

Оглавление

Об этом сборнике	10
I ML & функциональное программирование	11
1 Основы Standard ML © Michael P. Fourman	12
1.1 Введение	12
2 Programming in Standard ML'97	14
An On-line Tutorial © Stephen Gilmore	14
II Язык Prolog: логическое программирование и искусственный интеллект	15
3 Adventure in Prolog	16
Preface	16
Prolog tools	17
3.1 Getting Started	18
3.1.1 Jumping In	21
Appendix	22
4 Учебник Фишера	23
Введение	23
4.1 Установка и запуск Prolog-системы	25
4.2 Разбор примеров программ	29
4.2.1 Раскраска карт	29
4.2.2 Два определения факториала	33
4.2.3 Классическая задача “Ханойские башни”	37
4.2.4 Загрузка, редактирование, хранение программ	40
4.2.5 2.5 Negation as failure	42
4.2.6 2.6 Tree data and relations	42
4.2.7 2.7 Prolog lists and sequences	43
4.2.8 2.8 Change for a dollar	43
4.2.9 2.9 Map coloring redux	43

4.2.10	2.10 Simple I/O	43
4.2.11	2.11 Chess queens challenge puzzle	43
4.2.12	2.12 Finding all answers	43
4.2.13	2.13 Truth table maker	43
4.2.14	2.14 DFA parser	43
4.2.15	2.15 Graph structures and paths	44
4.2.16	2.16 Search	44
4.2.17	2.17 Animal identification game	44
4.2.18	2.18 Clauses as data	44
4.2.19	2.19 Actions and plans	44
4.3	Как работает <i>Prolog</i>	44
4.3.1	Деривационные деревья, выборы и унификация	44
	Унификация термов <i>Prologa</i>	47
4.3.2	3.2 Cut	49
4.3.3	3.3 Meta-interpreters in <i>Prolog</i>	54
4.4	4. Built-in Goals	54
4.4.1	4.1 Utility goals	54
4.4.2	4.2 Universals (true and fail)	54
4.4.3	4.3 Loading <i>Prolog</i> programs	54
4.4.4	4.4 Arithmetic goals	54
4.4.5	4.5 Testing types	54
4.4.6	4.6 Equality of <i>Prolog</i> terms, unification	54
4.4.7	4.7 Control	54
4.4.8	4.8 Testing for variables	54
4.4.9	4.9 Assert and retract	54
4.4.10	4.10 Binding a variable to a numerical value	54
4.4.11	4.11 Procedural negation, negation as failure	54
4.4.12	4.12 Input/output	54
4.4.13	4.13 <i>Prolog</i> terms and clauses as data	54
4.4.14	4.14 <i>Prolog</i> operators	54
4.4.15	4.15 Finding all answers	54
4.5	5. Search in <i>Prolog</i>	54
4.5.1	5.1 The A* algorithm in <i>Prolog</i>	54
4.5.2	5.2 The 8-puzzle	54
4.5.3	5.3 $\alpha\beta$ search in <i>Prolog</i>	54
4.6	6. Logic Topics	54
4.6.1	6.1 Chapter 6 notes	54
4.6.2	6.2 Positive logic	54
4.6.3	6.3 Convert first-order logic to normal form	54
4.6.4	6.4 A normal rulebase goal interpreter	54
4.6.5	6.5 Evidentiary soundness and completeness	54
4.6.6	6.6 Rule tree visualization using Java	54
4.7	7. Introduction to Natural Language Processing	54

4.7.1	7.1 Prolog grammar parser generator	54
4.7.2	7.2 Prolog grammar for simple English phrase structures	54
4.7.3	7.3 Idiomatic natural language command and question interfaces	54
4.8	8. Prototyping with Prolog	54
4.8.1	8.1 Action specification for a simple calculator	54
4.8.2	8.2 Animating the 8-puzzle (\$5.2) using character graphics	54
4.8.3	8.3 Animating the blocks mover (\$2.19) using character graphics	54
4.8.4	8.4 Java Tic-Tac-Toe GUI plays against Prolog opponent (\$5.3)	54
4.8.5	8.5 Structure diagrams and Prolog	54
	References	54
5	ASTLOG: Язык для анализа синтаксических деревьев	55
	Abstract	55
5.1	Introduction	56
5.1.1	The awk Approach	56
5.1.2	The Logic Programming Approach	58
5.2	Elements of ASTLOG	59
	Figure 1: Complete Syntax of ASTLOG	59
5.2.1	Objects	60
5.2.2	The Current Object	60
	Figure 2: Some core ASTLOG primitives	62
	Figure 3: Some primitive node and symbol predicates	63
5.2.3	Examples	63
	Figure 4: Actual ASTLOG code for follow_stmt	66
	Figure 5: Definition of flatten	67
	Figure 6: Parameterized version, flatten2	67
5.3	Higher order features	68
5.3.1	3.1 Lambdas and Applications	68
	Figure 7: Parameterized version of sametree	69
	Figure 8: Embedded Query State Primitives	70
5.3.2	Queries as Objects	70
	Figure 9: Query Accumulators qcount and qlist	71
5.4	Implementation	72
5.5	Conclusions and Future Work	73
5.6	Acknowledgements	75
	References	75
	Appendix	76
	Figure 10: Outline of astlog Operational Semantics	77
6	Warren's Abstract Machine	78
	Абстрактная машина Варрена	
	Предисловие к репринтному изданию	78
	Предисловие	79
	Реализация машины вывода на C ⁺	80

6.1	Введение	81
6.1.1	Существующая литература	81
6.1.2	Этот учебник	83
6.2	Унификация — ясно и просто	84
6.2.1	Представление термов	85
6.2.2	Компиляция \mathcal{L}_0 запросов	87
6.2.3	2.3 Compiling L programs	13
6.2.4	2.4 Argument registers	19
6.3	3 Flat Resolution 25	90
6.3.1	3.1 Facts	26
6.3.2	3.2 Rules and queries	27
6.4	4 Prolog 33	90
6.4.1	4.1 Environment protection	34
6.4.2	4.2 What's in a choice point	36
6.5	5 Optimizing the Design 45	90
6.5.1	5.1 Heap representation	46
6.5.2	5.2 Constants, lists, and anonymous variables	47
6.5.3	5.3 A note on set instructions	52
6.5.4	5.4 Register allocation	54
6.5.5	5.5 Last call optimization	56
6.5.6	5.6 Chain rules	57
6.5.7	5.7 Environment trimming	58
6.5.8	5.8 Stack variables	60
6.5.9	5.9 Variable classification revisited	69
6.5.10	5.10 Indexing	75
6.5.11	5.11 Cut	83
6.6	6 Conclusion 89	90
6.7	A Prolog in a Nutshell 91	90
6.8	B The WAM at a glance 97	90
6.8.1	B.1 WAM instructions	97
6.8.2	B.2 WAM ancillary operations	112
6.8.3	B.3 WAM memory layout and registers	117
7	An Efficient Unification Martelli/Montanary Algorithm	91
	Abstract	91
7.1	INTRODUCTION	92
7.2	UNIFICATION AS THE SOLUTION OF A SET OF EQUATIONS: A NONDETERMINISTIC ALGORITHM	93
III	Дурдом на дереве	96
8	The Tree Processing Language	97
	Defining the structure and behaviour of a tree	97

Abstract	97
Preface	98
8.1 Introduction	98
8.1.1 Compiler Construction and Abstract Syntax Trees	99
8.1.2 Problem Statement	100
8.1.3 Outline	100
IV Язык <i>bI</i>	101
9 DLR: Dynamic Language Runtime	102
10 Система динамических типов	105
10.1 sym : символ = Абстрактный Символьный Тип /AST/	105
10.2 Скаляры	107
10.2.1 str : строка	107
10.2.2 int : целое число	107
10.2.3 hex : машинное hex	107
10.2.4 bin : бинарная строка	107
10.2.5 num : число с плавающей точкой	107
10.3 Композиты	107
10.3.1 list : плоский список	107
10.3.2 cons : cons-пара и списки в <i>Lisp</i> -стиле	107
10.4 Функционалы	107
10.4.1 op : оператор	107
10.4.2 fn : встроенная/скомпилированная функция	107
10.4.3 lambda : лямбда	107
11 Программирование в свободном синтаксисе: FSP	108
11.1 Типичная структура проекта FSP: <i>lexical skeleton</i>	108
11.1.1 Настройки (g Vim	109
11.1.2 Дополнительные файлы	109
11.1.3 Makefile	110
12 Синтаксический анализ текстовых данных	111
12.1 Универсальный Makefile	111
12.2 C₊ интерфейс синтаксического анализатора	112
12.3 Минимальный парсер	112
12.4 Добавляем обработку комментариев	114
12.5 Разбор строк	115
12.6 Добавляем операторы	116
12.7 Обработка вложенных структур (скобок)	119

13 Синтаксический анализатор	121
13.1 lpp.lpp: лексер /flex/	121
13.2 урр.урр: парсер /bison/	123
V skelex: скелет программы в свободном синтаксисе	126
Структура проекта	127
Makefile	127
урр.урр: синтаксический парсер	128
lpp.lpp: лексер	130
hpp.hpp: хедеры	131
cpr.cpr: ядро интерпретатора	133
Тестирование интерпретатора	135
Комментарии	135
Скаляры и базовые композиты	135
Операторы	136
VI emLinux для встраиваемых систем	138
Структура встраиваемого микроЛinuxа	139
Процедура сборки	140
14 clock: коридорные электронные часы = контроллер умного дурдома	141
15 gambox: игровая приставка	142
VII GNU Toolchain и C₊⁺ для встраиваемых систем	143
16 Программирование встраиваемых систем с использованием GNU Toolchain [23]	144
16.1 Введение	144
16.2 Настройка тестового стенда	145
16.2.1 Qemu ARM	145
16.2.2 Инсталляция Qemu на Debian GNU/Linux	145
16.2.3 Установка кросс-компилятора GNU Toolchain для ARM	145
16.3 Hello ARM	146
16.3.1 Сборка бинарника	147
16.3.2 Выполнение в Qemu	150
16.3.3 Другие команды монитора	152
16.4 Директивы ассемблера	152
16.4.1 Суммирование массива	152
16.4.2 Вычисление длины строки	154

16.5 Использование ОЗУ (адресного пространства процессора)	155
16.6 Линкер	157
16.6.1 Разрешение символов	157
16.6.2 Релокация	159
16.7 Скрипт линкера	164
16.7.1 Пример скрипта линкера	165
16.7.2 Анализ объектного/исполняемого файла утилитой <code>objdump</code>	166
16.8 Данные в RAM, пример	167
16.8.1 RAM энергозависима (<i>volatile</i>)!	168
16.8.2 Спецификация адреса загрузки LMA	169
16.8.3 Копирование ‘.data’ в ОЗУ	170
16.9 Обработка аппаратных исключений	172
16.10 Стартап-код на Си	173
16.10.1 Стек	174
16.10.2 Глобальные переменные	176
16.10.3 Константные данные	176
16.10.4 Секция <code>.eeprom</code> (AVR8)	176
16.10.5 Стартовый код	176
16.11 Использование библиотеки Си	180
16.12 Inline-ассемблер	181
16.13 Использование ‘make’ для автоматизации компиляции	181
16.13.1 Выбор конкретной <i>цели</i>	182
16.13.2 Переменные	183
16.14 <i>Contributing</i>	183
16.15 <i>Credits</i>	183
16.15.1 <i>14.1. People</i>	183
16.15.2 <i>14.2. Tools</i>	183
16.16 <i>Tutorial Copyright</i>	183
16.17 A. ARM Programmer’s Model	183
16.18 B. ARM Instruction Set	183
16.19 C. ARM Stacks	183
17 Embedded Systems Programming in <i>C₊</i> [22]	184
18 Сборка кросс-компилятора GNU Toolchain из исходных текстов	185
APP/HW: приложение/платформа	186
Подготовка BUILD-системы: необходимое ПО	186
dirs: создание структуры каталогов	186
Сборка в ОЗУ на ramdiske	187
Пакеты системы кросс-компиляции	187
gz: загрузка исходного кода для пакетов	188
Макро-правила для автоматической распаковки исходников	189
Общие параметры для <code>./configure</code>	189
18.1 Сборка кросс-компилятора	190

18.1.1	cclibs0: библиотеки поддержки gcc	190
18.1.2	binutils0: ассемблер и линкер	191
18.1.3	gcc00: сборка stand-alone компилятора Си	193
18.1.4	newlib: сборка стандартной библиотеки libc	194
18.1.5	gcc0: пересборка компилятора Си/C ⁺	194
18.2	Поддерживаемые платформы	194
18.2.1	i386: ПК и промышленные PC104	194
18.2.2	x86_64: серверные системы	194
18.2.3	AVR: Atmel AVR Mega	194
18.2.4	arm: процессоры ARM Cortex-Mx	194
18.2.5	armhf: SoСи Cortex-A, PXA270,..	194
18.3	Целевые аппаратные системы	195
18.3.1	x86: типовой компьютер на процессоре i386+	195
19	Porting The GNU Tools To Embedded Systems	196
20	Оптимизация кода	197
20.1	PGO оптимизация	197
VIII	Микроконтроллеры Cortex-Mx	198
IX	os86: низкоуровневое программирование i386	199
	Специализированный GNU Toolchain для i386-pc-gnu	200
	MultiBoot-загрузчик	200
X	Спецификация MultiBoot	201
21	Introduction to Multiboot Specification	203
21.1	The background of Multiboot Specification	203
21.2	The target architecture	203
21.3	The target operating systems	204
21.4	Boot sources	204
21.5	Configure an operating system at boot-time	204
21.6	How to make OS development easier	204
21.7	Boot modules	205
	The definitions of terms used through the specification	205
22	The exact definitions of Multiboot Specification	207
22.1	OS image format	207
22.1.1	The layout of Multiboot header	207
22.1.2	The magic fields of Multiboot header	208
22.1.3	The address fields of Multiboot header	209

22.1.4 The graphics fields of Multiboot header	210
22.2 Machine state	210
22.3 Boot information format	211
Examples	216
History	216
Index	216

XI Технологии 217

XII Сетевое обучение 218

XIII Базовая теоретическая подготовка 219

23 Математика 220

23.1 Высшая математика в упражнениях и задачах [68]	220
Запуск Maxima и Octave в пакетном режиме	221
23.1.1 Аналитическая геометрия на плоскости	221

XIV Прочее 229

Ф.И.Атауллаханов об учебниках США и России	230
--	-----

24 Настройка редактора/IDE (g)Vim 231

24.1 для вашего собственного скриптового языка	231
--	-----

Книги 231

Книги must have любому техническому специалисту	231
Математика, физика, химия	231
Обработка экспериментальных данных и метрология	232
Программирование	232
САПР, пакеты математики, моделирования, визуализации	233
Разработка языков программирования и компиляторов	234
Lisp/Sheme	236
Haskell	236
ML	236
Электроника и цифровая техника	237
Конструирование и технология	238
Приемы ручной обработки материалов	238
Механообработка	238
Использование OpenSource программного обеспечения	239
\LaTeX	239

Математическое ПО: Maxima, Octave, GNUPLOT,	240
САПР, электроника, проектирование печатных плат	241
Программирование	241
GNU Toolchain	241
JavaScript, HTML, CSS, Web-технологии:	241
Python	241
Prolog и логическое программирование	242
Разработка операционных систем и низкоуровневого ПО	244
Базовые науки	244
Математика	244
Символьная алгебра	247
Численные методы	249
Теория игр	250
Физика	250
Химия	252
Задачники	253
Математика	253
Стандарты и ГОСТы	254
Индекс	254

Об этом сборнике

© Dmitry Ponyatov <dponyatov@gmail.com>

В этот сборник (блогбук) я пишу отдельные статьи и переводы, сортированные только по общей тематике, и добавляю их, когда у меняя очередной раз зачешется L^AT_EX.

Это сборник черновых материалов, которые мне лень компоновать в отдельные книги, и которые пишутся просто по желанию “чтобы было”. Заказчиков на подготовку учебных материалов подобного типа нет, большая часть только на этапе освоения мной самим, просто хочется иметь некое слабоупорядоченное хранилище наработок, на которое можно дать кому-то ссылку.

Сборник сверстан в микроформат¹ для просмотра на телефонах и мобильных девайсах, проверялось на удобство чтения на Alcatel Onetouch 4007D Pixi: в горизонтальной ориентации вполне читается в транспорте.

¹ А6 и менее

Часть I

ML & функциональное программирование

Глава 1

Основы Standard ML © Michael P. Fourman

<http://homepages.inf.ed.ac.uk/mfourman/teaching/mlCourse/notes/L01.pdf>

1.1 Введение

ML обозначает “MetaLanguage”: Метаязык. У Robin Milner была идея создания языка программирования, специально адаптированного для написания приложений для обработки логических формул и доказательств. Этот язык должен быть **метаязыком** для манипуляции объектами, представляющими формулы на логическом **объектном языке**.

Первый *ML* был метаязыком вспомогательного пакета автоматических доказательств Edinburgh LCF. Оказалось что метаязык Милнера, с некоторыми дополнениями и уточнениями, стал инновационным и универсальным языком программирования общего назначения. Standard ML (SML) является наиболее близким потомком оригинала, другой — CAML, Haskell является более дальним родственником. В этой статье мы представляем язык SML, и рассмотрим, как он может быть использован для вычисления некоторых интересных результатов с очень небольшим усилием по программированию.

Для начала, вы считаете, что программа представляет собой последовательность команд, которые будут выполняться компьютером. Это неверно! Представление последовательности инструкций является лишь одним из способов программирования компьютера. Точнее сказать, что **программа — это текст спецификации вычисления**. Степень, в которой этот текст можно рассматривать как последовательность инструкций, изменяется в разных языках программирования. В этих заметках мы будем писать программы на языке *ML*, который не является столь явно императивным, как такие языки, как Си и Паскаль, в описании мер, необходимых для выполнения требуемого вычисления. Во многих отношениях *ML* **проще** чем Паскаль и Си. Тем не менее, вам может потребоваться некоторое время, чтобы оценить это.

ML в первую очередь функциональный язык: большинство программ на *ML* лучше всего рассматривать как спецификацию **значения**, которое мы хотим вычислить, без явного описания примитивных шагов, необходимых для достижения этой цели. В частности, мы не будем описывать, и вообще беспокоиться о способе, каким значения, хранимые где-то в памяти, изменяются по мере выполнения программы. Это позволит нам сосредоточиться на **организации** данных и вычислений, не втягиваясь в детали внутренней работы самого вычислителя.

В этом программирование на *ML* коренным образом отличается от тех приемов, которыми вы привыкли пользоваться в привычном императивном языке. **Попытки транслировать ваши программистские привычки на *ML* непролетарны — сопротивляйтесь этому искушению!**

Мы начнем этот раздел с краткого введения в небольшой фрагмент на *ML*. Затем мы используем этот фрагмент, чтобы исследовать некоторые функции, которые будут полезны в дальнейшем. Наконец, мы сделаем обзор некоторых важных аспектов *ML*.

Крайне важно пробовать эти примеры на компьютере, когда вы читаете этот текст.¹

Примечание переводчика Для целей обучения очень удобно использовать онлайн среды, не требующие установки программ, и доступные в большинстве браузеров на любых мобильных устройствах. В качестве рекомендуемых online реализаций Standart ML можно привести следующие:

CloudML <https://cloudml.blechschmidt.saarland/>

описан в блогпосте B. Blechschmidt как онлайн-интерпретатор диалекта Moscow ML

TutorialsPoint SML/NJ http://www.tutorialspoint.com/execute_smlnj_online.php

Moscow ML (**offline**) <http://mosml.org/> реализация Standart ML

- Сергей Романенко, Келдышевский институт прикладной математики, РАН, Москва
- Claudio Russo, Niels Kokholm, Ken Friis Larsen, Peter Sestoft
- используется движок и некоторые идеи из Caml Light © Xavier Leroy, Damien Doligez.
- порт на MacOS © Doug Currie.

¹ Пользовательский ввод завершается точкой с запятой “;”. В большинстве систем, “;” должна завершаться нажатием [Enter]/[Return], чтобы сообщить системе, что надо послать строку в *ML*. Эти примеры тестировались на системе Abstract Hardware Limited’s Poly/ML. В **Poly/ML** запрос ввода символ > или, если ввод неполон — #.

Глава 2

Programming in Standard ML'97

<http://homepages.inf.ed.ac.uk/stg/NOTES/>

© Stephen Gilmore
Laboratory for Foundations of Computer Science
The University of Edinburgh

Часть II

Язык *Prolog*: логическое
программирование
и искусственный интеллект

Глава 3

Adventure in Prolog

1

© Published by: Amzi! inc.

Enjoy the adventure...

-Dennis Merritt

Adventure established the architecture of all fantasy computer games to follow. It was the first to create structures representing places and items and characters and to let the player explore and interact with the environment.

It was as totally addictive as today's wonderous graphical games that have built on the that very same architecture.

To learn more about the granddaddy of all such games, see: [Colossal Cave Adventure](#) and/or [google Adventure](#) and [Willie Crowther](#).

Preface

I was working for an aerospace company in the 1970s when someone got a copy of the original [Adventure](#) game and installed it on our mainframe computer. For the next month my lunch hours, evenings and weekends, as well as normal work hours, were consumed with fighting the fierce green dragon and escaping from the twisty little passages. Finally, with a few hints about the plover's egg and dynamite, I had proudly earned all the points in the game.

My elation turned to terror as I realized it was time for my performance review. My boss was a stern man, who was more comfortable with machines than with people. He opened up a large computer printout containing a log of the hours each of his programmers spent on the mainframe computer. He said he noticed that recently I

¹ © <http://www.amzi.com/AdventureInProlog/index.php>

had been working evenings and weekends and that he admired that type of dedication in his employees. He gave me the maximum raise and told me to keep up the good work.

Ever since I've had a warm spot in my heart for adventure games. Years later, when I got my first home computer, I immediately started to write my own adventure game in C#. First came the tools, a simple dynamic database to keep track of the game state and pattern matching functions to search that database. Then came a natural language parser for the front end. Functions implemented the various rules of the game.

At around the same time I joined the Boston Computer Society and attended a lecture of the newly formed Artificial Intelligence group. The lecture was about *Prolog*. I was amazed — here was a language that included all of the tools needed for building adventure games and more.

It had a much richer dynamic database and more powerful pattern matcher than the one I had written, plus its syntax was rules, which are much more natural for coding the specification of the game. It had a built-in search engine and, to top it all off, had tools for natural language processing.

I learned *Prolog* from the classic Clocksin and Mellish [31] text and started writing adventure games anew.

I went on to use *Prolog* for a number of expert system applications at my then current job, including a mainframe database performance tuning system and installation expert. This got others interested in the language and I began teaching it as well.

While the applications we were using Prolog for were serious and performed a key role in improving technical support for the growing company, I still found the adventure game to be an excellent showcase for teaching the language.

This book is the result of that work. It takes a pragmatic, rather than theoretical, approach to the language and is designed for programmers interested in adding this powerful language to their bag of tools.

I offer my thanks to Will Crowther and Don Woods for writing the first (and in my opinion still the best) adventure game and to the Boston Computer Society for testing the ideas in the book. Thanks also to Ray Reeves, who speaks fluent *Prolog*, and Nancy Wilson, who speaks fluent English, for their careful reading of the text.

© Dennis Merritt
Stow, Massachusetts, April 1996

Prolog tools

For either learning or deploying Prolog we recommend:

[Amzi! Prolog + Logic Server](#)
FREE (IDE only)

The Amzi! Prolog IDE, with its source code debugger, is an excellent tool for getting a solid understanding of Prolog's dynamic variable binding and built-in search.

The Amzi! Logic Server provides the tools for deploying Prolog with other development tools.

The full Amzi! Prolog + Logic Server package is available on either:

- an individual basis or
- an institutional site license.

Download

Other resources for learning Prolog

Checkout the numerous articles on Prolog:

Prolog Articles

and the archives from the years when Amzi! was editor of the AI Expert magazine. Many of the AI techniques are illustrated with Prolog code.

AI Newsletter

3.1 Getting Started

Prolog stands for PROgramming in LOGic. It was developed from a foundation of logical theorem proving and originally used for research in natural language processing. Although its popularity has sprung up mainly in the artificial intelligence (AI) community where it has been used for applications such as expert systems, natural language, and intelligent databases, it is also useful for more conventional types of applications. It allows for more rapid development and prototyping than most languages because it is semantically close to the logical specification of a program. As such, it approaches the ideal of executable program specifications².

Programming in Prolog is significantly different from conventional procedural programs and requires a readjustment in the way one thinks about programming. Logical relations are asserted, and Prolog is used to determine whether or not certain statements are true, and if true, what variable bindings make them true. This leads to a very declarative style of programming.

In fact, the term program does not accurately describe a Prolog collection of executable facts, rules and logical relationships, so you will often see term *logicbase* used in this book as well.

While Prolog is a fascinating language from a purely theoretical viewpoint, this book will stress Prolog as a practical tool for application development.

² and fast prototyping and RAD

Much of the book will be built around the writing of a short adventure game. The adventure game is a good example since it contains mundane programming constructs, symbolic reasoning, natural language, data, and logic.

Through exercises you will also build a simple expert system, an intelligent genealogical logicbase, and a mundane customer order entry application.

You should create a source file for the game, and enter the examples from the book as you go. You should also create source files for the other three programs covered in the exercises. Sample source code for each of the programs is included in the appendix [3.1.1](#).

