. Фамилия)

(Подпись, дата)

РЕФЕРАТ

РПЗ 44 с., 4 рис., 5 табл., 15 источн., 1 прил.

**СТЕК, КОМПИЛЯТОР, СТЕКОВЫЙ ЯЗЫК, ЯЗЫК ПРОГРАММИРОВАНИЯ,
ОБРАТНАЯ ПОЛЬСКАЯ ЗАПИСЬ**

Объектом анализа являются стековые языки программирования.

Предметом анализа являются аспекты построения и применения стековых языков программирования, а также их компиляторов.

Цель работы – проанализировать и сравнить свойства существующих стековых языков программирования, сделать вывод о тенденциях и применимости стековых языков.

В результате работы составлены аналитический обзор и сравнение таких свойств стековых языков как: область применения, модель исполнения, используемые синтаксические конструкции и их семантика, типизация, статический и динамический анализ программ, возможные оптимизации, работа с памятью и подходы к построению стандартной библиотеки. Сформулирована область применения и выделены достоинства стековых языков.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	7
1 Аналитический обзор существующих языков	9
1.1 Язык программирования Forth	9
1.2 Язык программирования Joy	12
1.3 Язык программирования Cat	15
1.4 Язык программирования Factor	18
1.5 Язык программирования Wasm	21
2 Сравнительный анализ	24
ЗАКЛЮЧЕНИЕ	35
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	36
ПРИЛОЖЕНИЕ А ТЕХНИЧЕСКОЕ ЗАДАНИЕ	38

ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

Алгоритм Хиндли-Милнера – алгоритм вывода типов в полиморфных системах типов, автоматически определяющий типы выражений на основе их структуры и правил типизации без явных аннотаций (при соблюдении условий применимости системы)

Виртуальная машина – программная (либо программно-аппаратная) среда, реализующая абстрактную вычислительную модель и обеспечивающая выполнение программ в заданном формате (байткод, промежуточное представление) независимо от конкретной аппаратной архитектуры

Гомоморфизм – отображение между алгебраическими структурами, сохраняющее заданные операции (структуру), то есть переводящее результат операции в образ результата при отображении

Комбинатор – функция высшего порядка, задающая способ комбинирования других функций или программных фрагментов

Компилятор – программа, выполняющая преобразование исходного текста программы на языке программирования в эквивалентную форму представления (как правило, объектный код, машинный код или промежуточное представление), пригодную для последующего исполнения или дальнейшей трансляции

Копирующий сборщик мусора – алгоритм автоматического управления памятью, при котором доступные (достижимые) объекты копируются в новую область памяти, а старая область очищается целиком

Лексема – минимальная значимая единица исходного текста программы, выделяемая на этапе лексического анализа (например, идентификатор, ключевое слово, литерал, оператор, разделитель)

Монoid – алгебраическая структура, состоящая из множества и одной ассоциативной бинарной операции, имеющей нейтральный элемент

Нетерминал – символ формальной грамматики, который может быть заменён последовательностью терминалов и/или нетерминалов по правилам вывода и используется для построения синтаксических конструкций языка

Нитевой код – способ представления программы в виде последовательности ссылок (адресов) на подпрограммы или примитивы исполнения, используемый для ускорения интерпретации и уменьшения объёма кода в стековых системах
Обратная польская запись – форма записи выражений, в которой операторы следуют за operandами (постфиксная нотация)

РБНФ – расширенная форма Бекуса–Наура, нотация задания формальных грамматик, расширяющая БНФ средствами записи повторений, альтернатив, optionalных элементов и группирования

Стек – структура данных, организованная по принципу «последним пришёл – первым вышел» (LIFO), поддерживающая операции добавления и извлечения элементов с одной стороны, называемой вершиной стека

Терминал – символ формальной грамматики, который не подлежит дальнейшему разложению по правилам вывода и входит в алфавит конечной строки языка

.NET CIL – .NET Common Intermediate Language, промежуточный язык (байткод) платформы .NET, в который транслируются программы и который исполняется средой выполнения (CLR) или преобразуется в машинный код

.NET CLR – .NET Common Language Runtime, виртуальная машина платформы .NET, исполняющая программы на языке .NET CIL

Mark-sweep-compact – алгоритм сборки мусора, включающий фазу пометки достижимых объектов (mark), последующую очистку неиспользуемых объектов (sweep) и уплотнение памяти (compact) для устранения фрагментации
S-выражение – скобочный формат записи, где каждая операция представлена скобками, внутри которых на первом месте стоит оператор, а затем перечисляются operandы

ВВЕДЕНИЕ

Стековые (или стек-ориентированные) языки программирования характеризуются применением стека данных в качестве основного механизма передачи информации и хранения результатов вычислений. Стек-ориентированность позволяет программам на таких языках компактно выглядеть и эффективно исполняться. Исторически первым стековым языком стал Forth, разработанный Чарльзом Муром в начале 1970-х годов. Язык Forth изначально создавался для системного и низкоуровневого программирования [1], но в целом позволяет писать достаточно выразительный и понятный высокоуровневый код.

Цель работы – выполнить аналитический обзор различных стековых языков программирования, выделить и сравнить свойства представленных языков, сделать выводы о существующих тенденциях в области стековых языков.

Модель исполнения стекового языка – это некая (чаще всего виртуальная) стековая машина, имеющая стек данных, операционный стек и набор команд для работы со стеком. Например, реализация Forth оперирует стеком данных для хранения аргументов и результатов и стеком возвратов (операционным стеком) для адресов возврата при вызове подпрограмм [1]. Программа представляет собой последовательность команд, которые могут быть либо встроенными примитивами (например, арифметические операции, операции над стеком; интроспекциими), либо определенными пользователем субпрограммами. В терминах стековых языков команды принято называть словами, или комбинаторами, или функциями. Каждое слово либо модифицирует состояние стека, либо изменяет поток исполнения программы. Большинство инструкций в такой машине берёт необходимые операнды с вершины стека, выполняет вычисление и кладёт полученный результат обратно на стек. Когда выполнение программы завершено, результат обычно считается находящимся на вершине стека, откуда его можно извлечь для дальнейшего использования.

Поток исполнения программы контролируется при помощи счетчика команд (program counter), который указывает на текущую команду. Для большинства команд поток исполнения линеен: после выполнения одной команды машина берет на исполнение следующую команду (то есть инкрементирует значение счетчика до следующей команды). Исключение составляют команды условного и безусловного переходов, в некоторых реализациях стековых машин также присутствуют команды циклов. Для таких команд машина может менять значение счетчика неким особым образом согласно семантике конкретной команды. Следует обратить внимание, что отсутствие команд циклов не означает невозможность итерирования: в таком случае цикличность исполнения может быть достигнута при помощи команд ветвления и рекурсивных вызовов.

Таким образом модель исполнения (машина) содержит в себе следующие компоненты:

- стек данных;
- стек возвратов;
- счетчик команд;
- словарь поддерживаемых команд (в число которых могут входить команды модификации стека данных, а также команды модификации стека возвратов и счетчика команд), который может быть расширен в процессе исполнения программы.

1 Аналитический обзор существующих языков

Начиная с семидесятых годов прошлого века было создано достаточно большое количество стековых языков. Далее рассмотрены несколько значимых языков, каждый из которых привнес важные нововведения в области стековых языков (речь о которых пойдет далее). Для языков рассматриваются аспекты синтаксиса, типизации, работы с памятью и стандартной библиотекой, особенности каждого языка. В обзор не включены, однако также заслуживают внимания:

- конкатенативный язык программирования Kitten, который продолжает идеи языка Cat в статическом анализе и представляет способ явного указания типов;
- .NET CIL и байткод Java, которые также являются языками под стековые виртуальные машины (.NET CLR и JVM соответственно).

1.1 Язык программирования Forth

Forth является одним из первых стековых конкатенативных языков программирования. Логотип языка представлен на рисунке 1. Forth создан Чарльзом Муром в начале 1970-х годов как сочетание расширяемого языка и интерактивной методологии разработки. Основная идея состоит в минимальном ядре и словаре «слов» (подпрограмм), через которые реализуется как высокоуровневая логика, так и низкоуровневое аппаратное управление. Новые слова добавляются прямо во время работы системы, что позволяет подстраивать язык под конкретную предметную область.

