Простой Рефал. Руководство пользователя

Оглавление

[1 История диалекта 3](#_Toc419396101)

[2 Начинающим 4](#_Toc419396102)

[3 Язык Простой Рефал: синтаксис и семантика 4](#_Toc419396103)

[3.1 Структура программы 4](#_Toc419396104)

[3.2 Программные элементы (объявления и определения) 5](#_Toc419396105)

[3.2.1 Основные понятия 5](#_Toc419396106)

[3.2.2 Объявления функций ($FORWARD и $EXTERN) 5](#_Toc419396107)

[3.2.3 Идентификаторы ($LABEL) 6](#_Toc419396108)

[3.2.4 Пустые функции ($ENUM и $EENUM) 7](#_Toc419396109)

[3.2.5 Статические ящики ($SWAP и $ESWAP) 8](#_Toc419396110)

[3.2.6 Регулярные функции 8](#_Toc419396111)

[3.3 Функции, семантика и сравнение с РЕФАЛом-5 9](#_Toc419396112)

[3.3.1 Синтаксис функций, общее с РЕФАЛом-5 9](#_Toc419396113)

[3.3.2 Типы данных Простого Рефала 10](#_Toc419396114)

[3.3.3 Чего нет в РЕФАЛе-5, но есть в Простом Рефале 12](#_Toc419396115)

[3.3.4 Что есть в РЕФАЛе-5, но нет в Простом Рефале 16](#_Toc419396116)

[3.3.5 Отличия от РЕФАЛа-5 на уровне библиотеки 16](#_Toc419396117)

[4 Библиотека функций 17](#_Toc419396118)

[4.1 Используемые обозначения 17](#_Toc419396119)

[4.2 Базовая библиотека (Library) 17](#_Toc419396120)

[4.2.1 Add, Sub, Mul, Div, Mod 17](#_Toc419396121)

[4.2.2 WriteLine 17](#_Toc419396122)

[4.2.3 ReadLine 17](#_Toc419396123)

[4.2.4 FOpen 18](#_Toc419396124)

[4.2.5 FClose 18](#_Toc419396125)

[4.2.6 FWriteLine 18](#_Toc419396126)

[4.2.7 FReadLine 18](#_Toc419396127)

[4.2.8 Arg 18](#_Toc419396128)

[4.2.9 ExistFile 18](#_Toc419396129)

[4.2.10 GetEnv 18](#_Toc419396130)

[4.2.11 Exit 18](#_Toc419396131)

[4.2.12 System 19](#_Toc419396132)

[4.2.13 IntFromStr 19](#_Toc419396133)

[4.2.14 StrFromInt 19](#_Toc419396134)

[4.2.15 Chr, Ord 19](#_Toc419396135)

[4.2.16 SymbCompare 19](#_Toc419396136)

[4.2.17 SymbType 19](#_Toc419396137)

[4.3 Расширенная библиотека (LibraryEx) 19](#_Toc419396138)

[4.4 LibraryEx: надстройки над функциями базовой библиотеки и другие полезные функции 20](#_Toc419396139)

[4.4.1 LoadFile 20](#_Toc419396140)

[4.4.2 SaveFile 20](#_Toc419396141)

[4.4.3 Inc, Dec 20](#_Toc419396142)

[4.4.4 FastIntFromStr 20](#_Toc419396143)

[4.4.5 ArgList 20](#_Toc419396144)

[4.4.6 Compare 20](#_Toc419396145)

[4.4.7 Compare-T 20](#_Toc419396146)

[4.4.8 Type, Type-T 21](#_Toc419396147)

[4.4.9 Trim 21](#_Toc419396148)

[4.5 LibraryEx: функции высших порядков 21](#_Toc419396149)

[4.5.1 Понятие функтора, Apply 21](#_Toc419396150)

[4.5.2 Map 22](#_Toc419396151)

[4.5.3 Reduce 22](#_Toc419396152)

[4.5.4 Fetch 23](#_Toc419396153)

[4.5.5 Y-комбинатор 24](#_Toc419396154)

[4.5.6 MapReduce 25](#_Toc419396155)

[4.5.7 UnBracket 26](#_Toc419396156)

[4.5.8 DelAccumulator 26](#_Toc419396157)

[4.5.9 Seq 26](#_Toc419396158)

[4.6 Пример кода, использующий функции высших порядков 27](#_Toc419396159)

[5 Компилятор Простого Рефала: реализация 29](#_Toc419396160)

[5.1 Ключи командной строки, вызов компилятора 29](#_Toc419396161)

[5.2 Макросы препроцессора C++ 30](#_Toc419396162)

[5.3 Утилита srmake 31](#_Toc419396163)

[6 Отладка программ на Простом Рефале 32](#_Toc419396164)

[6.1 Использование отладчика C++ 32](#_Toc419396165)

[6.2 Средства отладки рантайма 32](#_Toc419396166)

[6.3 Идиомы отладки 33](#_Toc419396167)

[7 Интерфейс с языком C++ 34](#_Toc419396168)

[7.1 Вычислительная модель 34](#_Toc419396169)

[7.2 Написание внешних функций 34](#_Toc419396170)

[7.2.1 Быстрый и грязный способ 34](#_Toc419396171)

[7.2.2 Написание функции вручную 34](#_Toc419396172)

[8 Установка компилятора 34](#_Toc419396173)

[8.1 Установка на Windows 34](#_Toc419396174)

[8.2 Установка на Linux (MacOS X) 34](#_Toc419396175)

[9 Структура каталогов дистрибутива 34](#_Toc419396176)

[10 Известные ошибки 35](#_Toc419396177)

[11 Список литературы 35](#_Toc419396178)

# История диалекта

РЕФАЛ — РЕкурсивный Функциональный АЛгоритмический язык[[1]](#footnote-1), язык функционального программирования, ориентированный на символьные вычисления, обработку и преобразование текстов. У данного языка есть много несовместимых диалектов, Простой Рефал — один из них.

*Примечание.* Устоявшегося способа написания названия языка РЕФАЛ нет, встречается запись как кириллицей, так и латиницей, целиком в верхнем регистре, с большой буквы или с маленькой. В этом документе, когда речь идёт о семействе языков, я буду использовать запись в верхнем регистре (РЕФАЛ), когда будет идти речь о конкретном диалекте, буду использовать запись, принятую в руководстве конкретного языка. Если слово «рефал» будет употребляться в роли прилагательного, то оно будет писаться с маленькой буквы: рефал-выражение, рефал-программа. Поскольку я автор Простого Рефала, то я выбрал ему имя и поэтому утверждаю: оба слова пишутся с большой буквы, название диалекта можно дословно переводить на другие языки, но при этом оба слова тоже должны быть с большой буквы. Например, по-английски он называется Simple Refal (не Easy Refal, см. далее, почему).

Язык создавался мною как исследовательский проект — хотел для себя понять, как осуществляется компиляция кода на РЕФАЛе в императивный код. Целью было написание минимального, но при этом алгоритмически полного компилятора диалекта Базисного РЕФАЛа, простота транслятора (например, однопроходность) была важнее удобства программирования на языке.[[2]](#footnote-2) Несмотря на это ограничение (а, возможно, благодаря ему), язык получился довольно целостным и согласованным. Компилятор оказался достаточно простым, благодаря чему в МГТУ имени Н. Э. Баумана на кафедре ИУ9 он использовался (и используется) как тестовый полигон для нескольких курсовых проектов. В частности, в настоящем дистрибутиве можно видеть результат одного из таких исследований — генерация интерпретируемого кода (о чём подробно будет сказано позже).

В процессе эволюции компилятор незначительно отошёл от своей первоначальной простоты. Во-первых, он стал использоваться как back-end для Модульного Рефала, для чего в него были добавлены такие возможности, как идентификаторы, абстрактные типы данных и статические ящики. Позже на нём исследовалась реализация вложенных функций — соответственно, язык пополнился и этим удобным средством. Однако, в обоих случаях концептуальная целостность и согласованность языка не пострадали (как мне кажется).

# Начинающим

Писать хороший учебник, объясняющий язык с нуля, содержащий примеры, задачи и контрольные вопросы, я здесь не буду. Во-первых, написание учебника, в отличие от справочника, требует гораздо большего времени и сил, а во-вторых, всё уже написано до нас. Есть хороший учебник (Турчин, 1989) по похожему диалекту РЕФАЛ-5, написанный самим Валентином Фёдоровичем Турчиным, автором РЕФАЛа. Поэтому рекомендуем сначала изучить именно его, поскольку дальнейшее изложение будет вестись в предположении, что читатель владеет диалектом РЕФАЛ-5. В частности, такие понятия, как объектное выражение, абстрактная рефал-машина и др., введённые в том учебнике, здесь определяться не будут.

Прочитали учебник? Напишите программу преобразующую арифметические выражения в инфиксной нотации в постфиксную (польскую инверсную запись). Получилось? Переходите к следующему разделу.

# Язык Простой Рефал: синтаксис и семантика

## Структура программы

Программа на Простом Рефале (далее для краткости — на Рефале) состоит из набора единиц трансляции — исходных файлов, написанных на Рефале (с расширением .sref) и C++ (с расширением .cpp), последние включают в себя библиотеки первичных функций и среду поддержки времени выполнения (run-time support library, далее для краткости — рантайм). Компилятор Простого Рефала транслирует исходные тексты на Рефале в исходные тексты на C++, получившийся набор файлов транслируется компилятором C++ (совместимым со стандартом ANSI/ISO C++ 1998 года).

Файл исходного текста на Рефале состоит из последовательности **программных элементов** — объявлений и определений программных сущностей, независимо транслируемых в соответствующие программные конструкты на C++.

Синтаксически это выглядит так:

Program = { ProgramElement }.

Программа пишется в свободном формате, то есть переводы строк приравнены к обычным пробельным символам, там, где допустим пробельный символ (пробел или табуляция), допустима вставка перевода строки. Пробельные символы (далее, пробелы) можно вставлять между двумя любыми лексемами языка. Пробелы обязательны в том случае, когда две лексемы, записываемые подряд, могут интерпретироваться как одна сплошная лексема (например, два числа, записанные слитно, будут интерпретироваться как одно число и т.д.). Пробелы недопустимы внутри идентификаторов, чисел, директив. Пробелы внутри цепочек литер интерпретируются как образы литер со значением «пробел».

В любом месте, где допусти́м пробел, можно вставить и **комментарий**. Комментарии могут быть двух видов: однострочные (в стиле C) и многострочные (в стиле C++). Многострочные комментарии начинаются с /\* и заканчиваются на \*/, могут пересекать границы текста. Последовательность символов /\* внутри многострочного комментария запрещена. Однострочные комментарии начинаются на // и заканчиваются в конце строки.

Каждая программная сущность имеет своё уникальное **имя**. Имя представляет собой последовательность латинских букв, цифр и знаков - (минус) и \_ (прочерк), начинающуюся с заглавной латинской буквы. Примеры имён: GN-Local, FuncArguments, Example12345, A-b\_c. Имена чувствительны к регистру. Символы - и \_ взаимозаменяемы, поэтому имена A-b\_c и A\_b-c эквивалентны. При трансляции в C++ имена программных конструкций в целевом коде получаются из имён в исходном коде путём замены знаков - на \_ (т.е. предыдущее имя после трансляции преобразится в A\_b\_c). Длина имён не ограничена, однако с очень длинными именами в целевом коде может отказаться работать нижележащий компилятор C++ (либо считать имена идентичными, если у них совпадают первые N символов, где N зависит от выбранного компилятора).