The adventure game is called Nani Search. Your persona as the adventurer is that of a three year old girl. The lost treasure with magical powers is your nani (security blanket). The terrifying obstacle between you and success is a dark room. It is getting late and you're tired, but you can't go to sleep without your nani. Your mission is to find the nani.



This is Nani

Nani Search is composed of

- A read and execute command loop
- A natural language input parser
- Dynamic facts/data describing the current environment
- Commands that manipulate the environment
- Puzzles that must be solved

You control the game by using simple English commands (at the angle bracket ($>$) prompt) expressing the action you wish to take. You can go to other rooms, look at your surroundings, look in things, take things, drop things, eat things, inventory the things you have, and turn things on and off.

Figure 1.1. A sample run of Nani Search

You are in the kitchen.

You can see: apple, table, broccoli

You can go to: cellar, office, dining room

```
> go to the cellar
```

You can't go to the cellar because it's dark in the cellar,
and you're afraid of the dark.

```
> turn on the light
```

You can't reach the switch and there's nothing to stand on.

```
> go to the office
```

You are in the office.

You can see the following things: desk

You can go to the following rooms: hall, kitchen

```
> open desk
```

The desk contains:

flashlight

crackers

```
> take the flashlight
```

You now have the flashlight

```
> kitchen
```

You are in the kitchen

```
> turn on the light
```

flashlight turned on.

...

Figure 1.1 shows a run of a completed version of Nani Search. As you develop your own version you can of course change the game to reflect your own ideas of adventure.

The game will be implemented from the bottom up, because that fits better with the order in which the topics will be introduced. Prolog is equally adept at supporting top-down or inside-out program development.

A Prolog logicbase exists in the listener's workspace as a collection of small modular units, called *predicates*. They are similar to subroutines in conventional languages, but on a smaller scale.

The predicates can be added and tested separately in a Prolog program, which

makes it possible to incrementally develop the applications described in the book. Each chapter will call for the addition of more and more predicates to the game. Similarly, the exercises will ask you to add predicates to each of the other applications.

We will start with the Nani Search logicbase and quickly move into the commands that examine that logicbase. Then we will implement the commands that manipulate the logicbase.

Along the way there will be diversions where the same commands are rewritten using a different approach for comparison. Occasionally a topic will be covered that is critical to Prolog but has little application in Nani Search.

One of the final tasks will be putting together the top-level command processor. We will finish with the natural language interface.

The goal of this book is to make you feel comfortable with

- The Prolog logicbase of facts and rules
- The built-in theorem prover that allows Prolog to answer questions about the logicbase (backtracking search)
- How logical variables are used (They are different from the variables in most languages.)
- Unification, the built in pattern matcher
- Extra-logical features (like read and write that make the language practical)
- How to control Prolog's execution behavior

3.1.1 Jumping In

As with any language, the best way to learn Prolog is to use it. This book is designed to be used with a Prolog listener, and will guide you through the building of four applications.

1. Adventure game
2. Intelligent genealogical logicbase
3. Expert system
4. Customer order entry business application

The adventure game will be covered in detail in the main body of the text, and the others you will build yourself based on the exercises at the end of each chapter.

There will be two types of example code throughout the book. One is code, meant to be entered in a source file, and the other is interactions with the listener. The listener interactions are distinguished by the presence of the question mark and dash (-) listener prompt.

Here is a two-line program, meant to help you learn the mechanics of the editor and your listener.

```
mortal(X) :- person(X).  
person(socrates).
```

In the **Amzi! Eclipse IDE**, first create a project for your source files. Select **File** **New** **Project** on the main menu, then click on **Prolog** and **Project**, and enter the name of your project, **adventure**. Next, create a new source file. Select **File** **New** **File**,

and enter the name of your file, **mortal.pro**. Enter the program in the edit window, paying careful attention to upper and lowercase letters and punctuation. Then select **File > Save** from the menu.

Next, start the Prolog listener by selecting **Run > Run As > Interpreted Project**. Loading the source code in the **Listener** is called *consulting*. You should see a message indicating that your source file, **mortal.pro**, was consulted. This message is followed by the typical listener prompt.

```
?-
```

Entering the source code in the Listener is called *consulting*. Select **Listener > Consult** from the main menu, and select **mortal.pro** from the file menu. You can also consult a Prolog source file directly from the listener prompt like this.

```
?- consult(mortal).  
yes
```

See the documentation and/or online help for details on the Amzi! listener and Eclipse IDE.

In all the listener examples in this book, you enter the text after the prompt (?), the rest is provided by Prolog. When working with Prolog, it is important to remember to include the final period **.** and to press the **return** key. If you forget the period (and you probably will), you can enter it on the next line with a **.** **return**.

Once you've loaded the program, try the following Prolog queries.

```
?- mortal(socrates).  
yes  
?- mortal(X).  
X = socrates.
```

Appendix

Глава 4

Учебник Фишера

© J.R.Fisher 's *PrologTutorial* ¹

Введение

Prolog — язык декларативного логического программирования. Детально рассматривая его имя, получаем что это сокращение от PROGramming in LOGic: логическое программирование. Наследие *Prologa* включает исследования в области *автоматического доказательства теорем* и других *дедуктивных систем*, разработанных в 1960-70х гг. *Механизм вывода Prologa* базируется на принципе разрешения Робинсона (1965) и механизмах вывода ответов, приложенных Грином (1968). Эти идеи используются вместе с процедурой линейного разрешения. Процедуры точного целевого линейное разрешения, такие как методы Kowalski / Kuehner (1971) и Kowalski (1974), дали толчок к разработке систем логического программирования общего назначения. “Первым” *Prologом* был “Марсельский *Prolog*”, реализация которого основана на работе Colmerauer (1970). Первым делательным описанием языка *Prolog* было руководство к интерпретатору Marseille Prolog (Roussel, 1975). Другим сильным влиянием на природу этого первого *Prologa* была адаптация этого интерпретатора к задачам *обработки естественных языков*.

Prolog является наиболее часто упоминаемым примеров языков программирования четвертого поколения, которые поддерживают парадигму **декларативного программирования**. Японский проект Fifth-Generation Computer Project², анонсированный в 1981, применял *Prolog* как язык разработки, и сосредоточивал значительные усилия на языке и его возможностях. Программы в этом учебнике написаны на “стандартном” *Prologе* Эдинбургского университета³, как это сделано в классической книге по *Prologу* под авторством Clocksin и Mellish (1981,1992).

¹ © https://www.cpp.edu/~jrfisher/www/prolog_tutorial/contents.html

² компьютерный проект пятого поколения

³ University of Edinburgh Prolog

Другой заметной версией *Prologa* является семейство реализаций *PrologII*, которые являются ответственными за Марсельского *Prologa*. Справочник Giannesini, et.al. (1986) использует версию *PrologII*. Есть некоторые различия между этими двумя вариантами *Prologa*; часть различий в синтаксисе, и часть в семантике. Тем не менее, студенты изучавшие одну из версий, впоследствии могут легко адаптировать к другой.

Цель этого учебника — помочь изучить самые необходимые, базовые концепции языка *Prolog*. Примеры программ были особенно аккуратно выбраны для иллюстрации программирования искусственного интеллекта на *Prolog*. *Lisp* и *Prolog* наиболее часто используемые языки символьного программирования для приложений искусственного интеллекта. Они часто упоминаются как великолепные языки для “исследовательского” и “прототипного программирования”.

Раздел 4.1 рассматривает среду программирования на *Prolog* для начинающих.

Раздел 4.2 объясняет синтаксис *Prologa* и многие аспекты программирования на нем через реализацию аккуратно выбранных программ-примеров. Эти примеры организованы так, чтобы студент обучался через реализацию *Prolog*-программ “сверху вниз” в декларативном стиле. Были приняты меры к рассмотрению техник программирования на *Prolog*, которые очень важны для курса искусственного интеллекта. Фактически, **этот учебник может служить удобным, маленьkim, кратким введением в применение Prologa в приложениях искусственного интеллекта**. Аспекты семантики языка *Prolog* рассматриваются с самого начала с точки зрения концепции дерева условий программы, которое используется для определения последовательностей спецификаций *Prolog*-программы в абстрактном виде. Автор надеется что этот подход позволит рассмотреть базовые принципы формальной верификации программ при программировании на *Prolog*. Последняя секция этого раздела приводит пример, который показывает что *Prolog* может быть эффективно использован для аккуратной, точной спецификации программных систем, несмотря на его репутацию трудно документируемого языка, так что *Prolog* легко использовать как средство прототипирования.

Раздел 4.3 рассматривает работу внутренних механизмов *Prolog*-движка. Раздел 4.3 рекомендуется просмотреть сразу после того, как студент изучил 2-3 примера программ из раздела 4.2. Последняя секция этого раздела рассматривает **мета-интерпретаторы Prologa**.

Раздел 4.4 дает краткий обзор основных встроенных предикатов, многие из которых используются в разделе 4.2..

Раздел 4.5 рассматривает разработку программ A*-поиска на *Prolog*. Раздел 4.5.3 содержит программу $\alpha\beta$ -поиска для игры tic tac toe.

Раздел 4.6 представляет уникальное и обширное описание логического мета-интерпретатора для нормальных логических баз правил.⁴

Раздел 4.7 представляет введение во встроенный в *Prolog* генератор парсеров

⁴ Замечание от 9/4/2006: Я значительно отредактировал этот раздел, и обновил все ссылки на секции.

грамматики, и дает общий обзор приемов, с помощью которых *Prolog* может быть использован для разбора выражений натурального языка (английского). Также эта секция описывает построение программных интерфейсов, использующих идеоматически-простые натуральные языки.

Раздел 4.8 показывает приемы реализации различных *Prolog*-прототипов. Новый раздел 4.8.4 раскрывает интерактивную связку между машиной вывода *Prolog* и Java GUI для игры tic tac toe. Рассмотренная простая модель связки легко адаптируемая и применима.

Ранние версии частей этого учебника датируются 1988 годом. Вводный материал изначально использовался для объяснения работы интерпретатора *Prologa*, разработанного автором⁵ для применения в учебном процессе. Автор надеется что вводный материал, собранный в форме этого учебника, может быть очень полезным для студентов, которые хотят быстрое, но при этом хорошо сбалансированное, введение в программирование на *Prolog*.

Для дальнейшего изучения *Prologa* можно посоветовать книги Clocksin и Melliss (1981,1992), O'Keefe (1990), Clocksin (1997, 2003), или Sterling и Shapiro (1986).

Подробные заметки по истории *Prolog* и обработке натуральных языков с его использованием содержатся в работе Pereira and Shieber (1987).

© Помона, Калифорния
1988-2015

4.1 Установка и запуск *Prolog*-системы

Примеры этого учебника *Prologa* были подготовлены с использованием

- Quintus Prolog на компьютерах Digital Equipment Corporation MicroVAXes (далекая история)
- SWI Prolog на Sun Spark (давным давно)
- персональных компьютерах с *Windows*
- или OS X на Macах

Другие заметные *Prolog*-системы (Borland, XSB, LPA, Minerva . . .) использовались для разработки и тестирования последние 25 лет. В этом учебнике запланирован новый раздел, в котором описано использование любых *Prolog*-систем в общем, но пока этот раздел недоступен.

Сайт SWI-Prolog содержит много информации, ссылки на загрузку, и документацию:

<http://www.swi-prolog.org/>

Особо следует отметить возможность попробовать SWI Prolog on-line без регистрации и SMS: <http://swish.swi-prolog.org/>. Этот вариант особенно удобен, так как не требует никакой установки ПО, административных прав, вы можете работать с этим учебником даже в интернет-кафе.

⁵ сейчас недоступен

Примеры в этом учебнике используют упрощенную форму взаимодействия в типичным *Prolog*-интерпретатором, так что программы должны работать похоже в любой *Prolog*-системе эдинбургского типа или интерактивном компиляторе.

Если в вашей UNIX-системе уже установлен SWI-Prolog, запустите окно терминала, и начните интерактивную сессию командной:

```
user@computer$ swipl
```

Мы не будем использовать команду запуска именно в такой форме все время: при запуске могут быть указаны дополнительные параметры командной строки, которые можно использовать в определенных случаях. Читатель должен рассмотреть эту возможность после освоения базовых приемов работы, чтобы получить больше возможностей.

Если вы хотите установить SWI Prolog под Debian *Linux*, выполните команду:

```
sudo apt install swi-prolog
```

Под *Windows* инсталлятор SWI-Prolog добавляет иконку запуска интерпретатора, который вы можете запустить простым двойным щелчком по иконке. При запуске интерпретатор создает свое собственное командное окно.

После запуска интерпретатора обычно появляется сообщение о версии, лицензии и авторах, а затем выводится приглашение ввода *цели* типа

```
?- _
```

Интерактивные *цели* в *Prolog* вводятся пользователем за приглашением `?-.`

Многие *Prolog*-системы поддерживают предоставление документации по запросу из командной строки. В SWI Prolog встроена подробная система помощи. Документация индексирована, и помогает пользователю в процессе работы. Попробуйте ввести

```
?- help(help).
```

Обратите внимание что должна быть введены все символы, и ввод завершен возвратом каретки.

Для иллюстрации нескольких приемов взаимодействия с *Prolog* рассмотрим следующий пример сессии. Если приведена ссылка на файл, предполагается что это локальный файл в пользовательском каталоге, который был создан пользователем, получен копированием из другого публично доступного источника, или получен сохранением текстового файла из веб-браузера. Способ достижения последнего — следователь URL-ссылке на файл и сохранить его, или выбрать кусок текста из онлайн-учебника *Prologa*, скопировать его, вставить в текстовый редактор, и сохранить полученный файл из текстового редактора. Комментарии вида `/*...*/` после целей используются для описания этих целей.

Листинг 1: Лог типичной Prolog-сессии

```
?- ['2_2.pl'].          /* 1. Загрузка программы из локального файла */
true.

?- listing(factorial/2). /* 2. Вывод листинга программы на экран */

factorial(0,1).

factorial(A,B) :-
    A > 0,
    C is A-1,
    factorial(C,D),
    B is A*D.

true.

?- factorial(10,What). /* 3. Вычислить 10! (в переменную) */
What=3628800 .           /* нажмите Enter */

?- ['2_7.pl'].            /* 4. Загрузить другую программу */

?- listing(takeout).

takeout(A,[A|B],B).
takeout(A,[B|C],[B|D]) :-
    takeout(A,C,D).

true.

?- takeout(X,[1,2,3,4],Y). /* 5. Взять X из списка [1,2,3,4] */
X=1  Y=[2,3,4] ;           /* Prolog ждет ... нужно ввести ';' и Enter */
X=2  Y=[1,3,4] ;           /* следующий ... */
X=3  Y=[1,2,4] ;           /* следующий ... */
X=4  Y=[1,2,3] ;           /* следующий ... */
false.                      /* Обозначает: больше нет ответов. */

?- takeout(X,[1,2,3,4],_), X>3. /* 6. Конъюнкция целей */
X=4 ;
false.

?- halt.                  /* 7. Выход из интерпретатора в OS */
```

Комментарии в правой части были добавлены в текстовом редакторе. Они отмечают некоторые вещи, перечисленные ниже:

1. Определение *цели* Prologа завершается точкой . . В этом случае цель бы-

ла загружена в внешнего файла с исходным тестом программы. Этот скобочный стиль записи программы унаследован из самых первых реализаций *Prologa*. Можно загрузить несколько файлов сразу, указав их имена в одиночных кавычках, разделяя запятыми. В нашем случае имя файла **2_2.pl**, программа содержит два программы на *Prolog* для вычисления факториала от положительного целого. Подробно эта программа описана в разделе [4.2.2](#). Файл программы ищется в текущем каталоге. Если поиск неуспешен, нужно явно указать полный путь обычным для вашей ОС способом.

2. Встроенный предикат *listing* выводит листинг программы из ОЗУ — в нашем случае программу вычисления факториала, загруженную ранее. Внешний вид этого листинга несколько отличается от исходного кода в файле из [4.2.2](#). Заметим, что **Quintus Prolog** компилирует программу, если отдельно не указано что определенные предикаты являются динамическими. Скомпилированные предикаты не могут быть выведены через *listing*, поэтому если у вас он не срабатывает, возможно требуется дополнить исходник декларацией динамического предиката, чтобы пример сработал. В **SWI Prolog** этот пример работает без модификации.
3. Эта цель *factorial(10,What)* говорит “факториал 10ти что?”. Слово *What* начинается с большой буквы, указывающей что это **логическая переменная**. *Prolog* удовлетворяет цель находя все возможные значения переменной *What*.
4. Теперь в памяти находятся обе программы из файлов **2_1.pl** и **2_7.pl**. Файл **2_7.pl** содержит несколько определений обработки списков (см. [4.2.7](#)).
5. Только что загруженная программа (**2_7.pl**) содержит определение предиката *takeout*. Цель *takeout(X, [1,2,3,4], Y)* запрашивает поиск всех таких *X* что значение взятое из списка **[1,2,3,4]** оставляет остаток в переменной *Y*, для всех возможных случаев. Существует четыре способа сделать это, как показано в результате. Предикат *takeout* обсуждается в разделе [4.2.7](#). Таким образом, **в Prolog заложен поиск всех возможных ответов**: после того как будет выведен очередной ответ, *Prolog* ожидает реакции пользователя мигая курсором в конце строки с ответом. Если пользователь нажмет **;**, *Prolog* будет выполнять поиск следующего ответа. Если пользователь просто нажмет **Enter**, *Prolog* остановит поиск.
6. Составная, или **конъюнктивная цель**, определяет одновременное удовлетворение **двух** отдельных целей. Отметим что используется арифметическая цель (встроенное отношение) *X>3*. *Prolog* будет пытаться удовлетворить эти цели **слева направо**, в порядке чтения. В нашем случае существует единственный ответ. Отметим использование в цели **анонимной переменной _**, **биндинг (привязка)** для которой не выводится (переменная “не важно”).

7. Цель `halt` всегда успешна и завершает работу интерпретатора.

4.2 Разбор примеров программ

В этом разделе мы рассмотрим несколько специально подобранных примеров программ на *Prolog*. Порядок примеров специально выбран от наиболее простых до более сложных. Ключевая цель — показать основные приемы *представления знаний* и методов декларативного программирования.

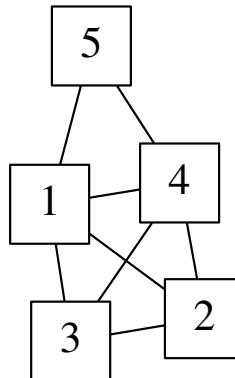
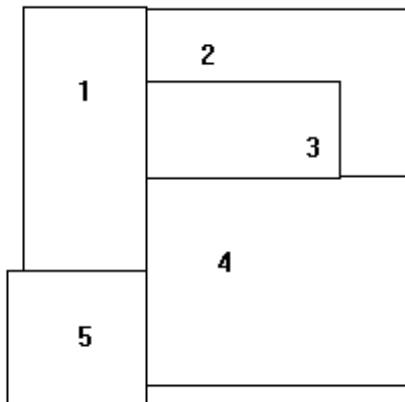
4.2.1 Раскраска карт

Этот раздел использует известную математическую проблему — **раскраска географических карт** — в качестве иллюстрации применения набора фактов и логических правил. Рассмотренная *Prolog*-программа показывает представление смежных регионов карты, ее раскраски, и определение конфликтов раскраски: когда **два смежных региона имеют одинаковый цвет**. Секция также показывает применение концепции **семантического дерева** и его применение в логическом программировании.

Согласно формулировке известной математической задачи по раскраске смежных плоских регионов⁶, необходимо подобрать минимум цветов раскраски, и цвета регионов, так что никакие два смежных региона не имеют один цвет. Два региона являются смежными, если они имеют некоторый общий сегмент границы, например⁷. По данным численным именам регионов строим представление в виде *графа смежности*:

⁶ таких как географические карты

⁷ упрощенно, только прямоугольные области



Мы удалили все границы, и нарисовали дугу между именами каждой двух смежных областей. Фактически граф смежности содержит полную оригинальную информацию о смежности областей. Для представления информации о смежности в синтаксисе *Prologa* запишем следующее:

adjacent (1 ,2).	adjacent (2 ,1).
adjacent (1 ,3).	adjacent (3 ,1).
adjacent (1 ,4).	adjacent (4 ,1).
adjacent (1 ,5).	adjacent (5 ,1).
adjacent (2 ,3).	adjacent (3 ,2).
adjacent (2 ,4).	adjacent (4 ,2).
adjacent (3 ,4).	adjacent (4 ,3).
adjacent (4 ,5).	adjacent (5 ,4).

это набор выражений устанавливает факт смежности $A \rightarrow B$: `adjacent(A,B)`.

Если загрузить этот файл в *Prolog*-систему, можно проверить работу целей:

```
?- adjacent(2,3).
true .
?- adjacent(5,3).
false .
?- adjacent(3,R).
R = 1 ;
R = 2 ;
R = 4 ;
false .
```

Аналогично можно задать два набора раскраски регионов используя единичные заключения: вариант **a** и вариант **b**:

```
color(1, red , a).      color(1, red , b).
color(2, blue , a).     color(2, blue , b).
color(3, green , a).    color(3, green , b).
color(4, yellow , a).   color(4, blue , b).
color(5, blue , a).     color(5, green , b).
```

в форме

```
<имя отношения:color> (
  <номер зоны/узла графа>,
  <присвоенный цвет>,
  <имя раскраски>
).
```

Что обозначает **факт**: “имеется отношение color между номером узла, цветом и именем раскраски”⁸.

Теперь мы хотим написать *Prolog*-определение конфликта раскрасок, имея в виду совпадение цветов для двух регионов, например:

```
conflict(Coloring) :-  
  adjacent(X,Y),  
  color(X, Color , Coloring),  
  color(Y, Color , Coloring).
```

Например,

```
?- conflict(a).  
false .  
?- conflict(b).  
true .  
?- conflict(Which).  
Which = b .
```

Запрашивая отношение с неким значением-константой, или переменной⁹ (последний случай), мы получаем от *Prolog*-системы заключение: выполняется ли запрошенное отношение-*целк* и при каких значениях переменных, имея в виду ранее

⁸ причем не указывается какой элемент главный или подчиненный, все элементы отношения равноправны

⁹ имя с большой буквы

определенный *набор фактов и отношений*¹⁰. В случае использования переменной *Prolog* выдаст нам **все** значения переменных, для которых запрос истинен.

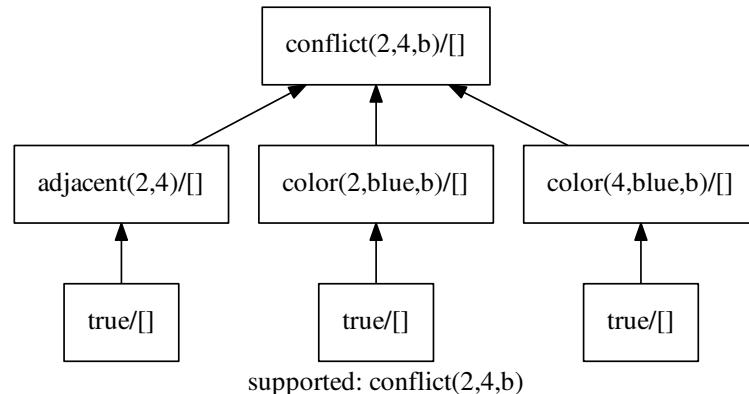
Можно определить другое отношение с тем же именем **conflict** но с другим количеством логических параметров:

```
conflict(R1,R2,Coloring) :-  
    adjacent(R1,R2),  
    color(R1,Color,Coloring),  
    color(R2,Color,Coloring).
```

Prolog позволяет отличать два отношения с одинаковым именем: одно имеет один параметр **conflict/1**, а другой — **conflict/3**.¹¹

```
?- conflict(R1,R2,b).  
R1 = 2    R2 = 4  
?- conflict(R1,R2,b),color(R1,C,b).  
R1 = 2    R2 = 4    C = blue
```

Последняя *цель* значит что регионы 2 и 4 связаны (*adjacent*) и оба синие (*blue*). *Обоснованные* случаи, такие как **conflict(2,4,b)**, называются **консеквенцией** или **выводом** *Prolog*-программы. Один из способов демонстрации консеквенции — нарисовать **дерево заключений**, которое имеет консеквенцию в корне дерева, используя заключения программы для обхода дерева, получая в результате конечное дерево, в котором все листья имеют истинное значение. Например следующее дерево заключений может быть построено используя полностью обоснованные заключения программы без переменных:



Нижняя левая ветка дерева соответствует unit clause:

¹⁰ которые являются *базой знаний*, или *экспертной системой*

¹¹ /цифра имеет название *арифтост*

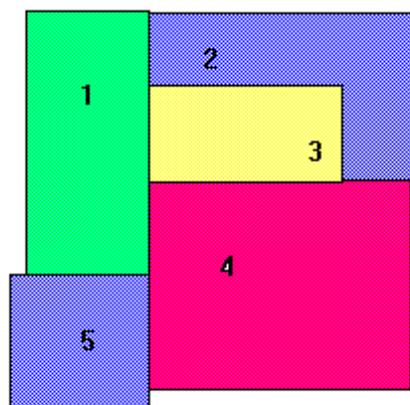
`adjacent(2,4).`

которая в *Prolog* эквивалента clause

`adjacent(2,4) :- true.`

С другой стороны `conflict(1,3,b)` не является consequence в *Prolog*-программе так как невозможно construct finite clause tree используя grounded clauses P содержащие все `true` листья. Аналогично `conflict(a)` не консеквенция, как можно ожидать. В последующих секциях clause деревья в subsequent sections описаны более подробно.