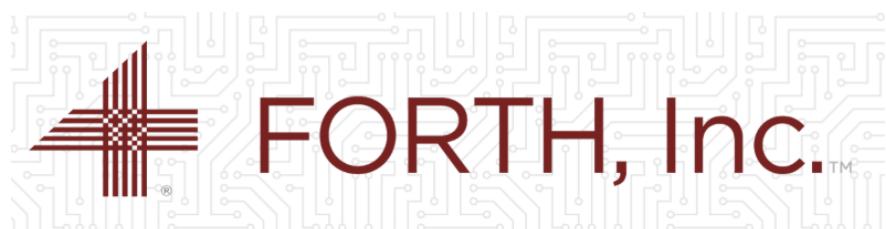


Рисунок 1 — Логотип Forth

Синтаксис Forth основан на обратной польской нотации: программа представляет собой последовательность слов, разделённых пробелами. Интерпретатор последовательно читает токены, ищет их в словаре и либо выполняет связанный код, либо интерпретирует токен как число и помещает

его на стек. Определения новых слов задаются конструкцией вида «: NAME ... ;», при этом большинство лексем — от переменных до управляющих конструкций — оформляются как слова. Грамматика языка показана в листинге .

Стандартной библиотеки в привычном понимании в Forth нет: расширение возможностей осуществляется за счёт подключения и использования дополнительных наборов слов (word sets), входящих в конкретную реализацию среды.

Листинг 1 — РФБН Forth

```
program      ::= { item } ;
item         ::= word | number | string | comment |
definition ;

definition   ::= ":" name { item } ";" ;

word         ::= name ;
name         ::= letter { letter | digit | "-" | "_" } ;

number       ::= [ "+" | "-" ] digit { digit } ;
string       ::= ".\"{ char - "\"" } "\" ;
comment     ::= "(" { char - ")" } ")" | "\" " { char -
EOL } ;

letter       ::= "A".."Z" | "a".."z" ;
digit        ::= "0".."9" ;
```

С точки зрения семантики типов Forth рассматривается как «безтиповый» [2]: данные представляются машинными словами фиксированной разрядности, а язык не навязывает проверку согласованности типов — ответственность за корректность интерпретации содержимого стека возлагается на разработчика. Такой подход обеспечивает гибкость в описании программ и предсказуемое временное поведение (так как любая команда), но затрудняет статический анализ и контроль.

Forth является интерактивной средой, в которой процесс исполнения программы можно разделить на 2 составляющие: состояние интерпретации и состояние компиляции. В состоянии интерпретации среда занимается исполнением примитивных слов. В то же время, если среда в процессе встречает некоторый набор слов (например, «:» — создание нового

слова), то она переходит в состояние компиляции и осуществляет разбор подпрограмм, встречаемых синтаксических конструкций. Выполнить переключение состояния также можно при помощи специальных слов «[« – переход в состояние интерпретации и «]» – переход в состояние компиляции. Определения новых слов задаются конструкцией вида «: NAME ... ;», причём большинство лексем – от переменных до управляющих конструкций – являются словами. Стандартный «core word set» ANS Forth [2] включает базовые стековые операции («DUP», «DROP», «SWAP», «OVER»), арифметику («+ - * /»), сравнения, примитивы работы с памятью («@» – чтение по адресу, «!» – запись по адресу), конструкции ветвления и циклов («IF ... ELSE ... THEN», «DO ... LOOP»), а также определяющие слова «CREATE», «VARIABLE», «CONSTANT» и др. Остальные word set'ы стандарта являются опциональными расширениями. Почти все управляющие слова в Forth реализуются при помощи понятия слов времени компиляции, то есть конструкций, которые во время компиляции разворачиваются в более примитивные последовательности слов. Например, конструкция с ветвлением «... DUP 6 < IF DROP 5 ELSE 1 - THEN ...» разворачивается в последовательность слов «... DUP LIT 6 < ?BRANCH 5 DROP LIT 5 BRANCH 3 LIT 1 - ...» – конструкция условного перехода заменена на слова условного и безусловного переходов [1].

В листинге 2 определяется слово «HELLO», которое затем вызывается и выводит в консоль текст. Круглые кавычки используются для комментариев.

Листинг 2 — Определение и использование слова «HELLO»

```
: HELLO CR ." Hello, World!" ;
HELLO ( выводит Hello, World! в консоль на следующей строке
после вызова слова )
```

Forth применяется во встраиваемых и ресурсно-ограниченных системах, прошивках, системах реального времени и космической технике. В частности, он используется в контроллерах космических аппаратов и приборов NASA, а также в реализациях Open Firmware для платформ Apple, IBM и Sun. К характерным особенностям практического использования относятся малый размер полной среды (компилятор, интерпретатор и редактор могут

размещаться в памяти 8-битных систем), а также единство языка и среды исполнения: Forth одновременно выступает командной оболочкой, компилятором и минимальной операционной системой, при этом стандарт описывает набор слов и модель стека, оставляя свободу построения предметно-ориентированных расширений и диалектов [1].

1.2 Язык программирования Joy

В 2001 году Манфред фон Тун из университета Ла Троба представил язык Joy как попытку формализации логики стековых языков. Joy представляет собой функциональный стековый конкатенативный язык программирования [3]. В отличие от традиционных функциональных языков, которые строятся вокруг операции применения функции к аргументу, Joy использует операцию композиции функций [4]. Автор вводит формальную основу: каждая команда в Joy обозначает унарную функцию вида «stack -> stack»: на вход подаётся состояние (стек данных), в процессе применения команды происходит его модификация и передача следующей команде; значения и подпрограммы при этом также передаются через стек.

Язык с точки зрения математики формализуется следующим образом: множество всех программ образует синтаксический моноид по операции конкатенации (ассоциативная операция «склейки» программ и пустая программа как нейтральный элемент), а множество функций «stack -> stack» – семантический моноид по операции композиции и тождественной функции в роли нейтрального элемента [4]. Отображение, которое каждой программе сопоставляет её «смысл» (как некоторую функцию над стеком), является гомоморфизмом моноидов, то есть сохраняет операцию: смысл «P Q» равен композиции смыслов «P» и «Q», а смысл пустой программы – тождественная функция [5].

Как в случае с Forth, синтаксис Joy основан на постфиксной записи [3]. Программа представляет собой последовательность слов, разделённых пробелами, исполнение происходит слева направо. Каждое слово выполняет некоторое преобразование над стеком данных. Операторы снимают одно или несколько значений с вершины стека и помещают результат обратно. Язык

определяет способ записи отложенных вычислений при помощи механизма цитат, которая содержит фрагмент программы как данные [3]. Для представления структур данных используются литералы: цитаты программ записываются в квадратных скобках «[...]», множества — в фигурных скобках, строки — в кавычках. Определения новых слов записываются в виде равенств, например «square == dup * .». Язык при этом остаётся функционально чистым: стандартные операции не изменяют скрытое глобальное состояние, а только преобразуют стек, что позволяет рассматривать программу как математическую функцию «stack -> stack» без побочных эффектов [4]. Листинг 3 иллюстрирует грамматику языка.

Листинг 3 — РБНФ грамматики Joy

```

program      ::= { term } ;
term        ::= word | literal | quote | set | string |
definition ;

definition   ::= name "==" program "." ;

quote        ::= "[" program "]" ;
set          ::= "{" { term } "}" ;
string       ::= "\"" { char - "\"" } "\"" ;

literal      ::= number ;
word         ::= name ;
name         ::= letter { letter | digit | "_" | "-" } ;
number       ::= ["+" | "-"] digit { digit } ;

```

При построении идиоматических программ на Joy активно используются комбинаторы. В контексте языка это стековые функции, которые принимают одну или несколько цитат программ (списков слов) и управляют их исполнением различными способами [3]. Например, функция-комбинатор «i» («interpret») берет вершину стека и исполняет список слов, которые лежали внутри вершины. С точки зрения синтаксиса это эквивалентно опусканию скобочек: «[1 print] i» значит то же самое, что и «1 print». Для ветвлений используется комбинатор «ifte», для циклических алгоритмов — комбинаторы различных схем рекурсии «primrec», «linrec», «binrec». Программист волен создавать и свои функции-комбинаторы на основе уже существующих через механизм определения новых слов.