## Программные элементы (объявления и определения)

### Основные понятия

**Программные элементы** делятся на две категории: объявления и определения программных сущностей. **Программные сущности** делятся на функции и идентификаторы. **Объявления** служат для уведомления компилятора о том, что где-то (в текущем файле ниже по тексту программы или в другой единице трансляции) определена программная сущность с заданным именем и заданными свойствами (вид сущности, тип доступа). **Определения**, собственно, определяют данную программную сущность. **Функции**, как правило, содержат исполнимый код. **Идентификаторы** пополняют пространство атомов (рефал-символов) определёнными пользователем атомами-идентификаторами. В единице трансляции может быть определено не более одной функции и одного идентификатора с заданным именем.

Каждая функция имеет тип доступа. Функции с **локальным** типом доступа видимости (локальные функции) доступны только в том файле, где они определены, в различных единицах трансляции могут существовать локальные функции с одинаковыми именами. На функции с **глобальным** типом доступа видимости (т. н. **entry-функции**) можно ссылаться из других единиц трансляции при помощи директивы (ключевого слова) $EXTERN. Entry-функции должны иметь уникальные имена в рамках всей программы. *Примечание:* имя локальной функции в одной из единиц трансляции может совпадать с именем entry-функции из другой единицы при условии, что в первой из них entry-функция не импортируется директивой $EXTERN. Для идентификаторов понятие типа доступа не определено: в различных единицах трансляции могут быть определения одноимённых идентификаторов (это роднит идентификаторы с локальными функциями), но при этом все эти определения определяют один и тот же идентификатор на всю программу (это роднит их с entry-функциями), т. е. определения идентификаторов можно трактовать как их объявления идентификаторов, определённых в некоторой вымышленной единице трансляции, содержащей все возможные идентификаторы.

В целевом коде на C++ локальные функции соответствуют функциям, определённым с модификатором static, entry-функции — функциям без этого модификатора. Идентификаторы компилируются в довольно хитрую конструкцию, о которой будет рассказано в соответствующем разделе.

Под **пространством имён** будем понимать множество имён функций (с атрибутами области типа доступа и определённости) и идентификаторов, доступных в данной точке программы.

Синтаксис программного элемента:

ProgramElement =

FunctionDeclaration

| IdentifierDefinition

| EnumDefinition

| SwapDefinition

| FunctionDefinition

| ";".

*Примечание.* Таким образом, синтаксис Простого Рефала допускает на верхнем уровне, помимо объявлений и определений, любое количество точек с запятой. Программисты на Рефале-6 при желании могут после каждого блока функции ставить точку с запятой.

### Объявления функций ($FORWARD и $EXTERN)

Синтаксис:

FunctionDeclaration = ( "$FORWARD" | "$EXTERN" ) NameList.

NameList = NAME { "," NAME } ";".

Объявление функций помещает перечисленные имена в пространство имён как локальные функции (если указана директива $FORWARD), либо как entry-функции (если указана директива $EXTERN). Если в пространстве имён уже присутствует одно из указанных имён как имя функции, но имеет другой тип доступа, то сигнализируется синтаксическая ошибка.

Директива $FORWARD необходима при определении взаимно рекурсивных локальных функций. Кроме того, читаемость программ часто улучшается, если определение функции не предшествует, а следует за её использованием (особенно, при написании парсера методом рекурсивного спуска). Данная директива позволяет придерживаться такого стиля программирования.

Директива $EXTERN почти всегда используется для ссылки на entry-функцию, определённую в другой единице трансляции. Однако может быть полезна при обращении к entry-функции, определённой в той же единице трансляции, но ниже по тексту программы.

Объявления функций на Рефале компилируются в соответствующие объявления функций на C++, $FORWARD — с модификатором static, $EXTERN — с модификатором extern.

Пример:

$FORWARD A, B, C;

$EXTERN E, F, G;

Компилируется в:

static refalrts::FnResult A(refalrts::Iter arg\_begin, refalrts::Iter arg\_end);

static refalrts::FnResult B(refalrts::Iter arg\_begin, refalrts::Iter arg\_end);

static refalrts::FnResult C(refalrts::Iter arg\_begin, refalrts::Iter arg\_end);

extern refalrts::FnResult E(refalrts::Iter arg\_begin, refalrts::Iter arg\_end);

extern refalrts::FnResult F(refalrts::Iter arg\_begin, refalrts::Iter arg\_end);

extern refalrts::FnResult G(refalrts::Iter arg\_begin, refalrts::Iter arg\_end);

### Идентификаторы ($LABEL)

Синтаксис:

IdentifierDefinition = "$LABEL" NameList.

Определение идентификатора помещает перечисленные имена в область видимости как имена идентификаторов. Если в области видимости уже присутствует одно из указанных имён как имя идентификатора, то сигнализируется синтаксическая ошибка.

Определение идентификаторов компилируется в набор конструкций вида

//$LABEL *ИмяИдентификатора*

template <typename T>

struct *ИмяИдентификатора* {

static const char \*name() {

return "*ИмяИдентификатора*";

}

};

по одной для каждого из имён после директивы $LABEL. Обращение к идентификатору осуществляется конструкцией вида & *ИмяИдентификатора*<int>::name.

Как это работает. Определение идентификатора есть определение шаблонного класса, обращение к нему — конкретизация шаблона некоторым типом (тип не имеет значения, важно только всегда использовать один и тот же, в компиляторе используется int). При конкретизации шаблона класса компилятор C++ порождает код метода этого класса. В практике программирования на C++ библиотеки шаблонов размещаются в заголовочных файлах, в результате чего скомпилированный код для одних и тех же шаблонов с одинаковыми параметрами-типам многократно дублируется в объектных файлах. Компоновщик C++ обнаруживает такие случаи и в целевой исполнимый файл помещает только единственный экземпляр данного кода. Возвращаясь к Рефалу. Если в нескольких разных единицах трансляции присутствует определение одного и того же идентификатора, то компилятор С++ породит в соответствующих объектных файлах идентичный код для функции ИмяИдентификатора<int>::name(), компоновщик устранит дублирующийся код и в целевой исполнимый файл будет помещён единственный код для этой функции. Соответственно, указатели на эту функцию из других функций из разных единиц трансляции будут идентичны.

Если вы ничего не поняли, относитесь к этому как к магии.

### Пустые функции ($ENUM и $EENUM)

Синтаксис:

EnumDefinition = ( "$ENUM" | "$EENUM" ) NameList.

Определение пустых функций помещает перечисленные имена в пространство имён как определённые имена локальных (директива $ENUM) или entry-функций (директива $EENUM — **e**ntry **enum**). Если в пространстве имён уже присутствует одно из этих имён как определённая функция или как функция с другим типом доступа, то сигнализируется синтаксическая ошибка.

Функция, определённая таким образом, является пустой, её тело не имеет ни одного предложения (см. далее про синтаксис функций). Вызов такой функции всегда приводит к авосту (аварийному останову) — невозможности сопоставления.

Код

$ENUM L;

$EENUM E;

идентичен следующему (см. далее синтаксис функций):

L { /\* пусто \*/ }

$ENTRY E { /\* пусто \*/ }

Данная конструкция устаревшая и не рекомендуется для дальнейшего применения (**deprecated**) в качестве замены идентификаторов. Первоначально в Простом Рефале отсутствовали идентификаторы, вместо них использовались пустые функции. Они предназначались для той же цели, что и идентификаторы, а именно создание новых атомов с описательными именами — в качестве таких атомов использовались пустые функции. Сейчас есть более удобное средство для этой цели — собственно сами идентификаторы, рекомендуем использовать именно их. Однако, пустые функции встречаются в старом коде (в том числе и в исходном тексте самого компилятора), а также в стандартной библиотеке (которую надо в этом плане актуализировать), поэтому знать эту конструкцию надо. По сравнению с идентификаторами, пустые функции имеют следующие недостатки:

1. При использовании $EENUM необходимо следить, чтобы функция была определена только в одной единице трансляции, а в других она импортировалась как $EXTERN. Данная асимметрия затрудняет программирование.
2. При использовании транслятора MS Visual C++ 2013 и установки наивысшей оптимизации компилятор устраняет дублирующиеся определения функций, даже с разными именами. В результате все атомы пустых функций (поскольку их тело полностью идентично) становятся неразличимыми, что противоречит их предназначению. (Возможно, эта ошибка будет исправлена в дальнейшем.)

Единственное оправдание использования пустых функций — как метки для абстрактных типов данных (см. далее).

*Примечание.* Идея пустых функций заимствована из РЕФАЛа-2 (Алешин, Красовский, Романенко, & Шерстнев, 1991).

### Статические ящики ($SWAP и $ESWAP)

Синтаксис:

SwapDefinition = ( "$SWAP" | "$ESWAP" ) NameList.

Определение статических ящиков помещает перечисленные имена в пространство имён как определённые имена локальных (директива $SWAP) или entry-функций (директива $ESWAP). Если в пространстве имён уже присутствует одно из этих имён как определённая функция или как функция с другим типом доступа, то сигнализируется синтаксическая ошибка.

Статические ящики представляют собой функции с состоянием. Функция-статический ящик при вызове с любым аргументом возвращает аргумент своего предыдущего вызова, при первом вызове возвращает пустое выражение. Таким образом, данная функция может выполнять роль глобальной переменной в программе, которая может хранить некоторое значение и обменивает его на свой аргумент. Для чтения значения из статического ящика необходимо сначала вызвать ящик с любым аргументом, запомнить то, что он возвратил, снова вызвать с тем, что он возвратил за первый вызов и отбросить результат второго вызова — первый вызов извлекает хранимое значение, второй его восстанавливает. Для записи следует вызвать статический ящик с новым значением и отбросить результат вызова.

*Примечание.* Знатоки РЕФАЛа-5 заметят, что статический ящик ведёт себя также, как и функция Rp.

*Примечание 2.* Если есть статические ящики, значит должны быть и нестатические? Они есть, но не здесь. В диалекте РЕФАЛ-2 (Алешин, Красовский, Романенко, & Шерстнев, 1991) существуют как статические ящики (аналогичные описанным здесь), так и динамические, которые создаются в куче во время выполнения программы. Простой Рефал не поддерживает динамические ящики, однако, за неимением лучшего заимствует данную терминологию.

### Регулярные функции

Синтаксис:

FunctionDefinition = [ "$ENTRY" ] NAME Block.

Определение регулярной функции (далее для краткости — просто функции) помещает имя функции в пространство имён как определённое имя локальной (без директивы $ENTRY) или entry-функции (c директивой $ENTRY). Если в пространстве имён уже присутствует данное имя как определённая функция или как функция с другим типом доступа, то сигнализируется синтаксическая ошибка. Синтаксис блока и семантика выполнения функций будут рассмотрены в последующих разделах.

Все разновидности функций (пустые функции, статические ящики, регулярные функции) компилируются в функции C++ со следующей сигнатурой:

static refalrts::FnResult *ЛокальнаяФункция*(  
 refalrts::Iter arg\_begin, refalrts::Iter arg\_end  
) {

…

}

refalrts::FnResult *EntryФункция*(refalrts::Iter arg\_begin, refalrts::Iter arg\_end) {

…

}

Смысл возвращаемого значения и параметров arg\_begin и arg\_end будет рассмотрен в разделе, посвящённом интерфейсу с языком C++.

## Функции, семантика и сравнение с РЕФАЛом-5

Простой Рефал можно охарактеризовать как расширенное подмножество Базисного РЕФАЛА диалекта РЕФАЛ-5 (главы 1 и 2 руководства (Турчин, 1989)). Основными отличиями от РЕФАЛа-5 являются используемые типы данных, совсем другая стандартная библиотека, отсутствие средств Расширенного РЕФАЛа диалекта РЕФАЛ-5 (главы 4 и 6 в том же руководстве), специфические языковые особенности (предобъявления: любая программная сущность должна быть описана перед использованием; незначительные отличия на уровне синтаксиса) и расширения языка, отсутствующие в РЕФАЛе‑5 (самое важное из таких расширений — вложенные функции).