Мы повторно рассмотрим проблему раскраски карт в ??, где мы разработали *Prolog*-программу которая генерирует все возможные схемы раскраски¹². Известная Гипотеза Четырех Цветов гласит что любая плоская карта требует для раскраски не более 4x цветов. Это было доказано в работе Appel и Haken (1976). Решение использовало компьютерную программу¹³ для проверки всевозможных карт, с целью выявить возможные проблемные случаи. Следующая схема раскраски например требует не менее 4x цветов:



Упражнение 2.1 Если карта имеет N регионов, определите сколько вычислений должно быть выполнено для определения есть ли конфликт раскраски. Аргументируйте используя program clause дерева.

4.2.2 Два определения факториала

Этот раздел вводит в вычисления математических функций используя *Prolog*. Обсуждаются различные встроенные арифметические операции. Также обсуждается концепция derivation дерева, и как derivation деревья связаны с трассировкой в *Prolog*.

¹² given colors to color with

¹³ не на *Prolog*

В файле **2_2.pl** находятся два определения предикатов, являющиеся определением функции вычисления факториала:

первый вариант

```
factorial(0,1).
```

```
factorial(N,F) :-
```

```
    N>0,
```

```
    N1 is N-1,
```

```
    factorial(N1,F1),
```

```
    F is N * F1.
```

Эта программа состоит из двух clauses. Первое заключение — формулировка **факта** (unit clause) **без тела**. Второе заключение — **правило**, так как **у него есть тело**. Тело второго заключения находится после `:-`, которое можно читать как “если”. Тело содержит литералы, разделенные запятыми, каждую запятую можно читать как “и”. **Заголовок правила** — весь текст **факта** или часть текста до `:-` в правиле. Рассматривая текст как декларативную программу, первое (фактическое) предложение читается как “факториал 0 есть 1”¹⁴, и второе предложение заявляет что “факториал N есть F¹⁵ если N>0 и N1 есть N-1 , и факториал N1 есть F1, и F есть N*F1.

Prolog-цель (goal) для вычисления факториала от 3 дает ответ в W — **переменной цели**:

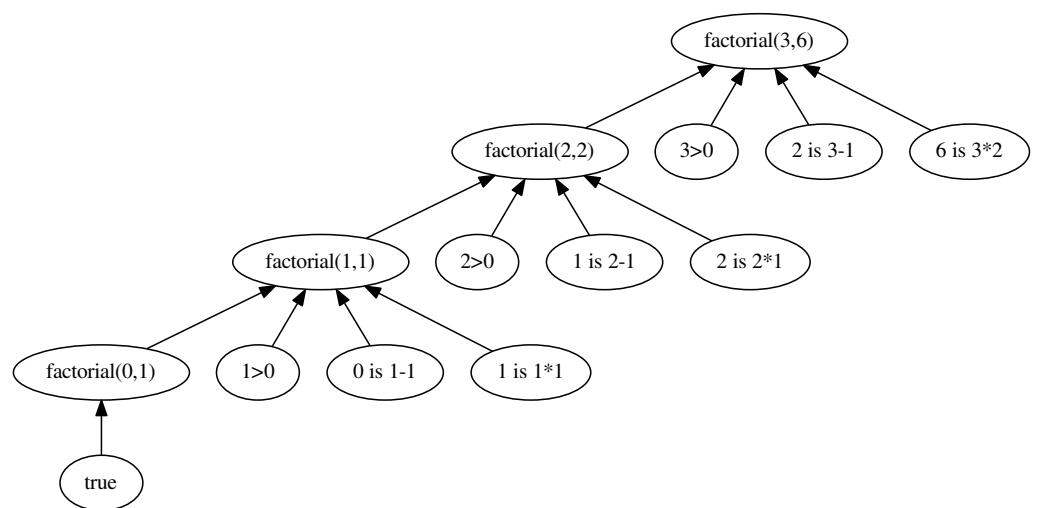
```
?- factorial(3,W).
```

```
W=6 .
```

Рассмотрим следующее clause дерево сконструированное для литерала `factorial(3,W)`. Как описано в предыдущей секции, clause дерево не содержит никаких свободных переменных, вместо этого включает непосредственно их значения. Каждое ветвление под узлом определяется clause оригинальной программы, используя непосредственно вхождения значений переменных; узел задается заголовком правила, а литералы тела становятся дочерними узлами.

¹⁴ или: 0 и 1 **связаны отношением** “факториал”, но у объектов одновременно могут быть и другие отношения, например биты(0,1) и целые(0,1)

¹⁵ точнее: N и F связаны отношением “факториал”



Все арифметические листья |true| при исполнении¹⁶, и самая нижняя связь в дереве соответствует самому первому clause в программе вычисления факториала. Первый clause может быть записан как:

```
factorial(0,1) :- true.
```

и фактически `?- true.`. *Prolog*-цель которая всегда успешна¹⁷. Для краткости, мы не отрисовали `true` для всех листьев, являющихся арифметическими литералами.

Программное clause дерево показывает значение цели в корне дерева. Так, `factorial(3,6)` является консеквенцией *Prolog*-программы, так как существует clause дерево с корнем `factorial(3,6)`, все листья которого `true`. С другой стороны литерал `factorial(5,2)` не консеквенция, так как такого дерева для него нет, а значением программы для литерала `factorial(5,2)` является `false`:

```
?- factorial(3,6).
true .
?- factorial(5,2).
false .
```

как и следовало ожидать. Clause-деревья также называются AND-деревьями, так как чтобы корень был консеквенцией программы, все его поддеревья также должны быть консеквенциями. Позже clause деревья будут рассмотрены подробнее. Мы отметили что **clause дерево описывает семантику (значение) программы**. В разделе 4.6 мы рассмотрим другой подход к семантике программ. Clause-деревья предоставляют интуитивный и корректный подход к описанию семантики.

¹⁶ в соответствии с предполагаемой интерпретацией

¹⁷ `true` встроенный предикат

Нам нужно отличать clause деревья программы и **деревья вывода**. Clause-деревья статичны, и могут быть нарисованы для программы или цели через механизм удовлетворения частичных (под)целей, как описано выше. Грубо говоря, clause-деревья соответствуют декларативному чтению программы.

Деревья вывода наоборот, имеют в виду механизм привязки переменных *Prolog* и порядок в котором удовлетворяются вложенные частичные цели. Подробнее деревья вывода описаны в разделе [4.3.1](#), но тем не менее посмотрите анимацию, предоставляемую динамическим отладчиком, как описано ниже.

Трассировка исполнения *Prolog*-программы также показывает как переменные привязываются при удовлетворении целей. Следующий пример показывает включение/выключение трассировки в типичной *Prolog*-системе.

```
?- trace.  
% The debugger will first creep -- showing everything (trace).  
  
true .  
[trace]  
?- factorial(3,X).  
 (1) 0 Call: factorial(3,_8140) ? [Enter] creep  
 (1) 1 Head [2]: factorial(3,_8140) ? [Enter] creep  
 (2) 1 Call (built-in): 3>0 ? creep  
 (2) 1 Done (built-in): 3>0 ? creep  
 (3) 1 Call (built-in): _8256 is 3-1 ? creep  
 (3) 1 Done (built-in): 2 is 3-1 ? creep  
 (4) 1 Call: factorial(2, _8270) ? creep  
 ...  
 (1) 0 Exit: factorial(3,6) ?  
X=6 .  
[trace]  
?- notrace.  
% The debugger is switched off  
  
true .
```

The animated tree below gives another look at the derivation tree for the *Prolog* goal `factorial(3,X)`. To start (or to restart) the animation, simply click on the **Step** button.

Заголовок этого раздела говорит “**Два** определения факториала”, вот второй вариант, использующий три переменных:

второй вариант

```
factorial(0,F,F).
```

```
factorial(N,A,F) :-  
    N > 0,  
    A1 is N*A,  
    N1 is N -1,  
    factorial(N1,A1,F).
```

Для этой версии используйте следующую цель-запрос:

```
?- factorial(5,1,F).  
F=120 .
```

Второй параметр в определении называется *параметр-аккумулятор*, который также хорошо известен в функциональном программировании. Эта версия факториала определена с использованием *хвостовой рекурсии*. Важно чтобы вы выполнили следующие упражнения:

Упражнение 4.2.2.1 Используя первый вариант программы факториала, четко покажите что не существует clause-дерева с корнем `factorial(5,2)`, имеющего все true листья.

Упражнение 4.2.2.2 Нарисуйте clause-дерево для цели `factorial(3,1,6)` со всеми true-листьями, в виде аналогичном ранее описанному дереву для `factorial(1,1,1)`. Покажите, чем отличаются два варианта программы в процессе вычисления факториала? Также, протрассируйте цель `factorial(3,1,6)` используя Prolog-систему.

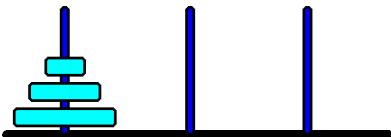
4.2.3 Классическая задача “Ханойские башни”

Показано формулирование и решение классической задача на *Prolog*. Рассмотрены декларативные и процедурные подходы к программированию. Решение задачи выводится на экран.

Цель известной головоломки — переместить N дисков с левого штыря на правый, используя центральный штырь как дополнительное хранилище. Требование: **нельзя класть больший диск на мénьший**. Следующая диаграмма показывает начальное положение для $N=3$ дисков.

Рекурсивная программа на *Prolog*, решающая головоломку, состоит из двух утверждений:

Ханойские башни



```
1 move(1 ,X,Y,_ ) :-  
2   write ( 'Move_top_disk_from_') ,  
3   write(X) ,  
4   write ( '_to_') ,  
5   write(Y) ,  
6   nl .  
7 move(N,X,Y,Z) :-  
8   N>1,  
9   M is N-1,  
10  move(M,X,Z,Y) ,  
11  move(1 ,X,Y,_ ) ,  
12  move(M,Z,Y,X) .
```

Переменная `_` (или любое другое имя начинающееся с подчеркивания) — переменные **don't-care** (не важно). *Prolog* позволяет использовать эти перемененные как обычные в любых структурах, но для них **не выполняется привязка**.

Вот что выводится при решении задачи при $N=3$:

```
?- move(3, left, right, center).  
Move top disk from left to right  
Move top disk from left to center  
Move top disk from right to center  
Move top disk from left to right  
Move top disk from center to left  
Move top disk from center to right  
Move top disk from left to right  
true .
```

Первое предложение программы описывает перемещение одного диска. Второе предложение описывает как можно получить решение рекурсивно. Например, декларативное чтение второго предложения для случая $N=3$, $X=left$, $Y=right$, и $Z=center$ приводит к следующему:

```
move(3, left, right, center) если  
  move(2, left, center, right) и ] *  
  move(1, left, right, center) и  
  move(2, center, right, left). ] **
```

Это декларативное чтение очевидно правильно. Процедурное чтение тесно связано с декларативной интерпретацией рекурсивного утверждения, оно должно выглядеть как-то так:

удовлетворить цель `?-move(2, left, center, right)`, и потом

удовлетворить цель `?-move(1, left, right, center)`, и потом
удовлетворить цель `?-move(2, center, right, left)`.

Аналогично мы можем записать декларативное прочтение для случая N=2:

```
move(2, left, center, right) если ] *
move(1, left, right, center) и
move(1, left, center, right) и
move(1, right, center, left).
move(2, center, right, left) если ] **
move(1, center, left, right) и
move(1, center, right, left) и
move(1, left, right, center).
```

Теперь подставим содержимое последних двух implications и увидим решение которое генерирует *Prolog*:

```
move(3, left, right, center) если
move(1, left, right, center) и
move(1, left, center, right) и *
move(1, right, center, left) и
-----
move(1, left, right, center) и
-----
move(1, center, left, right) и
move(1, center, right, left) и **
move(1, left, right, center).
```

Процедурное прочтение последних двух больших implication должно быть очевидно. Этот пример показывает при основных операции *Prologa*:

1. Цели сопоставляются с головой правила, и
2. тело правила (с соответствующе привязанными переменными) становится новой последовательностью целей; процесс повторяется
3. пока не будет удовлетворена основная цель или условие, или не будет выполнено простое действие, например выведен текст.

Процесс сопоставления переменных с образцом (variable matching)
называется **унификацией**.

Упражнение 4.2.3.1 Нарисуйте clause-дерево для цели `move(3, left, right, center)` и покажите что это конвенция программы. Как полученное дерево связано с процессом подстановки, поисанным выше ?

Exercise 4.2.3.2 Попробуйте *Prolog*-цель `?-move(3, left, right, left)`. Что не так? Предложите способ исправления, и проследите процесс работы исправления.

4.2.4 Загрузка, редактирование, хранение программ

Примеры показывают различные способы хранения и загрузки *Prolog*-программ, и пример вызова системного редактора. Читателю предлагается предварительно заглянуть в разделы 4.3.1, 4.3.2 чтобы иметь представление о том, как работает *Prolog*.

Стандартные предикаты для загрузки программ это `consult`, `reconsult`, и скобочная нотация загрузки `[...]`. Например цель `?- consult('lists.pro')`. открывает файл `lists.pro` и загружает из него предложения в память.

Существует два способа, которыми *Prolog*-программа может быть неправильна:

1. исходный код имеет синтаксические ошибки, в этом случае при загрузке будут выводиться сообщения об ошибках, и
2. в программе есть какие-то логические ошибки, которые программист должен найти через тестирование программы.

Программа в ее текущей версии должна рассматриваться как прототип корректной версии в будущем, и принятая обычная практика редактирования текущей версии, и ее перезагрузка с повторным тестированием. Существуют хорошие приемы быстрого прототипирования, чтобы программист уделял все время и усилия на анализ проблемы. Интересно что если подход быстрого прототипирования кажется ошибочным, это отличный сигнал взять ручку и бумагу, еще раз проанализировать требования, и начать сначала!

Мы можем вызывать редактор непосредственно в *Prolog*:

```
?- edit('lists.pro'). %% редактор определенный пользователем, см. ниже ..
```

и после возврата из редактора¹⁸ использовать цель

```
? reconsult('lists.pro').
```

для перезагрузки утверждений программы в память, автоматически замещая предыдущие определения. Если использовать `consult` вместо `reconsult`, старая¹⁹ версия утверждений программы останется в памяти наряду с новыми определениями²⁰.

Если в память было загружено несколько файлов, и требуется перезагрузить только один, используйте `reconsult`. Если перегружаемый файл определяет предикаты, которые не определяются в остальных файлах, перезагрузка не повлияет на кляузы, которые были загружены в остальных файлах.

Скобочная нотация очень удобна, например

¹⁸ предполагается что новая версия файла была сохранена в том же файле

¹⁹ и скорее всего неправильная

²⁰ это поведение зависит от конкретной версии *Prolog*-системы

```
?- ['file1.pro',file2.pro',file3.pro'] .
```

загрузит (точнее `reconsult`) все три файла в память *Prolog*-системы.

Многие *Prolog*-системы оставляют программисту определение любимого текстового редактора. Здесь описан пример программы, которая вызывает **TextEdit** на Mac(OSX)²¹.

```
edit(File) :-  
    name(File,FileString),  
    name('open -e ', TextEditString), %% укажите ваш любимый редактор  
    append(TextEditString,FileString,EDIT),  
    name(E,EDIT),  
    shell(E).
```

Для использования этого редактора, этот код должен быть загружен²²

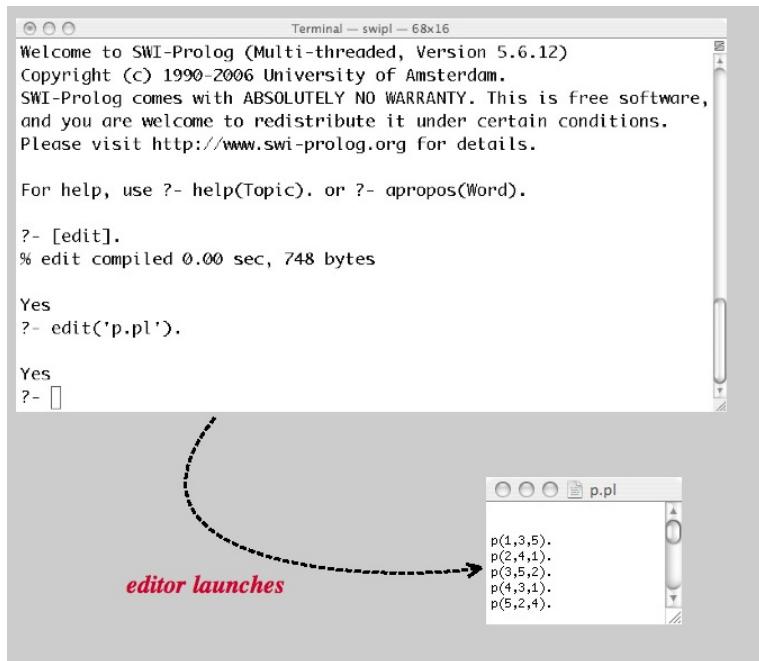
```
?- [edit].
```

```
yes
```

и цель `edit` может быть использована²³

```
?- edit('p.pl').
```

```
{ TextEdit запускается с файлом, редактируйте его...}  
{ и сохраните измененную программу с тем же именем файла ... }
```



²¹ это просто пример; мы не используем конкретно **TextEdit**

²² предполагаем локальную *Prolog*-сессию

²³ опять же предполагаем что файл для редактирования локален для сессии

Вызов внешнего редактора

После редактирования и сохранения мы можем перезагрузить новую версию

```
?- reconsult('p.pl').
```

{ наша prolog-сессия перезагружает программу для тестирования ...}

Для редактирования утверждений, введенных пользователем интерактивно, используйте цели

```
?-consult(user).  
?-reconsult(user).  
?- [user].
```

Пользователь вводит предложения интерактивно, используя символ останова . в конце набора утверждений, и сочетание клавиш **Ctrl**+**D** для окончания ввода.

Упражнение 4.2.4 Проанализируйте как работает редактирование программы. Сначала попробуйте цели

```
?-name('name',NameString).
```

и

```
?- name(Name,"name").
```

`name/2` описана в разделе [4.4.13](#).

Теперь хороший момент для читателя немного заглянуть вперед и попробовать почитать первые две секции из раздела [4.3](#) “Как работает Prolog”, и затем вернуться к остальным примерам программ. Необходимо чтобы вы понимали как работают машина вывода *Prologa*, чтобы понять как конструируются следующие примеры программ.

4.2.5 2.5 Negation as failure

The section gives an introduction to *Prolog's* negation-as-failure feature, with some simple examples. Further examples show some of the difficulties that can be encountered for programs with negation as failure.

4.2.6 2.6 Tree data and relations

This section shows *Prolog* operator definitions for a simple tree structure. Tree processing relations are defined and corresponding goals are studied.

4.2.7 2.7 Prolog lists and sequences

This section contains some of the most useful Prolog list accessing and processing relations. Prolog's primary dynamic structure is the list, and this structure will be used repeatedly in later sections.

4.2.8 2.8 Change for a dollar

A simple change maker program is studied. The important observation here is how a *Prolog* predicate like 'member' can be used to generate choices, the choices are checked to see whether they solve the problem, and then backtracking on 'member' generates additional choices. This fundamental generate and test strategy is very natural in *Prolog*.

4.2.9 2.9 Map coloring redux

We take another look at the map coloring problem introduced in Section 2.1. This time, the data representing region adjacency is stored in a list, colors are supplied in a list, and the program generates colorings which are then checked for correctness.

4.2.10 2.10 Simple I/O

This section discusses opening and closing files, reading and writing of *Prolog* data.

4.2.11 2.11 Chess queens challenge puzzle

This familiar puzzle is formulate in *Prolog* using a permutation generation program from Section 2.7. Backtracking on permutations produces all solutions.

4.2.12 2.12 Finding all answers

Prolog's 'setof' and 'bagof' predicates are presented. An implementation of 'bagof' using 'assert' and 'retract' is given.

4.2.13 2.13 Truth table maker

This section designs a recursive evaluator for infix Boolean expressions, and a program which prints a truth table for a Boolean expression. The variables are extracted from the expression and the truth assignments are automatically generated.

4.2.14 2.14 DFA parser

A generic DFA parser is designed. Particular DFAs are represented as *Prolog* data.

4.2.15 2.15 Graph structures and paths

This section designs a path generator for graphs represented using a static *Prolog* representation. This section serves as an introduction to and motivation for the next section, where dynamic search grows the search graph as it works.

4.2.16 2.16 Search

The previous section discussed path generation in a static graph. This section develops a general *Prolog* framework for graph searching, where the search graph is constructed as the search proceeds. This can be the basis for some of the more sophisticated graph searching techniques in A.I.

4.2.17 2.17 Animal identification game

This is a toy program for animal identification that has appeared in several references in some form or another. We take the opportunity to give a unique formulation using *Prolog* clauses as the rule database. The implementation of verification of askable goals (questions) is especially clean. This example is a good motivation for expert systems, which are studied in Chapter 6.

4.2.18 2.18 Clauses as data

This section develops a *Prolog* program analysis tool. The program analyses a *Prolog* program to determine which procedures (predicates) use, or call, which other procedures in the program. The program to be analyzed is loaded dynamically and its clauses are processed as first-class data.

4.2.19 2.19 Actions and plans

An interesting prototype for action specifications and plan generation is presented, using the toy blocks world. This important subject is continued and expanded in Chapter 7.

4.3 Как работает *Prolog*

4.3.1 Деривационные деревья, выборы и унификация

Для иллюстрации того, как *Prolog*-программа создает ответы, рассмотрим следующую простую программу регистрации данных (это не функции):

Листинг:

```
/* program P                                cause # */  
p(a).  
p(X) :- q(X), r(X).  
                                  /* #1 */  
                                  /* #2 */
```

```

p(X) :- u(X).          /* #3 */

q(X) :- s(X).          /* #4 */

r(a).                  /* #5 */
r(b).                  /* #6 */

s(a).                  /* #7 */
s(b).                  /* #8 */
s(c).                  /* #9 */

u(d).                  /* #10 */

```

Упражнение 4.3.1.1 Загрузите программу **P** в *Prolog* и посмотрите что случится при вводе цели `?-p(X)`. Когда будет выведен ответ, нажмайте чтобы *Prolog* продолжил трассировку и нашел все ответы.

Упражнение 4.3.1.2 Загрузите программы, включите трассировку, и посмотрите что происходит при вводе той же цели. Нажмайте **Enter** в каждой строке трассировки, и в конце строки ответа, чтобы найти все ответы. Используйте `?-help(trace)` если необходимо.

Листинг 2: Трассировка

```

?- trace.
true.

[trace] ?- p(X).
Call: (6) p(_G2873) ? [Enter] creep
Exit: (6) p(a) ? [Enter] creep
X = a ; []
Redo: (6) p(_G2873) ? creep
Call: (7) q(_G2873) ? creep
Call: (8) s(_G2873) ? creep
Exit: (8) s(a) ? creep
Exit: (7) q(a) ? creep
Call: (7) r(a) ? creep
Exit: (7) r(a) ? creep
Exit: (6) p(a) ? creep
X = a ;
Redo: (8) s(_G2873) ? creep
Exit: (8) s(b) ? creep

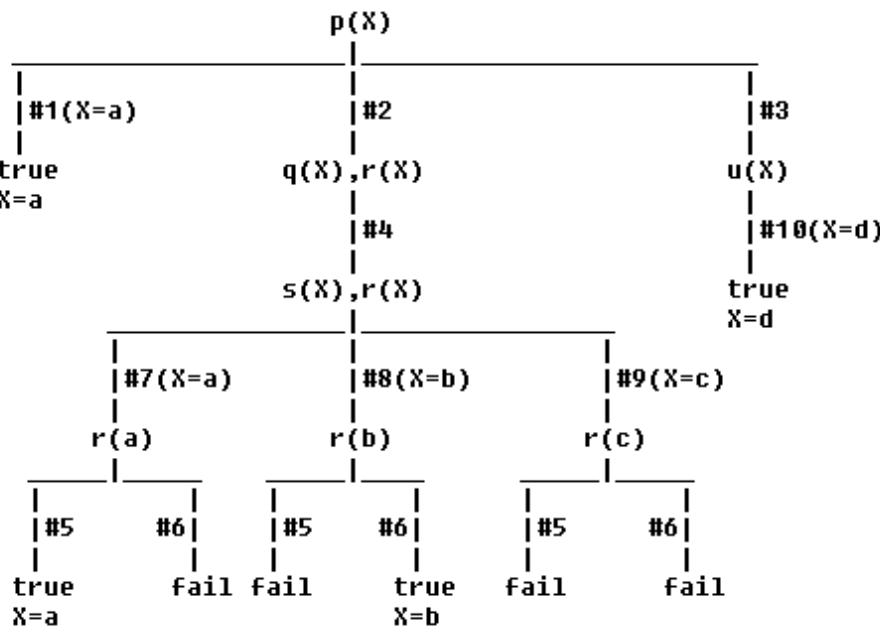
```

```

Exit: (7) q(b) ? creep
Call: (7) r(b) ? creep
Exit: (7) r(b) ? creep
Exit: (6) p(b) ? creep
X = b .

```

Следующая диаграмма показывает полное **дерево вывода** для цели $?-p(X)$. Ребра помечены номером утверждения в исходном файле программы **P**, которое было использовано для подмены цели подцелями. Прямые потомки под каждой (под)целью в дереве вывода соответствуют **вариантам выбора**. Например корневая цель $p(X)$ **унифицируется** заголовками утверждений #1, #2, #3, порождая три выбора.