Благодаря возможности определять новые слова, которые могут выступать как обычные функции или как комбинаторы, Joy располагает достаточно разветвлённой стандартной библиотекой. В стандартной поставке присутствует несколько библиотек, сгруппированных по областям применения. Так, библиотеки «agplib» и «seqlib» содержат обобщённые операции над агрегатами (неупорядоченными множествами, строками и списками) и последовательностями, «numlib» включает числовые функции и численные методы, а «mtrlib» предназначена для работы с матрицами [6].

Joy является динамически типизированным языком. В язык встроены числовые типы (целые и вещественные числа), агрегатные типы (строки, списки и неупорядоченные множества), а также тип цитат [3]. Типизация стека на этапе компиляции не контролируется, и проверка типов выполняется в момент применения операции. При несовпадении ожидаемого и фактического типов возникает ошибка времени выполнения [6]. Для уменьшения числа подобных ошибок рекомендуется сопровождать определения функций комментариями в нотации эффектов стека, где указывается, какой стек функция принимает и какой возвращает, например «`dup : A -> A A`» или «`+ : Int Int -> Int`». Такая нотация позволяет рассматривать программы как формулы в рамках алгебры стеков и служит основой для рассуждений о корректности и эквивалентности программ [5], что можно рассматривать как предпосылку к статическому анализу.

Работа с памятью в Joy в значительной степени скрыта от программиста. Базовая реализация языка представляет собой интерпретатор на языке С с автоматической сборкой мусора, что позволяет безопасно работать с динамическими структурами данных, такими как списки, строки и цитаты, без явного управления временем их жизни [6].

Язык Joy имеет преимущественно академическое и экспериментальное применение, так как используется для демонстрации и исследования идей функционального программирования в стековой модели исполнения. В результате Joy можно рассматривать как компактную и хорошо формализованную модель стекового языка высокого уровня, в которой

математические понятия монида и гомоморфизма служат точным описанием связи между текстом программы и её поведением при выполнении.

1.3 Язык программирования Cat

Логическим продолжением Joy является экспериментальный язык Cat (логотип языка показан на рисунке 2), по которому в 2006 году Кристофером Диггинсом опубликовано несколько статей [7,8]. Cat наследует идею Joy: любая программа рассматривается как функция вида «stack \rightarrow stack», а основная операция — композиция таких функций, реализуемая простой конкатенацией лексем в исходном тексте программы. Cat вводит статическую систему типов поверх этой модели, что позволяет компилятору на этапе компиляции вычислить размер стека данных и его типизацию для любого момента времени исполнения программы и верифицировать программу на соответствие ожидаемым и фактическим типам [8]. Таким образом, за счёт статического анализа возможно предотвратить ошибки времени выполнения, связанные с несовпадением типов, опустошением или переполнением стека; кроме того, статический анализ открывает путь к агрессивным оптимизациям, затрагивающим текст программы и промежуточные представления.



Рисунок 2 — Логотип Cat

Аналогично Joy, программы на языке Cat опираются на обратную польскую запись с последовательным перечислением слов слева направо [8]. Работа с потоком исполнения организована при помощи комбинаторов «if» и «while», соответствующих ветвлению и циклу по условию: комбинаторы принимают набор цитат (которые, аналогично Joy, записываются в квадратных скобках; также цитаты называют замыканиями) и организуют поток управления. Определение новых слов возможно при помощи слова «define», определения являются глобальными и единственными (множественные

определения запрещены). РБНФ грамматики синтаксиса Cat показана в листинге 4.

Стандартная библиотека Cat небольшая. В множество команд языка доступны арифметические и логические операции, операции сравнения, операции манипуляции стеком и комбинаторы (или, как привычнее для Cat, функции высшего порядка). Также существуют примитивы для работы с последовательностями и рекурсивными алгоритмами.

Листинг 4 — РБНФ грамматики Cat

```
program      ::= { term } ;
term        ::= word | literal | quote | definition ;

definition   ::= "define" name program ";" ;      (*
глобальное и единственное определение *)
quote       ::= "[" program "]" ;                  (* цитата/
замыкание *)

literal      ::= number ;
word         ::= name ;
name         ::= letter { letter | digit | "_" | "-" } ;
number       ::= ["+" | "-"] digit { digit } ;
```

Наибольшего внимания заслуживает система типов Cat, благодаря которой строится статический анализ. Система типов описывает поведение функций в терминах потребления и производства элементов стека. Тип функции определяется тем, что функция ожидает на стеке перед началом выполнения, и тем, что она оставляет на стеке после выполнения. Тип записывается как круглые скобки с стрелкой внутри между конфигурациями стека [7], например «(int int -> int)» для операции сложения или «('S 'a -> 'S 'a 'a)» для дублирования вершины стека. Типы с апострофом («'a», «'b») обозначают типовые переменные, а специальные «типовыe векторы», обозначаемые заглавными буквами («'A», «'B»), представляют произвольные последовательности типов, то есть «хвост» стека. Такое представление соответствует row-полиморфизму: каждая функция не только определяет, какие значения она снимает и кладёт на вершину стека, но и неявно пропускает через себя остальную часть стека, обозначаемую элементов типового вектора; это позволяет типизировать локальные преобразования независимо от длины

и состава хвоста стека. Диггинс разделяет хвост (типовыи вектор) и голову (типовыи переменные) стека через различие между «родами» (kinds) для обычных типов и для стеков значений: типы значений и типы стеков образуют две категории, для которых определяются функции «stack \rightarrow stack», а типизация выражается в виде правил натурального вывода. Вывод типов в Cat базируется на вариации алгоритма Хиндли–Милнера, обобщённой на стековые типы и row-полиморфизм. В статье [8] описывается пример вывода типов композиции нескольких комбинаторов (которая сама также является комбинатором). Конструкции языка допускают полиморфизм более высоких рангов (в контексте Cat это означает, что позволяет выводить типы для функций, которые манипулируют другими полиморфными функциями сколь угодной степени вложенности), а аннотации типов («stack diagrams») используются как документация и как данные для статического анализа. В статье [8] описывается чистое подмножество языка Cat с правилами вывода, показанными в листинге 5: алгоритм проходит по словам, находит тип каждой композиции, после чего выполняет унификацию типов согласно алгоритму Хиндли–Милнера.

Листинг 5 — Правила вывода основных слов в Cat

```

pop : ('a  $\rightarrow$  )
dup : ('a  $\rightarrow$  'a 'a)
swap: ('a 'b  $\rightarrow$  'b 'a)
if  : ('A bool ('A  $\rightarrow$  'B) ('A  $\rightarrow$  'B)  $\rightarrow$  'B)
eval: ('A ('A  $\rightarrow$  'B)  $\rightarrow$  'B)
while: ('A ('A  $\rightarrow$  'A) ('A  $\rightarrow$  'A bool)  $\rightarrow$  'A)

```

Статический анализ в Cat не ограничивается проверкой согласованности типов. Поскольку тип функции фиксирует не только типы элементов, но и форму стека до и после применения функции, система гарантирует отсутствие недопустимых ситуаций, таких как недостаток аргументов на стеке или несовместимость конфигураций стека в ветвлении [7]. Для условного оператора «if» тип требует, чтобы обе ветви, принимая на вход одну и ту же конфигурацию стека, возвращали стек одинаковой структуры; в противном случае выражение считается некорректным. Для цикла «while» тип требует, чтобы после выполнения цикла конфигурация стека оставалась такой же, как

и на момент начала цикла (как это и показано в листинге 5). Дополнительно вводится различие между чистыми и побочными функциями при помощи двух видов стрелок (\rightarrow и \rightsquigarrow): функция считается имеющей побочный эффект, если в её реализации используется хотя бы одна побочная операция (такая операция, которая меняет не только переданные данные, то есть стек данных, но и окружение; например, операция ввода или вывода). Эта информация фиксируется в типах и может использоваться оптимизатором и верификатором [7].