### Синтаксис функций, общее с РЕФАЛом-5

Общие черты языков Простой Рефал и РЕФАЛ-5 (подмножество Базисного РЕФАЛа):

* Функции внешне имеют сходный синтаксис (есть небольшие различия, позже на них остановимся).
* Функции оперируют **объектными выражениями**: принимают в качестве аргумента одно объектное выражение и возвращают его в качестве результата.
* Семантика выполнения программы тоже определяется в терминах абстрактной рефал-машины (см. 1.5 в (Турчин, 1989)).
* Идентичная семантика сопоставления с образцом. Структура результатного выражения незначительно отличается.
* *(Относится не к языку, а к реализации)* Затраты времени на операции. Реализация Простого Рефала использует классическое списковое представление, поэтому особенности стоимости отдельных операций (сопоставление с открытыми и повторными переменными, стоимость конкатенации и копирования) те же, что и в реализации РЕФАЛа-5 (версия PZ Oct 29 2004).

Синтаксис функции:

FunctionDefinition = [ "$ENTRY" ] NAME Block.

Block = "{" { Sentence } "}".

Sentence = Pattern "=" Result ";".

Pattern = { PatternTerm }.

Result = { ResultTerm }.

PatternTerm =

CommonTerm | "(" Pattern ")" | "[" NAME Pattern "]" | RedefinitionVariable.

RedefinitionVariable = VARIABLE "^".

ResultTerm =

CommonTerm | "(" Result ")" | "[" NAME Result "]" | "<" Result ">" | Block.

CommonTerm = CHAR | NUMBER | NAME | VARIABLE | "#" NAME.

Глобальная регулярная функция (та, которая не вложенная) представляет собой именованный блок, который может предваряться ключевым словом $ENTRY. **Блок**, в свою очередь, содержит последовательность нуля или более предложений (в отличие от РЕФАЛа-5, где в любой функции должно быть как минимум одно предложение).

**Предложение** состоит из двух частей: образцовой части и результатной части, разделённых знаком «=». Каждое предложение заканчивается точкой с запятой (в отличие от РЕФАЛа-5, где после последнего предложения точка с запятой не обязательно).

**Образцовая часть** (синонимы: левая часть, образцовое выражение, образец) состоит из последовательности образцовых термов, среди которых могут быть литералы атомарных термов (см. далее), скобочные термы двух видов (структурные круглые скобки и именованные квадратные скобки, т. н. абстрактные типы данных) и переменные, некоторые из которых могут быть помечены знаком переопределения ^ (о них будет рассказано при описании вложенных функций).

**Результатная часть** (синонимы: правая часть, результатное выражение, результат) состоит из последовательности результатных термов, среди которых могут быть литераты атомарных термов, пассивные скобочные термы тех же двух видов (круглые и абстрактные скобки), скобки активации и блоки — литералы вложенных функций. Одно из отличий от РЕФАЛа-5 заключается в том, что после открывающей скобки активации синтаксически не обязано находиться имя функции, проверка того, что за ним находится вызываемая сущность (функция или замыкание) осуществляется во время выполнения программы.

**Переменные**, как и в РЕФАЛе‑5, могут быть трёх видов: s‑переменные — могут сопоставляться только с атомами, t‑переменные — могут сопоставляться с любым термом и e‑переменные — могут сопоставляться с любым объектным выражением. Область видимости переменной, объявленной в некотором образцовом выражении — образцовое и результатное выражение данного предложения, а также все вложенные функции, определённые внутри результатного выражения. Однако, внутри вложенных функций переменные могут скрывать другие одноимённые переменные из внешней области видимости, если они помечены знаком переопределения ^ (подробнее в разделе 3.3.3.3). Переменные записываются как вид.индекс, где вид может быть одной из букв s, t или e, индекс — любая непустая последовательность из латинских букв, цифр и знаков - (дефис) и \_ (прочерк), как и в случае с именами, последние два символа эквивалентны, индексы чувствительны к регистру. В одной области видимости не может быть двух переменных с одинаковым индексом, но разного вида.

### Типы данных Простого Рефала

**Объектное выражение** в Простом Рефале представляет собой последовательность **объектных термов**. Объектные термы относятся к одной из двух категорий: атомарные термы и составные термы.

**Атомарные термы** (атомы) представляют собой элементы данных, которые невозможно разбить на составные части механизмом сопоставления с образцом. Традиционно они называются символами (symbols), но, чтобы не возникало путаницы с литеральными символами (characters), я решил отступить от сложившейся терминологии. Виды атомов:

* **Литеральные символы** (ASCII-символы, characters) Один атом представляет собой 8-битный символ, хранимый в переменной типа char языка C++. Поддержки Уникода и кодовых страниц на уровне рантайма и стандартной библиотеки языка на данный момент нет — ввод-вывыд в текстовые файлы и на консоль осуществляется побайтово средствами стандартной библиотеки C++ (stdio.h).

**Синтаксис.** Литеральные символы записываются в одинарных кавычках, последовательность нескольких литеральных символов подряд можно записывать внутри одной пары кавычек. Для записи одинарной кавычки она удваивается. Также можно использовать escape-последовательности: \n, \r, \t, \\, \' с тем же смыслом, что и в языке C++, \ddd — символ *с десятичным* кодом ddd.

**Примеры:**

|  |  |  |
| --- | --- | --- |
| В тексте программы | Что из себя представляет | Примечание |
| 'A' 'B' 'C' | ABC |  |
| 'ABC' | ABC | несколько символов можно записывать слитно |
| ' ' |  | пробел |
| '' | ' | одинарная кавычка |
| '\'' | ' | тоже самое |
| '''''' | ''' | три кавычки |
| '''A''B' | 'A'B |  |
| '\r\n' | CR+LF | перевод строки в dos/windows |
| '\013\010' | CR+LF | запись символов в десятичном коде |

* **Целые числа.** Представляют собой десятичную запись беззнаковых целых чисел в диапазоне от 0 до 2N−1, где N=32 при компиляции (компилятором C++) в 32-разрядный код и N=64 при компиляции в 64-разрядный код. При синтаксическом анализе целочисленное переполнение не проверяется, поэтому в программе можно записать число хоть из 100 десятичных цифр, программа успешно скомпилируется, но это число будет проинтерпретировано как некоторое другое десятичное число, входящее в допустимый диапазон. Следствие: если вы планируете выполнять вычисления с большими числами на 64-разрядной платформе, убедитесь, что компилятор также собран под 64-разрядную платформу.

**Примеры:** 10, 00010 (лидирующие нули игнорируются), 4294967295 (максимально допустимое число на 32-разрядной платформе), 18446744073709551615 (максимально допустимое число на 64-разрядной платформе).

* **Замыкания** (указатели на функции). Замыкание, создаваемое из глобальной функции (**указатель на** глобальную **функцию**) записывается просто как имя функции, может присутствовать как в левой, так и в правой части. Замыкание, создаваемое из вложенной функции, записывается как вложенный блок в результатной части, соответственно, может присутствовать тоже только в правой части. Подробнее о них будет рассказано в разделах 3.3.3.1 и 3.3.3.3.

Если имя функции в месте записи указателя на функцию отсутствует в пространстве имён, сигнализируется синтаксическая ошибка.

**Примеры:** Fab (указатель на функцию Fab), { t.B = (t.A t.B); } (замыкание из вложенной функции).

* **Идентификаторы.** Запись атома-идентификатора представляет собой имя идентификатора, предварённое символом #. Между # и именем идентификатора допусти́м пробел. Имя идентификатора в месте записи идентификатора должно находиться в пространстве имён.

**Примеры:** # True, #False, # NotFound, # /\* комментарий \*/ TokenString — в последнем случае между знаком # и именем находится комментарий, запись синтаксически корректна, но лучше так не пишите.

* **Дескрипторы файлов** (указатели). Данные атомы служат для представления некоторого низкоуровневого ресурса, представляемого указателем C++ на некие данные. Записываться как литералы в тексте программы не могут, их могут порождать во время выполнения внешние функции, записанные на C++. В стандартной библиотеке (Library) атомы этого типа используются для хранения указателей FILE\*, за что и получили своё название (см. функцию FOpen). Сторонние библиотеки могут хранить в них любые другие типы ресурсов (например, сокеты из winsock). На печать выводятся (средствами Library, либо в дампе поля зрения) как \*XXX…X, где XXX…X — шестнадцатеричная запись указателя.

**Составные термы** бывают двух видов:

* **Структурные** (круглые) **скобки.** Полностью аналогичны структурным скобкам РЕФАЛа‑5.
* **Абстрактные типы данных,** они же АТД-скобки (АТД-термы), они же ADT-скобки (ADT-термы), они же квадратные скобки, они же именованные скобки. Записываются как [ИМЯ выражение], где ИМЯ — имя некоторой функции из пространства имён в данной точке, выражение — объектное выражение. При отсутствии имени функции после открывающей скобки сигнализируется синтаксическая ошибка. Подробнее об абстрактных типах данных написано в разделе 3.3.3.2.

Между открывающей квадратной скобкой и именем допустим пробел.

### Чего нет в РЕФАЛе-5, но есть в Простом Рефале

#### Функции как атомы и как объекты первого класса

В РЕФАЛе‑5 атомы могут быть только трёх разных видов[[3]](#footnote-3): идентификаторы, литеральные символы и целые числа (там они называются макроцифрами). Простой Рефал к этому набору добавляет ещё и указатели на функции (а также и вложенные функции, но о них позже). Кроме того, если РЕФАЛ‑5 требует, чтобы после < всегда следовало имя функции, определённой в текущей единице трансляции, то Простой Рефал снимает это ограничение. Наличие атома-функции за открывающей скобкой вызова проверяется лишь на этапе выполнения.

Что нам это даёт? Чуть меньше надёжности и гораздо больше гибкости! Теперь можно записывать и такие конструкции:

<s.Func некий аргумент>

<<GetFunc аргумент1> аргумент2>

Здесь предполагается, что s.Func содержит корректный указатель на функцию, а <GetFunc…> указатель на функцию возвращает. Знатоки РЕФАЛа‑5 могут меня спросить: а чем отличается эта конструкция от функции Mu? Ответ: функция Mu работает в статической области видимости, т.е. вызывает функцию, имя которой совпадает с идентификатором, являющимся первым термом результата, т.е. функция Mu не способна вызвать функции, находящиеся в других единицах трансляции, если, конечно, эти функции не были импортированы директивой $EXTERN. Это, как можно догадаться, затрудняет написание библиотеки функций высшего порядка, поскольку, к примеру, функция Map, определённая в файле MyLib.ref следующим образом:

\* файл MyLib.ref

$ENTRY Map {

s.Func t.Next e.Tail =

<Mu s.Func t.Next> <Map s.Func e.Tail>;

s.Func /\* пусто \*/ = /\* всё \*/;

}

сможет преобразовывать выражение только встроенными функциями, либо функциями определёнными (либо импортированными) в MyLib.ref. А это снижает полезность библиотеки: чтобы выполнить некое новое преобразование, необходимо либо добавить соответствующую функцию в MyLib.ref, либо в ней же её импортировать. В обоих случаях приходится модифицировать файл библиотеки.

Более того, функция Map, в том виде, в каком она описана, имеет уязвимость. В качестве s.Func можно передать любой идентификатор, и, если он совпадёт с именем одной из локальных функций библиотеки, эта функция будет вызвана. Таким образом, вызов функции Mu с аргументом, полученным от вызывающего кода, может нарушить инкапсуляцию.

Аналогичная библиотека на Простом Рефале (описанная в файле, к примеру, MyLib.sref) с аналогичной функцией Map:

// файл MyLib.sref  
$ENTRY Map {

s.Func t.Next e.Tail =

<s.Func t.Next> <Map s.Func e.Tail>;

s.Func /\* пусто \*/ = /\* всё \*/;  
}

сможет преобразовывать объектное выражение любыми пользовательскими функциями, доступными в других единицах трансляции, поскольку косвенный вызов функции выполняется не по имени, а по указателю. Также устраняется уязвимость, связанная с вызовом по имени локальных функций — локальные функции, определённые в некоей единице трансляции, в Простом Рефале не доступны извне, только если сам код явно не вернёт указатель на такую функцию.