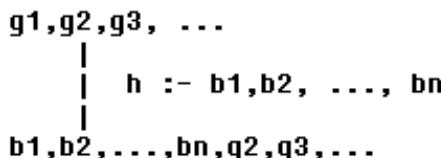


Трассировка упражнения 4.3.1.2 для цели $?-p(X)$ соответствует обходу дерева вывода вглубь. Каждый узел дерева вывода *Prologa* в определенный момент времени становится текущей целью. Аналогично каждый узел — последовательность субцелей. Ребра сразу ниже узла соответствуют доступным выборам замены для текущего узла. Текущий side clause, номер которого отмечает дугу в дереве вывода²⁴, тестируется следующим способом: если самая левая подцель текущего узла²⁵ унифицируется головой side clause²⁵, затем самая левая подцель заменяется телом side clause²⁶. Графически мы можем это показать вот так:

²⁴ отмечена как **g1** в небольшой диаграмме ниже

²⁵ отмечена как **h** в диаграмме

²⁶ **b1,b2,...,bn**



Одна важная вещь не показана в диаграмме — логические переменные в результирующей цели $b_1, b_2, \dots, b_n, g_2, g_3, \dots$ были привязаны в результате унификации, и *Prolog* требует отслеживать эти унифицирующие подстановки, в процессе роста дерева вывода вниз, во всех ветках.

Итак, обход дерева вывода вглубь значит что альтернативные варианты выбора будут проверены тогда, когда поиск возвратиться в точку ветвления, содержащую этот выбор. Процесс называется ***backtracking***.

Естественно, если хвост цели пуст, самая левая подцель эффективно удаляется. Если все подцели могут быть удалены по одному из путей дерева вывода, то находится ответ, и возвращается результат ***true***. В этой точке привязки переменных могут быть использованы длядачи ответа на оригинальный запрос.

Унификация термов *Prologa*

Prolog unification matches two Prolog terms T1 and T2 by finding a substitution of variables mapping M such that if M is applied T1 and M is applied to T2 then the results are equal.

For example, Prolog uses unification in order to satisfy equations (T1=T2) ...

```
?- p(X,f(Y),a) = p(a,f(a),Y).
X = a    Y = a
```

```
?- p(X,f(Y),a) = p(a,f(b),Y).
```

No

In the first case the successful substituton is {X/a, Y/b}, and for the second example there is no substitution that would result in equal terms. In some cases the unification does not bind variables to ground terms but result in variables sharing references ...

```
?- p(X,f(Y),a) = p(Z,f(b),a).
X = _G182    Y = b    Z = _G182
```

In this case the unifying substitution is {X/_G182, Y/b, Z/_G182}, and X and Z share reference, as can be illustrated by the next goal ...

```
?- p(X,f(Y),a) = p(Z,f(b),a), X is d.
X = d    Y = b    Z = d
```

{X/_G182, Y/b, Z/_G182} was the most general unifying substitution for the previous goal, and the instance {X/d, Y/b, Z/d} is specialized to satisfy the last goal.

Prolog does not perform an occurs check when binding a variable to another term, in case the other term might also contain the variable. For example (SWI-Prolog) ...

```
?- X=f(X).  
X = f(**)
```

The circular reference is flagged (**) in this example, but the goal does succeed {X/f(f(f(...)))}. However ...

```
?- X=f(X), X=a.  
No
```

The circular reference is checked by the binding, so the goal fails. "a canNOT be unified with f(_Anything)" ...

```
?- a \=f(_).  
Yes
```

Some Prologs have an occurs-check version of unification available for use. For example, in SWI-Prolog ...

```
?- unify_with_occurs_check(X,f(X)).  
No
```

The Prolog goal satisfaction algorithm, which attempts to unify the current goal with the head of a program clause, uses an instance form of the clause which does not share any of the variables in the goal. Thus the occurs-check is not needed for that.

The only possibility for an occurs-check error will arise from the processing of Prolog terms (in user programs) that have unintended circular reference of variables which the programmer believes should lead to failed goals when they occur . Some Prologs might succeed on these circular bindings, some might fail, others may actually continue to record the bindings in an infinite loop, and thus generate a run-time error (out of memory). These rare situations need careful programming.

It is possible to mimic the general unification algorithm in Prolog. But here we present a specialized version of unification, whose computational complexity is linear in the size of the input terms, and simply matches terms left-to-right. The match(+General predicate attempts to match its first argument (which may contain variables) against its second argument (which must be grounded). This little program should be considered just as an illustration, or a programming exercise, although we do know of cogent applications for the LR matching algorithm in situations where general unification is not needed. We would not use match, however, in a Prolog application because built-in unification would be so much faster; we would simply have to ensure that the

presuppositions for match are appropriately checked when built-in unification is used. The reference Apt and Etalle (1993) discusses the question in general regarding how much of general unification is actually NOT needed by Prolog.

```
%%%%%%%%%%%%%%%%
%% match(U,V) :
%%   U may contain variables
%%   V must be ground
%%%%%%%%%%%%%%%
% match a variable with a ground term
match(U,V) :-  
    var(U),  
    ground(V),  
    U = V. % U assigned value V

% match grounded terms
match(U,V) :-  
    ground(U),  
    ground(V),  
    U == V.

% match compound terms
match(U,V) :-  
    \+var(U),  
    ground(V),  
    functor(U,Functor,Arity),  
    functor(V,Functor,Arity),  
    matchargs(U,V,1,Arity).

% match arguments, left-to-right
matchargs(_,_,N,Arity) :-  
    N > Arity.
matchargs(U,V,N,Arity) :-  
    N =< Arity,  
    arg(N,U,ArgU),  
    arg(N,V,ArgV),  
    match(ArgU,ArgV),  
    N1 is N+1,  
    matchargs(U,V,N1,Arity).
```

4.3.2 3.2 Cut

The Prolog **cut** predicate, or **!**, eliminates choices is a Prolog derivation tree. To illustrate, first consider a cut in a goal. For example, consider goal `?-p(X),!.` for the

The cut goal succeeds whenever it is the current goal, AND the derivation tree is trimmed of all other choices on the way back to and including the point in the derivation tree where the cut was introduced into the sequence of goals.

For the previous derivation tree, this means that the branches labeled #2 and #3 are eliminated, and hence the entire subtrees below these two edges are also cut off. This then corresponds to only producing the first answer X=a:

```
?- p(X),!.
X=a ;
no
```

Here we tried to get Prolog to find some more answers using ';' but they have already been cut off. Consider also:

```
?- r(X),!,s(Y).
X=a Y=a ;
X=a Y=b ;
X=a Y=c ;
no
```

Note that there is no backtracking into that first goal. Also,

```
?- r(X), s(Y), !.
X=a Y=a ;
no
```

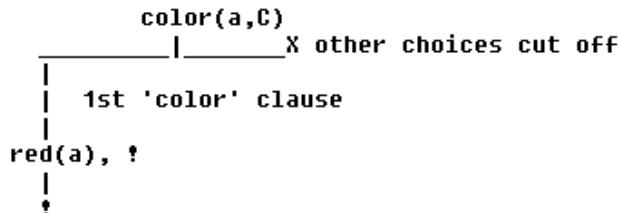
as expected.

Suppose that a cut occurs in the body of the program. The cut rule (above) still applies when the cut appears as a called subgoal. The cut is used in the body of a given clause so as to avoid using clauses appearing after the given clause in the program. To illustrate, consider the following program:

```
part(a). part(b). part(c).
red(a). black(b).
color(P,red) :- red(P),!.
color(P,black) :- black(P),!.
color(P,unknown).
```

The intention is to determine a color for a part based upon specific stored information or else conclude that the color is 'unknown' otherwise.

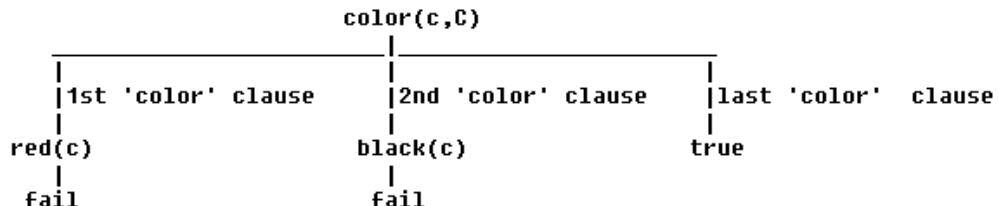
A derivation tree for goal ?- color(a,C) is



which corresponds with

```
?- color(a,C).
C = red
```

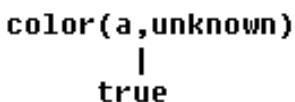
and a derivation tree for goal ?- color(c,C) is



which corresponds with

```
?- color(c,C).
C = unknown
```

The Prolog cut is a procedural device for controlling goal satisfaction. The use of cut affects the meanings of programs. For example, in the 'color' program, the following program clause tree says that 'color(a,unknown)' should be a consequence of the program:



4.3.3 3.3 Meta-interpreters in *Prolog*

4.4 4. Built-in Goals

- 4.4.1 4.1 Utility goals
- 4.4.2 4.2 Universals (true and fail)
- 4.4.3 4.3 Loading *Prolog* programs
- 4.4.4 4.4 Arithmetic goals
- 4.4.5 4.5 Testing types
- 4.4.6 4.6 Equality of *Prolog* terms, unification
- 4.4.7 4.7 Control
- 4.4.8 4.8 Testing for variables
- 4.4.9 4.9 Assert and retract
- 4.4.10 4.10 Binding a variable to a numerical value
- 4.4.11 4.11 Procedural negation, negation as failure
- 4.4.12 4.12 Input/output
- 4.4.13 4.13 *Prolog* terms and clauses as data
- 4.4.14 4.14 *Prolog* operators
- 4.4.15 4.15 Finding all answers

4.5 5. Search in *Prolog*

- 4.5.1 5.1 The A* algorithm in *Prolog*
- 4.5.2 5.2 The 8-puzzle
- 4.5.3 5.3 $\alpha\beta$ search in *Prolog*

4.6 6. Logic Topics

- 4.6.1 6.1 Chapter 6 notes
- 4.6.2 6.2 Positive logic
- 4.6.3 6.3 Convert first-order logic to normal form
- 4.6.4 6.4 A normal rulebase goal interpreter
- 4.6.5 6.5 Evidentiary soundness and completeness

Глава 5

ASTLOG: Язык для анализа синтаксических деревьев

¹ © Roger F. Crew <rfc@microsoft.com>
Microsoft Research Microsoft Corporation Redmond, WA 98052

Abstract

We desired a facility for locating/analyzing syntactic artifacts in abstract syntax trees of Си/ C_+^+ programs, similar to the facility **grep** or **awk** provides for locating artifacts at the lexical level. *Prolog*, with its implicit pattern-matching and backtracking capability is a natural choice for such an application. We have developed a *Prolog* variant that avoids the overhead of translating the source syntactic structures into the form of a *Prolog* database; this is crucial to obtaining acceptable performance on large programs. An interpreter for this language has been implemented and used find various kinds of syntactic bugs and other questionable constructs in real programs like **Microsoft SQL server** (450Klines) and **Microsoft Word** (2Mlines) in time comparable to the runtime of the actual compiler.

The model in which terms are matched against an implicit current object, rather than simply proven against a database of facts, leads to a distinct “inside-out functional” programming style that is quite unlike typical *Prolog*, but one that is, in fact, well-suited to the examination of trees. Also, various second-order *Prolog* set-predicates may be implemented via manipulation of the current object, thus retaining an important feature without entailing that the database be dynamically extensible as the usual implementation does.

¹ © <http://www.cs.nyu.edu/~lharris/papers/crew.pdf>

5.1 Introduction

Tools like **grep** and **awk** are useful for finding and analyzing lexical artifacts; e.g., a one-line command locates all occurrences of a particular string. Unfortunately, many simple facts about programs are less accessible at the character/token level, such as the locations of assignments to a particular C_+^+ class member. In general, reliably extracting such syntactic constructs requires writing a parser or some fragment thereof. And after writing one's twenty-seventh parser fragment, one might begin to yearn for a more general tool capable of operating at the syntax-tree level.

Even given a compiler front-end that exposes the abstract syntax tree (AST) representation for a given program, there remains the question of what exactly to do with it. To be sure, supplying a Си programmer with a sufficiently complete interface to this representation generally solves any problem one might care to pose about it. One may just as easily say that all problems at the lexical level may be solved via proper use of the UNIX standard IO library `<stdio.h>`, a true, but utterly trivial and unsatisfying statement. The question is rather that of building a simpler, more useful and flexible interface: one that is less error-prone, more concise than writing in Си, and more directly suited to the task of exploring ASTs. We first consider a couple of prior approaches.

5.1.1 The **awk** Approach

One of the more popular approaches is to extend the **awk** [?] paradigm. An **awk** script is a list of pairs, each being a regular-expression with an accompanying statement in a C-like imperative language. For each line in the input file, we consider each pair of the script in turn; if the regular-expression matches the line, the corresponding statement is executed.

Extending this to the AST domain is straightforward, though with numerous variations. One defines a regular-expression-like language in which to express tree patterns and an **awk**-like imperative language for statements. The tree nodes of the input program are traversed in some order (e.g., preorder), and for each node the various pairs of the script are considered as before.

We have two objections to this approach, the first having to do with the hardwired framework that usually implicit. In some cases (e. g., **TAWK** [?]), the traversal order for the AST nodes is essentially fixed; using a different order would be analogous to attempting to use plain **awk** to scan the lines of a text file in reverse order. In **A*** [?], while the user may define a general traversal order, only one traversal method may be defined/active at any given time, making difficult any structure comparisons between subtrees or other applications that require multiple concurrent traversals. Since the imperative language is quite general in both cases, little is deffinitively impossible, however for some applications one may be little better off than when programming in straight Си.

The second objection has to do with the kinds of pattern-abstraction available. Inevitably there exist simply-described patterns that are a poor fit to a regular-

expression-like syntax. This tends to happen when said simple descriptions are in terms of the idioms of a particular programming language; most of the various tree-**awk** pattern languages tend to be designed with the intent of being language independent.

Suppose one wishes to find all consecutive occurrences of one statement immediately preceding another, e. g., places where a given system call `syscall()`; is followed immediately by an `assert()`; ². A tree-regular-expression pattern of the form

```
<syscall() pattern>; <assert() pattern>
```

(where ; is the regular-expression sequence operator) finds all instances of the two calls occurring consecutively within a single block, but it misses instances like

```
syscall();  
{  
    assert();  
    ...  
}
```

and

```
if (...) {  
    syscall();  
}  
else {  
    ...  
}  
assert();
```

While the tree-**awk** languages allow one to write patterns to match each of these cases, without a pattern-abstraction facility, we may be back at square one when it comes time to look for some **different** pair of consecutive function calls. We prefer to write a single consecutive-statement pattern constructor **once** and then be able to use it for a variety of cases where we need to find pairs of consecutive statements satisfying certain criteria, invoking it as

```
follow_stmt(<syscall() pattern>, <assert() pattern>)
```

for the above problem, or, if we instead want to be finding all of the places where a C switch-case falls through, as

```
follow_stmt(not(<unconditional-jump pattern>),  
            <case-labeled stmt pattern>)
```

² on the theory that testing of outcomes of system calls should be done in production code rather than just debugging code

One solution, used by **TAWK**, is to use **cpp**, the C preprocessor, to preprocess the script, allowing for pattern-abstractions to be expressed as `#define` macros whose invocations are then expanded as needed. This is unsatisfactory in a number of ways, whether one wants to consider the problem of recursively-defined patterns, macros with large bodies that result in a corresponding blow-up in the size of the script, or the difficulty of tracing script errors that resulted from complex macro-expansions.

Another way out is to fall back on the procedural abstraction available in the imperative language that the patterns invoke. One essentially uses a degenerate pattern that always matches and then allows the imperative code to test whether the given node is in fact the desired match, defining functions to test for particular patterns. Once again, it seems we are back to programming in straight C and not deriving as much benefit from having a pattern language available as we could be.

In general, the philosophical underpinning of the **awk** approach is that the designer has already determined the kinds of searches the user will want to do; the effort is put towards making those particular searches run efficiently. There is also an assumption that the underlying imperative language for the actions has all the abstraction facilities one will ever need, so that if the pattern language is lacking in various ways, this is not deemed a serious problem. While this is not an unreasonable approach, we have less confidence of having identified all of the reasonable search possibilities, and thus would prefer instead to make the pattern language more flexible and extensible, being willing to sacrifice some efficiency to do so.

5.1.2 The Logic Programming Approach

Another common approach is to run an inference engine over a database of program syntactic structures [?, ?, ?]. *Prolog* [?] is a convenient language for this sort of application. Backtracking and a form of pattern matching are built in, the abstraction mechanisms to build up complex predicates exist at a fundamental level, and finally, *Prolog* allows for a more declarative programming style.

The problems with using *Prolog* are two-fold. First there is the issue of efficiency. Second, we must represent the AST for our source program in the *Prolog* database. Large programs ($10^5..10^6$ lines) will result in correspondingly large *Prolog* databases, most likely with a significant performance penalty.

We finesse the second problem by not attempting to import the source program's AST at all, instead opting to modify the interpretation of the predicates and queries of *Prolog* so as to be applicable to external objects rather than just facts provable in the existing database. Removing reasons that require the database to grow beyond the initial script creates significant opportunities for optimization. This, however, requires removing primitives like `assert()` and `retract()` that allow for the dynamic (re)definition or removal of predicates, which in turn removes many higher-order logical features that are defined in terms of them. Fortunately, some of the more essential ones can be restored at relatively little cost.

5.2 Elements of ASTLOG

Section 5.2 gives the complete syntax for our language, ASTLOG. The ASTLOG interpreter reads a script of user-defined predicate operator definitions and then runs one or more queries.

As in *Prolog*, the definition of a user-defined predicate operator is composed of one or more *clauses*. A compound term `opname(term, ...)` appearing at top level in a clause body is interpreted as a predicate, whether `opname` be primitive or user-defined. In the latter case, the script is searched for a defining clause whose head terms successfully unify with the respective operand terms of the given compound term, variables are bound accordingly, and the terms of the clause body are likewise interpreted. The clause **succeeds** (i. e., is found to be true) if all of its body terms succeed. Whenever a clause head fails to unify, or a clause body term **fails** (i. e., is found to be false), or any primitive term fails by the rules of evaluation of that primitive, we backtrack to the last point where there was a choice (e. g., of clauses to try for a given compound term) and continue.

A *query* is a clause whose head terms are all variables. Ultimately, whenever all terms of a query body succeed, the bindings of any variables listed in the query head (*qhead*) are reported. Otherwise, we report failure. Thus far, this is all exactly like *Prolog*.

Figure 1: Complete Syntax of ASTLOG

script	::= named-clause*	script file syntax
query	::= imports? (varname*) clause-body ;	query syntax
imports	::= { varname+ }	
named-clause	::= opname anon-clause	
anon-clause	::= (term*) clause-body? ;	
clause-body	::= <- term+	
<hr/>		
Essential Term Syntax		
term	::= literal	reference to denotable object
	::= varname	
	::= opname (term*)	compound term
	::= FN imports? (anon-clause+)	anonymous predicate-operator-valued (“lambda”) term
	::= ’ opname arity-spec?	named predicate-operator-valued (“function quote”) term
	::= (term)(term*)	“application” term

Gratuitous Term Syntax

<code>::= # constname</code>	named constant (\equiv corresponding literal number)
<code>::= [term*]</code>	<code>[]</code> \equiv nil(), <code>[term]</code> \equiv cons(term; nil()), etc..
<code>::= [term+ term]</code>	<code>[term1 term2]</code> \equiv cons(term1,term2), etc.
arity-spec	<code>::= / integer</code>

5.2.1 Objects

ASTLOG refers to external objects. Given a Cи/C₊⁺ compiler front end that provides a (C₊⁺) interface to the syntactic/semantic data structures built during the parse of a given program, it is simple to graft this onto the core of ASTLOG so that it may recognize object references corresponding to

- whole C/C++ programs,
- single files,
- symbols,
- AST nodes (including statements, expressions, and declarations), and
- Cи/C₊⁺ type descriptions.

For the purposes of ASTLOG, an *object* is simply some external entity that is significant for its identity and for the primitive predicates that it may satisfy. To simplify the language we regard the traditional constants (integers, floats, and strings) to be references to “external” objects as well, though one could just as easily take the converse view in which the universe of object references is just a (very large) pool of constants³.

In any case, object references are terms in ASTLOG. Only references to equal objects can unify, equality meaning numeric equality for numbers, same-sequence-of-characters for strings, and identity for all other classes of objects. Only objects that have denotations (numbers, strings and the unique `null object*`) can find their way into scripts.

5.2.2 The Current Object

The first significant departure from the *Prolog* model is that a query or predicate term always evaluates under an ambient *current object*. Every query and every term being evaluated as a predicate is not so much a standalone statement that may or may not be intrinsically true (i. e., provable from the “facts” in the script) as it is a specification that may or may not be satisfied by the current object, or, alternatively, a *pattern* that may or may not *match* the current object. For example, in *Prolog*

```
odd(3)
```

always succeeds by virtue of 3 being odd or because the “fact” `odd(3)` exists in the script somewhere. By contrast, in ASTLOG

³ “atoms” in the usual *Prolog* terminology

`odd()`

succeeds if the current object happens to be the integer 3, fails if the current object is 4, and raises an error if the current object is the string "Hi mom". Another way to view this is that every predicate term takes an extra, hidden current-object operand.

While one normally only expects to see compound (and application) terms in predicate position, ASTLOG allows variables and object references there as well. The rules for matching are as follows:

- An object reference matches the current object, if it references an equal object.
- A bound variable matches according as whatever term it is bound to.
- An unbound variable gets bound to reference the current object (and thus automatically matches it).
- A compound term whose operator is defined via clauses matches if there exists a clause whose head operands unify with the term operands and whose body terms themselves all match the current object.

Section 5.3.1 describes the operator-valued and application terms.

The evaluation rules for compound terms having primitive operators are widely varied, however the operands are usually treated one of two ways:

1. (`foo-pred`) requiring the operand to be match some object⁴, not necessarily the same current object as that which the full term is being matched against. For example, the operand of `strlen` (see 5.2.2) and the second operand of `with` are treated this way.
2. (`foo`) requiring the operand be an object reference, whether this be a literal or an object-reference-bound variable. The operands of `re`, `gt`, and the first operand of `with` are treated this way.

Most primitives also expect a current object to be of a particular kind and raise an error if confronted with something different.

The use of an implicit current object is not by itself an increase in expressivity. If we had, in a *Prolog* database, terms representing the various AST nodes, there would be a fairly straightforward translation of ASTLOG terms into *Prolog* terms, one in which we simply modify all terms to make the current object an explicit operand.

Nevertheless, ASTLOG programs exhibit a distinct style of programming. Consider as an example that we might, in a typical functional language, write a function call

`strlen(string)`

⁴ which becomes the current object for that evaluation

to find the length of the string returned by the expression `string`. Here the length result is implicitly returned to the context of the call. In *Prolog*, the natural style would be to express this as a relation

```
strlen(string, length)
```

which stipulates that `length` is in fact the length of `string`. In ASTLOG, we would write

```
strlen(length-pred)
```

where now it is the `string` argument that is implicitly supplied **as the current object** by the context while the length result is returned *to* the subterm `length-pred`, which in turn can be some arbitrary term expecting a numeric current object as its implicit argument. For example, given an `odd()` predicate as above, the term `strlen(odd())` would match any string consisting of an odd number of characters. It is this “inside-out functional” evaluation strategy that makes ASTLOG well-suited to constructing anchored patterns to match tree-like structures.

Figure 2: Some core ASTLOG primitives

- `and(object-pred, ...)`
The current object satisfies every `object-pred` operand.
- `or(object-pred, ...)`
The current object satisfies some `object-pred` operand.
- `if(object-pred, then-pred, else-pred)`
The current object satisfies `then-pred` or `else-pred` according as it satisfies or fails to satisfy `object-pred` (once; if `object-pred` matches but `then-pred` does not, we do not retry `object-pred`).
- `not(object-pred)`
 $= \text{if}(\text{object-pred}, \text{or}(), \text{and}())$
- `with(object, object-pred)`
`object` satisfies `object-pred` (outer current object is ignored).
- `strlen(integer-pred)`
The current string object has length satisfying `integer-pred`.
- `re(string)`
The regular expression `string` matches the current string.
- `gt(integer)`
The current integer is greater than `integer`.

- `minus(integer-pred, integer)`
`integer-pred` matches the current integer + `integer`.
- `minus(integer, integer-pred)`
`integer-pred` matches `integer` — the current integer.
(An error is raised if neither operand of a minus term is an integer object reference.)
- `plus(integer-pred, integer)`
`integer-pred` matches the current integer — `integer`

Figure 3: Some primitive node and symbol predicates

- `parent(ast-pred)`
This AST node is not a root node and its parent satisfies `ast-pred`.
- `kid(integer-pred; ast-pred)`
This AST node has a child satisfying `ast-pred` whose (0-based) index satisfies `integer-pred`.
- `kidcount(integer-pred)`
The number of children of this AST node satisfies `integer-pred`.
- `op(integer-pred)`
The opcode of this AST node satisfies `integer-pred`.
- `atype(type-pred)`
This AST node has a return type satisfying `type-pred`.
- `asym(symbol-pred)`
This AST node is a symbol satisfying `symbol-pred`.
- `aconst(const-pred)`
This AST node is a constant (integer, float or string) satisfying `const-pred`.
- `sname(string-pred)`
This symbol's name satisfies `string-pred`.