Статический анализ открывает путь к агрессивным оптимизациям, которые могут изменять текст программы и промежуточные представления для повышения производительности. Поскольку типы фиксируют форму и размер стека, оптимизатор может опираться на результаты вывода типов как на формализованный контракт корректности [7,8]: форма и размер стека должны остаться прежними после оптимизаций с точки зрения глобального наблюдателя.

Область применения Cat складывается из двух направлений. Во-первых, язык используется как исследовательская и учебная платформа для изучения конкатенативного программирования, типовых систем с выводом типов и row-полиморфизма в стековой модели; работы по типизации функциональных стековых языков опираются на Cat и рассматривают его как пример статически типизированного конкатенативного языка. Во-вторых, Cat разрабатывается как промежуточный язык для компиляторов и инструментов анализа, также используется для верификации и оптимизации программ под платформу .NET CIL [8]. Язык повлиял на дальнейшие разработки в области конкатенативных языков и типовых систем. В итоге Cat формирует пример полноценно типизированного стекового языка, в котором идея «программа как функция stack \rightarrow stack, конкатенация как композиция» соединяется с сильной статической системой типов, ориентированной на анализ и оптимизацию, что дает хорошие возможности для развития новых прикладных решений.

1.4 Язык программирования Factor

Среди стековых языков Factor занимает нишу языков общего назначения [9]. Первая версия языка была выпущена в 2003 году Славой Пестовым. Логотип языка показан на рисунке 3. Factor заимствует многие элементы из языка Joy: конкатенативный синтаксис, использование комбинаторов для управления потоком исполнения, а также аналогичный способ создания собственных комбинаторов при помощи цитат.



Рисунок 3 — Логотип Factor

Синтаксис Factor является конкатенативным: программа записывается как последовательность слов, применяемых слева направо, а для управления потоком исполнения используются комбинаторы и цитаты. Пример определения нового слова приведён в листинге 6; в нём показано использование нотации «stack effects», а также комбинатора «if» и цитат. В определении грамматики видно в листинге 7, что описание stack effect встроено в синтаксис.

Листинг 6 — Определение слова вычисления факториала

```
: factorial ( n -- n! )
dup 0 =
[ drop 1 ]
[ dup 1 - factorial * ]
if ;
```

Листинг 7 — РБНФ Factor

```
program      ::= { term } ;
term        ::= word | literal | quotation | definition ;

definition   ::= ":" name [ stack_effect ] program ";" ;
stack_effect ::= "(" { token } "--" { token } ")" ;

quotation    ::= "[" program "]" ;
```

Продолжение листинга 7

```
literal      ::= number | string ;
string      ::= "\"" { char - "\"" } "\" ;
word        ::= name ;
name        ::= letter { letter | digit | "_" | "-" } ;
number      ::= ["+" | "-"] digit { digit } ;
```

Прикладным язык считается за счёт большой стандартной библиотеки, которая позволяет решать широкий спектр прикладных задач. В отличие от Joy, стандартная библиотека Factor обладает существенно большими размерами и количеством возможностей. Например, в рамках стандартной библиотеки поставляется веб-фреймворк, XML-парсер, компоненты для построения приложений с графическим интерфейсом. Также для языка существует большое количество различных инструментов разработки, объединённых в IDE: интерактивный отладчик, браузер документации, инспектор объектов, механизм модификации программы без её перезапуска [10].

Как и Joy, Factor является динамически типизированным языком [9]. Документация и сообщество активно используют нотацию «stack effects», с помощью которой описывается, как то или иное слово при применении модифицирует стек. Оптимизирующий компилятор может проверить объявленные эффекты и отклоняет определения, для которых эффект не удаётся вывести, трактуя это как ошибку компиляции. Внутренний механизм этой проверки описывается как абстрактная интерпретация программы: stack-checker симулирует эффекты слов на абстрактном стеке, при встрече ветвлений анализирует обе ветви и унифицирует состояния, а несовместимость высоты или формы стека превращается в ошибку времени компиляции. Такая «проверка стека» играет роль статического анализатора, который обнаруживает классы ошибок, типичные для стековых языков, и одновременно создаёт основу для агрессивных оптимизаций [9]. Прикладным язык считается также за счёт статического анализа (stack-checker) и динамических проверок времени исполнения, которые обеспечивают верификацию программ и повышают применимость языка для продуктовых задач.

В целом язык предлагает множество высокоуровневых и низкоуровневых оптимизаций. Оптимизирующий компилятор состоит из фронтенда, который строит промежуточное представление языка, используемое для работы stack-checker и высокоуровневых оптимизаций, а также бекенда, который генерирует исполняемый машинный код и выполняет низкоуровневые оптимизации [9].

Для работы с памятью язык использует сборщик мусора, встроенный в виртуальную стековую машину. Для молодых объектов применяется копирующий сборщик, для старшего поколения используется алгоритм «mark-sweep-compact» [9]. Для работы с неуправляемыми ресурсами используются библиотека «destructors» и комбинатор «with-destructors». Использование умного сборщика мусора также повышает степень пригодности языка к решению задач широкого спектра.

Язык Factor относится к языкам общего назначения и ориентирован на решение прикладных задач за счёт обширной стандартной библиотеки, развитого инструментария и оптимизирующей компиляции [9,10].

1.5 Язык программирования Wasm

В 2017 году группа компаний, в которую входят W3C, Mozilla, Google, Microsoft и Apple, представила язык WebAssembly (Wasm). Логотип Wasm показан на рисунке 4. Wasm представляет собой низкоуровневый переносимый байткод и архитектуру набора команд под виртуальную стековую машину, запускаемую преимущественно (но не только) в веб-браузерах наравне с движками ECMAScript для исполнения клиентского кода [11]. Это первый из представленных в обзоре языков, который является целью компиляции, а не сам компилируется во что-либо. Поддержка компиляции в Wasm имеется для множества высокоуровневых языков, в первую очередь для C/C++/Rust, которые имеют небольшой (или вовсе не имеют) рантайм. Особенность области использования wasm – веб-браузер не может считать никакой получаемый код доверенным, в результате чего к языку и машине Wasm предъявляются парадоксальные требования: виртуальная машина wasm должна дополнительно верифицировать код, порожденный тем или иным компилятором, за минимальное время с целью скорейшего

непосредственного запуска программы. Хотя для других описанный в обзоре языков такая проверка была уместна, так как тексты программ пишутся человеком, которому свойственно ошибаться, для Wasm подобная проверка остается также уместной. Таким образом, в исходных целях проектирования Wasm фиксируются требования к компактности двоичного представления, быстрой однопроходной валидации и компиляции, а также к «песочнице» с предсказуемой семантикой, пригодной для запуска недоверенного кода [11].

Байткод Wasm в основном поставляется в виде бинарного представления, которое при необходимости может быть транслировано в текстовое человекочитаемое представление. Текстовый вариант основан на записи S-выражений [12] и называется WAT. Листинг 8 демонстрирует определение функции сложения числа с самим собой в человеко-читаемом виде.



Рисунок 4 — Логотип Wasm

Листинг 8 — Определение функции сложения числа с самим собой

```
(func (param $p i32)
      (result i32)
      local.get $p
      local.get $p
      i32.add
)
```

Поскольку язык является низкоуровневым, понятие стандартной библиотеки к нему не применяется. Вместо встроенных API используется модель импортов: окружение предоставляет функции и объекты, а модуль экспортит точки входа.