В расширенной библиотеке LibraryEx уже имеется функция Map, причём она более мощная, нежели определена здесь.

*Примечание.* Конструкции <10 …>, <'A' …> или <> являются синтаксически корректными (компилятор даже не выдаёт на них предупреждения), однако, при передаче управления на них, они приводят к ошибке времени выполнения.

#### Абстрактные типы данных

Язык РЕФАЛ‑5 реализует инкапсуляцию на уровне функций — функции, не помеченные как $ENTRY, недоступны из других единиц трансляции (однако, смотри оговорку про функцию Mu в предыдущем разделе). Данные же никак не защищены. Если функции модуля оперируют с неким термом, как абстрактным типом данных, то этот терм может быть либо скобочным термом, либо атомом — в любом случае реализация доступна вызывающему коду. Копилка также доступна для модификации любому коду в программе: функция Dgall извлекает всё содержимое копилки.

Простой Рефал реализует инкапсуляцию на уровне данных. Во-первых, как говорилось выше, можно объявить функцию-статический ящик локальной, а это значит, что внешний код не будет иметь данным в ящике никакого доступа (в противоположность доступности копилки). Во-вторых…

Во-вторых, в язык встроены абстрактные типы данных! Можно описать составной терм, доступ к содержимому которого есть только в той единице трансляции, где он создаётся. Для пользователей библиотеки он будет похож на атом — невозможно будет написать образец, который разобьёт этот терм на составные части. (но атомом не будет: сопоставляться с s‑переменными он не сможет, копирование будет выполняться за время пропорциональное размеру терма).

Для создания такого терма необходимо сначала описать некоторую локальную функцию (проще всего это сделать с помощью директивы $ENUM), она будет служить тегом АТД-терма, а затем создать терм вида [ИмяФункции некое-закрытое-содержимое], где ИмяФункции — тег АТД, некое-закрытое-содержимое — данные, подлежащие защите. Поскольку функция локальная, то сослаться из внешнего кода на неё невозможно, а значит и невозможно написать образец (вне единицы трансляции, где локальная функция описана), который бы получал доступ ко внутреннему содержимому.

В качестве тега можно использовать и имя entry-функции, тогда код из любой единицы трансляции сможет получить доступ к содержимому терма (импортировав функцию директивой $EXTERN и указав её имя после открывающей квадратной скобки). Такой терм не будет обеспечивать инкапсуляцию, однако может повысить надёжность программы (для обращения к таким данным надо будет писать образец вида [Имя …], попытка сопоставления с (e.Any) сразу выявит ошибку в программе).

*Примечание.* Код на C++ имеет полный доступ к содержимому абстрактных типов данных, созданных в других единицах трансляции, например, базовая библиотека (Library) может выводить их содержимое в консоль или в текстовый файл.

*Примечание 2.* Абстрактные типы данных не полностью закрыты для кода на Рефале из других единиц трансляции — их можно сравнивать на равенство путём сопоставления с повторными t-переменными. Истинные абстрактные типы данных должны были бы сами определять операцию сравнения на равенство.

#### Вложенные функции

В работе (Скоробогатов & Чеповский, 2006) был предложен диалект Refal-7, первый диалект РЕФАЛа, в котором поддерживались вложенные функции. Данная идея меня заинтересовала, и я тоже решил добавить вложенные функции себе в язык (см. doc/historical/note003.txt в каталоге дистрибутива). Однако, в отличие от Refal-7, где вложенные функции могут быть как именованные, так и безымянные, в Простом Рефале поддерживаются только безымянные функции.

Функции в программе на Простом Рефале существуют в двух ипостасях. Во-первых, функция — это некий исполнимый блок кода, выполняющий некоторые действия, при передаче управления на него. Блок кода порождается из соответствующего фрагмента текста программы. Во-вторых, функция — это некий объект, присутствующий в поле зрения, с которым программа может производить некие манипуляции: копировать, сравнивать на равенство, вызывать. Функция во второй ипостаси в разделе 3.3.3.1 была названа указателем на функцию. Можно сказать, что отношения между функцией и указателем на функцию такие же, как между классом и объектом в ООП: функции существуют только на стадии компиляции, по ним порождаются экземпляры функций, существующие во время выполнения.

Функция состоит из имени функции и блока (неявно создаются блоки у статических ящиков и пустых функций). Блок функции содержит исполнимый код, имя служит для ссылки на этот код из других мест программы. Однако, часто пишутся функции, на которые ссылка осуществляется только один раз — почему бы для них вместо отдельного определения функции и однократного использования имени не записать сразу блок в месте использования? При трансляции такую конструкцию можно трактовать как некоторую глобальную функцию с уникальным именем и ссылку на неё в правой части предложения. При выполнении соответствующего предложения на месте текстуальной записи такой *вложенной функции* будет помещаться указатель на эту функцию. Получается как-то так.

Вместо

UnBracket { (e.X) = e.X; }

Plain {

e.Brackets =

<Map UnBracket e.Brackets>;

}

Можно записать

Plain {

e.Brackets =

<Map { (e.X) = e.X; } e.Brackets>;

}

У внимательного читателя должен возникнуть вопрос: а что делать, если внутри вложенной функции используются переменные с теми же именами, что и во внешней функции? Ответ: в Простом Рефале такие переменные получают те значения, которые были с ними связаны во внешней функции. Таким образом, возможен случай, когда в результатной части предложения вложенной функции присутствует переменная, которой нет в образцовой части — такая переменная была связана уровнем выше.

Пример. Рассмотрим функцию, вычисляющую декартово произведение двух множеств.

// Функция CardProd вычисляет декартово произведение двух множеств  
// <CartProd ('a' 'b' 'c') (1 2)>[[4]](#footnote-4)  
// == ('a' 1) ('a' 2) ('b' 1) ('b' 2) ('c' 1) ('c' 2)

$ENTRY CartProd {

(e.SetA) (e.SetB) =

<Map

{ // № 1

t.A =

<Map

{ t.B = (t.A t.B); } // № 2

e.SetB

>;

}

e.SetA

>;

}

В этом примере вложенная функция, помеченная как № 1, содержит ссылку на переменную e.SetB, связанную внутри функции CartProd. Функция, помеченная как № 2, ссылается на переменную t.A, связанную в функции № 1.

Очевидно, что объект, порождаемый из такой вложенной функции во время выполнения, будет гораздо сложнее, чем просто указатель на функцию в смысле языка C++ — он, помимо указателя на код, выполняющий вычисления, должен хранить и значения связанных переменных. Такой объект в дальнейшем будем называть *замыканием* (а указатели на глобальные функции будем рассматривать как частный случай замыкания).

Введём определение. *Контекстом* замыкания вложенной функции называется множество переменных, связанных снаружи и используемых внутри функционального блока. Для примера выше контекст первой функции — переменная e.SetB, второй — t.A.

Замыкания являются атомами, а значит и объектными выражениями. Их, как и другие объектные выражения, можно копировать и сравнивать на равенство (путём использования одноимённых переменных). Опишем характерные свойства замыканий.

1. Замыкания, как и другие атомы, копируются за константное время.
2. Вызов замыкания может выполняться как за константное время, так и за время пропорциональное размеру контекста. Последнее возможно, если в поле зрения присутствует несколько копий данного замыкания, поэтому для вызова каждого из экземпляров требуется своя копия контекста — такова списковая реализация.
3. Два экземпляра одного замыкания, полученные путём копирования одного атома, равны.
4. Два замыкания, построенные из текстуально разных функциональных блоков, не равны[[5]](#footnote-5).
5. Два замыкания, имеющие разное содержимое элементов контекста, не равны.

Пункты 3—5 определяют понятие равенства двух замыканий, которое справедливо, в том числе, и для указателей на глобальные функции. Отношение равенства двух замыканий, не покрываемое пунктами 3—5, не определено и зависит от реализации. Так например, если вложенная функция имеет пустой контекст (т.е. не ссылается на переменные, объявленные вне её), то два экземпляра данной функции, построенные независимо, в текущей реализации окажутся равными, т.к. будут представлять собой один и тот же указатель на неявно сгенерированную глобальную функцию (см. пример с функцией Plain ранее по тексту).

Внутри вложенных функций некоторые переменные в образце можно пометить знаком ^, что будет обозначать, что они не включаются в контекст, а скрывают одноимённую переменную во внешней области видимости. На данный момент эта функциональность не отлажена, может приводить к генерации синтаксически неверного кода на C++.

#### Другие отличия

Просто кратко перечислим:

* Пустые функции. В отличие от РЕФАЛа-5, функция на Простом Рефале может иметь пустое тело — такая функция приводит к аварийному останову программы при любом аргументе. Для пустых функций есть специальный синтаксис (см. 3.2.4).
* Статические ящики — см. 3.2.5.

### Что есть в РЕФАЛе-5, но нет в Простом Рефале

* **Условия.** Их нет. Будут в следующих версиях.
* **Блоки.** Их нет, и они не нужны. Функциональность блоков реализуется при помощи вложенных функций и библиотечной функции Fetch.
* **Копилка.** Есть близкое по функциональности средство — статические ящики. Несложно на основе статических ящиков реализовать библиотеку, реализующую возможности копилки.
* **Метавычисления.** Вместо метафункции Mu можно использовать косвенный вызов функций (см. 3.3.3.1), функциональность, связанная с метакодом, пока не реализована в библиотеке Простого Рефала.

### Отличия от РЕФАЛа-5 на уровне библиотеки

Встроенных в язык функций в Простом Рефале нет. Есть лишь внешние первичные функции, которые пишутся вручную на C++. Одна из библиотек таких первичных функций входит в дистрибутив Простого Рефала и располагается в файле Library.cpp. Можно сравнить библиотеку Library с библиотекой встроенных функций РЕФАЛа-5:

* Обе библиотеки предоставляют самые базовые средства программирования: преобразование атомов (литера ↔ её ASCII-код, строка цифр ↔ число), арифметические действия, ввод-вывод на консоль и в файлы.
* Преобразование идентификатора в строку литер Простым Рефалом пока не поддерживается, преобразование строки литер в идентификатор не поддерживается принципиально (все идентификаторы литерально задаются во время компиляции).
* Функция сравнения Compare возвращает не знаки арифметических действий, а знаки отношений ('<', '=', '>'), определена для любых двух термов.
* В РЕФАЛе-5 есть ограниченное число дескрипторов файлов, задаваемое целыми числами от 1 до 20. Функции открытия файлов связывают конкретные файлы с этими дескрипторами. В Простом Рефале функции открытия файлов создают новые дескрипторы, их можно создать столько, сколько позволит операционная система.
* Простой Рефал не поддерживает ввод с перфокарт (Card) и печать (Print, Prout) на АЦПУ[[6]](#footnote-6). Для чтения со стандартного ввода и вывода на стандартный вывод используются функции с нейтральными именами ReadLine и WriteLine.
* Сокращения <+ …>, <− …>, </ …> и т.д. для <Add …>, <Sub …>, <Div …> и т.д. не поддерживаются — синтаксис языка регулярнее РЕФАЛа-5.
* Арифметические функции не поддерживают неограниченный размер аргументов, атомы-числа не являются макроцифрами. При арифметическом переполнении (как и при делении на 0) результат не определён.

# Библиотека функций

В дистрибутив Простого Рефала входит стандартная библиотека языка, представленная двумя единицами трансляции — Library.cpp (далее на неё будем ссылаться как Library) и LibraryEx.sref (далее — LibraryEx). Первая содержит набор первичных функций, т. е. функций, которые невозможно выразить на Рефале: преобразования атомов, арифметические функции, ввод-вывод, вторая — набор удобных функций, написанных на Рефале, и использующих библиотеку Library. Прежде чем переходить к рассмотрению функций, входящих в библиотеку, сначала введём нотацию для описания форматов функций (про форматы функций см. (Турчин, 1989)).