There are named constants available for designating the opcodes of various kinds of nodes for use in `op()` terms, and the indices of particular children for use in `kid()`.

5.2.3 Examples

Given the set of AST node primitives in Figure 3, we could write

```
and(op(#=), kid(#LEFT, asym(sname("foo"))))
```

which would be satisfied by any AST node that is an assignment (=) expression whose left-hand side is itself a symbol expression where the symbol name is "foo". Here, #= and #LEFT are numeric literals for the assignment node opcode and the assignment target's childindex, respectively.

To define a predicate `assignment/2` to match assignment nodes, a script could include the clause

```
assignment(target, value)
  <- op(#=),
      kid(#LEFT, target),
      kid(#RIGHT, value);
```

which would then allow writing the previous term as

```
assignment(asym(sname("foo")), _)
```

As in *Prolog*, the underscore (_) is “wild-card” variable, i.e., one that is internally given a distinct identity so as not to be conflated with any other instances of _. Such a variable, being guaranteed to be unbound, will match any object or unify with any term.

Defining a general purpose node-traversal predicate is also straightforward

```
somenode(pred)
  <- or(pred, kid(_, somenode(pred))));
```

Given this definition, an attempt to match `somenode(test)` to a given node will create an instance of the defining clause of `somenode/1` above with `pred` bound to `test`. Satisfying the clause body requires that either `pred` match the current node, or, if (when) that fails, that `kid(_, somenode(pred))` match the current node. The latter in turn will attempt to match the variable `_` with 0 (easy) and the term `somenode(pred)` with the first child, and, when that fails, `_` with 1 and `somenode(pred)` with the second child, etc... Making the interpreter fail and backtrack after each hit (in the usual manner of *Prolog*) eventually causes `test` to be matched with the original node and all of its descendants.

So, if we issue the query

```
(v) <- somenode(
  assignment(asym(sname("foo")), v)
);
```

on the root node of some function’s AST, we obtain, via the successive bindings reported for `v` on each hit, all of the expressions assigned to variables named "foo" within that function.

As an example that makes less trivial use of backtracking, consider the problem of whether two trees have the same structure (i.e., root nodes have the same opcode and all corresponding children have the same structure).

```
sametree(node)
<- op(nodeop),
  with(node, op(nodeop)),
  not(and(with(node, kid(n, nkid)),
    kid(n, not(sametree(nkid)))));
```

This defines a predicate `sametree(node)` that holds if `node` is a reference to an AST node with the same structure as the current object. The first line of the clause body binds the current node's opcode to `nodeop`, the second line compares that to the opcode of `node`, while the remaining lines search for children whose subtrees have distinct structure. The term `kid(n, nkid)` will match each child of `node`, since both variables are initially unbound. If `sametree(nkid)` happens to be true of the corresponding child of the current node, the inner `not` fails and we go back and try another child of `node`. If `sametree(nkid)` happens to be true of **every** corresponding child of the current node, then the enclosing `not` and thus the outer `sametree(node)` invocation succeeds.

The preceding version of `sametree/1` is a purely structural comparison; there is no attempt to take account of the commutativity/associativity of the various operators, e. g., `a + b` and `b + a` are not considered the same. If, say, we **did** want to consider commutativity, we could define

```
csametree(node)
<- op(nodeop),
  with(node, op(nodeop)),
  kidcount(if(with(nodeop, commutes()),
    any_perm(perm),
    id_perm(perm))),
  not(and(with(node, kid(corresp(perm, n),
    nkid)),
    kid(n, not(csametree(nkid)))));
```

along with suitable definitions of

`commutes()`

the current integer is the opcode of a commutative operator,

`any_perm(perm)`

`perm` is any permutation of the sequence
`0, ... , (<current-object> - 1),`

`id_perm(perm)`

`perm` is the identity permutation of the sequence
`0, ... , (<current-object> - 1),`

`corresp(perm, n)`

permutation `perm` takes the current integer to something matching `n`.

Here, permutations can be represented by list terms. Note that since all of the commutative C_+^+ operators are, in fact, binary, this all simplifies significantly.

It should, incidentally, be clear that there is nothing about the core language that is specifically tailored for the examination of compiler-produced ASTs, let alone ASTs for a given language. The language in fact lends itself to the examination of a wide variety of external structures, e. g., hierarchical file systems, or collections of web pages. All that is needed is a suitable collection of primitive ASTLOG predicates for querying said structures.

Figure 4: Actual ASTLOG code for follow_stmt

Actual ASTLOG code for `follow_stmt` and how one uses it to find case statement fallthroughs. The `cond` operator is an if-then-elseif- construct, that is, `cond(p1, e1, p2)` is equivalent to `if(p1, e1, if(p2, e2, ..., e))`. `sfa(emit(string))` always succeeds and, as a side-effect, emits the source location of the current AST node in grep-output form.

```
follow_stmt.astlog
// FOLLOW_STMT(P1 P2)
//      <=> P1 and P2 are true of consecutive statements in this AST

follow_stmt(p1, p2)
<- if(op(#FUNCTION),
      kid(#FUNCTION/BODY, follow_stmt(p1, p2, *)),
      follow_stmt(p1, p2, *));

follow_stmt(p1, p2, after)
<- cond(op(#BLOCK), follow_block_stmt(p1, p2, after),
       op(#IF), kid(not(#IF/PRED), follow_stmt(p1, p2, after)),
       op(#SWITCH), kid(#SWITCH/BODY, follow_stmt(p1, p2, after)))

       op(#WHILE), follow_iter_stmt(#WHILE/BODY, p1, p2, after),
       op(#DO), follow_iter_stmt(#DO/BODY, p1, p2, after),
       op(#FOR), follow_iter_stmt(#FOR/BODY, p1, p2, after),

       or(op(#LABEL), op(#CASE), op(#DEFAULT)),
           kid(#LABELSTMT/STMT, follow_stmt(p1, p2, after)),

       follow_simple_stmt(p1, p2, after));

follow_simple_stmt(p1, p2, after)
<- with(after, not(*)), p1, with(after, first_stmt(p2));

follow_iter_stmt(nbody, p1, p2, after)
<- or(follow_simple_stmt(p1, p2, after),
      and(this, kid(nbody, follow_stmt(p1, p2, this))));
```

```

follow_block_stmt(p1, p2, after)
<- and(kid(minus(next,1), first),
       if(kid(next, second),
          with(first, follow_stmt(p1, p2, second)),
          with(first, follow_stmt(p1, p2, after))));

first_stmt(p)
<- if(op(#BLOCK),
      kid(0, first_stmt(p)),
      stmt);

// CASEFALL()
// emits all locations of switch-case fallthroughs in this AST tree
casefall()
<- follow_stmt(and(not(op(or(#BREAK,#CONTINUE,#GOTO,#RETURN))),
                   op(#CASE)),
               with(first, sfa(emit("Fall through to next case."))));


```

Figure 5: Definition of flatten

```

flatten(test, lst)
<- flatten(test, lst, []);

flatten(test, head, tail)
<- if(test,
      first(head, hrest),
      unify(head, hrest)),
      flattenkids(test, 0, hrest, tail);

flattenkids(test, n, head, tail)
<- if(kid(n, flatten(test, head, mid)),
      and(with(n, minus(nplus1,1)),
          flattenkids(test, nplus1,
                      mid, tail)),
      unify(head, tail));

first([o|rest],rest) <- o;
unify(x,x);

```

Figure 6: Parameterized version, flatten2

```

flatten2(test, lst)
<- flatten2(test, lst, []);

```

```

flatten2(test, head, tail)
<- if((test)(value),
      unify(head, [value|hrest]),
      unify(head, hrest)),
   flatten2kids(test, 0, hrest, tail);

flatten2kids(test, n, head, tail)
<- if(kid(n, flatten2(test, head, mid)),
      and(with(n, minus(nplus1,1)),
          flatten2kids(test, nplus1,
                      mid, tail)),
      unify(head, tail));

unify(x,x);

```

5.3 Higher order features

We have already included some of the non-1st-order features of *Prolog*, notably “cut” (in the form of `if()`) and the corresponding notion of negation, `not()`. There are others that turn out to be essential as well.

5.3.1 3.1 Lambdas and Applications

One may observe that, in `somenode(test)`, because this is an existential query, it does not matter that we are matching the same term `test` to every node of the tree. If variables in `test` get bound as a result of matching a given node, those bindings will be undone prior to advancing to the next node.

If one instead wants to write a conjunctive predicate over all tree nodes, say

```
flatten(test, list)
```

which holds if `list` is a list of **all** descendant nodes satisfying `test`, — we give a definition in Figure 5 — this will not work correctly if `test` contains any variables that are bound during the course of matching any node; said variables will **stay** bound for the duration of the `flatten` evaluation.

Even in an existential query, there is the possibility that the `test` being passed in will itself need to take a parameter. For example, one might imagine defining a version of `sametree` that also requires an additional user-specified `test` to hold at each corresponding pair of nodes. If `test` is a mere compound term, it can be matched against one of the nodes, but not both.

Thus we introduce **“application” terms** and operator-valued **“lambda” terms**. For an application `(fterm)(term;...)` to match the current object, the term `fterm` must be (or be a variable bound to) a predicate-operator-valued term, which will either be

- a reference, `'foo/3` to a named predicate operator, in which case the application evaluates exactly as the corresponding compound term would, or
- an anonymous predicate operator `FN{importvars ... } (anon-clauses ...)`, in which case the application evaluates **almost** exactly as if there were a named predicate-operator defined by the given clauses and this were a compound term on that operator. The difference is that any variables of those clauses that are also on the `{importvars... }` list are identified with the correspondingly-named variables in the clause where the `FN` term occurs lexically.

An `FN` term with imports can be thought of as a kind of ***closure***.

The parameterized version of flatten, namely

```
flatten2(test, list)
```

which holds iff `list` is a list of all `x` corresponding to descendants that `(test)(x)` matches, is defined in Figure ??.

Figure 7: Parameterized version of sametree

```
sametree(node, equiv)
<- unify(same,
FN{same,equiv}
((node)
<- op(nodeop),
with(node,op(nodeop)),
(equiv)(node),
not(and(with(node,kid(n,nkid)),
kid(n,not((same)(nkid)))))),
(same)(node);
```

The parameterized version of `sametree` is invoked as

```
sametree(node, equiv)
```

which holds iff `node` is a reference to an AST node with the same tree structure as the current node and, for every descendant `n` of `node`, the corresponding node in the current tree satisfies `equiv(n)`; this predicate is defined in Figure 5.3.1. This definition demonstrates the use of import lists, both to define a recursive anonymous predicate, and to make `equiv` available at once to all evaluations of that predicate. Given that definition, the following

```
sametree(node,
FN((n) <- if(aconst(c),
with(n, aconst(c)),
and());))
```

would then test whether the current tree has the same structure as underneath `node` and such that all corresponding constants are the same.

Figure 8: Embedded Query State Primitives

query(fterm; query-pred)

The embedded query state object created from fterm satisfies query-pred.

qnext(pred; thisquery-pred; nextquery-pred)

If the current embedded query state is a failure, pred is true, otherwise the current object satisfies this query-pred and, after the embedded query is advanced to the next hit or to failure, the resulting query state satisfies nextquery-pred.

qget(object-pred;:::)

Each object-pred matches the object bound to the corresponding variable in the head of the embedded query corresponding to the current query state object. An error will be raised if the embedded query has failed or if any head variable is not bound to an object.

5.3.2 Queries as Objects

Sometimes one wishes to build a collection or some other kind of aggregate of all objects found by a query. Unfortunately, when backtracking to get to the next hit, information about the previous hit will generally be lost. One solution is to rewrite the query into a conjunctive form, as we did in the previous section converting writing flatten as a conjunctive version of somenode (see Figure 5.2.3). We can already see that even in simple cases this process can be non-trivial and is not readily generalized.

It may also be the case for some conjunctive queries that they require memory proportional to the size of the data structure being searched, instead of merely memory proportional to the depth of the data structure. Judicious use of if() | astlog's moral equivalent of the cut operator | can avoid this, but this is sometimes cumbersome to get right.

As it happens, Prolog provides a number of setpredicates for accumulating query results. For example,

bagof(x, term, list)

binds list to a list of the bindings of x corresponding to each instance where term holds true. Unfortunately, this is usually implemented in terms of assert and retract, meaning we would have to abandon the idea of keeping our script small and fixed. Even just adding this as a new primitive is dubious if we have to add, say, another new primitive to merely count query hits, and yet more new primitives for each accumulation method anyone dreams up.

The key observation is that the execution model of astlog allows for the possibility of treating some subset of its own internal structures as "external" objects which can then serve as the current object of various kinds of queries. To be sure, some care needs to be exercised, since the internal structures of astlog are not static the way the program asts are. We can however, take a query whose hits we wish to accumulate, and

encapsulate its state after a given hit as an astlog object. Such an embedded query in a given state can now be the current object for the evaluation of some other predicate term. We thus only need to provide suitable primitive predicates applicable to query-state objects that may be used in such a term. Figure 5.3.1 lists these primitives.

Figure 9: Query Accumulators qcount and qlist

```

qcount(n) <- qcount(0, n);
qcount(sofar, return)
<- qnext(unify(sofar, return),
with(sofar, minus(sofarp1,1)),
qcount(sofarp1, return));
qlist(lst)
<- qnext(unify(lst, []), 
qget(first(lst,rest)),
qlist(rest));
// utilities
first([o|rest],rest) <- o;
unify(x,x);

```

Using this mechanism, it is then possible to define a wide variety of accumulators of query results. Given an ast node, and a query to see if there exists a descendant satisfying `test(x)`

```
() <- somenode(test(x));
```

the corresponding query to count the number of descendants satisfying `test(x)` would be

```
(n) <- query(FN(() <- somenode(test(x)); ),
qcount(n));
```

where `qcount/1` is defined as in Figure 5.3.2. Evaluating the `query()` term starts an embedded query corresponding to the first operand and builds a query state object representing the resulting first state (first hit or failure). This object then becomes the current object to which we try to match `qcount(n)`. It is the `qnext()` term therein that does the actual work. If the query-state is a success state, we increment the count of hits thus far (`sofar`), advance the embedded query, and recursively try to match a `qcount` term to the new state. If the query-state is a failure, we unify the count of hits thus far with the `return` variable.

To build a list of bindings for `x` corresponding to the query hits, we can do

```
(list) <- query(FN((x) <- somenode(test(x)); ),
qlist(list));
```

which is essentially the same as before except that now `qlist(list)` uses `qget` to examine the query state. Since the embedded query has only one head variable `x`, the `qget` term must likewise have at most one operand.

Some care is required when using embedded queries to phrase them so that the head variables will always be bound to objects. `qget()` will in fact raise an error if a head variable is not bound to an object. This requirement is crucial since, with a non-object term, there is no guarantee that said term will remain intact when the embedded query backtracks to the next state. Better to keep terms constructed by an embedded query from polluting the outer world.

The mechanism is also somewhat impure in that evaluating a `qnext` on a given query state object essentially destroys that object. Subsequent attempts to match additional terms against that query state will raise an error since the state of a query is lost once we advance it.

5.4 Implementation

`astlog` has been implemented as an interpreter in roughly 11,000 lines of C_+^+ for the core `astlog` interpreter and supporting utilities. Another 1100 lines define the roughly 60 primitives and supporting structures to invoke the various functions of the AST library. Coverage of the library API is in not entirely complete, but it is sufficient to perform various interesting tasks:

- Finding all instances of a simple assignment expression (`=`) occurring in any boolean context, for example,

```
if ((major == SORTM)
|| (major == MEMORYM)
|| ((major == BUFFERM)
&& (minor = B_NOIO)))
```

- Finding all instances of an equality-test (`==`) or dereference expression occurring in any void context (i. e., where results are discarded); the converse to the previous problem.
- Finding all case statement fall-throughs, i. e., where the preceding statement is not a `break`.
- Finding various patterns of irreducible control-flow in functions.
- Obtaining all static call-graph edges.
- Computing the McCabe cyclomatic complexity [?] of a function. Our code to do so looks like

```

mccabe(n) <- query(
FN(()<- somenode(
op(or(#IF,#FOR,#DO,
#WHILE,#CASE,
#?,#||,#&&)));
qcount(minus(n,1))
);

```

which might be compared with the 17-line version in Aria [?]. Admittedly, fairness would probably entail including the definitions of somenode and qcount as well.

- Finding gaps (unused space due to alignment rules) in structure definitions; this is a matter of traversing C_{II} type structures rather than asts.

A typical running time (on a 200MHz Pentium P6 with 64meg of RAM) for a one-pass search that evaluates a simple predicate on every ast node in Microsoft **SQLserver** (roughly 450,000 lines, 4300 functions) is roughly 10 minutes, of which 7.5 minutes are taken up by the ast library building the actual trees. For Microsoft Word (roughly 2,000,000 lines) the corresponding times are 45-60 minutes of which about 30 minutes is taken up by the tree builder.

Though this dreadfully slow in comparison with grep, these times are arguably acceptable in comparison with the times taken by the actual compiler | what one might expect for a tool that requires the use of compiler's data structures. One is, of course, free to write arbitrarily non-linear programs in astlog, so there are no guarantees. In any case we would doubtless see a certain amount of speedup if we actually were to attempt some kind of compilation of the astlog code.

5.5 Conclusions and Future Work

We have described a language for doing syntax-level analysis for C/C++ programs, though the core language is, in fact, adaptable to many other kinds of structures. As with previous such tools, the utility to users who are thus no longer required to write their own parse/semantic-analysis phase is apparent. The contribution here is a pattern language sufficiently powerful to provide traversal possibilites beyond what is naturally available in prior awk-like frameworks while avoiding some of the inefficiencies of importing the entire program structure into a logical inference engine. The Pan work [?] stressed the need to partition code and data; this we have done in a rather straightforward way. The surprise is that the *Prolog* with-an-ambient-current-object model turns out to be so well suited to analyzing treelike structures.

To be sure, there are various rough edges:

1. As already noted, embedded queries are slightly unsafe; there may exist a more robust set of primitives to use. Some form of type inference to detect unsafe uses of qnext may also be worth considering. More generally, there is the issue

of typing of astlog expressions to reduce the incidence of unbound variables or objects of the wrong type appearing as operands where object-references of a particular type are required.

2. Occasionally, we run up against the generally cumbersome nature of arithmetic in Prolog, which is arguably worse in astlog. The “inside-out functional” nature of astlog may be good for ast patterns, but it can make arithmetic operations like

```
with(n; divide(minus(x; 1); 2))
```

downright unreadable. Algebraic syntax could help, e. g.,

```
with(n; (x - 1)=2)
```

but even so, one must stare at this pretty hard to realize that n is being multiplied by 2 and then incremented by 1.

One possibility is to complicate the language by introducing actual “forward” functional operator definitions. For example, with such forward operators for addition and multiplication, one could then write

```
with(2 n + 1; x)
```

where the appearance of the + (plus) term in a slot normally requiring an object reference invokes the forward return-value-to-context definition of the operator + to sum its operands rather than the usual “backward” return-value-to-operand definition (see Figure 2) in which one operand is treated as a predicate.

3. Though there is a surprising amount of mileage to be had via instantiating terms with unbound variables in them, there are those occasions when a genuinely mutable data structure is required. Fortunately, given the strong partition between the script/database and the objects, having mutable objects exist and primitives that side-effect them when they match would not disrupt astlog’s execution model.
4. Currently, new primitives need to be manually written. Given the current collection of macros available, this is not actually an arduous task. Still, while language-independence was not one of our priorities, given that the core language is rather language-independent anyway, one would hope for a more automatic means of adapting astlog to work with other language parsers, perhaps by adapting GENII [?] or some similar tool to generate code for the basic primitive predicate operators for a fresh language.

5.6 Acknowledgements

ASTLOG would not have been possible without the existence of an ast library for C/C++ implemented by the members of Program Analysis group at Microsoft Research particularly Linda O’Gara, David Gay, Erik Ruf and Bjarne Steensgaard. I would also like to thank Bruce Duba, Michael Ernst, Chris Ramming, and the conference reviewers for much useful commentary and discussion.

References

- AKW86** A. V. Aho, B. W. Kernighan, and P. J. Weinberger. The AWK Programming Language. Addison Wesley, Reading, MA, 1986.
- BCD88** P. Borras, D. Clement, Th. Despeyroux, J. Incerpi, G. Kahn, B. Lang, and V. Pascual. Centaur: The system. In Proceedings of the SIGSOFT/SIGPLAN Software Engineering Symposium on Practical Software Development Environments, Boston, MA, 1988.
- BGV90** Robert A. Ballance, Susan L. Graham, and Michael L. Van De Vanter. The pan language-based editing system for integrated development environments. In Proceedings of the 4th ACM SIGSOFT Symposium on Software Development Environments, pages 77..93, Irvine, CA, 1990.
- CMR92** Mariano Consens, Alberto Mendelzon, and Arthur Ryman. Visualizing and querying software structures. In Proceedings of the Fourteenth International ACM Conference on Software Engineering, pages 138..156, 1992.
- Dev92** Premkumar T. Devanbu. Genoa - a customizable, language-and-front-end independent code analyzer. In Proceedings of the Fourteenth International ACM Conference on Software Engineering, pages 307..319. ACM Press, 1992.
- DR96** Premkumar T. Devanbu and David S. Rosenblum. Generating testing and analysis tools with aria. ACM Transactions on Software Engineering and Methodology, 5(1):42..62, January 1996.
- GA96** William G. Griswold and Darren C. Atkinson. Fast, exible syntactic pattern matching and processing. In Proceedings of the IEEE Workshop on Program Comprehension. ACM Press, 1996.
- LR95** David A. Ladd and J. Christopher Ramming. A*: A language for implementing language processors. IEEE Transactions on Software Engineering, 21(11):894..901, November 1995.
- McC76** T. McCabe. A complexity measure. IEEE Transactions on Software Engineering, 2(4):308..320, December 1976.

Appendix

For those who would prefer to see a slightly more formal description, we include a brief outline of an operational semantics for astlog in Figure 10, one that bears some resemblance to the actual implementation.

For any given term that is not an object reference, one may imagine there being numerous instances of that term in existence at any given time. We differentiate the various instances by assigning each a unique frame identifier (f) which is only significant for its identity. A variable v occurring within a given term t may, for a particular instance $hf ; [[t]]_i$ of that term, be bound to some object o or other term instance $hf_0 ; [[t_0]]_i$, this being indicated by having a binding, i.e., one of $hf ; [[v]]_i \text{ if } o$ or $hf ; [[v]]_i \text{ if } hf_0 ; [[t_0]]_i$ present in the current binding stack, which in turn is nothing more than a list of bindings. The semantic function $vlookup(B ; hf ; [[t]]_i)$ returns

- $hf ; [[t]]_i$ itself if t is not a variable.
- ? if the variable t is not bound in B .
- o if $hf ; [[t]]_i \text{ if } o$ is in B
- $vlookup(B ; hf_0 ; [[t_0]]_i) \text{ if } hf ; [[t]]_i \text{ if } hf_0 ; [[t_0]]_i$ is in B .

At any given time, the full state of our abstract machine is described by a failure of the form $B ' C :: F$ which consists of

- the current binding stack B ,
- the current continuation $C = (o; f; g; C_0)$, which in turn consists of a current object o , a current frame identifier f , a current goal, usually a term, but this can also be one of the auxiliary goals “apply(…)” or “cut(…),” and finally another continuation C_0 to which we advance if the goal succeeds
- the next failure F , to which we advance if the current goal fails.

Note that unlike the case where the goal succeeds, failure may involve undoing one or more bindings; thus, a failure (F) contains its own binding stack (a subset of B) whereas the continuations (C, C_0) do not.

The bottom half of Figure 10 (partially) defines a transition relation between states of the abstract machine. Given an initial current object o and a query $[[query]]$ with n head variables, we take the initial state to be

`F0 = [] ‘ (o; f0; apply(f0; [[query]]; [[v1;:::; vn]]); yes) :: no` If there is a sequence of transitions `F0 ! B1 ‘ yes :: F\verb` then we have a hit and the various query head bindings are available as `vlookup(B1; hf0; [[vi]]i)` for $i = 1$. Likewise, if `Fk ! Bk ‘ yes :: Fk+1` then we have a $|(k + 1)\text{th}$ hit.

When we have a $(k + 1)\text{th}$ hit. The semantic function `mgu(B; f; [[t1;:::;tn]]; [01;:::;t0n]])` returns an augmented binding stack that includes `B` together with those additional bindings that make up the most general unifier of the respective term instances `hf ; [[t1]]i` with `hf 0 ; [[t01]]i`, etc:::. If there is no most general unifier, `mgu()` returns `ufail`.

In the actual implementation, because the script is unified, we may precompute at load time mgus of all pairs of same-operator-and-arity compound terms occurring in the script, making clause invocation no more expensive than a function call in many cases. We also omit the “occurs check” [?] for the run-time portion of unification (i.e., where we’re transitively following variable bindings), with the usual increase in speed and infinite-loop risk. Thus far, unification has played a somewhat smaller role in astlog scripts than expected, so there’s some question whether we need to be doing even this much.

As noted above objects only unify with equal objects. The idea of allowing an object to unify with a compound predicate term that matches it has been considered, but rejected due to the significant complications it would introduce. Also, once one has subgoals being attempted during the course of unification, the user’s control over evaluation order is drastically reduced, something to be avoided if one is interested in having users being able to write efficient scripts.