Как стековый язык исполнения Wasm опирается на неявный операндный стек и структурированное управление потоком [11]. Инструкции потребляют значения с вершины стека и помещают результаты обратно, при этом реализация не обязана хранить «настоящий» стек: спецификация описывает

интерпретацию стека как набора анонимных регистров, а статическая типовая проверка делает высоту стека известной на этапах валидации и компиляции. Типизация в Wasm является статической и рассчитана на проверку до исполнения: функции имеют явные типы параметров и результатов, инструкции типизируются через эффекты над стеком, а модуль проходит валидацию, гарантирующую корректность применения инструкций к операндам нужных типов и отсутствие underflow-ошибок. Верификация модуля выполняется за один проход, что соответствует целям языка по скорости исполнения [11]. Управление потоком задаётся структурными конструкциями (block, loop, if) и переходами к меткам (br, br_if, br_table). Единица трансляции Wasm называется модулем [11].

В качестве оптимизационных целей Wasm фиксируются компактность двоичного представления и быстрая однопроходная валидация и компиляция [11]. Низкоуровневый стековый байткод занимает меньше пространства и может исполняться с меньшими накладными расходами по сравнению с традиционным для браузеров кодом на ECMAScript, поскольку исполняется в специализированной виртуальной машине и компилируется в машинный код средствами рантайма.

Управление памятью осуществляется в ручном режиме: так как Wasm является целью компиляции, ответственность за гарантии при работе с памятью ложится на средства исходного языка [11] (например, статические проверки Rust, сборщики мусора, либо, в случае C/C++, на разработчика программного обеспечения).

Язык и виртуальная машина используются в качестве безопасной «песочницы» с производительностью низкоуровневого кода, в основном исполняемой в веб-браузерах. По сравнению с ECMAScript, байткод Wasm обычно требует меньших накладных расходов при выполнении (в частности, по сравнению с исполнением через движки ECMAScript, например V8), что делает его удобной целью компиляции для переносимых модулей клиентского кода.

2 Сравнительный анализ

Для обеспечения полноты работы выполнен сравнительный анализ синтаксических свойств рассмотренных языков.

С точки зрения синтаксиса Forth, Joy, Cat и Factor относятся к конкатенативным языкам: программа представляется как последовательность «слов» (токенов), разделённых пробелами, а композиция вычислений выражается простой конкатенацией лексем. При этом языки различаются механизмами построения составных выражений:

- Forth опирается на режимы состояния;
- Joy, Cat и производные активно используют цитаты (фрагменты программы как данные);
- Factor дополняет язык соглашением «stack effects».

Wasm занимает отдельное место. Основной формат является бинарным, однако предусмотрено человеко-читаемое представление, основанное на S-выражениях. Поэтому по «форме записи» Wasm отличается от конкатенативных языков высокого уровня: в WAT структура задаётся скобочной формой, хотя сами инструкции по смыслу остаются стековыми (так как операндный стек неявен) [11,12].

Таблица с сравнительным анализом языков по синтаксису представлена в таблице 1.

Таблица 1 — Сравнение синтаксиса языков (сводная таблица)

Язык	Базовая форма записи	Ключевые особенности синтаксиса
Forth	Постфиксная	Управляющие конструкции реализуются как слова времени компиляции
Joy	Постфиксная	Цитаты используются как данные, управление через комбинаторы («i», «ifte», «primrec/linrec/binrec»)

Продолжение таблицы 1

Язык	Базовая форма записи	Ключевые особенности синтаксиса
Cat	Постфиксная	Синтаксис близок к Joy
Factor	Постфиксная	Обязательная документация эффектов стека
Wasm	S-выражения	Повышенная человеко-читаемость за счет иной грамматики записи

В рассматриваемых языках подходы к построению стандартной библиотеки (подключению дополнительной функциональности) существенно отличаются и напрямую связаны с назначением языка и моделью исполнения. Таблица 2 описывает принципы подключения дополнительной функциональности. В Forth «стандартная библиотека» в привычном смысле отсутствует: стандарт задаёт наборы слов (word sets), а расширение функциональности выполняется подключением дополнительных наборов слов и формированием предметно-ориентированных словарей поверх минимального ядра. В Joy стандартная поставка включает набор библиотек, сгруппированных по областям (операции над агрегатами и последовательностями, численные методы и т.п.), что соответствует учебно-исследовательскому характеру языка. В Cat библиотека остаётся небольшой и ориентирована на ядро примитивов и комбинаторов, так как основной акцент сделан на типовую систему и статический анализ. В Factor есть большая развитая стандартная библиотека и богатый набор инструментов. В WebAssembly аналогично Forth стандартной библиотеки нет (что можно рассматривать как признак низкоуровневости): взаимодействие с внешним миром осуществляется через импорты, а набор доступных функций задаётся окружением исполнения, что поддерживает переносимость модулей и отделяет семантику языка от API платформы.

Можно выделить закономерность. Если язык предназначен для некоего высокоуровневого взаимодействия, то он скорее имеет библиотеку, пусть и

не самую большую (Joy [3], Cat [8], Factor [9]). Если же речь идет про язык близкий к машине, то стандартной библиотеки у него нет, но есть иной способ подключать дополнительную функциональность (Forth [2], Wasm [11]).

Таблица 2 — Сравнение подходов к подключению дополнительной функциональности

Язык	Средства подключения дополнительной функциональности
Forth	Стандарт задаёт «word sets». Расширение через подключаемые наборы слов и предметно-ориентированные словари.
Joy	Набор библиотек по областям (агрегаты, последовательности, численные методы, матрицы) в стандартной поставке.
Cat	Небольшая библиотека: примитивы стека, базовые операции и комбинаторы; акцент на типизации и анализе.
Factor	Обширная стандартная библиотека и интегрированные инструменты разработки, ориентированные на прикладные задачи.
Wasm	Доступ к функциональности через импорты окружения; модель модулей.

В рассмотренных языках наблюдается широкий спектр подходов к типизации – от отсутствия обязательной типовой дисциплины до строгой статической валидации как части спецификации. Сравнение приведено в таблице 3. В Forth язык обычно рассматривается как «безтиповый»: элементы стека представлены машинными словами фиксированной разрядности, а корректность их интерпретации обеспечивается соглашениями программиста. Joy является динамически типизированным языком: операции проверяют типы значений в момент применения, и при несоответствии ожидаемого и фактического типов возникает ошибка времени выполнения [6]. Factor также динамически типизирован, однако в компиляторе реализована обязательная проверка согласованности «эффектов на стек» (stack-checker), которая статически подтверждает корректность формы и размера стека для определений [9]. В Cat типизация является статической и описывает поведение

слов в терминах преобразования конфигураций стека. В WebAssembly типизация является статической: функции имеют типы параметров и результатов, инструкции типизируются через эффекты над стеком, а модуль проходит валидацию до исполнения [11].

Анализ корректности в указанных языках реализуется различными средствами. В Forth контроль в основном возлагается на разработчика и среду исполнения. В Joy основным механизмом остаётся динамическая проверка во время выполнения, однако используется дисциплина документирования эффектов стека, позволяющая рассматривать программу как композицию преобразований и служащая предпосылкой к статическим рассуждениям [5]. В Factor статический анализ реализован через абстрактную интерпретацию эффектов слов на абстрактном стеке: анализатор проверяет согласованность высоты и структуры стека по всем ветвям управления и отклоняет некорректные определения на этапе компиляции [9]. В Cat статический анализ является частью системы вывода типов: типы фиксируют форму стека до и после применения слов и комбинаторов, что обеспечивает верификацию согласованности ветвлений и инвариантов циклов [7]. В WebAssembly анализ представлен в форме однопроходной валидации модуля, гарантирующей корректность типов операндов и структур управления до запуска кода [11].

Таблица 3 — Сравнение типизации и анализа

Язык	Тип типизации	Используемый анализ
Forth	Безтиповой	Проверки типовой согласованности не навязываются; корректность интерпретации содержимого стека обеспечивается соглашениями и выявляется на этапе исполнения.
Joy	Динамическая	Динамическая проверка типов при выполнении

Продолжение таблицы 3

Язык	Тип типизации	Используемый анализ
Factor	Динамическая	Stack-checker как статический анализ: абстрактная интерпретация эффектов слов на абстрактном стеке и проверка согласованности ветвей; ошибки выявляются при компиляции.
Cat	Статическая (эффекты стека)	Вывод типов и верификация формы стека, включая row-полиморфизм; гарантии согласованности ветвлений и инвариантов циклов.
Wasm	Статическая (нормативная)	Однопроходная валидация модуля по типам и структурированному управлению; предотвращение underflow и несогласованности типов до исполнения.