## Используемые обозначения

Формат функции будем описывать следующим образом:

<ИмяФункции аргументы>

== вариант результата 1

== вариант результата 2

Где ИмяФункции — имя функции, аргументы — запись формата аргумента в виде некоего результатного выражения с переменными, но без вызовов других функций, вариант результата — один из возможных форматов результатного выражения. Переменные в формате имеют имена, описывающие смысл данной части аргумента/результата. Некоторые фрагменты результата или

В записи результата может использоваться знак |, обозначающий альтернативу, знак \* обозначающий повторение конструкции любое количество раз, знак + — один и более раз, ? — опциональный элемент. Кроме того, для пояснения структуры отдельных переменных, входящих в формат аргумента или результата функции, может использоваться следующий вариант БНФ:

переменная ::= некоторое выражение | некоторое выражение 2 | …

## Базовая библиотека (Library)

### Add, Sub, Mul, Div, Mod

<Add s.X s.Y>

<Sub s.X s.Y>

<Mul s.X s.Y>

<Div s.X s.Y>

<Mod s.X s.Y>

== s.Result

s.X, s.Y, s.Result ::= number

Арифметические функции. Выполняют, соответственно, сложение, вычитание, умножение, деление, вычисление остатка двух чисел. Возвращаемый результат — атом-число. При арифметическом переполнении или делении на ноль результат не определён.

### WriteLine

<WriteLine e.Expr> == пусто

Функция осуществляет запись в стандартный вывод некоторого результатного выражения, после которого выводит символ перевода на новую строку. Внутри e.Expr допустимы термы любого типа. Литеральные символы выводятся как есть, числа выводятся в десятичной системе счисления, для указателей на функции и идентификаторов выводится имя данной сущности, причём после последних трёх принудительно ставится пробел, скобки (круглые и квадратные) выводятся как есть, файл выводится как \*XXXX…X, где XXXX…X — шестнадцатеричное представление указателя, для замыканий в фигурных скобках выводится указатель на сгенерированную глобальную функцию + контекст данной функции.

### ReadLine

<ReadLine> == e.Chars 0?

Функция считывает символы из стандартного ввода, пока не обнаружит перевод строки, либо конец файла. Перевод строки функция не возвращает. Если обнаружился конец файла, функция завершает строку числом 0.

### FOpen

<FOpen s.Mode e.FileName> == s.FileHandle

s.Mode ::= 'r' | 'w'

Данная функция пытается открыть файл с данным именем в указанном режиме (соответственно, чтение либо запись). Если e.FileName пустое, либо содержит что-либо отличное от литеральных символов, происходит аварийный останов. Если файл открыть не удалось (нижележащий fopen() вернул NULL), то тоже происходит аварийный останов.

### FClose

<FClose s.FileHandle> == пусто

Закрывает указанный дескриптор файла. Если нижележащий fclose() вернул EOF, то происходит аварийный останов.

### FWriteLine

<FWriteLine s.FileHandle e.Expr> == s.FileHandle

Аналогична функции WriteLine, только запись производит в s.FileHandle.

### FReadLine

<FReadLine s.FileHandle> == s.FileHandle e.Chars 0?

Аналогична функции ReadLine, только чтение производит из s.FileHandle.

### Arg

<Arg s.Number> == пусто | char+

Возвращает n-й аргумент командной строки. Если номер аргумента превышает число аргументов, то возвращается пустое выражение.

*Примечание.* Библиотека Library не даёт возможность определить число аргументов командной строки, т. к. пустой аргумент — вполне законная конструкция как на Windows, так и на Linux. Можно вызвать программу с командной строкой prog 1 "" 3 — три аргумента, второй из них пустой.

### ExistFile

$EENUM True, False;

<ExistFile e.FileName> == True | False

Проверяет существование данного файла. Проверка на существование осуществляется путём попытки вызова fopen(filename, "r"), если функция вернула не NULL, то возвращается True, в противном случае — False.

### GetEnv

<GetEnv e.VarName> == e.VarValue

Функция возвращает значение переменной среды с именем e.VarName. Если такая переменная не определена, возвращается пустое выражение.

### Exit

<Exit s.RetCode>

Функция завершает программу с кодом возврата s.RetCode. Очевидно, ничего не возвращает.

### System

<System e.Command> == пусто

Функция выполняет указанную команду оболочки.

### IntFromStr

$EENUM Success, Fails;

<IntFromStr e.Text>  
 == Success s.Number e.Rest  
 == Fails e.Text

Функция пытается прочитать от начала текста e.Text последовательность литер, представляющих собой десятичную запись числа. Если e.Text начинается с последовательности десятичных цифр, то функция возвращает Success, целое число, соответствующее данной последовательности и остаток строки. Если же e.Text не начинается с последовательности десятичных цифр, то функция возвращает Fails и свой аргумент.

### StrFromInt

<StrFromInt s.Number> == e.Text

Функция преобразует число в последовательность десятичных цифр.

### Chr, Ord

<Chr s.Number> == s.Char

<Ord s.Char> == s.Number

Две взаимно-обратные функции. Первая возвращает литеральный символ с ASCII-кодом s.Number, вторая — для данного символа возвращает его ASCII-код.

### SymbCompare

<SymbCompare s.Left s.Right> == '<' | '=' | '>'

Функция сравнивает два атома и возвращает результат сравнения. Атомы различных типов ранжируются следующим образом: число > литеральный символ > функция > идентификатор > файл. Замыкания с контекстом не поддерживаются. Числа сравниваются естественным образом, для символов сравниваются их ASCII-коды (в зависимости от компилятора C++, где тип char может быть как signed, так и unsigned, результат сравнения этой функцией может быть разным), функции и идентификаторы сравниваются по именам и упорядочиваются в лексиконографическом порядке, файлы сравниваются как указатели в C++ (сравнение для двух указателей типа void стандартом C++ разрешается, но зависит от реализации).

### SymbType

$EENUM TypeNumber, TypeCharacter, TypeFunction, TypeFile, TypeIdentifier;

<SymbType s.Atom>  
 == TypeNumber | TypeCharacter | TypeFunction | TypeIdentifier | TypeFile

Возвращает тип атома. Попытка вызвать функцию с замыканием приведёт к аварийному останову.

## Расширенная библиотека (LibraryEx)

Эта библиотека содержит набор функций, написанных на Рефале и предназначенных для упрощения процесса программирования. В неё входит ряд функций высшего порядка (типа Map или Reduce), а также удобные обёртки над функциями библиотеки Library (такие как Inc или LoadFile).

Для пояснения семантики ряда функций помимо формата будет приводиться и исходный код самих функций.

## LibraryEx: надстройки над функциями базовой библиотеки и другие полезные функции

### LoadFile

<LoadFile e.FileName> == (e.Line)\*

Загружает файл в поле зрения, каждая строка файла заворачивается в отдельный скобочный терм. Если файл заканчивается на пустую строку, она игнорируется. Для чтения файла используется функция FReadLine.

### SaveFile

<SaveFile (e.FileName) (e.Line)\*> == пусто

Сохраняет последовательность строк (e.Line)\* в файле с заданным именем. Для записи строк используется функция FWriteLine.

### Inc, Dec

<Inc s.Number> == s.Number  
<Dec s.Number> == s.Number

$ENTRY Inc {

s.Num = <Add s.Num 1>;

}

$ENTRY Dec {

s.Num = <Sub s.Num 1>;

}

Функции увеличивают и уменьшают число на единицу. Запись <Inc s.Number> компактнее и выразительнее, чем <Add s.Number 1>.

### FastIntFromStr

<FastIntFromStr e.Text> == s.Number

Преобразует последовательность десятичных цифр в число. Если e.Text не является последовательностью десятичных цифр, происходит аварийный останов.

### ArgList

<ArgList> == (e.Arg)+

Возвращает аргументы командной строки <Arg 0> <Arg 1> … до первого пустого значения, возвращённого функцией Arg. Если среди аргументов программы присутствует пустой аргумент, аргументы после него будут проигнорированы.

### Compare

<Compare t.Left t.Right> == '<' | '=' | '>'

Сравнивает два терма. Термы ранжируются следующим образом: скобочный терм > число > литеральный символ > функция > идентификатор > файл. Замыкания с контекстом и абстрактные типы данных не поддерживаются. Скобочные термы сравниваются рекурсивно в лексиконографическом порядке. Атомы между собой сравниваются также, как и для функции SymbCompare.

### Compare-T

<Compare-T t.Left t.Right>  
 == s.Relation t.Left t.Right  
s.Relation ::= '<' | '=' | '>'

Сравнивает два терма и возвращает помимо отношения термов, также сам аргумент (суффикс -T означает, что функция прозрачная (transparent)). Сравнение выполняет аналогично Compare.

### Type, Type-T

$EENUM TypeBracket;

<Type t.Term> == s.Type  
<Type-T e.Expr> == s.Type e.Expr  
  
s.Type ::= TypeNumber | TypeCharacter | TypeFunction | TypeIdentifier | TypeFile  
 | TypeBracket | '\*'

Функция Type определяет тип терма. Замыкания с контекстом и абстрактные типы данных не поддерживаются. Функция Type-T определяет тип первого терма выражения e.Expr, если выражение пустое, то в качестве типа возвращает '\*'.

### Trim

<Trim e.String> == e.String’

Функция удаляет пробелы и символы табуляции в начале и в конце строки.

## LibraryEx: функции высших порядков

### Понятие функтора, Apply

Большинство функций высших порядков, представленных в библиотеке, оперирует таким объектом, как функтор:

t.Functor ::= s.Closure | (t.Functor e.BoundedArgs)

Функтор может быть либо замыканием, либо скобочным термом, первым элементом которого является функтор. Вызов функтора осуществляется функцией Apply, которая разворачивает скобочные термы до тех пор, пока первым термом не будет атом, который затем и вызывает:

<Apply t.Function e.Args> ≈≈ <t.Function e.Args>

t.Function ::=

s.Closure | (t.Function e.BoundedArgs)

$ENTRY Apply {

s.Fn e.Argument = <s.Fn e.Argument>;

(t.Closure e.Bounded) e.Argument =

<Apply t.Closure e.Bounded e.Argument>;

}

Функции высших порядков (кроме функции Y), определённые ниже, используют функцию Apply для вызова функторов.

Исторически функторы использовались для имитации функциональности замыкания с контекстом: писалась глобальная функция, указатель на неё «связывался» с некоторым значением, которое играло роль контекста. Функция CardProd, рассмотренная ранее, с использованием глобальных функций и функторов выглядела бы так:

// Функция CardProd вычисляет декартово произведение двух множеств  
// <CartProd ('a' 'b' 'c') (1 2)>[[7]](#footnote-7)  
// == ('a' 1) ('a' 2) ('b' 1) ('b' 2) ('c' 1) ('c' 2)

$FORWARD CartProd-By-tA, CartProd-By-tB;

$ENTRY CartProd {

(e.SetA) (e.SetB) =

<Map (CardProd-By-tA e.SetB) e.SetA>;

}

CartProd-By-tA {

e.SetB t.A =

<Map (CartProd-By-tB t.A) e.SetB>;

}

CartProd-By-tB {

t.A t.B = (t.A t.B);

}

Как правило, удобнее использовать вложенные функции вместо функторов, однако функторы могут оказаться удобными в тех случаях, когда действие сводится исключительно к вызову *имеющейся* функции с фиксированной частью аргумента. Например, вызов

<Map (Add 3) e.Numbers>

лаконичнее и читабельнее, чем

<Map { s.Number = <Add 3 s.Number>; } e.Numbers>

Другой пример будет приведён в примере с функцией Fetch.