Figure 10: Outline of astlog Operational Semantics

Глава 6

Warren's Abstract Machine Абстрактная машина Варрена

¹

© Hassan Aït-Kaci <hak@cs.sfu.ca>
© David H. D. Warren

Предисловие к репринтному изданию

Этот документ — репринтное издание книги имеющей то же название, которая была опубликована MIT Press, в 1991 году с кодом ISBN 0-262-51058-8 (мягкая обложка) and ISBN 0-262-01123-9 (тканый переплет). Редакция книги MIT Press сейчас не передается, и права на издание были переданы автору. Оригинальная версия² была бесплатно доступна всем, кто хочет ее использовать в некоммерческих целях, с веб-сайта автора:

<http://www.isg.sfu.ca/~hak/documents/wam.html>

Сейчас ссылка недоступна, книга переехала на <http://wambook.sourceforge.net/>

Если вы используете ее, пожалуйста дайте мне знать кто вы и для каких целей хотите ее использовать.

Thank you very much.

Hassan Aït-Kaci
Burnaby, BC, Canada
May 1997

¹ © <http://wambook.sourceforge.net/>

² английская <http://wambook.sourceforge.net/>

Предисловие

Язык *Prolog* был задуман в начале 1970х Alain Colmerauer и его коллегами из Марсельского университета. Его реализация языка была первым практическим воплощением концепции *логического программирования*, предложенной Robert Kowalski. Ключевая идея логического программирования — вычисления могут быть выражены в виде контролируемого вывода (дедукции) из набора декларативных утверждений. Несмотря на то что эта область значительно развилась за последнее время, *Prolog* остается наиболее фундаментальным и широко известным языком логического программирования.

Первой реализацией *Prologa* был интерпретатор, написанный на Фортране членами группы Colmerauer. Несмотря на очень ущербную в некотором смысле реализацию, эта версия считается в некотором смысле первым камнем: она доказала жизнеспособность *Prologa*, помогла распространению языка, и заложила основные принципы реализаций *Prologa*. Следующим шагом возможно была *Prolog*-система для PDP-10, разработанная в Университете Эдинбурга мной и коллегами. Эта система построена на базе техник Марсельской реализации, с добавлением понятия компиляции *Prologa* в низкоуровневый язык (в случае PDP-10 это машинный код), а также различные техники экономии памяти. Позже я уточнил и абстрагировал принципы реализации *Prolog DEC-10* в то, что я называю **WAM** (Warren Abstract Machine).

WAM — абстрактная (виртуальная) машина с архитектурой памяти и набором команд, заточенных под язык *Prolog*. Она может быть эффективно реализована на широком наборе аппаратных архитектур, и служить целевой платформой для переносимых компиляторов *Prologa*. Сейчас она принимается как стандартный базис при реализации *Prologa*. Это конечно лично приятно, но неудобно в том, что WAM слишком легко принимается как стандарт. Несмотря на то что WAM явилась результатом длительной работы и большого опыта в реализации *Prologa*, это отнюдь не единственно возможный подход. Например, в то время как WAM применяет *копирование структуры*³ для представления *термов Prologa*, метод *общих структур*⁴, использованный в Марсельской и DEC-10 реализациях, все еще можно рекомендовать к применению. Как бы то ни было, я считаю WAM хорошей отправной точкой для изучения технологий реализации *Prolog*-машины.

К сожалению до сих пор не было хорошей книги для ознакомления с внутренним устройством WAM. Мой оригинальный технический отчет слишком сложен, содержит только скелетное описание *Prolog*-машины, и написан для опытного читателя. Другие работы обсуждают WAM с различных точек зрения, но все же не могут быть использованы в качестве хорошего вводного руководства.

Поэтому очень приятно видеть появление этого прекрасного учебника, написанного Hassan Aït-Kaci. Эту книгу приятно читать. Она объясняет WAM с большой ясностью и элегантностью. Я думаю что читатели, интересующиеся информа-

³ structure copying

⁴ structure sharing

тикой, найдут эту книгу очень стимулирующим введением в увлекательную тему — реализацию *Prologa*. Я очень благодарен Хассану за донесение моей работы до широкой аудитории.

© David H. D. Warren
Бристоль, UK
Февраль 1991

Реализация машины вывода на C_+^+

В перевод книги Варрена мной⁵ добавлен пример реализации виртуальной машины вывода на C_+^+ . Исходные тексты находятся в каталоге *prolog/warren/*. Для вставки отдельных частей исходника по ходу книги полные файлы *hpp.hpp* и *cpr.cpp* разделены на отдельные небольшие фрагменты в каталогах *hpp/* и *cpr/*. Файл сборки *prolog/warren/Makefile* содержит не только

Makefile: запуск программы и генерация *log.log*

```
log.log: ./exe.exe src.src
    ./exe.exe < src.src > log.log && tail $(TAIL) log.log
```

Makefile: типовой блок компиляции лексической программы

```
C = cpp.cpp ypp.tab.cpp lex.yy.c
H =_hpp.hpp ypp.tab.hpp
CXXFLAGS = -std=gnu++11 -DMODULE=\"$(notdir $(CURDIR))\" \
./exe.exe: $(C) $(H) Makefile
$(CXX) $(CXXFLAGS) -o $(C)
```

Makefile: генерация кода синтаксического анализатора

```
ypp.tab.cpp: ypp.ypp
bison $<
```

Makefile: генерация кода лексера

```
lex.yy.c: lpp.lpp
flex $<
```

но и скрипты склейки файлов исходников из частей в каталогах *hpp/* *cpp/* *mk/*.

hpp.hpp

```
#ifndef _H_WARREN
#define _H_WARREN
```

⁵ <dponyatov@gmail.com>

hpp.hpp: типовые #include

```
#include <iostream>
#include <sstream>
#include <cstdlib>
#include <vector>
#include <map>
using namespace std;
```

hpp.hpp: базовый класс для структур WAM

```
struct WAM {
    string val;
    WAM(string);
    virtual string dump(int=0);
};
```

hpp.hpp

```
#endif // _H_WARREN
```

cpp.cpp

```
#include "hpp.hpp"
int main() { return yyparse(); }
```

6.1 Введение

В 1983 году Дэвид Варрэн разработал абстрактную машину для реализации языка *Prolog*, содержащую специальную архитектуру памяти и набор инструкций [?]. Эта разработка стала известна как Warren Abstract Machine (WAM) и стала стандартом де-факто для реализаций компиляторов *Prologa*. В [?] Варрэн описан WAM в минималистичном стиле, который слишком сложен для понимания неподготовленным читателем, даже заранее знакомым в операциями *Prologa*. Слишком многое было несказаным, и very little is justified in clear terms⁶. Это привело к очень скучному количеству поклонников WAM, которые могли был похвастаться пониманием деталей ее работы. Обычно это были реализаторы *Prologa*, которые решили уделить необходимое время для обучения через делание и кропотливого достижения просветления.

6.1.1 Существующая литература

Свидетельством недостатка понимания может служить тот факт, что за первые шесть лет было крайне мало публикаций о WAM, не говоря о том чтобы формально доказать ее корректность. Кроме оригинального герметического доклада

⁶ David H. D. Warren поделился в частной беседе что он “чувствовал что WAM важна, но к деталям ее реализации вряд ли будет широкий интерес, поэтому он использовал стиль личных заметок”

Варрена [?], практически не было никаких официальных публикаций о WAM. Несколько лет спустя группой Аргонской Национальной Лаборатории был выпущен единственный черновой стандарт [?]. Но следует отметить что этот манускрипт был еще менее понятен, чем оригиналный отчет Варрена. Его недостатком была цель описать готовую WAM как есть, а не как пошагово трансформируемый и оптимизируемый проект.

Стиль пошагового улучшения фактически был использован в публикации David Maier и David S. Warren⁷ [?]. В этой работе можно найти описание техник компиляции *Prologa* похожие на принципы WAM⁸. Тем не менее мы считаем что эта похвальная попытка все еще страдает от нескольких недостатков, если его рассматривать как окончательный учебник. Прежде всего эта работа описывает собственный достаточно близкий вариант WAM, но строго говоря не ее саму. Так что описаны не все особенности WAM. Более того, объяснения ограничены иллюстративными примерами, и редко четко и исчерпывающие очерчивают контекст, в котором применяются некоторые оптимизации. Во-вторых, часть посвященная компиляции *Prologa*, идет очень поздно — в предпоследней главе, полагаясь в деталях реализации на свердетализированные процедуры на Паскалье, и структуры данных, последовательно улучшаемые в течение предыдущих разделов. Мы чувствуем что это уводит и запутывает читателя, интересующегося абстрактной машиной. Наконец, несмотря на то что публикация содержит серию последовательно улучшаемых вариантов реализации, этот учебник не отделяет независимые части *Prologa* в процессе. Все представленные версии — полные *Prolog*-машины. В результате, читатель интересующийся выбором и сравнением отдельных техник, которые он хочет применить, не может различить отдельные техники в тексте. По всей справедливости, книга Майера и С.Варрена имеет амбиции быть первой книгой по логическому программирования. Так что они совершили подвиг, охватывая так много материала, как теоретического так и практического, и даже включили техники компиляции *Prologa*. Более важно, что их книга была первой доступной официальной публикацией, содержащей реальный учебник по техникам WAM.

After the preliminary version of this book had been completed, another recent publication containing a tutorial on the WAM was brought to this author's attention. It is a book due to Patrice Boizumault [?] whose Chapter 9 is devoted to explaining the WAM. There again, its author does not use a gradual presentation of partial *Prolog* machines. Besides, it is written in French — a somewhat restrictive trait as far as its readership is concerned. Still, Boizumault's book is very well conceived, and contains a detailed discussion describing an explicit implementation technique for the `freeze` meta-predicate⁹.

Even more recently, a formal verification of the correctness of a slight simplification

⁷ Это другой человек, а не разработчик WAM, работа которого вдохновила S.Warren на исследования. В свою очередь достаточно интересно что David H. D. Warren позже работал над параллельной архитектурой реализации *Prologa*, поддерживая некоторые идеи, независимо предложенные David S. Warren.

⁸ chap.9

⁹ chap.10

of the WAM was carried out by David Russinoff [?]. That work deserves justified praise as it methodically certifies correctness of most of the WAM with respect to *Prolog*'s SLD resolution semantics. However, it is definitely not a tutorial, although Russinoff defines most of the notions he uses in order to keep his work self-contained. In spite of this effort, understanding the details is considerably impeded without working familiarity with the WAM as a prerequisite. At any rate, Russinoff's contribution is nevertheless a **première** as he is the first to establish rigorously something that had been taken for granted thus far. Needless to say, that report is not for the fainthearted.

6.1.2 Этот учебник

1.2.1 Disclaimer and motivation 5

The length of this monography has been kept deliberately short. Indeed, this author feels that the typical expected reader of a tutorial on the WAM would wish to get to the heart of the matter quickly and obtain complete but short answers to questions. Also, for reasons pertaining to the specificity of the topic covered, it was purposefully decided not to structure it as a real textbook, with abundant exercises and lengthy comments. Our point is to make the WAM explicit as it was conceived by David H. D. Warren and to justify its workings to the reader with convincing, albeit informal, explanations. The few proposed exercises are meant more as an aid for understanding than as food for further thoughts.

The reader may find, at points, that some design decisions, clearly correct as they may be, appear arbitrarily chosen among potentially many other alternatives, some of which he or she might favor over what is described. Also, one may feel that this or that detail could be “simplified” in some local or global way. Regarding this, we wish to underscore two points: (1) we chose to follow Warren's original design and terminology, describing what he did as faithfully as possible; and, (2) we warn against the casual thinking up of alterations that, although that may appear to be “smarter” from a local standpoint, will generally bear subtle global consequences interfering with other decisions or optimizations made elsewhere in the design. This being said, we did depart in some marginal way from a few original WAM details. However, where our deviations from the original conception are proposed, an explicit mention will be made and a justification given.

Our motivation to be so conservative is simple: our goal is not to teach the world how to implement Prolog optimally, nor is it to provide a guide to the state of the art on the subject. Indeed, having contributed little to the craft of Prolog implementation, this author claims glaring incompetence for carrying out such a task. Rather, this work's intention is to explain in simpler terms, and justify with informal discussions, David H. D. Warren's abstract machine **specifically** and **exclusively**. Our source is what he describes in [?, ?]. The expected achievement is merely the long overdue filling of a gap so far existing for whoever may be curious to acquire **basic** knowledge of Prolog implementation techniques, as well as to serve as a spring board for the expert eager to contribute further to this field for which the WAM is, in fact, just

the tip of an iceberg. As such, it is hoped that this monograph would constitute an interesting and self-contained complement to basic textbooks for general courses on logic programming, as well as to those on compiler design for more conventional programming languages. As a stand-alone work, it could be a quick reference for the computer professional in need of direct access to WAM concepts.

1.2.2 Organization of presentation 6

Our style of teaching the WAM makes a special effort to consider carefully each feature of the WAM design in isolation by introducing separately and incrementally distinct aspects of Prolog. This allows us to explain as limpidly as possible specific principles proper to each. We then stitch and merge the different patches into larger pieces, introducing independent optimizations one at a time, converging eventually to the complete WAM design as described in [?] or as overviewed in [?]. Thus, in 6.2, we consider unification alone. Then, we look at flat resolution (that is, Prolog without backtracking) in 6.3. Following that, we turn to disjunctive definitions and backtracking in 6.4. At that point, we will have a complete, albeit naïve, design for pure Prolog. In 6.5, this first-cut design will be subjected to a series of transformations aiming at optimizing its performance, the end product of which is the full WAM. We have also prepared an index for quick reference to most critical concepts used in the WAM, something without which no (real) tutorial could possibly be complete.

It is expected that the reader already has a basic understanding of the operational semantics of *Prolog* — in particular, of unification and backtracking. Nevertheless, to make this work also profitable to readers lacking this background, we have provided a quick summary of the necessary *Prolog* notions in 6.7. As for notation, we implicitly use the syntax of so-called Edinburgh Prolog (see, for instance, [?]), which we also recall in that appendix. Finally, 6.8 contains a recapitulation of all explicit definitions implementing the full WAM instruction set and its architecture so as to serve as a complete and concise summary.

6.2 Унификация — ясно и просто

Напомним что терм (первого порядка) — **переменная** (задается большой буквой в начале имени), **константа** (задается маленькой буквой в начале имени) или **терм** — структура вида $f(t_1, \dots, t_n)$, где f символ называемый **функцией** (записывается аналогично константе, с маленькой буквы), а элементы t_i тоже термы первого порядка — **субтермы**. Число субтермов для данного функциона предопределено, и называется **арностью** функциона. Для обеспечения возможности использовать один и тот же символ с разной арностью, мы должны использовать запись f/n , что обозначает функцию f с арностью n . Таким образом, два функциона равны только в том случае, если они имеют **одинаковые символ f и арность n**. Разрешая случай $n = 0$ можно рассматривать константу как особый случай терма: константе c соответствует функция $c/0$ с нулевой арностью.

Мы рассмотрим очень простой низкоуровневый¹⁰ язык \mathcal{L}_0 . На этом языке мы можем описать два вида объектов: **терм программы** и **терм запроса**. Оба этих вида запросов являются термами первого порядка, но не переменными. Семантика \mathcal{L}_0 равносильна вычислению самого общего унификатора программы или запроса. Что касается синтаксиса, \mathcal{L}_0 будет описывать программу как t и запрос как $?-t$ где t является термом. Область видимости переменных ограничена термом программы/запроса. Таким образом, **значение программы не зависит от имен ее переменных**. Интерпретатор для \mathcal{L}_0 будет использовать определенное представление данных для термов и использовать алгоритм унификации для ее операционной семантики. Затем мы опишем $\mathcal{M}_0 = (\mathcal{D}_0, \mathcal{I}_0)$, дизайн абстрактной машины для \mathcal{L}_0 содержащий представление данных \mathcal{D}_0 , над которыми выполняется множество \mathcal{I}_0 машинных инструкций.

Идея достаточно проста: имея определенных программный терм p , мы можем выполнить любой запрос $?-q$, и выполнение запроса завершится с ошибкой если p и q не унифицируются, или будет успешным с привязкой переменных в q полученной при унификации запроса с p .

6.2.1 Представление термов

Для начала давайте определим внутреннее представление термов в языке \mathcal{L}_0 . Мы будем использовать глобальный блок хранения данных в форме адресуемой **кучи** который мы назовем **HEAP**: массив ячеек данных. Адресом ячейки в куче является индекс элемента массива **HEAP**.

Для представления произвольных термов в **HEAP** будет достаточно закодировать переменные и “структуры” имеющие форму $f(@_1, \dots, @_n)$ где f/n функтор и $@_i$ ссылки на адреса кучи для n субтермов. Таким образом существует два вида данных, хранимых в куче: переменные и структуры термов. Явно заданные **тэги**, появляющиеся как часть внутреннего формата ячеек кучи, будут использоваться для различия между этими двумя типами данных.¹¹

Переменная будет идентифицироваться как указатель, и представляться как одна ячейка кучи, так что мы должны говорить о **ячейках переменных**. Ячейка переменной отмечается тэгом **REF**, и обозначается как $\langle \text{REF}, k \rangle$ где k адрес хранения, т.е. индекс в **HEAP**. Этот механизм предназначен для облегчения связывания переменных через установление ссылки на терм в переменной, которая связывается с этим термом. Таким образом при связывании переменной адресная часть **REF**-ячейки получает значение соответствующего адреса терма. Соглашение о представлении **несвязанной переменной** — адресная часть **REF**-ячейки указывает на саму переменную. Таким образом **несвязанные переменные представляются REF-ячейкой со ссылкой на саму себя**.

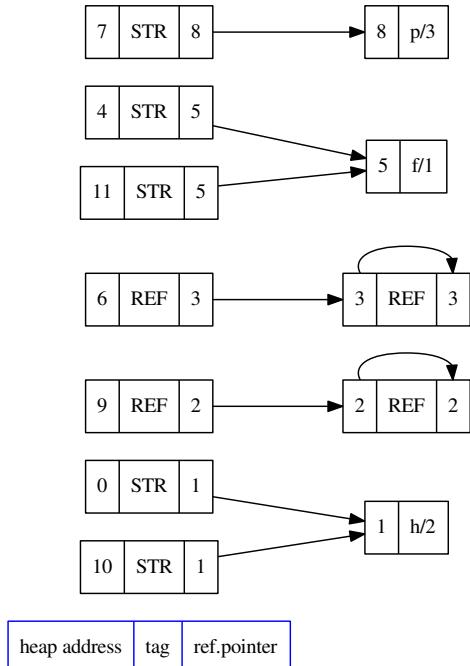
Структуры — термы не являющиеся переменными. Формат кучи для представления структуры $f(t_1, \dots, t_n)$ содержит $n + 2$ ячеек кучи. Первые две ячейки

¹⁰ IL – Intermediate Language

¹¹ интересно рассмотреть расширение тэгирования для реализации ООП и динамического контроля типов

не обязательно смежные. По сути первая из этих двух ячеек выступает в роли сортированного указателя на вторую ячейку, и в то же время сама выступает как представление функтора f/n .¹² Остальные n ячеек предназначены для упорядоченного хранения ссылок на корни соответствующих n субтермов.

Детальнее, первая из $n + 2$ ячеек представляющих терм $f(t_1,..,t_n)$ форматирована как тэгированная **структурная ячейка**, которую можно записать как $<STR, k>$, содержит тэг STR и указатель k на **ячейку функтора**, хранящую представление функтора f/n . Важно отметить что **непосредственно за ячейкой функтора в смежных адресах всегда следуют n структурных ячеек, представляющих каждый из t_i субтермов**. Так что если $HEAP[k] = f/n$ то $HEAP[k+1]$ будет ссылаться на первый субтерм t_1 , а $HEAP[k+n]$ будет ссылаться на последний субтерм t_n .



Фиг. 2.1: Представление кучи для терма $p(Z, h(Z, W), f(W))$

¹² причина использования этой кажущейся странной косвенной адресации — реализация разделяемых структур (structure sharing) — будет вскоре ясна

```

0  STR  1
1  h/2
2  REF  2
3  REF  3
4  STR  5
5  f/1
6  REF  3
7  STR  8
8  p/3
9  REF  2
10 STR  1
11 STR  5

```

Например, рассмотрим представление кучи для терма $p(Z, h(Z, W), f(W))$, начальная ячейка которого находится по адресу 7 (иллюстрация 6.2.1). Отметим что **для каждой** непривязанной переменной существует только одно вхождение, представленное как **REF**-ячейка, в то время как другие ее вхождения в исходный терм представляются как ссылки на первое вхождение ($Z = \text{HEAP}[2]$, $W = \text{HEAP}[3]$). Также обратите внимание что за структурными ячейками по адресам 0, 4 и 7 **сразу** следуют их ячейки функторов, но это не так для адресов 10 и 11.

6.2.2 Компиляция \mathcal{L}_0 запросов

Согласно операционной семантике \mathcal{L}_0 обработка запроса состоит из подготовке в решению уравнения с одной стороны. А именно, терм запроса q транслируется в последовательность инструкций, целью которой является построение экземпляра q на куче из текстового представления q . Таким образом, из-за древовидной структуры терма и множественных вхождениях переменных, необходимо, чтобы при обработке части терма где-то временно сохранялись части терма, которые еще предстоит обработать, или переменные которые могут встретиться еще раз далее по ходу работы. Для этой цели виртуальная машина M_0 наделена достаточным количеством (изменяемых) **регистров** X_1, X_2, \dots которые используются для временного хранения данных кучи по мере создания промежуточных термов. Таким образом, содержимое каждого регистра должно иметь формат ячейки кучи. Эти изменяемые регистры выделяются для терма по мере доступности, так что (1) регистр X_1 всегда распределяется для охватывающего терма, и (2) тот же регистр распределяется для всех вхождений определенной переменной. Например

регистры для переменных терма $p(Z, h(Z, W), f(W))$ распределяются

$$X_1 = p(X_2, X_3, X_4)$$

$$X_2 = Z$$

$$X_3 = h(X_2, X_5)$$

$$X_4 = f(X_5)$$

$$X_5 = W$$

Это равносильно тому что терм рассматривается как сплющенный конъюктивный набор уравнений в форме $X_i = X$ или $X_i = f(X_{i_1},..,X_{i_n})$, ($n \geq 0$) , где члены X_i различные новые имена переменных. Есть два последствия распределения регистров: (1) все внешние имена переменных (такие как Z and W в нашем примере) могут быть забыты; и (2) терм запроса может быть трансформирован в его **сплющенную форму**, т.е. последовательность назначений регистров только в форме $X_i = f(X_{i_1},..,X_{i_n})$. Эта форма — то, что контролирует построение представления терма в куче. Таким образом, чтобы генерация кода слева направо была хорошо обоснована, необходимо сформировать сплющенный терм запроса, так чтобы гарантировать что **имена регистров не могут использоваться в правых частях присвоений (например как субтерм) до их инициализации**¹³. Например сплющенная форма терма запроса $p(Z, h(Z, W), f(W))$ это последовательность $X_3 = h(X_2, X_5)$, $X_4 = f(X_5)$, $X_1 = p(X_2, X_3, X_4)$ ¹⁴.

Сканируя сплющенный терм запроса слева направо, каждый компонент в форме $X_i = f(X_{i_1},..,X_{i_n})$ токенизируется в последовательность $X_i = f/n, X_{i_1},..,X_{i_n}$ такую что после регистра ассоциированного с n -арным функтором идет последовательность n имен регистров. Так что в потоке таких токенов полученных в результате токенизации полного сплющенного терма, существует три вида элементов для обработки:

1. регистр ассоциированный со структурным функтором;
2. регистр-аргумент который не был нигде ранее встречен в потоке;
3. регистр-аргумент который уже был упомянут в потоке.

Из такого потока легко получить представление кучи используя метод управляемого потоком токенов синтеза. Для реализации этого нужно выполнить соответствующие действия для каждого типа токенов:

1. создать на куче новую ячейку STR (и примыкающий функтор) и скопировать эту ячейку в указанный регистр;
2. создать на куче новую ячейку REF содержащую собственный адрес, и скопировать ее в указанный регистр;
3. создать на куче новую ячейку и копировать в нее значение регистра.

¹³ if it has one (viz., being the lefthand side)

¹⁴ исключена привязка переменных на регистры X_2 , X_5

6.2.3	2.3 Compiling L programs	13
6.2.4	2.4 Argument registers	19
6.3	3 Flat Resolution 25	
6.3.1	3.1 Facts	26
6.3.2	3.2 Rules and queries	27
6.4	4 Prolog 33	
6.4.1	4.1 Environment protection	34
6.4.2	4.2 What's in a choice point	36
6.5	5 Optimizing the Design 45	
6.5.1	5.1 Heap representation	46
6.5.2	5.2 Constants, lists, and anonymous variables	47
6.5.3	5.3 A note on set instructions	52
6.5.4	5.4 Register allocation	54
6.5.5	5.5 Last call optimization	56
6.5.6	5.6 Chain rules	57
6.5.7	5.7 Environment trimming	58
6.5.8	5.8 Stack variables	60
5.8.1	Variable binding and memory layout	62
5.8.2	Unsafe variables	64
5.8.3	Nested stack references	67
6.5.9	5.9 Variable classification revisited	69
6.5.10	5.10 Indexing	75
6.5.11	5.11 Cut	83
6.6	6 Conclusion 89	
6.7	A Prolog in a Nutshell 91	

Глава 7

An Efficient Unification Martelli/Montanary Algorithm

¹

© ALBERTO MARTELLI Consiglio Nazionale delle Ricerche
and
UGO MONTANARI Universita di Pisa ²

Abstract

The unification problem in first-order predicate calculus is described in general terms as the solution of a system of equations, and a nondeterministic algorithm is given. A new unification algorithm, characterized by having the acyclicity test efficiently embedded into it, is derived from the nondeterministic one, and a PASCAL implementation is given. A comparison with other well-known unification algorithms shows that the algorithm described here performs well in all cases.