Оптимизации в стековых языках можно условно разделить на две группы: (1) преобразования представления и диспетчеризации команд (ускоряют интерпретатор/виртуальную машину) и (2) оптимизации, использующие результаты анализа семантики (стек-эффекты, типы, структура управления), что позволяет убирать лишние проверки и преобразовывать программу без нарушения корректности.

Для Forth базовым приёмом ускорения является компиляция в нитевой код: программа представляется как последовательность ссылок на подпрограммы (слова), что уменьшает размер кода и удешевляет интерпретацию. Дополнительно важную роль играет то, что многие управляющие конструкции реализованы как слова времени компиляции и разворачиваются в более примитивные переходы и последовательности слов, что снижает накладные расходы управляющих конструкций и упрощает исполнение [1].

В Cat оптимизации тесно связаны со статической типизацией стека: типы фиксируют не только типы значений, но и форму стека до и после выполнения фрагмента программы. Это позволяет ещё на этапе компиляции исключать некорректные программы (underflow, несовместимость ветвей), а также использовать результаты вывода типов как формальный контракт при преобразованиях: оптимизатор может менять промежуточные представления и фрагменты кода, сохраняя наблюдаемую форму и типы стека [7] [8].

Factor сочетает динамическую типизацию с развитой инфраструктурой статического анализа стек-эффектов. Stack-checker выполняет абстрактную интерпретацию, проверяет согласованность ветвлений и строит промежуточное представление, на котором становятся возможны агрессивные высокоуровневые оптимизации (в том числе разворачивание макросов и инлайнинг цитат), а далее – низкоуровневые оптимизации на этапе генерации машинного кода [9].

Для Wasm ключевые оптимационные цели заложены в самом формате: компактное бинарное представление, однопроходная валидация и предсказуемая семантика обеспечивают быстрый переход от модуля к исполнимому коду. Типизация и структурированное управление потоком позволяют рантайму быстро проверять корректность и эффективно компилировать модуль, при этом многие оптимизации выполняются либо компилятором исходного языка до генерации Wasm, либо JIT/AOT-компилятором окружения при загрузке и выполнении [11].

Отдельный класс оптимизаций для стековых виртуальных машин – stack caching: хранение части виртуального стека (обычно верхних элементов) в регистрах реальной машины, чтобы сократить число загрузок/выгрузок и обновлений указателя стека. В работах М. А. Эрлтля предложены и подробно исследованы статический подход (компилятор отслеживает состояние кэша как автомат) и динамический подход (несколько версий интерпретатора для разных состояний кэша), а также показано, что stack caching хорошо сочетается с динамическими суперинструкциями, которые уменьшают долю времени на диспетчеризацию команд [13] [14] [15]. Эти техники применимы

не к одному конкретному языку, а к широкому классу реализаций стековых интерпретаторов и виртуальных машин, в том числе к реализациям, близким по модели исполнения к Forth, Factor и другим стековым системам.

Сводка основных техник приведена в таблице 4.

Таблица 4 — Сравнение основных оптимизационных техник в стековых языках

Оптимизация	Описание
Нитевой код (Forth)	Представление программы как последовательности ссылок на слова; снижает размер кода и удешевляет интерпретацию.
Разворачивание слов времени компиляции (Forth)	ТриPLICATION управляющих конструкций в примитивные переходы и последовательности слов, уменьшая накладные расходы управления.
Переписывание и специализация без побочных эффектов (Joy)	Конкатенация цитат, перестановка инструкций с сохранением формы и размера стека с точки зрения внешнего наблюдателя
Инлайнинг цитат и макрорасширение (Factor)	Снижение накладных расходов вызовов и диспетчирования за счёт разворачивания и специализации на этапе компиляции.
Stack caching (общая техника)	Кэширование верхних элементов стека в регистрах, уменьшение обращений к памяти и обновлений указателя стека.
Динамические суперинструкции (общая техника)	Объединение частых последовательностей инструкций в одну “суперинструкцию” для уменьшения диспетчирования.

Работа с памятью в рассматриваемых языках определяется прежде всего назначением языка и типичной средой исполнения. В Forth память является частью базовой модели языка: программист напрямую работает с адресами и примитивами чтения/записи, а управление размещением данных в значительной степени задаётся соглашениями и библиотечными словами конкретной реализации, что соответствует системному и встраиваемому профилю языка [2]. В Joy активное использование динамических структур

данных (списки, строки, цитаты) приводит к тому, что в базовой реализации управление временем жизни объектов скрыто от программиста и обеспечивается автоматической сборкой мусора [6]. В Cat модель памяти в данном контексте определяется целевой платформой (.NET CIL): для практического исполнения типично использование управляемой среды. В Factor управление памятью встроено в виртуальную машину и реализовано сборщиком мусора (с разделением по поколениям), а для неуправляемых ресурсов предусмотрены отдельные средства, что соответствует ориентации языка на прикладное программирование общего назначения [9]. В WebAssembly управление памятью в базовом виде остается низкоуровневым: доступ к памяти осуществляется через модель линейной памяти и инструкции загрузки/сохранения, а гарантии корректности работы с памятью делегируются исходному языку и его рантайму (или разработчику), поскольку Wasm является целью компиляции [11]. Различия приведены в таблице 5.

В целом же можно отметить взаимосвязь: для низкоуровневых языков управление памятью остается за пределами их ответственности (и ложится на плечи программиста или реализуемого рантайма). Для высокоуровневых же типично использование сборщиков мусора.

Таблица 5 — Сравнение подходов к управлению памятью

Язык	Способ управления памятью
Forth	Явная работа с адресами и памятью через примитивы чтения/записи; управление размещением определяется реализацией и соглашениями.
Joy	Автоматическая сборка мусора в базовой реализации; управление временем жизни объектов скрыто от программиста.
Cat	Опора на целевую управляемую среду (.NET CIL); акцент на верификации стековых преобразований, а не на собственной модели памяти.

Продолжение таблицы 5

Язык	Способ управления памятью
Factor	Сборщик мусора во встроенной виртуальной машине (поколения); отдельные средства для неуправляемых ресурсов.
Wasm	Низкоуровневая линейная память и явные load/store; гарантии корректности и управление временем жизни задаются исходным языком и его рантаймом.

Области применения рассмотренных языков в значительной степени определяются тем, на какие ограничения и требования изначально ориентирована их модель исполнения. Forth применяется во встраиваемых и ресурсно-ограниченных системах и в системах реального времени, где наиболее характерны малый размер среды, предсказуемость поведения и возможность низкоуровневого управления аппаратурой через минимальное ядро и расширяемый словарь слов [1]. Joy имеет преимущественно академическое и экспериментальное применение. Cat также относится к исследовательским языкам и изучает возможности верификации и оптимизации программ .NET CIL. Factor занимает нишу языка общего назначения: для него наиболее характерны развитая стандартная библиотека, инструменты разработки и оптимизирующая компиляция, при этом сохранён конкатенативный стиль и используется статический контроль формы стека (stack-checker) как средство повышения надёжности и производительности [9] [10]. WebAssembly используется как переносимый низкоуровневый формат для безопасного исполнения недоверенного кода, для него наиболее характерна предсказуемая семантика и совместимость с различными исходными языками и рантаймами [11].