### Map

<Map t.Func t.Elem\*> == <t.Func t.Elem>\*

$ENTRY Map {  
 t.Fn t.Next e.Tail = <Apply t.Fn t.Next> <Map t.Fn e.Tail>;  
  
 t.Fn = ;  
}

Функция Map последовательно применяет функтор к каждому терму своего аргумента.

### Reduce

<Reduce t.Func t.Acc t.Elem\*> == t.Acc”,  
<t.Func t.Acc t.Elem> == t.Acc’

$ENTRY Reduce {  
 t.Fn t.Acc t.Next e.Tail =  
 <Reduce  
 t.Fn <Apply t.Fn t.Acc t.Next> e.Tail  
 >;  
  
 t.Fn t.Acc = t.Acc;  
}

Функция Reduce сворачивает каждый элемент последовательности (в порядке слева-направо) с термом-аккумулятором с применением данного функтора. Пример:

<Reduce Add 0 e.Numbers> // вычисляет сумму чисел в e.Numbers  
<Reduce Max 0 e.Numbers> // вычисляет максимальное число  
 // (определение Max см. в примере к Fetch)

### Fetch

<Fetch e.Argument t.Func> == <t.Func e.Argument>

$ENTRY Fetch {  
 e.Argument t.Function =  
 <Apply t.Function e.Argument>;  
}

Функция служит для повышения читабельности вызовов вложенных функций. Пример:

|  |  |
| --- | --- |
| Вызов с Fetch  Max {  t.X t.Y =  <Fetch  <Compare t.X t.Y> {  '<' = t.Y;  s.Other = t.X;  }  >; } | Вызов без Fetch  Max {  t.X t.Y =  <  {  '<' = t.Y;  s.Other = t.X;  }  <Compare t.X t.Y>  >; } |

Видно, что применение функции Fetch упрощает восприятие кода. Сначала описываются данные, которые требуется обработать (результат сравнения двух термов), а затем обработка этих данных. Без функции Fetch сначала описывается некая функция, и уже потом, когда взгляд переходит к аргументу, удаётся понять, для чего она описывалась.

Функцию Fetch можно сравнить с операторами if и switch-case императивных языков программирования: сначала вычисляется «выбирающее» выражение, затем, в зависимости от результата, выполняется та или иная ветвь.

Знатоки РЕФАЛа‑5 заметят, что функция Fetch эквивалентна блоку, with-конструкции, с той лишь разницей, что блок в РЕФАЛе‑5 является частью синтаксиса предложения, а Fetch — идиома, построенная на основе других средств языка (вложенные функции, косвенный вызов) и находится внутри результатной части. Та же функция Max на РЕФАЛе‑5 выглядеть будет вот так:

\* РЕФАЛ‑5  
Max {  
 s.X s.Y, <Compare s.X s.Y>: {  
 '−' = s.Y;  
 s.Other = s.X;  
 }  
}

*Примечание.* В РЕФАЛе‑5 функция Compare может применяться только к числам и возвращает знак разности первого и второго аргумента: '−', '0' или '+'.

Вторым аргументом Fetch является не замыкание, а функтор, поэтому можно писать, например, такой код:

// Прочитать файл и удалить начальные и конечные пробелы.  
<Fetch  
 <LoadFile e.FileName>  
 (Map {  
 (e.Line) = (<Trim e.Line>);  
 })  
>

Особую мощь функция Fetch приобретает в сочетании с функцией Seq (см. далее).

### Y-комбинатор

<Y { s.Loop = { рекурсивная функция с вызовом s.Loop }; }>  
 == результат рекурсивной функции

$ENTRY Y {  
 s.Func = { e.Arg = <<s.Func <Y s.Func>> e.Arg>; };  
}

Прежде чем перейти к описанию функции Y, необходимо сделать небольшое лирическое отступление. Вложенные функции позволяют записать некоторую служебную функцию в том месте, где она используется. Однако, эта служебная функция не может быть рекурсивной — чтобы вызвать себя, надо сослаться на себя, а имени у неё нет. Как же быть? Если пополнить язык именованными вложенными функциями (по аналогии с Refal‑7, (Скоробогатов & Чеповский, 2006)), тогда Простой Рефал перестанет быть Простым. Есть другой способ — передать внутрь замыкания само это замыкание. Функция Y именно это и делает.

Своё название функция получила из мира математики — из комбина́торной логики. В рамках этого раздела математики вводится набор особых функций высшего порядка — комбинаторов, с применением которых можно задавать любой вычислительный процесс. Один из таких комбинаторов — комбинатор неподвижной точки (Y-комбинатор) служит для превращения нерекурсивной функции в рекурсивную. Фактически функция Y и реализует данный комбинатор.[[8]](#footnote-8) Далее по тексту функцию Y будем называть Y‑комбинатором.

Y‑комбинатор принимает в качестве аргумента функцию, принимающую единственный аргумент — s‑переменную и возвращающую замыкание, которое следует сделать рекурсивным, возвращает функцию, которая является рекурсивным эквивалентом функции, возвращаемой аргументом. При вычислении Y‑комбинатора функция-аргумент комбинатора вызывается и получает аргументом функцию, которую возвращает в результатном выражении, но уже сделанную рекурсивной — функционально то же самое, что и возвращает Y‑комбинатор.

Поясню на примере, т. к. без примера здесь не разобраться. Допустим, мы хотим написать функцию, итеративно вычисляющую (т.е. с применением хвостовой рекурсии) факториал:

$FORWARD DoFact;

Fact {  
 0 = 1;  
 s.N = <DoFact s.N 1 1>;  
}

DoFact {  
 s.N s.Prod s.N = <Mul s.Prod s.N>;  
 s.N s.Prod s.K = <DoFact s.N <Mul s.Prod s.K> <Inc s.K>>;  
}

Здесь функция DoFact является служебной, однако просто так её записать внутри правой части функции Fact нельзя — функция является рекурсивной, а вложенные функции должны быть безымянными.

Если бы программа писалась на Refal‑7, то можно было бы записать вот так:

Fact {  
 0 = 1;  
  
 s.N =  
 <Fetch  
 1 1  
 $func Loop {  
 s.Prod s.N = <Mul s.Prod s.N>;  
 s.Prod s.K = <Loop <Mul s.Prod s.K> <Inc s.K>>;  
 }  
 >;  
}

Вложенная функция имеет имя Loop, поэтому может вызывать себя рекурсивно. На Простом Рефале на помощь приходит Y‑комбинатор:

Fact {  
 0 = 1;  
  
 s.N =  
 <Fetch  
 1 1  
 <Y  
 { // № 1  
 **s.Loop** = { // № 2  
 s.Prod s.N = <Mul s.Prod s.N>;  
 s.Prod s.K = <**s.Loop** <Mul s.Prod s.K> <Inc s.K>>;  
 };  
 }  
 >  
 >;  
}

Y‑комбинатор превращает функцию, помеченную как № 2 в рекурсивную и её же возвращает (эту функцию вызывает функция Fetch). Внутри функции № 2 к себе рекурсивной можно обращаться через переменную s.Loop, которая выполняет роль имени функции.

*Примечание.* Y‑комбинатор — единственная функция высшего порядка в библиотеке, которая не работает с функторами.

*Примечание 2.* Если вам непонятно, как работает Y‑комбинатор, относитесь к нему как к ~~магии~~ особому синтаксису для записи рекурсивных вложенных функций.

### MapReduce

<MapReduce t.Func t.Acc t.Elem\*> == t.Acc” e.Mapped\*,  
где <t.Func t.Acc t.Elem> == t.Acc’ e.Mapped

Функция MapReduce сочетает в себе свойства функций Map (преобразование каждого элемента последовательности функтором) и Reduce (свёртка каждого элемента последовательности с аккумулятором). Функтор принимает два терма: текущее значение аккумулятора и очередной элемент, возвращает модифицированное значение аккумулятора и некоторое объектное выражение — результат трансформации элемента. Функция MapReduce применяет функтор последовательно к каждому элементу и очередному значению аккумулятора, возвращает результирующее значение аккумулятора и результаты трансформации всех элементов. Приведём пример — напишем функцию, которая загружает файл и нумерует в нём строчки:

LoadNumerated {  
 e.FileName =  
 <DelAccumulator // См. далее  
 <MapReduce  
 {  
 s.Number (e.Line) =  
 <Inc s.Number> (s.Number e.Line);  
 }  
 1 // начальное значение счётчика  
 <LoadFile e.FileName>  
 >  
 >;  
}

Вложенная функция принимает два терма: номер текущей строки и строку из файла, возвращает номер следующей строки (на единицу больше) и преобразованную строку файла, первым термом содержащую число — её номер. Значение аккумулятора после анализа нам не нужно, оно отбрасывается при помощи функции DelAccumulator.

### UnBracket

<UnBracket (e.Expr)> == e.Expr

$ENTRY UnBracket {  
 (e.Expr) = e.Expr;  
}

Снимает скобки с выражения. Может быть полезна в сочетании с другими функциями высшего порядка.

### DelAccumulator

<DelAccumulator t.Acc e.Data> == e.Data

$ENTRY DelAccumulator {  
 t.Acc e.Tail = e.Tail;  
}

Функция DelAccumulator отбрасывает первый терм от своего аргумента, используется в сочетании с функцией MapReduce — когда значение аккумулятора по завершении вычислений стало уже не нужным.

### Seq

<Seq t.Func1 t.Func2 … t.FuncN> ==  
 { e.Arg = <t.FuncN … <t.Func2 <t.Func1 e.Arg>>…>; }

$ENTRY Seq {  
 t.Func = t.Func;  
  
 t.Func e.Funcs =  
 {  
 e.Arg = <Fetch <Apply t.Func e.Arg> <Seq e.Funcs>>;  
 };  
  
 /\* пусто \*/ = { e.Arg = e.Arg; };  
}

Функция Seq возвращает композицию функторов, переданных в качестве аргумента. Как правило, используется в сочетании с функцией Fetch следующим образом:

<Fetch  
 аргумент  
 <Seq  
 {  
 образец =  
 результат 1;  
 }  
 {  
 образец 1 =  
 результат 2;  
 }  
 {  
 образец 2 =  
 результат 3;  
 }  
 {  
 образец 3 =  
 результат цепочки  
 }  
 >  
>

Одинаковыми цветами здесь показаны выражения с общим форматом — видно, что они располагаются по соседству. При выполнении цепочки функций, перечисленных в аргументе Seq, выражение как бы «просеивается» через эти функторы.

## Пример кода, использующий функции высших порядков

Следующая функция осуществляет синтаксический анализ .ini-файла. Даётся в иллюстративных целях, к примеру, вопросы корректности имён ключей и содержимого параметров не рассматриваются. Определения идентификаторов и директивы $EXTERN для внешних функций опущены.

Полный текст примера можно найти в каталоге doc/example настоящего дистрибутива.