Categories and Subject Descriptors: F.2.2 [Analysis of Algorithms and Problem Complexity]: Nonnumerical Algorithms and Problems—complexity of proof procedures; F.4.1 [Mathematical Logic and Formal Languages]: Mathematical Logic—mechanical theorem proving; I.2.3 [Artificial Intelligence]: Deduction and Theorem Proving—resolution

¹ © <http://www.nsl.com/misc/papers/martelli-montanari.pdf>

² Authors' present addresses: A. Martelli, Istituto di Scienze della Informazione, Università di Torino, Corso M. d'Aeglio 42, 1-10125 Torino, Italy; U. Montanari, Istituto di Scienze della Informazione, Università di Pisa, Corso Italia 40, 1-56100 Pisa, Italy.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association for Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.

© 1982 ACM 0164-0925/82/0400-0258 \$00.75

ACM Transactions on Programming Languages and Systems, Vol. 4, No. 2, April 1982, Pages 258-282.

7.1 INTRODUCTION

In its essence, the unification problem in first-order logic can be expressed as follows: Given two terms containing some variables, find, if it exists, the simplest substitution (i.e., an assignment of some term to every variable) which makes the two terms equal. The resulting substitution is called the *most general unifier* and is unique up to variable renaming.

Unification was first introduced by Robinson [17, 18] as the central step of the inference rule called resolution. This single, powerful rule can replace all the axioms and inference rules of the first-order predicate calculus and thus was immediately recognized as especially suited to mechanical theorem provers. In fact, a number of systems based on resolution were built and tried on a variety of different applications [5]. Even though further research made it apparent that resolution systems are difficult to direct during proof search and thus are often prone to combinatorial explosion [6], new impetus to the research in this area was given by Kowalski's idea of interpreting predicate logic as a programming language [10]. Here predicate logic clauses are seen as procedure declarations, and procedure invocation represents a resolution step. From this viewpoint, theorem provers can be regarded as interpreters for programs written in predicate logic, and this analogy suggests efficient implementations [3, 25].

Resolution, however, is not the only application of the unification algorithm. In fact, its pattern matching nature can be exploited in many cases where symbolic expressions are dealt with, such as, for instance, in interpreters for equation languages [4, 11], in systems using a database organized in terms of productions [19], in type checkers for programming languages with a complex type structure [14], and in the computation of critical pairs for term rewriting systems [9].

The unification algorithm constitutes the heart of all the applications listed above, and thus its performance affects in a crucial way the global efficiency of each. The unification algorithm as originally proposed can be extremely inefficient; therefore, many attempts have been made to find more efficient algorithms [2, 7, 13, 15, 16, 22]. Unification algorithms have also been extended to the case of higher order logic [8] and to deal directly with associativity and commutativity [20]. The problem was also tackled from a computational complexity point of view, and linear algorithms were proposed independently by Martelli and Montanari [13] and Paterson and Wegman [15].

In the next section we give some basic definitions by representing the unification problem as the solution of a system of equations. A nondeterministic algorithm, which comprehends as special cases most known algorithms, is then defined and proved correct. In Section 3 we present a new version of this algorithm obtained by grouping together all equations with some member in common, and we derive from it a first version of our unification algorithm.

In Sections 4 and 5 we present the main ideas which make the algorithm efficient,

and the last details are described in Section 6 by means of a PASCAL implementation.

Finally, in Section 7, the performance of this algorithm is compared with that of two well-known algorithms, Huet's [7] and Paterson and Wegman's [15]. This analysis shows that our algorithm has uniformly good performance for all classes of data considered.

7.2 UNIFICATION AS THE SOLUTION OF A SET OF EQUATIONS: A NONDETERMINISTIC ALGORITHM

In this section we introduce the basic definitions and give a few theorems which are useful in proving the correctness of the algorithms. Our way of stating the unification problem is slightly more general than the classical one due to Robinson [18] and directly suggests a number of possible solution methods.

Let

$$A = \bigcup_{i=0,1,\dots} A_i \quad (A_i \cap A_j = \emptyset, i \neq j)$$

be a ranked alphabet, where A_i contains the i -adic function symbols (the elements of A_0 are constant symbols). Furthermore, let V be the alphabet of the variables. The *terms* are defined recursively as follows:

- (1) constant symbols and variables are terms;
- (2) if t_1, \dots, t_n ($n \geq 1$) are terms and $f \in A_n$, then $f(t_1, \dots, t_n)$ is a term.

A *substitution* ϑ is a mapping from variables to terms, with $\vartheta(x) = x$ almost everywhere. A substitution can be represented by a finite set of ordered pairs $\vartheta = (t_1, x_1), (t_2, x_2), \dots, (t_m, x_m)$ where t_i are terms and x_i are distinct variables, $i = 1, \dots, m$. To apply a substitution ϑ to a term t , we simultaneously substitute all occurrences in t of every variable x_i in a pair (t_i, x_i) of ϑ with the corresponding term t_i . We call the resulting term t_ϑ .

For instance, given a term $t = f(x_1, g(x_2, a))$ and a substitution $\vartheta = (h(x_2), x_1), (b, x_2)$ we have $t_\vartheta = f(f(x_2), g(b), a)$ and $t_{\vartheta\vartheta} = f(h(b), g(b), a)$.

The standard unification problem can be written as an equation

$$t' = t''$$

A solution of the equation, called a *unifier*, is any substitution ϑ , if it exists, which makes the two terms identical. For instance, two unifiers of the equation $f(x_1, h(x_1), x_2, f(g(x_3), x_4, x_3))$ are $\vartheta_1 = (g(x_3), x_1), (x_3, x_2), (h(g(x_3)), x_4)$ and $\vartheta_2 = (g(a), x_1), (a, x_2)$.

In what follows it is convenient also to consider sets of equations

$$t'_j = t''_j, \quad j = 1, \dots, k$$

Again, a *unifier* is any substitution which makes all pairs of terms t'_j, t''_j identical simultaneously.

Now we are interested in finding transformations which produce *equivalent* sets of equations, namely, transformations which preserve the sets of all unifiers. Let us introduce the following two transformations:

(1) Term Reduction. Let

$$f(t'_1, t'_2, \dots, t'_n) = f(t''_1, t''_2, \dots, t''_n), \quad f \in A_n \quad (7.1)$$

be an equation where both terms are not variables and where the two root function symbols are equal. The new set of equations is obtained by replacing this equation with the following ones:

$$t'_1 = t''_1 \quad (7.2)$$

$$t'_2 = t''_2 \quad (7.3)$$

$$\dots \quad (7.4)$$

$$\dots \quad (7.5)$$

$$\dots \quad (7.6)$$

$$t'_n = t''_n \quad (7.7)$$

If $n = 0$, then f is a constant symbol, and the equation is simply erased.

(2) Variable Elimination. Let $x = t$ be an equation where x is a variable and t is any term (variable or not). The new set of equations is obtained by applying the substitution $\vartheta = (t, x)$ to both terms of all other equations in the set (without erasing $x = t$).

We can prove the following theorems:

THEOREM 2.1. *Let S be a set of equations and let $f'(t'_1, \dots, t'_n) = f''(t''_1, \dots, t''_n)$ be an equation of S . If $f' \neq f''$, then S has no unifier. Otherwise, the new set of equations S' , obtained by applying term reduction to the given equation, is equivalent to S .*

PROOF. If $f' \neq f''$, then no substitution can make the two terms identical. If $f' = f''$, any substitution which satisfies 7.2 also satisfies 7.1, and conversely for the recursive definition of term. \square

THEOREM 2.2. *Let S be a set of equations, and let us apply variable elimination to some equation $x = t$, getting a new set of equations S' . If variable x occurs in t (but t is not x), then S has no unifier; otherwise, S and S' are equivalent.*

PROOF. Equation $x = t$ belongs both to S and to S' , and thus any unifier ϑ (if it exists) of S or of S' must unify x and t ; that is, x_ϑ and t_ϑ are identical. Now let $t_1 = t_2$ be any other equation of S , and let $t'_1 = t'_2$ be the corresponding equation in S' . Since t'_1 and t'_2 have been obtained by substituting t for every occurrence of x in t_1 and t_2 , respectively, we have $t_{1\vartheta} = t'_{1\vartheta}$ and $t_{2\vartheta} = t'_{2\vartheta}$. Thus, any unifier of S is also a unifier of S' and vice versa. Furthermore, if variable x occurs in t (but t is not x), then no substitution ϑ can make x and t identical, since x_ϑ becomes a subterm of t_ϑ , and thus S has no unifier. \square

A set of equations is said to be *in solved form* iff it satisfies the following conditions:

(1) the equations are $x_j = t_j, j = 1, \dots, k;$

(2) every variable which is the left member of some equation occurs only there.

A set of equations in solved form has the obvious unifier

$$\vartheta = (t_1, x_1), (t_2, x_2), \dots, (t_k, x_k)$$

Часть III

Дурдом на дереве

Глава 8

The Tree Processing Language Defining the structure and behaviour of a tree

¹ © E. Papegaaij <e.papegaaij@alumnus.utwente.nl>
Supervisors dr. ir. Theo C. Ruys
 ir. Philip K.F. Hözespies
 dr. ir. Arend Rensink
Institute University of Twente
Chair Formal Methods and Tools

Enschede, March 7, 2007

Abstract

Tree structures are commonly used in many applications. One of these is a compiler, in which the tree is called an abstract syntax tree (AST). Different techniques have been developed for building and working with ASTs. However, many of these techniques are limited in their applicability, require major effort to implement or introduce maintenance problems in an evolving application.

This thesis introduces the Tree Processing Language, a language for defining the structure of a tree and adding functionality to this tree. The compiler **TPLc** is used to produce the actual class hierarchy implementing the specified tree. TPL provides a clear separation between the structure of a tree, a **tree definition**, and behaviour of a tree, **logic specifications**. Different aspects of the behaviour of a tree can be provided in separate logic specifications, allowing a clear separation of concerns.

TPLc generates a heterogeneous tree structure with strictly typed children. Functionality in a logic specification is specified using the inheritance pattern. To allow different inheritance trees in different logic specifications, the inheritance pattern is enhanced

with multiple inheritance. For languages that do not support multiple inheritance, the inheritance pattern with composition is developed.

To prove the applicability of TPL, **TPLc** is written in TPL. When compared with an implementation in Java, this implementation provides a better separation of concerns and is easier to maintain.

Preface

Compiler construction has always been one of my favourite fields of software engineering. In the past few years I've written several parsers and compilers. Of these compilers, the compiler for the functional programming language Tina has been the most challenging. I used a hand-written heterogeneous abstract syntax tree as underlying data structure. The most important algorithm applied onto this AST, the transformation of Tina into a core lambda expression language, was written as part of these AST node classes. However, the overwhelming number of AST classes (almost 100) made this approach increasingly difficult to maintain when other algorithms (such as a lambda lifter) were added. At that moment, it became clear that a more structured approach was required. To keep the development of an application, based on a heterogeneous tree, maintainable, different algorithms needed to be separated in different files. The development of tpl is an attempt to provide such an environment.

When I first approached Theo C. Ruys, my premiere supervisor, for an assignment, I had no idea I would be solving this problem, which had bothered me for a long time. At first, ambitious as I was, I proposed to design and implement a completely new parser generator. Luckily, Theo slowed me down a bit and directed me to focus on the real problem: the heterogeneous AST.

For his help in concreting the features of tpl, reading and correcting this thesis and his patience during the endless discussions we had last year, I would like to thank Theo C. Ruys, my premiere supervisor. His guiding helped me structure my thoughts, to be able to write them down. I would also like to thank Philip K.F. Hözzenpies for his help in writing and formatting this thesis. His knowledge of the English language has proven to be far better than mine. Last, but not least, I would like to thank Arend Rensink for having taken the time to examine this thesis.

Emond Papegaaij
Enschede, March 7, 2007

8.1 Introduction

Tree structures have been, and probably will be for a considerable time in the future, a widely used way of organising and working with data. Tree structures are used to represent the structure of an input file², user interface components, the representation

² concrete and abstract syntax trees

of HTML pages³, XML and many more. Due its wide acceptance, extensive research has been spent on working with tree structures.

This thesis is placed in the context of working with tree structures in an object-oriented programming environment. The main focus is on defining the runtime organisation of the tree and applying algorithms on this structure. The origin of the tree — the system responsible for constructing the tree structure — and the actual construction of the tree are discussed, but fall outside the main research area.

In this chapter, an introduction on compiler construction is given, in [8.1.1](#). This section shows how an abstract syntax tree is acquired, and what the typical operations are that need to be performed on an AST. [8.1.2](#) describes the problem statement of this thesis. Finally, the outline of this thesis is given in [8.1.3](#).

8.1.1 Compiler Construction and Abstract Syntax Trees

A multi-pass compiler performs the compilation of a source file in several stages. These stages will be discussed in this section. Compilation starts with reading a source file, and recognising the syntax of the input. Next, an abstract representation of this input is constructed. This is the abstract syntax tree. This AST is used in subsequent phases to perform context checking and code generation. More complex compilers might have more phases, such as optimisers.

Abstract syntax trees are also commonly used in other disciplines, such as communication (eg. a web browser) and source code refactoring in an integrated development environment (IDE) [?]. It is also possible that the abstract syntax tree is not the result of a parser reading an input file, but from speech, or from a graphical programming language. However, the most common usage is a compiler, which reads an input language.

Lexical Analysis and Parsing

In the first stage, the lexical analysis, the compiler reads the input file and produces a stream of tokens. Every token corresponds to a fragment, or construct, found in the input file, such as identifiers, literals, operators and keywords. These tokens are fed to a parser, which discovers (and checks) the structure of the input.

Writing a lexer (or scanner) and parser by hand is tedious, difficult and error prone. Many programs have been developed, which assist the developer in writing the lexer and parser. These tools often take a syntax specification in (E)BNF, and generate a lexer and parser from this specification. Therefore, these tools are commonly called parser generators. Some of these tools are mentioned in ??.

Different strategies exist, on how a parser matches the input language, such as LALR and recursive descent parsing. However, a discussion of these is beyond the scope of this thesis⁴.

³ the document object model

⁴ An explanation of various parsing algorithms, such as LR(k) and LL(k), can be found in [?].

Construction of the AST

In a multi-pass compiler, the task of the parser is to record the structure of the parsed input in an abstract syntax tree. This tree contains all relevant information from the input. What exactly is relevant information, depends on the subsequent phases. Normally, tokens, such as comma's and brackets, are discarded. Also, nesting of parser production rules is removed.

AST construction is exemplified with the grammar presented in fragment 1.1. This grammar matches simple expressions with addition and multiplication. The actual values are represented by numbers and identifiers. Expressions can be nested with brackets.

This grammar matches sentences such as ‘1’, ‘1+1’ and ‘(1+a)*b’. The parse tree of the sentence ‘3+5*(a+b)’ is given in figure 1.1. This figure shows how the complete sentence is matched as an hexpressioni. The hexpressioni consists of a htermi, followed by the literal ‘+’, again followed by a htermi. The left htermi is a simple hatomi, which in turn is a hnumberi. The right htermi consists of two hatomis, separated by a ‘*’. This process is continued until all tokens (the bottom line of the figure) are matched.

The parse tree clearly shows the structure of the parsed text, but this structure is not very practical to work with. If an interpreter for this grammar is needed, a set of four constructs is sufficient: addition, multiplication, numbers and identifiers. The node adds the results of the left and right operands. This node is created when a ‘+’ is matched in hexpressioni. The node multiplies the left operand with the right. It is created when a ‘*’ is matched in htermi. A node is created when a hnumberi is matched, and yields the value of the number. Finally, the node, which is created when an hidentifieri is matched, resolves the value in a symbol table.

Context Checking and Code Generation

8.1.2 Problem Statement

8.1.3 Outline

Часть IV

Язык bI

Глава 9

DLR: Dynamic Language Runtime

DLR: Dynamic Language Runtime — может использоваться как runtime-ядро для реализации динамических языков, или только в качестве библиотеки хранилища данных

синтаксический парсер для разбора текстовых данных, файлов конфигурации, скриптов и т.п., необязателен. В результате разбора формируется синтаксическое дерево из динамических объектов DLR. По реализации может быть

конфигурируемым в runtime добавление/изменение/удаление правил привил грамматики в процессе работы программы
статическим неизменный синтаксис, реализация в виде внешнего модуля, в самом простом случае достаточно использования **flex/bison**

библиотека динамических типов данных выполняет функции хранения данных, может быть реализована

в *Lisp-стиле* базовый набор скаляров **10.2** (символы, строки и числа) и тип **cons-ячейка** позволяющий конструировать составные структуры данных

bI-стиль универсальный символьный тип **10.1**, позволяющий хранить как скаляры, так и вложенные элементы; в базовый тип **AST** заложено хранение типа данных **tag**, его значения **value**, и два способа вложенных хранилищ: плоский упорядоченный список **nest** и именованный неупорядоченный со строковыми ключами **pars**.

От базового символьного типа наследуются

скаляры символ, строка, несколько вариантов чисел (целые, плавающие, машинные, комплексные)¹

¹ критерием скалярности можно считать возможность распознавания элемента данных лексером

композиты структуры данных и объекты
функционалы объекты, для которых определен *оператор аппликации*

или

библиотека операций над данными для преобразования данных и символьных вычислений на списках, деревьях, комбинаторах и т.п.

Lisp стандартная библиотека функций языка *Lisp*

bI каждый тип данных имеет набор унарных и бинарных *операторов*, реализованных в виде виртуальных методов классов

подсистема ОП реализация механизмов ОП, наследования от класса и объекта, вывод типов, преобразование объектных моделей

реализация механизмов функциональных языков хвостовая рекурсия, pattern matching, динамическая компиляция, автоматическое распараллеливание на map/reduce

менеджер памяти со сборщиком мусора

динамический компилятор функциональных типов — через библиотеку JIT LLVM

статический компилятор

в **объектный код** через LLVM
кодогенератор C_+^+

Расширенный функционал

подсистема облачных вычислений и кластеризации расширение DLR на кластера: распределение объектов и процессов между вычислительными узлами. Варианты кластера с высокой связностью², Beowulf³ с постоянным составом, интернет-облака с переменным составом: узлы асинхронно подключаются/отключаются, гомо/гетерогенные: по аппаратной платформе узлов и ОС/среде на каждом узле. Распределение вычислений на одно- и многопроцессорных SMP-системах⁴

прикладные библиотеки GUI, CAD/CAM/EDA, численные методы, цифровая обработка сигналов, сетевые сервера и протоколы, . . .

подсистема крос-трансляции между ходовыми языками программирования (C_+^+ , *JavaScript*, *Python*, PHP, Паскаль) через связку: парсер входного языка → система типов DLR → кодогенератор выходного языка

² аппаратная разделяемая память через сеть InfiniBand — “Сергей Королев”

³ компьютеры общего назначения (офисные) с передачей сообщений по Gigabit Ethernet

⁴ многопоточные вычисления на одном многоядерном узле

интерактивная объектная среда а-ля *SmallTalk* с виджетами и функционалом GUI, CAD, IDE и визуализации данных

сервер приложений обслуживающий тонких браузерных клиентов по HTTP/JS

Глава 10

Система динамических типов

10.1 sym: символ = Абстрактный Символьный Тип / AST

Использование класса **Sym** и виртуально наследованных от него классов, позволяет реализовать на C_+^+ хранение и обработку **любых** данных в виде деревьев¹. Прежде всего этот **символьный тип** применяется при разборе текстовых форматов данных, и текстов программ. **Язык bI построен как интерпретатор AST, примерно так же как язык Lisp использует списки.**

```
// _____ = ABSTRACT SYMBOLIC TYPE
struct Sym {
    // _____ тип (класс) и значение элемента данных
    string tag;                                // data type / class
    string val;                                 // symbol value
    // _____ конструкторы (токен используется в лексере)
    // _____ с о
    Sym(string ,string );                      // <T:V>
    Sym(string );                            // token
```

Хранение вложенных элементов реализовано через указатели на базовый тип **Sym**. Благодаря виртуальному наследованию и использованию RTTI, этими указателями можно пользоваться для работы с любыми другими наследованными типами данных²

```
AST может хранить (и обрабатывать) вложенные элементы
// _____ nest [] e
```

¹ в этом АСТ близок к традиционной аббревиатуре AST: Abstract Syntax Tree

² числа, списки, высокоровневые и скомпилированные функции, элементы GUI,..

```
vector<Sym*> nest;
void push(Sym*);
void pop();
```

параметры (и поля класса)

```
// _____ pa
map<string ,Sym*> pars;
void par(Sym*); // add parameter
```

вывод дампа объекта в текстовом формате

```
// _____
virtual string dump(int depth=0); // dump symbol object as text
virtual string tagval(); // <T:V> header string
string tagstr(); // <T: 'V'> Str-like header s
string pad(int); // padding with tree decorat
```

Операции над **символами** выполняются через использование набора
операторов:

вычисление объекта

```
// _____ compute
virtual Sym* eval();
```

операторы

```
// _____
virtual Sym* str(); // str(A) string represent
virtual Sym* eq(Sym*); // A = B assignment
virtual Sym* inher(Sym*); // A : B inheritance
virtual Sym* member(Sym*); // A % B,C named member (cl
virtual Sym* at(Sym*); // A @ B apply
virtual Sym* add(Sym*); // A + B add
virtual Sym* div(Sym*); // A / B div
virtual Sym* ins(Sym*); // A += B insert
};
```

10.2 Скаляры

- 10.2.1 str: строка
- 10.2.2 int: целое число
- 10.2.3 hex: машинное hex
- 10.2.4 bin: бинарная строка
- 10.2.5 num: число с плавающей точкой

10.3 Композиты

- 10.3.1 list: плоский список
- 10.3.2 cons: cons-пара и списки в *Lisp*-стиле

10.4 Функционалы

- 10.4.1 op: оператор
- 10.4.2 fn: встроенная/скомпилированная функция
- 10.4.3 lambda: лямбда

Глава 11

Программирование в свободном синтаксисе: FSP

11.1 Типичная структура проекта FSP: *lexical skeleton*

Скелет файловой структуры FSP-проекта = lexical skeleton = skelex

Создаем проект **prog** из командной строки (*Windows*):

```
mkdir prog
cd prog
touch src.src log.log ypp.ypp lpp.lpp hpp.hpp cpp.cpp Makefile bat.bat
echo gvim -p src.src log.log ... Makefile bat.bat .gitignore >> bat.bat
```

Создали каталог проекта, сгенерили набор пустых файлов (см. далее), и запустили батник-hepler который запустит **(g)Vim**.

Для пользователей GitHub **mkdir** надо заменить на

```
git clone -o gh git@github.com:yourname/prog.git
cd prog
git gui &
...

```

src.src		исходный текст программы на вашем скриптовом языке
log.log		лог работы ядра <i>bI</i>
ypp.ypp	flex	парсер ??
lpp.lpp	bison	лексер ??
hpp.hpp	<i>C₊⁺</i>	заголовочные файлы ??
cpp.cpp	<i>C₊⁺</i>	код ядра ??
Makefile	make	зависимости между файлами и команды сборки (для <i>Linux</i>)
bat.bat	<i>Windows</i>	запускалка (g)Vim ??
.gitignore	git	список масок временных и производных файлов ??

11.1.1 Настройки (g)Vim

При использовании редактора/IDE (g)Vim удобно настроить сочетания клавиш и подсветку **синтаксиса вашего скриптового языка** так, как вам удобно. Для этого нужно создать несколько файлов конфигурации .vim: по 2 файла¹ для каждого диалекта скриптового языка², и привязать их к расширениям через dot-файлы (g)Vim в вашем домашнем каталоге. Подробно конфигурирование (g)Vim см. 24.

filetype.vim	(g)Vim	привязка расширений файлов (.sr
syntax.vim	(g)Vim	синтаксическая подсветка для скр
/vimrc	Linux	настройки для пользователя
/vimrc	Windows	
/.vim/ftdetect/src.vim	Linux	привязка команд к расширению .s
/vimfiles/ftdetect/src.vim	Windows	
/.vim/syntax/src.vim	Linux	синтаксис к расширению .src
/vimfiles/syntax/src.vim	Windows	

11.1.2 Дополнительные файлы

README.md	github	описание проекта для репозитория github
logo.png	github	логотип
logo.ico	Windows	
rc.rc	Windows	описание ресурсов: логотип, иконки приложения, меню



logo.png: Логотип

¹ (1) привязка расширения файла и (2) подсветка синтаксиса

² если вы пользуетесь сильно отличающимся синтаксисом, но скорее всего через какое-то время практики FSP у вас выработается один диалект для всех программ, соответствующий именно вашим вкусам в синтаксисе, и в этом случае его нужно будет описать только в файлах /.vim/(ftdetect|syntax).vim

11.1.3 Makefile

Для сборки проекта используем команду **make** или **ming32-make** для Windows/MacOS.