Таким образом, по областям применения можно выделить характерные черты областей применимости стековых языков: для встраиваемых и системных задач наиболее важны минимализм и прямой контроль над ресурсами (Forth); для учебных и исследовательских задач – формальная

простота модели и возможность строгих рассуждений о композиции программ (Joy); для задач верификации и статического контроля – типизация эффектов стека и вывод типов, позволяющие гарантировать корректность управления потоком и формы стека (Cat); для прикладного программирования общего назначения – сочетание богатой библиотеки и инструментария с компиляцией и анализом, уменьшающими типичные риски стековой модели (Factor); для переносимого и изолированного исполнения – стандартизованный байткод, однопроходная валидация и возможность эффективной компиляции в машинный код в рантайме (Wasm).

В целом же в пути развития стековых языков видна явная тенденция: изначально работающие на высоких скоростях и низких уровнях абстракций благодаря простоте, высокой переносимости и оптимизациям, доступным фактически бесплатно за счет структуры программ и модели исполнения (например, нитевой код Forth), стековые языки обзаводятся мощной формальной базой, которая позволяет точно описывать работу программ и проводить мощный статический анализ. В свою очередь статический анализ открывает для стековых языков новые возможности. Верификация кода и агрессивные оптимизации являются несомненно большими преимуществами для низкоуровневых языков.

Одной из вершин достижений стековых языков можно считать Wasm. Исходные цели проектирования [11] диктуют идеальные для использования стекового языка условия, на основе которых выросла система, сочетающая в себе сразу несколько преимуществ, изначально считающихся несочетаемыми: отличная портируемость под разные платформы, но при этом высокая производительность (например, в сравнении с традиционными движками ECMAScript) за счет низкоуровневости и мощных оптимизаций, а также верифицируемость (и оттого частично безопасность). Wasm вбирает в себя теоретические и практические наработки в области стековых языков, решая с помощью них сложную задачу исполнения прикладного кода с учетом заданных требований.

В результате анализа можно сделать вывод, что современная ниша стековых языков – это быстрые виртуальные машины, запускаемые на широком диапазоне устройств. Это утверждение также подтверждается фактом, что спецификации самых распространенных виртуальных машин (JVM, .NET CLR, Wasm) используют стековую модель исполнения кода.

ЗАКЛЮЧЕНИЕ

В ходе выполнения работы были рассмотрены несколько представителей семейства стековых языков, выделены возможности и свойства каждого. Затем проведен сравнительный анализ, который позволил сформулировать преимущества стековых языков, а также описать их области применения.

Проведенный обзор показал, что стековая модель вычислений сохраняет актуальность на протяжении длительного периода: от ранних системных решений, ориентированных на минимализм и прямой контроль над ресурсами, до современных переносимых форматов байткода и виртуальных машин. Развитие средств типизации и анализа демонстрируют важную тенденцию: повышение надежности стековых языков достигается за счет переноса значимой части проверок на этап компиляции и применения формальных методов верификации.

Сравнительный анализ выделил достоинства стековых языков: высокая степень переносимости, компактность представления программ, наличие ряда специализированных оптимизаций. Существенным ограничением классических стековых языков является сложность контроля корректности без дополнительных механизмов, однако современные подходы (вывод типов эффектов стека, валидация модулей, статическая проверка согласованности ветвлений и циклов) позволяют уменьшить число ошибок и расширить возможности оптимизации.

Полученные результаты позволяют сделать вывод, что стековые языки и стековые представления наиболее эффективны в тех областях, где требуется сочетание переносимости, предсказуемости исполнения и возможности формальной проверки корректности: во встраиваемых системах и системах реального времени, в инструментальных цепочках компиляции и анализаторах, а также в средах безопасного выполнения недоверенного кода. Сформулированные в работе характеристики и сравнительные критерии можно использовать при выборе модели исполнения и проектировании компилятора стекового языка программирования, включая обоснование требований к типизации, анализу и оптимизациям в разрабатываемой системе.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Forth documentation [Электронный ресурс]. URL: <https://www.forth.com/resources/forth-programming-language/> (дата обращения: 25.09.2025).
2. ANS Forth standard [Электронный ресурс]. URL: <https://forth-standard.org/standard/words> (дата обращения: 25.09.2025).
3. An informal tutorial on Joy [Электронный ресурс]. URL: <https://hypercubed.github.io/joy/html/j01tut.html> (дата обращения: 10.10.2025).
4. Mathematical foundations of Joy [Электронный ресурс]. URL: <https://hypercubed.github.io/joy/html/j02maf.html> (дата обращения: 10.10.2025).
5. The Algebra of Joy [Электронный ресурс]. URL: <https://hypercubed.github.io/joy/html/j04alg.html> (дата обращения: 10.10.2025).
6. The prototype implementation of Joy [Электронный ресурс]. URL: <https://hypercubed.github.io/joy/html/j09imp.html> (дата обращения: 10.10.2025).
7. Simple Type Inference for Higher-Order Stack-Oriented Languages [Электронный ресурс]. URL: <https://dcreager.net/remarkable/Diggins2008a.pdf> (дата обращения: 13.10.2025).
8. Typing Functional Stack-Based Languages [Электронный ресурс]. URL: <https://dcreager.net/remarkable/Diggins2008b.pdf> (дата обращения: 13.10.2025).
9. Factor: a dynamic stack-based programming language [Электронный ресурс]. URL: <https://factorcode.org/slava/dls.pdf> (дата обращения: 20.11.2025).
10. Factor wiki [Электронный ресурс]. URL: <https://concatenative.org/wiki/view/Factor> (дата обращения: 20.11.2025).
11. WebAssembly Specification 1.0 [Электронный ресурс]. URL: <https://webassembly.github.io/spec/versions/core/WebAssembly-1.0.pdf> (дата обращения: 11.11.2025).
12. Understanding WebAssembly text format [Электронный ресурс]. URL: https://developer.mozilla.org/en-US/docs/WebAssembly/Guides/Understanding_the_text_format (дата обращения: 11.11.2025).
13. Stack Caching for Interpreters [Электронный ресурс]. URL: <https://www2.cs.arizona.edu/~collberg/Teaching/553/2011/Resources/ertl94sc.pdf> (дата обращения: 10.12.2025).

14. Implementation of stack-based languages on register machines [Электронный ресурс]. URL: <https://repositum.tuwien.at/handle/20.500.12708/13464> (дата обращения: 10.12.2025).
15. Combining Stack Caching with Dynamic Superinstructions [Электронный ресурс]. URL: <https://dl.acm.org/doi/10.1145/1059579.1059583> (дата обращения: 10.12.2025).
16. Cat language repository [Электронный ресурс]. URL: <https://github.com/cdiggins/cat-language> (дата обращения: 13.10.2025).
17. A foundation for typed concatenative languages [Электронный ресурс]. URL: <https://www2.ccs.neu.edu/racket/pubs/dissertation-kleffner.pdf> (дата обращения: 15.10.2025).

ПРИЛОЖЕНИЕ А
ТЕХНИЧЕСКОЕ ЗАДАНИЕ

Листов 6

ПРИЛОЖЕНИЕ А

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

УТВЕРЖДАЮ

Заведующий кафедрой ИУ6

А.В. Пролетарский
«__» 2026 г.

ЧЕРНОВИК

КОМПИЛЯТОР СТЕКОВОГО ЯЗЫКА ПРОГРАММИРОВАНИЯ

Техническое задание

Листов 6

Студент

ИУ6-83Б
(Группа)

(Подпись, дата)

В.К. Залыгин
(И.О. Фамилия)

Руководитель

(Подпись, дата)

Б.И. Бычков
(И.О. Фамилия)

2026 г.

1 ВВЕДЕНИЕ

Настоящее техническое задание распространяется на разработку программы «Компилятор стекового языка программирования» [КСЯП], используемой для трансляции в байткод виртуальной машины WebAssembly (Wasm), а также верификации и оптимизации программ, написанных на конкатенативном Cat-подобном языке программирования.

Актуальность разработки обусловлена новизной выбора цели трансляции. С точки зрения языка, прямая трансляции в WebAssembly позволит избежать дополнительных накладных расходов, связанных с тяжеловесной машиной .NET CLR, увеличить скорость работы и снизить потребление памяти Cat-программ. С точки зрения виртуальной машины, текстовый формат программ которой основан на использовании S-выражений, WebAssembly получит инструмент описания программ при помощи конкатенативного языка, что является более распространенным для стековых языков.