/\*

<LoadIniFile e.FileName>

== t.IniFile t.ErrorMessage\*

t.IniFile ::= (t.Section\*)

t.Section ::= (s.LineNumber t.SectionName t.Parameter\*)

t.SectionName ::= #General | (e.Name)

t.Parameter = (s.LineNumber (e.Name) e.Value)

t.ErrorMessage ::=

(s.LineNumber #BadLine) |

(s.LineNumber #EmptyParameterName) |

(s.LineNumber #ReassignParameter e.ParameterName)

\*/

$ENTRY LoadIniFile {

e.FileName =

<Fetch

<LoadFile e.FileName>

<Seq

// Нумеруем строки файла  
 (MapReduce

{

s.LineNumber (e.Line) =

<Inc s.LineNumber> (s.LineNumber e.Line);

}

1

)

DelAccumulator

// Удаляем комментарии и начальные и концевые пробелы  
 (Map {

(s.LineNumber e.Line ';' e.Comment) = (s.LineNumber <Trim e.Line>);

(s.LineNumber e.Line) = (s.LineNumber <Trim e.Line>);

})

// Удаляем пустые строки  
 (Map {

(s.LineNumber) = /\* пусто \*/;

(s.LineNumber e.Line) = (s.LineNumber e.Line);

})

// Парсим строки  
 (Map {

(s.LineNumber '[' e.SectionName ']') =

(s.LineNumber #Section <Trim e.SectionName>);

(s.LineNumber e.Parameter '=' e.Value) =

(s.LineNumber #Parameter (<Trim e.Parameter>) <Trim e.Value>);

(s.LineNumber e.Other) = (s.LineNumber #BadLine);

})

// Группируем параметры в секции  
 (Reduce

{

(

e.Sections-B (s.SectionPos (e.Section) e.Params) e.Sections-E

(e.Errors)

)

(s.LineNumber #Section e.Section) =

(

e.Sections-B e.Sections-E (s.SectionPos (e.Section) e.Params)

(e.Errors)

);

(e.Sections (e.Errors)) (s.LineNumber #Section e.Section) =

(e.Sections (s.LineNumber (e.Section)) (e.Errors));

(e.Sections (e.Errors)) (s.LineNumber #Parameter () e.Value) =

(e.Sections

(e.Errors (s.LineNumber #EmptyParameterName))

);

(e.Sections

(s.SectionPos (e.LastSection)

e.Params-B (s.ParamPos (e.Param) e.OldValue) e.Params-E

)

(e.Errors)

)

(s.LineNumber #Parameter (e.Param) e.NewValue) =

(e.Sections

(s.SectionPos (e.LastSection)

e.Params-B (s.ParamPos (e.Param) e.NewValue) e.Params-E

)

(e.Errors (s.LineNumber #ReassignParameter e.Param))

);

(e.Sections (s.SectionPos (e.LastSection) e.Params) (e.Errors))

(s.LineNumber #Parameter (e.Param) e.Value) =

(e.Sections

(s.SectionPos (e.LastSection)

e.Params (s.LineNumber (e.Param) e.Value)

)

(e.Errors)

);

(e.Sections (e.Errors)) (s.LineNumber #BadLine) =

(e.Sections (e.Errors (s.LineNumber #BadLine)));

}

((1 #General) (/\* ошибки \*/))

)

// Формируем результат функции  
 {

(e.Sections (e.Errors)) =

(e.Sections) e.Errors;

}

>

>;

}

# Компилятор Простого Рефала: реализация

В дистрибутив компилятора входит довольно много файлов (их перечислению посвящён раздел 9), однако для разработки программ достаточно иметь только следующие пять: srefc.exe (на unix-like — srefc), refalrts.h, refalrts.cpp, Library.cpp и LibraryEx.sref:

* srefc.exe — исполнимый файл компилятора.
* refalrts.h — компонент рантайма. В refalrts.h находятся объявления функции и определения структур данных, используемых в сгенерированном коде на C++. Сгенерированный код всегда начинается с #include "refalrts.h", поэтому путь к этому файлу должен быть в списке путей поисках заголовочных файлов компилятора C++ (обычно, опция **-I** компилятора).
* refalrts.cpp — компонент рантайма. refalrts.cpp содержит систему поддержки времени выполнения: определения функций, объявленных в refalrts.h, структуру данных поля зрения, реализацию абстрактной рефал-машины, менеджер памяти и функцию main() языка C++.
* Library.cpp — библиотека Library, см. 4.2.
* LibraryEx.sref — библиотека LibraryEx, см. 4.3.

## Ключи командной строки, вызов компилятора

Синтаксис командной строки:

srefc [-c компилятор\_C++] [ -d путь\_поиска …] имя\_модуля[.sref|.cpp]…

srefc @конфигурационный\_файл

Опции:

* **-c компилятор\_C++** — префикс командной строки вызова компилятора C++. Если эта опция отсутствует, то компилятор просто произведёт компиляцию в код на C++. Если данная опция присутствует, то после завершения компиляции всех исходных текстов, будет вызван компилятор C++ путём приписывания к префиксу всех имён .cpp-файлов: как являющихся результатом трансляции из .sref-файлов, так и .cpp-файлов, указанных в командной строке компилятора.
* **-d путь\_поиска** (опция может повторяться) — задаёт путь поиска единиц трансляции, указанных в командной строке. По умолчанию поиск начинается с текущего каталога, затем во всех каталогах, указанных в опциях **-d**.
* **имя\_модуля[.sref|.cpp]** — задаёт имя одной из единиц трансляции. Если расширение .sref или .cpp отсутствует, то сначала проверяется наличие файла с данным именем и расширением .sref, если такой файл отсутствует, то с расширением .cpp. Если файл не найден, продолжается поиск в следующем пути поиска.
* **@конфигурационный\_файл** — файл, содержащий аргументы командной строки по одному на строчку. Поскольку опции **-c** и **-d** требуют после себя отдельного аргумента командной строки, они должны располагаться на отдельных строчках. Выглядит странно, возможно, в будущих версиях будет исправлено.

Рассмотрим пример. Пусть в текущем каталоге находятся файлы main.sref, http-parser.sref, sockets.cpp и build.prj, в каталоге C:\Library находятся файлы Library.cpp, LibraryEx.sref, refalrts.h и refalrts.cpp. Компилятор вызывается следующим образом:

srefc @build.prj

Файл build.prj содержит следующий текст:

-c  
g++ -IC:\Library -omyprog.exe -lwininet

-d  
C:\Library

main  
http-parser.sref  
sockets  
Library  
LibraryEx  
refalrts

Префикс командной строки содержит вызов компилятора g++, со следующими опциями: путь поиска заголовочных файлов C:\Library, компиляция в файл myprog.exe, подключение библиотеки wininet. Путь поиска единиц трансляции — C:\Library. Единицы трансляции: main, http-parser.sref, sockets, Library, LibraryEx, refalrts.

Рассмотрим, как будет выполняться этот вызов компилятора:

1. Сначала компилятор произведёт поиск всех единиц трансляции:

* main — расширения нет, поэтому будет осуществляться поиск файла main.sref или main.cpp в каждом из каталогов поиска. Первый каталог поиска — текущий каталог. Компилятор ищет main.sref в текущем каталоге и находит его.
* http-parser.sref — расширение есть, значит будет осуществляться поиск только файла http-parser.sref. Файл обнаруживается в первом каталоге поиска — в текущем каталоге.
* sockets — расширения нет. В первом каталоге поиска — текущем каталоге файл sockets.sref не обнаруживается, зато обнаруживается sockets.cpp. Простой Рефал компилировать его не будет, однако файл будет передан компилятору C++.
* Library — расширения нет. В текущем каталоге файлы Library.sref и Library.cpp отсутствуют. В следующем каталоге поиска — в C:\Library отсутствует файл Library.sref, но присутствует Library.cpp. Файл C:\Library\Library.cpp будет передан компилятору C++.
* LibraryEx — расширения нет. В текущем каталоге файлы LibraryEx.sref и LibraryEx.cpp отсутствуют. В следующем каталоге поиска — в C:\Library присутствует файл LibraryEx.sref. Файл C:\Library\LibraryEx.sref будет откомпилирован Простым Рефалом, получившийся файл будет передан компилятору C++.
* refalrts — его поиск аналогичен поиску Library.

1. Затем компилятор переведёт файлы main.sref, http-parser.sref и C:\Library\LibraryEx.sref на С++.
2. На третьем этапе файлы sockets.cpp, C:\Library\Library.cpp, C:\Library\refalrts.cpp и файлы, получившиеся в результате трансляции на предыдущем этапе, будут переданы компилятору C++ путём формирования командной строки из префикса и имён файлов.

## Макросы препроцессора C++

Опции командной строки, перечисленные в предыдущем разделе, не допускают тонкой настройки кодогенерации и поведения во время выполнения (типа оптимизаций или средств отладки). Однако, некоторые такие возможности реализация предоставляет — для управления ими используются макросы препроцессора C++. Для глобальной установки макроса компиляторы C++, как правило, используют опцию -D, после которой указывается либо просто имя определяемого макроса (если макрос затем проверяется директивой #ifdef), либо имя и значение, разделённые знаком =. Пример:

g++ -otest.exe -DTARGET\_WINDOWS -DPAGE\_SIZE=4096 test.cpp

Здесь устанавливаются макросы TARGET\_WINDOWS и PAGE\_SIZE, причём последний получает значение 4096.

Сгенерированный код поддерживает только один макрос — **INTERPRET**. Код построения правых частей предложений транслируется в C++ двумя разными способами: в режиме прямой кодогенерации (код представляет собой последовательность вызовов функций и проверок возвращаемого значения) и интерпретации (для правой части генерируется массив команд и вызов функции интерпретации этого массива). В целевом файле присутствуют оба варианта, по умолчанию выполняется код, построенный первым способом, для компиляции второго варианта следует использовать уже упомянутый макрос INTERPRET. Исполнимый файл, скомпилированный в режиме интерпретации, выполняется медленнее приблизительно на 15 % и имеет размер приблизительно на 30 % меньше, чем в режиме прямой кодогенерации.

Рантайм поддерживает конфигурацию несколькими макросами, в основном это включение/выключение отладочных функций. Здесь они будут только перечислены, об их использовании будет рассказываться в следующем разделе:

* **DUMP\_FILE** — определяет имя файла, в который будет осуществляться отладочный вывод, параметр должен быть строковым литералом языка C. Пример:

-DDUMP\_FILE="\_\_dump.txt"

* **SHOW\_DEBUG** — целочисленный параметр, определяет номер шага, после которого на каждом шаге рефал-машины выполнять дамп поля зрения.
* **DUMP\_FREE\_LIST** — когда определён, печатает не только дамп поля зрения, но и дамп списка свободных узлов. Может быть полезен при отладке внешних функций и рантайма.
* **MEMORY\_LIMIT** — целочисленный параметр. Если определён, то при попытке распределить память для узлов больше, чем указано в параметре, рефал-машина аварийно остановится с сообщением о нехватке памяти.
* **MODULE\_REFAL** — этот же рантайм используется компилятором Модульного Рефала, однако, используемые Модульным Рефалом структуры данных незначительно отличаются от структур Простого Рефала. При компиляции в Простом Рефале определять не нужно.
* **DONT\_PRINT\_STATISTICS**. При завершении программа по умолчанию выдаёт (на stderr) статистику о времени своей работы: общие затраты времени и памяти, а также затраты времени на выполнение отдельных видов вычислений (внешние функции, левые, правые части и т.д.). Макрос отключает вывод этой статистики.

## Утилита srmake

Для компиляции довольно больших проектов довольно утомительно перечислять в командной строке (либо в файле .prj) компилятора все необходимые единицы трансляции. В частности, при добавлении нового файла также придётся править сценарий сбоки, либо файл с опциями командной строки. А хотелось бы иметь возможность обнаруживать все требуемые файлы автоматически.

Для этой цели была написана утилита srmake — она в качестве параметра получает имя одного из исходных файлов, находит все зависимые и вызывает компилятор Простого Рефала для сборки всего проекта. Но для этого в файлах проекта надо указывать, от каких других файлов они зависят — для этого используется специальный комментарий //FROM. Синтаксис следующий:

//FROM ИмяФайла

Расширение у файла (.cpp или .sref) можно опустить. Утилита srmake сканирует предложенный файл в поисках таких комментариев, затем ищет файлы, указанные после //FROM, рекурсивно сканирует и их, до тех пор, пока все необходимые единицы трансляции не найдутся. Утилита srmake сканирует не только файлы на Простом Рефале, но и файлы на C++, поэтому допустимо писать комментарий //FROM и внутри них.