2 ОСНОВАНИЯ ДЛЯ РАЗРАБОТКИ

Компилятор для стекового языка программирования разрабатывается в соответствии с тематикой кафедры «Компьютерные системы и сети».

3 НАЗНАЧЕНИЕ РАЗРАБОТКИ

Основное назначение компилятора – преобразование (трансляция) программ, написанных на исходном языке в байткод, описанный спецификацией виртуальной машины WebAssembly. Также компилятор может осуществлять верификацию и оптимизацию программы при помощи статического анализа.

4 ИСХОДНЫЕ ДАННЫЕ, ЦЕЛИ И ЗАДАЧИ

4.1 Исходные данные

4.1.1 Исходными данными для разработки являются следующие материалы:

4.1.1.1 Simple Type Inference for HigherUOrder StackUOriented Languages [Электронный ресурс]. URL: <https://dcreager.net/remarkable/Diggins2008a.pdf> (дата обращения: 13.10.2025).

4.1.1.2 Cat language repository [Электронный ресурс]. URL: <https://github.com/cdiggins/catlanguage> (дата обращения: 13.10.2025).

4.1.1.3 WebAssembly Specification 1.0 [Электронный ресурс]. URL: <https://webassembly.github.io/spec/versions/core/WebAssemblyU1.0.pdf> (дата обращения: 11.11.2025).

4.1.1.4 Understanding WebAssembly text format [Электронный ресурс]. URL: https://developer.mozilla.org/en-US/docs/WebAssembly/Guides/Understanding_the_text_format (дата обращения: 11.11.2025).

4.2 Цель работы

Целью работы является компилятор стекового языка программирования.

4.3 Решаемые задачи

4.3.1 Анализ синтаксиса и семантики различных стековых языков программирования, выделение и сравнение их особенностей.

4.3.2 Анализ средств статического анализа языка программирования Cat.

4.3.3 Анализ устройства байткода и модели исполнения WebAssembly.

4.3.4 Проектирование программных компонентов: парсера, транслятора, статического анализатора.

4.3.5 Выбор технологии и средств разработки для программного обеспечения (далее ПО).

4.3.6 Разработка структуры программного обеспечения и его компонентов.

4.3.7 Разработка алгоритмов трансляции и статического анализа.

4.3.8 Реализация компонентов ПО с использованием выбранных средств разработки.

4.3.9 Проведение сборки и комплексного тестирования ПО.

4.3.10 Разработка технологии многопроходной оптимизации\генерации webassembly байткода\в процессе реализации станет более ясно.

5 ТРЕБОВАНИЯ К ПРОГРАММНОМУ ИЗДЕЛИЮ

5.1 Требования к функциональным характеристикам

5.1.1 Выполняемые функции:

– генерация байткода WebAssembly для переданной программы на исходном языке;
– верификация и анализ переданной программы на исходном языке с точки зрения корректности использования стека данных;

– оптимизация переданной программы на исходном языке.

5.1.2 Исходные данные

Исходные данные должны быть представлены в виде текста программы на исходном языке и управляющими флагами, указывающими на тип проводимого действия (генерация байткода и/или верификация и/или оптимизация).

Исходных язык должен быть подмножеством языка программирования Cat.

5.1.2 Результаты:

– в случае успешной операции – файл с байткодом WebAssembly или сообщение, подтверждающее корректность программы;

– в случае неуспешной операции – сообщение с описанием ошибки.

5.2 Требования к надежности

5.2.1 Предусмотреть контроль консистентности флагов, передаваемых при вызове компилятора.

5.2.2 Предусмотреть контроль синтаксической корректности текста на исходном языке.

5.2.3 Предусмотреть контроль корректности работы программ с стеком данных.

5.3 Условия эксплуатации

5.3.1 Условия эксплуатации в соответствие с СанПиН 2.2.2/2.4.1340-03.

5.4 Требования к составу и параметрам технических средств

Минимальная конфигурация технических средств, на которых может быть запущен компилятор:

- количество ядер процессора – 1 шт;
- объем ОЗУ – 1 Гб.

5.5 Требования к информационной и программной совместимости

5.5.1 Программа должна иметь интерфейс командной строки.

5.5.2 Программа должна работать под управлением операционных систем семейств Linux.

5.6 Требования к маркировке и упаковке

Требования к маркировке и упаковке не предъявляются.

5.7 Требования к транспортированию и хранению

Требования к транспортировке и хранению не предъявляются.

5.8 Специальные требования

Сгенерировать установочную версию ПО.

6 ТРЕБОВАНИЯ К ПРОГРАММНОЙ ДОКУМЕНТАЦИИ

6.1 Разрабатываемые программные модули должны быть самодокументированы, т.е. тексты программ должны содержать все необходимые комментарии.

6.2 В состав сопровождающей документации должны входить:

6.2.1 Расчетно-пояснительная записка на 55-60 листах формата А4 (без приложений).

6.2.2 Техническое задание (Приложение А).

6.2.3 Руководство программиста (Приложение Б).

6.3.4 Исходный текст программного компонента статического анализа (Приложение В).

6.4 Графическая часть должна быть выполнена на 6 листах формата А1 (копии формата А3/А4 включить в качестве приложений к расчетно-пояснительной записке):

6.3.1 Схема структурная ПО.

6.3.2 Схемы алгоритмов разбора текста передаваемой программы и статического анализа.

6.3.3 Функциональная диаграмма ПО.

6.3.4 Результаты сравнительного анализа стековых языков.

6.3.5 Диаграмма вариантов использования.

6.3.6 Синтаксическая диаграмма консольного интерфейса ПО в виде РБНФ.

7 ТЕХНИКО-ЭКОНОМИЧЕСКИЕ ПОКАЗАТЕЛИ

Требования не предъявляются.

8 СТАДИИ И ЭТАПЫ РАЗРАБОТКИ

Этапы разработки курсовой работы указаны в таблице А.1.

Таблица А.1 — Этапы разработки

№	Название этапа	Срок даты, %	Отчетность
1	Разработка технического задания	9.02.2026 – 28.02.2026, 5%	Утвержденное техническое задание и задание на выпуск-ную квалификацион-ную работу
2	Анализ требований и уточнение спецификаций (эскизный проект)	28.02.2026 – 14.03.2026, 20%	Спецификации ПО.
3	Проектирование структуры программной системы, проектирование компонентов (технический проект)	14.03.2026 – 01.04.2026, 25%	Схема структурная системы. Проектная документация: диаграмма вариантов использования, диаграмма вариантов использования, функциональная диаграмма, схемы алгоритмов разбора текста передаваемой программы и статического анализа, синтаксическая диаграмма консольного интерфейса ПО в виде РБНФ.
4	Реализация компонентов и автономное тестирование компонентов. Сборка и комплексное тестирование.	01.04.2026 – 14.05.2026, 25%	Тексты программных компонентов. Тесты, результаты тестирования.

Продолжение таблицы А.1

5	Разработка документации.	14.05.2026 – 25.05.2026, 15%	Расчетно-пояснительная записка.
6	Прохождение нормоконтроля, проверка на антиплагиат, получение рецензии, подготовка доклада и предзащита.	25.05.2026 – 6.06.2026, 5%	Иллюстративный материал, доклад, рецензия, справки о нормо-контrole и проценте плагиата.
7	Защита выпускной квалификационной работы.	1.06.2026 – 04.07.2026, 5%	–

9 ПОРЯДОК КОНТРОЛЯ И ПРИЕМА

9.1 Порядок контроля

Контроль выполнения осуществляется руководителем еженедельно.

9.2 Порядок защиты

Защита осуществляется перед государственной экзаменационной комиссией (ГЭК).

9.3 Срок защиты

Срок защиты определяется в соответствии с планом заседаний ГЭК.

10 ПРИМЕЧАНИЕ

В процессе выполнения работы возможно уточнение отдельных требований технического задания по взаимному согласованию руководителя и исполнителя.