Хорошей практикой является помещение комментария //FROM перед директивой $EXTERN — не только сообщает утилите srmake о зависимости, но и повышает читабельность кода:

//FROM Library

$EXTERN WriteLine;

//FROM LibraryEx

$EXTERN Inc, Dec, LoadFile;

//FROM socket

$EXTERN Accept, Bind, Connect, Listen;

В LibraryEx есть ссылка на Library, в Library есть ссылка на refalrts, поэтому если используются стандартные библиотеки, то модуль рантайма также будет найден. Если ваш проект прямо или косвенно не использует Library, то ссылку на refalrts нужно будет добавить вручную.

Синтаксис командной строки утилиты srmake напоминает синтаксис самого компилятора srefc:

srmake [-s компилятор\_srefc] -c компилятор\_C++ [ -d путь\_поиска … ] имя\_модуля[.sref|.cpp]

Опции:

* **-s компилятор\_srefc** — путь к компилятору Простого Рефала. Если опция не указана, то вызывается srefc, т. е. предполагается, что компилятор имеет имя srefc.exe (srefc на unix-like) и находится в текущем каталоге (только на Windows) или указан в переменной PATH.
* **-с компилятор\_C++** — смысл тот же, что и для компилятора Простого Рефала.
* **-d путь\_поиска** — смысл тот же, что и для компилятора Простого Рефала.
* **имя\_модуля[.sref|.cpp]** — имя одного из файлов исходного текста, параметр должен быть единственным.

# Отладка программ на Простом Рефале

Программ без ошибок не бывает, поэтому хорошая реализация языка программирования должна иметь средства, облегчающие поиск и исправление ошибок. В этом разделе мы расскажем о способах отладки программ на Простом Рефале.

## Использование отладчика C++

Компилятор Простого Рефала порождает код на C++, поэтому первое, что может прийти в голову программисту — использовать отладчик для C++. Такой вариант возможен, но он имеет ряд недостатков:

* Отладка ведётся на низком уровне — на уровне сопоставления отдельных элементов выражения в образце (атомов, скобок), построения отдельных элементов результата.
* Поле зрения представляет собой двусвязный список, поэтому исследовать его в отладчике очень неудобно.
* Сгенерированные функции имеют большой объём, что снижает наглядность при чтении кода.
* Программисту необходимо понимать, как компилятор преобразует код на Рефале в C++.

Однако, этот метод имеет свою сферу применения: отладка внешних функций и рантайма.

## Средства отладки рантайма

Пошагового отладчика для Простого Рефала нет (возможно, пока нет). Зато есть посмертный отладчик и трассировщик.

При аварийном останове программы (при невозможности сопоставления, либо при недостатке памяти) рантайм осуществляет выдачу дампа всего поля зрения, по умолчанию вывод осуществляется в stderr. В этом случае выводятся:

* причина ошибки:
  + RECOGNITION IMPOSSIBLE — невозможность сопоставления,
  + NO MEMORY — недостаточно памяти,
* номер шага, на котором произошла ошибка,
* вызов функции, приведший к ошибке,
* дамп всего поля зрения.

Наиболее распространённая ошибка в программе — это невозможность сопоставления и этого отладочного вывода, как правило, достаточно для локализации и исправления ошибки. Однако, если по-прежнему остаётся непонятно, каким образом сформировалось это поле зрения, можно пройти по шагам и получить последовательность дампов до аварийного останова. Для этого надо установить макрос препроцессора SHOW\_DEBUG, присвоив ему целое число — номер шага, начиная с которого надо выводить дамп.

Дамп поля зрения, как правило, достаточно объёмен, и читать его с консоли неудобно. Можно, конечно, его перенаправить в файл средствами оболочки (prog 2>err.txt), но, чтобы постоянно так не писать, можно установить другой макрос — DUMP\_FILE, присвоив ему имя файла для отладочного вывода.

Другой вид проблем — исчерпание свободной памяти. Оно может произойти по двум причинам: либо программа требует очень много памяти (более 2 Гбайт на 32‑разрядных системах и более суммарного объёма ОЗУ и файла подкачки на 64‑разрядных системах), либо программа зацикливается и на каждой итерации добавляет новые данные в поле зрения. В первом случае нужно либо поискать более оптимальный алгоритм, либо не подавать на вход такие объёмные задачи. Во втором случае — очевидно, исправлять ошибку.

При невозможности выделить память рантайм, как и в случае невозможности сопоставления, аварийно прерывает программу и выводит дамп. Но лучше до этого не доходить. Во-первых, из-за потребления памяти компьютер начнёт тормозить и свопить, что только замедлит процесс, во-вторых, дамп огромного поля зрения будет огромен и выводиться будет очень долго.

Для удобства отладки памяти предусмотрен макрос MEMORY\_LIMIT, которому следует присвоить целое число — максимально допустимый объём памяти в узлах. Один узел поля зрения хранит в себе либо один атом, либо одну скобку; на 32‑разрядной системе узел имеет объём 20 байт, на 64‑разрядной — 40 байт. При установке этого макроса программа будет падать не при исчерпании свободной памяти (когда malloc() вернёт NULL), а при попытке распределить узлов больше, чем значение макроса. На практике MEMORY\_LIMIT в несколько миллионов хватает за глаза (компилятор Модульного Рефала откомпилирован с -DMEMORY\_LIMIT=7654321).

## Идиомы отладки

//написать про отладочную печать и чёткие форматы функций

# Интерфейс с языком C++

## Вычислительная модель

## Написание внешних функций

### Быстрый и грязный способ

### Написание функции вручную

# Установка компилятора

## Установка на Windows

## Установка на Linux (MacOS X)

# Структура каталогов дистрибутива

**От:** [Александр Коновалов](mailto:a.v.konovalov87@mail.ru)  
**Отправлено:** ‎пятница‎, ‎3‎ ‎октября‎ ‎2014‎ г. ‎10‎:‎46  
**Кому:** [Constantine Belev](mailto:const.belev@yandex.ru)

Thu, 02 Oct 2014 23:49:29 +0400 от Constantine Belev <const.belev@yandex.ru>:

Здравствуйте! Прочитал статью Сергея Юрьевича на тему модернизированного векторного представления объектных выражений. Там он также расписывает структуры данных, а также набор инструкций (это, как я понимаю, то, во что будет компилироваться Рефал-программа в данном случае). Посмотрел слайды по простому рефалу, про библиотеку времени выполнения, которая реализует объектные выражения на двунаправленных списках, и компилируется в с++ код. Насколько я понял, моя задача как раз и состоит в написании похожей библиотеки, только с rope-ами. Так вот, я пока пытаюсь понять, где что находится у компилятора Простого Рефала (Кстати, а под unix-системами он запустится? ;-) ). Пока пара вопросов нарисовалась.

В статье Скоробогатова нам интересна только первая часть — сравнение различных промежуточных представлений.  
Набор инструкций довольно сильно отличается от нашего, поскольку он написан для диалекта Refal-7 (<http://iu9.bmstu.ru/science/refal.php>).

Где что находится.

* **bootstrap/** — скомпилированные исходные тексты на Си++, откомпилировав их, можно получить exe-шник.
* **Compiler/** — исходные тексты самого компилятора.
  + srefc.sref — главный файл, содержащий функцию Go.
  + Lexer.sref — лексический анализатор.
  + Driver.sref — синтаксический анализатор.
  + Algorithm.sref — генерирует последовательность императивных команд для предложения.
  + Generator.sref — генерирует код на Си++ по последовательности императивных команд.
  + SymTable.sref, Context.sref — таблицы имён функций и переменных.
  + Прочие файлы выполняют служебные функции.
* **LexGen/** — исходные тексты генератора лексера (он нам не интересен).
* **SRLib/** — библиотеки:
  + времени выполнения — refalrts.h, refalrts.cpp,
  + примитивных функций (ввод-вывод, арифметика и др.) — Library.cpp,
  + полезных функций на Рефале — LibraryEx.sref (LibraryEx.cpp).
* **SRMake/** — утилита, подбирающая все файлы исходного текста и запускающая их компиляцию (нам не интересна).
* **srprep/** — препроцессор, отображающий Простой Рефал в Модульный Рефал,  
  использовался на нулевом шаге раскрутки, нам не нужен.
* **Тесты и заготовки/** — тесты, использовавшиеся при разработке и заготовки для примитивных функций.
* **note00?.txt** — внутренняя документация, почитать полезно.

То, что нам потребуется изменять, я подчеркнул.

Под UNIX он запускается. Проверялось под QNX и разными дистрибутивами Linux (для архитектур x86 и x64).  
Для этого надо собрать exe-шник из исходных файлов в каталоге bootstrap.

Первое - важно ли то, на чем будет написана библиотека времени выполнения (я к вопросу о том, что можно попробовать написать её на Scala/Go)? Поправьте меня, если это так. Может быть, там вообще компилятор на Рефале как раз-таки и написан, и мне для моего же блага писать нужно будет на нём =)

Не важно. Библиотека времени выполнения переписывалась и на C#. Scala я не знаю, поэтому не могу сказать,  
насколько удобно будет писать генератор кода и библиотеку для неё. Для Go можно попробовать.  
Рефал — малоизвестный язык, вряд ли на этих языках или для этих языков есть его реализации,  
по крайней мере, я о них ничего не слышал.

Насчет второго, уже не припомню, пока буду читать документацию и мануалы, так просветление быстрее должно наступить.  
  
С уважением, Константин Белёв.

=================================  
Александр Коновалов aka Маздайщик

# Известные ошибки

# Список литературы

Алешин, А. Ю., Красовский, А. Г., Романенко, С. А., & Шерстнев, В. Ю. (1991). *Система программирования РЕФАЛ-2 для IBM PC, PDP-11 и VAX-11. Руководство пользователя.* Москва.

Немытых, А. П. (2014). Лекции по языку программирования РЕФАЛ. В *Cборник трудов по функциональному языку программирования Рефал* (Т. 1, стр. 118-165). Переславль-Залесский: Изд-во "Сборник". Получено из http://refal.botik.ru/library/refal2014\_issue-I.pdf

Скоробогатов, С. Ю., & Чеповский, А. М. (2006). Язык Refal с функциями высших порядков. *Информационные технологии*(№ 9). Получено из http://iu9.bmstu.ru/science/refal.pdf

Турчин, В. Ф. (1989). *Фрейм: РЕФАЛ-5. Руководство и справочник.* Получено 16 апреля 2015 г., из Содружество «РЕФАЛ/Суперкомпиляция»: http://refal.ru/rf5\_frm.htm

1. Так расшифровывает эту аббревиатуру Немытых А. П. (Немытых, 2014), я с ним согласен. [↑](#footnote-ref-1)
2. Подробнее о целях проекта и достигнутых результатах можно прочитать в файле doc/historical/note001.txt настоящего дистрибутива. [↑](#footnote-ref-2)
3. В руководстве (Турчин, 1989) утверждается, что язык поддерживает и вещественные числа, но по факту в реализации версии PZ Oct 29 2004 вещественные числа отсутствуют. [↑](#footnote-ref-3)
4. Нотация, используемая в комментарии, объясняется в разделе 4.1. [↑](#footnote-ref-4)
5. См., однако, в 3.2.4, замечание, касающееся особенностей некоторых версий Visual Studio и ключей оптимизации. [↑](#footnote-ref-5)
6. Алфавитно-цифровое печатающее устройство [↑](#footnote-ref-6)
7. Нотация, используемая в комментарии, объясняется в разделе 4.1. [↑](#footnote-ref-7)
8. Подробнее про Y-комбинатор можно прочитать по следующим ссылкам:

   <https://ru.wikipedia.org/wiki/Комбинатор_неподвижной_точки>

   [http://www.wikiznanie.ru/ru-wz/index.php/Комбинатор\_неподвижной\_точки](http://www.wikiznanie.ru/ru-wz/index.php/Комбинатор_неподвижной_точкиD%D0%B5%D0%BF%D0%B1%82%D0%BE%D1%87%D0%BA%D0%B8)

   <http://habrahabr.ru/post/79713/> [↑](#footnote-ref-8)