Прописываем в **Makefile** зависимости:

универсальный Makefile для fp-sp-проекта

```
log.log: ./exe.exe src.src
    ./exe.exe < src.src > $@ && tail $(TAIL) $@
C = cpp.cpp ypp.tab.cpp lex.yy.c
H =.hpp.hpp ypp.tab.hpp
CXXFILES += -std=gnu++11
./exe.exe: $(C) $(H) Makefile
    $(CXX) $(CXXFILES) -o $@ $(C)
ypp.tab.cpp: ypp.ypp
    bison $<
lex.yy.c: lpp.lpp
    flex $<
```

./exe.exe

префикс `./` требуется для правильной работы **ming32-make**, поскольку в *Linux* исполняемый файл может иметь любое имя и расширение, можем использовать `.exe`.

Для запуска транслятора используем простейший вариант — перенаправление потоков `stdin/stdout` на файлы, в этом случае не потребуется разбор параметров командной строки, и получим подробную трассировку выполнения трансляции.

переменные `C` и `H` задают набор исходных файлов ядра транслятора на C_+ :

cpp.cpp реализация системы динамических типов данных, наследованных от символьного типа AST [10.1](#). Библиотека динамических классов языка *bI IV* компактна, предоставляет достаточных набор типов данных, и операций над ними. При необходимости вы можете легко написать свое дерево классов, если вам достаточно только простого разбора.

hpp.hpp заголовочные файлы также используем из *bI IV*: содержат декларации динамических типов и интерфейс лексического анализатора, которые подходят для всех проектов

ypp.tab.cpp **ypp.tab.hpp** C_+ код синтаксического парсера, генерируемый утилитой **bison 13.2**

lex.yy.c код лексического анализатора, генерируемый утилитой **flex 13.1**
`CXXFLAGS += gnu++11` добавляем опцию диалекта C_+ , необходимую для компиляции ядра *bI*

Глава 12

Синтаксический анализ текстовых данных

12.1 Универсальный Makefile

Универсальный Makefile сделан на базе 11.1.3, с добавлением переменной APP указывающей какой пример парсера следует скомпилировать и выполнить.

Для хранения (и возможной обработки) отпарсенных данных используем ядро языка *bi* 10 — используем файлы *../bi.hpp.hpp* и *../bi/cpp.cpp*. Ядро **очень компактно**, но умеет работать со скалярными, составными и функциональными данными, и содержит минимальную реализацию *ядра динамического языка*.

Универсальный Makefile

```
APP = minimal
$(APP).log: ./$(APP).exe $(APP).src
    ./$(APP).exe < $(APP).src > $@ && tail $(TAIL) $@
C = ../bi/cpp.cpp ypp.tab.cpp lex.yy.c
H = ../bi.hpp.hpp ypp.tab.hpp
CXXFILES += -I../bi -I. -std=gnu++11
./$(APP).exe: $(C) $(H) minimal.mk
    $(CXX) $(CXXFILES) -o $@ $(C)
ypp.tab.cpp: $(APP).ypp
    bison -o $@ $<
lex.yy.c: $(APP).lpp
    flex -o $@ $<

.PHONY: src
src: minimal.src comment.src string.src ops.src brackets.src

minimal.src: ../bi/cpp.cpp
    head -n11 $< > $@
comment.src: ../bi/cpp.cpp
    head -n11 $< > $@
string.src: ../bi/cpp.cpp
```

```
head -n11 $< > $@  
ops.src: .. / bi / cpp .cpp  
head -n5 $< > $@  
brackets.src: .. / bi / cpp .cpp  
head -n5 $< > $@
```

12.2 C_+^+ интерфейс синтаксического анализатора

```
extern int yylex(); // получить код следующего токена, и увл  
extern int yylineno; // номер текущей строки файла исходника  
extern char* yytext; // текст распознанного токена, asciz  
#define TOC(C,X) { yyval.o = new C(yytext); return X; }  
  
extern int yyparse(); // отпарсить весь текущий входной поток  
extern void yyerror(string); // callback вызывается при синтаксической  
#include "ypp.tab.hpp"
```

12.3 Минимальный парсер

Рассмотрим минимальный парсер, который может анализировать файлы текстовых данных (например исходники программ), и вычленять из них последовательности символов, которые можно отнести к **скалярам** символ, строка и число.

¹

Лексер **minimal.lpp** /flex/

```
%{  
#include "hpp.hpp"  
%}  
%option noyywrap  
%option yylineno  
%%  
[a-zA-Z0-9_.]+ TOC(Sym,SYM)  
%%
```

(.. / bi /) **hpp.hpp** содержит определения интерфейса лексера 12.2, и ядра языка
bI 10 для хранения результатов разбора текстовых данных

noyywrap выключает использование функции **yywrap()**

yylineno включает отслеживание строки исходного файла, используется при выводе сообщений об ошибках. В минимальном парсере не используется, но требуется для сборки *bI*-ядра.

¹ эти три типа можно назвать атомами computer science

`%%.%` набор правил группировки отдельных символов в элементы данных — **токены**, правила задаются с помощью *регулярных выражений*

`TOC(Sym, SYM)` единственное правило, распознающее любые группы символов как класс **bi::sym**: латинские буквы, цифры и символы `_` и `.` (точка)²

Парсер `minimal.ypp /bison/`

```
%{  
#include "hpp.hpp"  
%}  
%defines %union { Sym*o; }           /* use universal bI abstract type */  
%token <o> SYM STR NUM            /* symbol 'string' number */  
%type <o> ex scalar              /* expression scalar */  
%%  
REPL : | REPL ex { cout << $2->tagval(); } ;  
scalar : SYM | STR | NUM ;  
ex : scalar ;  
%%
```

`hpp.hpp` заголовок аналогичен лексеру [12.3](#)

`%defines %union` указывает какие типы данных могут храниться в узлах разобранного **синтаксического дерева**. Поскольку мы используем *bI*-ядро, нам будет достаточно пользоваться только классами языка *bI*, прежде всего универсальным символьным типом AST [10.1](#) и его производными классами.

`%token` описывает токены, которые может возвращать лексер `??`, причем набор токенов должен быть согласованным между лексером и парсером³

`%type` описывает типы синтаксических выражений, которые может распознавать **грамматика** синтаксического анализатора,

`REPL` выражение, описывающее грамматику, аналогичную простейшему варианту цикла `REPL`: Read Eval Print Loop⁴. В нашем случае часть вычисления `Eval` не выполняется⁵, а часть `Print` выполняется через метод `Sym.tagval()`, возвращающий короткую строку вида `<класс:значение>` для найденного токена.

`ex` (`expression`) универсальное символьное выражение языка *bI*, в нашем случае оно должно представлять только `scalar`

² точка добавлена, так часто используется в именах файлов

³ определение токенов генерируется в файл `ypp.tab.hpp`

⁴ чтение/вычисление/вывод/повторить

⁵ разобранное выражение не вычисляется, хотя используемое ядро *bI* и поддерживает такой функционал

`scalar` выражение, представляющее только распознаваемые скаляры:

`SYM` символ,

`STR` строку [или](#)

`NUM` число⁶

В качестве тестового исходника возьмем C_+^+ код ядра языка bI : `../bi/cpp.cpp`:

minimal.src: Тестовый исходник

minimal.log: Результат прогона

```
#<sym: include> "<sym: hpp . hpp>"  
#<sym: define> <sym: YYERR> "\<sym: n>\<sym: n><<<sym: yylineno><<"<<<  
<sym: void> <sym: yyerror>(<sym: string> <sym: msg>) { <sym: cout><<<sym: Y  
<sym: int> <sym: main>() { <sym: return> <sym: yyparse>(); }  
  
<sym: Sym>::<sym: Sym>(<sym: string> <sym: T>, <sym: string> <sym: V>) { <sym:  
<sym: Sym>::<sym: Sym>(<sym: string> <sym: V>);<sym: Sym>(" <sym: sym> ", <sym:  
  
<sym: string> <sym: Sym>::<sym: tagval>() { <sym: return> "<" + <sym: tag> +  
<sym: string> <sym: Sym>::<sym: tagstr>() { <sym: return> "<" + <sym: tag> +  
<sym: string> <sym: Sym>::<sym: pad>(<sym: int> <sym: n>) { <sym: string> <sym:  
<sym: string> <sym: Sym>::<sym: dump>(<sym: int> <sym: depth>) { <sym: str  
    <sym: return> <sym: S>; }  
  
<sym: Sym>*<sym: Sym>::<sym: eval>() { <sym: return> <sym: this>; }
```

Как видно по логу **minimal.log**, все группы символов, соответствующих правилу лексера **SYM**^{12.3}, распознались как объекты bI , остальные остались символами и попали в лог без изменений.

12.4 Добавляем обработку комментариев

В тестах программ и файлов конфигурации очень часто используются [комментарии](#). В языке *Python*, bI и UNIX shell комментарием является все от символа `#` до конца строки.

Для обработки таких [строчных комментариев](#) достаточно добавить одно правило лексера, [обязательно первым правилом](#):

Лексер со строчными комментариями

```
%{  
#include "hpp . hpp"  
%}
```

⁶ числа в грамматике языка bI по типам не делятся, токен соответствует как `int`, так и `num`

```
%option noyywrap
%option yylineno
%%
#[^\\n]*          {}
[a-zA-Z0-9_.]+    TOC(Sym,SYM)
%%
```

Группа символов, начинающаяся с символа #, затем идет ноль или более []* любых символов не равных ^ концу строки \n. Пустое тело правила: C_+^+ код в {} скобках — выполняется и ничего не делает.

Тело правила SYM — вызов макроса TOC(C,X) 12.2, наоборот, при своем выполнении создает токен, и возвращает код токена =SYM.

comment.log: Результат прогона

```
<sym: void> <sym: yyerror>(<sym: string> <sym: msg>) { <sym: cout><<<sym:>
<sym: int> <sym: main>() { <sym: return> <sym: yyparse>(); }

<sym: Sym>::<sym: Sym>(<sym: string> <sym: T>, <sym: string> <sym: V>) { <sym:>
<sym: Sym>::<sym: Sym>(<sym: string> <sym: V>):<sym: Sym>("<sym: sym>", <sym:>

<sym: string> <sym: Sym>::<sym: tagval>() { <sym: return> "<" + <sym: tag> +
<sym: string> <sym: Sym>::<sym: tagstr>() { <sym: return> "<" + <sym: tag> +
<sym: string> <sym: Sym>::<sym: pad>(<sym: int> <sym: n>) { <sym: string> <
```

Как видно из лога, из вывода исчезли первые 2 строки, начинающиеся на #, причем концы этих строк остались (но не были как-либо распознаны).

12.5 Разбор строк

Для разбора строк необходимо использовать лексер с применением **состояний**. Строки имеют сильно отличающийся от основного кода синтаксис, и для его обработки нужно **переключать набор правил лексера**.

Лексер с состоянием для строк

```
%{
#include "hpp.hpp"
string LexString; /* string parser buffer */
%}
%option noyywrap
%option yylineno
%lex lexstring
%%
#[^\\n]*          {}
\\n                {BEGIN(lexstring); LexString="";}
%%
```

```
<lexstring>\\"          {BEGIN(INITIAL); yyval.o = new Str(LexString);
<lexstring>\n          {LexString+=yytext[0];}
<lexstring>.           {LexString+=yytext[0];}
```

```
[a-zA-Z0-9_.]+        TOC(Sym,SYM)
%%
```

string LexString строковая буферная переменная, накапливающая символы строки

%x lexstring создание отдельного состояния лексера lexstring

INITIAL основное состояние лексера

<lexstring>\n правило конца строки позволяет использовать многострочные строки⁷

<lexstring>. любой символ в состоянии <lexstring>

Лог разбора со строками

```
<sym: void> <sym: yyerror>(<sym: string> <sym: msg>) { <sym: cout><<sym:>
<sym: int> <sym: main>() { <sym: return> <sym: yyparse>(); }
```

```
<sym: Sym>::<sym: Sym>(<sym: string> <sym: T>, <sym: string> <sym: V>) { <sym:>
<sym: Sym>::<sym: Sym>(<sym: string> <sym: V>):<sym: Sym>(<str: 'sym'>,<sym:>
```

```
<sym: string> <sym: Sym>::<sym: tagval>() { <sym: return> <str:'>+<sym:>
<sym: string> <sym: Sym>::<sym: tagstr>() { <sym: return> <str:'>+<sym:>
<sym: string> <sym: Sym>::<sym: pad>(<sym: int> <sym: n>) { <sym: string> <sym:>
```

Обратите внимание, что ранее попадавшие в лог строки в двойных кавычках, типа "]\n\n", стали распознаваться как строковые токены <str:']\n\n'>.⁸

12.6 Добавляем операторы

Для разбора языков программирования необходима поддержка операторов, включим общепринятые одиночные операторы, операторы C^+ и bI . **Скобки различного вида тоже будет рассматривать как операторы.** Операторы реализованы в ядре bI как отдельный класс **op**, зададим пачку правил разбора операторов, создающих токены **TOC(Op,XXX)**:

⁷ символ конца строки не распознается метасимволом . (точка) в регулярном выражении, и требует явного указания

⁸ использованы 'одинарные кавычки' как в *Python/bI*

Лексер с операторами

```
%{  
#include "hpp.hpp"  
string LexString; /* string parser buffer */  
%}  
%option noyywrap  
%option yylineno  
%x lexstring  
%%  
#[^\\n]* { /* # line comment */  
  
\" {BEGIN(lexstring); LexString=""};  
<lexstring>\" {BEGIN(INITIAL); yyval.o = new Str(LexString);  
<lexstring>\\n {LexString+=yytext[0];}  
<lexstring>. {LexString+=yytext[0];}  
  
[a-zA-Z0-9_.]+ TOC(Sym,SYM) /* symbol */  
  
\( TOC(Op,LB) /* brackets */  
\) TOC(Op,RB)  
\[ TOC(Op,LQ)  
\] TOC(Op,RQ)  
\{ TOC(Op,LC)  
\} TOC(Op,RC)  
  
\+ TOC(Op,ADD) /* typical arithmetic operators */  
\- TOC(Op,SUB)  
\* TOC(Op,MUL)  
\/ TOC(Op,DIV)  
\^ TOC(Op,POW)  
  
\= TOC(Op,EQ) /* bi language specific */  
\@ TOC(Op,AT) /* assign */  
\~ TOC(Op,TILD) /* apply */  
\: TOC(Op,COLON) /* quote */  
/* inheritance */  
  
%%
```

Парсер с операторами

```
%{  
#include "hpp.hpp"  
%}  
%defines %union { Sym*o; } /* use universal bi abstract type */  
%token <o> SYM STR NUM /* symbol 'string' number */  
%token <o> LB RB LQ RQ LC RC /* brackets: () [] {} */  
%token <o> ADD SUB MUL DIV POW /* arithmetic operators: + - * / ^ */  
%token <o> EQ AT TILD COLON /* bi specific operators: = @ ~ : */  
%type <o> ex scalar /* expression scalar */
```

```
%type <o> bracket operator
%%
REPL : | REPL ex { cout << $2->dump(); } ;
scalar : SYM | STR | NUM ;
ex : scalar | operator ;
bracket : LB | RB | LQ | RQ | LC | RC ;
operator :
    bracket
    | ADD | SUB | MUL | DIV | POW
    | EQ | AT | TILD | COLON
;
%%
```

Лог уже стал нечитаем, переключаемся на древовидный вывод через метод `Sym.dump()`.

Разбор с операторами

```
<sym: void>
<sym: yyerror>
<op:(>
<sym: string>
<sym: msg>
<op:)>
<op:{>
<sym: cout><<
<sym: YYERR>;
<sym: cerr><<
<sym: YYERR>;
<sym: exit>
<op:(>
<op:->
<sym:1>
<op:)>;
<op:{>

<sym: int>
<sym: main>
<op:(>
<op:)>
<op:{>
<sym: return>
<sym: yyparse>
<op:(>
<op:)>;
<op:{>
```

12.7 Обработка вложенных структур (скобок)

Обработка вложенных структур возможна только парсером, лексер оставляем без изменений. Хранение вложенных структур в виде дерева — главная фича типа *bI AST* 10.1. Заменяем грамматическое выражение **bracket** на отдельные выражения для скобок:

Парсер со скобками

```
%{
#include "hpp.hpp"
%}
%defines %union { Sym*o; }      /* use universal bI abstract type */
%token <o> SYM STR NUM        /* symbol 'string' number */
%token <o> LB RB LQ RQ LC RC  /* brackets: () [] {} */
%token <o> ADD SUB MUL DIV POW /* arithmetic operators: + - * / ^ */
%token <o> EQ AT TILD COLON    /* bi specific operators: = @ ~ : */
%token <o> SCOLON GR LS
%type <o> ex scalar           /* expression scalar */
%type <o> operator
%%
REPL : | REPL ex { cout << $2->dump(); } ;
scalar : SYM | STR | NUM ;
ex :
    ex ex                  { $$=$1; $$->push($2); }
    | scalar | operator
    | LB ex RB              { $$=new Sym("(")); $$->push($2); }
    | LB RB                 { $$=new Sym("()"); }
    | LQ ex RQ              { $$=new Sym("["); $$->push($2); }
    | LC ex RC              { $$=new Sym("]"); $$->push($2); }
;
operator :
    ADD | SUB | MUL | DIV | POW
    | EQ | AT | TILD | COLON
    | SCOLON | GR | LS
;
%%
```

Разбор со скобками

```
<sym: void>
<sym: yyerror>
<sym: ()>
<sym: string>
```

```
<sym : msg>
<sym: {}>
    <sym : cout>
        <op:<>
            <op:<>
                <sym : YYERR>
                    <op:;>
                        <sym : cerr>
                            <op:<>
                                <op:<>
                                    <sym : YYERR>
                                        <op:;>
                                            <sym : exit>
                                                <sym: ()>
                                                    <op:->
                                                        <
                                                            <op:>
<sym : int>
    <sym : main>
        <sym: ()>
            <sym: {}>
                <sym : return>
                    <sym : yyparse>
                        <sym: ()>
                            <op:;>
```

Глава 13

Синтаксический анализатор

Синтаксис языка *bI* был выбран алголо-подобным, более близким к современным императивным языкам типа C_+^+ и *Python*. Использование типовых утилит-генераторов позволяет легко описать синтаксис с инфиксными операторами и скобочной записью для композитных типов 10.3, и не заставлять пользователя закапываться в клубок *Lisp*овских скобок.

Инфиксный синтаксис **для файлов конфигурации** удобен неподготовленным пользователям, а возможность определения пользовательских функций и объектная система,строенная в ядро *bI*, дает богатейшие возможности по настройке и кастомизации программ.

Единственной проблемой с точки зрения синтаксиса для начинающего пользователя *bI* может оказаться отказ от скобок при вызове функций, применение оператора явной аппликации \mathfrak{C} , и функциональные наклонности самого *bI*, претендующего на звание универсального **объектного метаязыка** и **языка шаблонов**.

13.1 lpp.lpp: лексер /flex/

lpp.lpp

```
%{
#include "hpp.hpp"
string LexString;                                     // string pa
void incLude(Sym*inc) {                                // .include
    if (!(yyin = fopen((inc->val).c_str(),"r")) ) yyerror("");
    yypush_buffer_state(yy_create_buffer(yyin,YY_BUF_SIZE));
}
%}
%option noyywrap
%option yylineno
%x lexstring docstring
S [\\-\\+]??
N [0-9]+
```

```

%%
#[^\\n]*
{ }                                /* == line comment == */

^\\.end                               /* == . directive */
^\\.inc [ \\t]+[^\\n]+                  /* .end */
^\\.\\.[a-z]+[^\\n]*                   /* .include */
TOC(Directive,DIR)                   /* .directive */

/* 'string' */
<lexstring>'                         /* BEGIN( lexstring ); LexString="" */
<lexstring>\\t                        /* BEGIN(INITIAL); yyval.o=new Str(LexString); ret */
{LexString+=\\t;}                      /* LexString+='t'; */
{LexString+=\\n;}                      /* LexString+='n'; */
{LexString+=yytext[0];}                /* LexString+=yytext[0]; */
{LexString+=yytext[0];}                /* LexString+=yytext[0]; */

/* "docstring" */
<docstring>\"                         /* BEGIN( docstring ); LexString="" */
<docstring>\\t                        /* BEGIN(INITIAL); yyval.o=new Str(LexString); ret */
{LexString+=\\t;}                      /* LexString+='t'; */
{LexString+=\\n;}                      /* LexString+='n'; */
{LexString+=yytext[0];}                /* LexString+=yytext[0]; */
{LexString+=yytext[0];}                /* LexString+=yytext[0]; */

/* == numbers == */
TOC(Num,NUM)                          /* floating point */
TOC(Num,NUM)                          /* exponential */
TOC(Int,NUM)                           /* integer */
TOC(Hex,NUM)                           /* machine hex */
TOC(Bin,NUM)                           /* bin string */

/* == symbol == */
TOC(Sym,SYM)                          /* symbol */

/* == brackets == */
TOC(Op,LP)                            /* [ */
TOC(Op,RP)                            /* ] */
TOC(Op,LQ)                            /* { */
TOC(Op,RQ)                            /* } */
TOC(Op,LC)                            /* < */
TOC(Op,RC)                            /* > */
TOC(Op,LV)                            /* <vector> */
TOC(Op,RV)                            /* > */

/* == operators == */
TOC(Op,INS)                           /* + */
TOC(Op,DEL)                           /* - */

/* == operators == */
TOC(Op,EQ)                            /* = */
TOC(Op,AT)                            /* @ */
TOC(Op,TILD)                           /* ~ */
TOC(Op,COLON)                          /* : */

```

```

\%          TOC(Op,PERC)
\.
\,          TOC(Op,DOC)
\|          TOC(Op,COMMA)

\+          TOC(Op,ADD)
\-
\*          TOC(Op,SUB)
\/
\^          TOC(Op,MUL)
\/
\^          TOC(Op,DIV)
\^          TOC(Op,POW)

[ \t\r\n]+    {}                                /* == drop spaces == */

<<EOF>>    { yydrop_buffer_state(); }           /* end of .include */
if (!YY_CURRENT_BUFFER)
    yyterminate();
%%
```

13.2 yacc.ypp: парсер /bison/

ypp.ypp

```

%{
#include "hpp.hpp"
%}
%defines %union { Sym*o; }                      /* universal bI abstract symbolic
%token <o> SYM STR NUM DIR DOC                /* symbol 'string' number . direction
%token <o> LP RP LQ RQ LC RC LV RV             /* () [] {} ◇
%token <o> EQ AT TILD COLON                   /* = @ ~ :
%token <o> DOT COMMA PERC                     /* . , %
%token <o> ADD SUB MUL DIV POW                 /* + - * / ^
%token <o> INS DEL                            /* += insert -= delete
%token <o> MAP                                /* |
%type <o> ex scalar list lambda               /* expression scalar [ list ] {lambda
%type <o> vector cons op bracket              /* <vector> co_ns operator bracket

%left INS
%left DOC
%left EQ
%left ADD SUB
%left MUL DIV
%left POW
%right AT
%right COMMA
%left PFX
%left TILD
```

```

%left PERC
%left COLON
%left DOT
%%
REPL : | REPL ex
;
ex      : scalar | DIR
| ex DOC
| LP ex RP
| LQ list RQ
| LC lambda RC
| LV vector RV
| TILD ex
| TILD op
| cons
| ADD ex %prec PFX
| SUB ex %prec PFX
| ex EQ ex
| ex AT ex
| ex COLON ex
| ex DOT ex
| ex PERC ex
| ex ADD ex
| ex SUB ex
| ex MUL ex
| ex DIV ex
| ex POW ex
| ex INS ex
| ex DEL ex
| ex MAP ex
;
op      : bracket |EQ |AT |TILD |COLON |DOT |COMMA |ADD |SUB |MUL |D
bracket : LP |RP |LQ |RQ |LC |RC |LV |RV ;
scalar  : SYM | STR | NUM ;
;
cons    : ex COMMA ex      { $$=new Cons($1,$3); } ;
list    :          | list ex { $$=new List(); }
|           { $$=$1; $$->push($2); }
;
lambda  :          | lambda SYM COLON { $$=$1; $$->par($2); }
| lambda ex   { $$=$1; $$->push($2); }
;
vector  :          | vector ex { $$=new Vector(); }
| vector ex  { $$=$1; $$->push($2); }
;
%%

/* REPL with full parse/eval logging */
{ cout << $2->dump();
cout << "\n-----";
cout << $2->eval()->dump();
cout << "\n-----\n"; } ;

```

В качестве типа-хранилища для узлов синтаксического дерева идеально подходит базовый символьный тип *bI* 10.1, причем его применение в этом качестве рассматривалось как основное: гибкое представление произвольных типов данных. Собственно его название намекает.

В качестве токенов-скаляров логично выбираются SYMвол, STRока и число NUM¹. Надо отметить, что в принципе можно было бы обойтись единственным SYM, но для дополнительного контроля грамматики полезно выделить несколько токенов: это позволит гарантировать что в определении класса ?? вы сможете использовать в качестве суперкласса и имен полей только символы. По крайне мере до момента, когда в очередном форке *bI* не появится возможность наследовать любые объекты.

¹ их можно вообще рассматривать как элементарные частицы Computer Science, правда к ним еще придется добавить PTR: божественный указатель

Часть V

skelex: скелет программы в
свободном синтаксисе

В этом разделе описана общая структура любого проекта, использующего принципы *программирования в свободном синтаксисе*, в виде примера определения синтаксиса и семантики языка *bI*.

Материал дублирует другие разделы, но может быть использован как вариант **минимизированного** языкового ядра FSP-проекта: нет комментариев, лишних классов, подробного описания работы ядра и т.п., **только краткие пояснения и минимальный код**.

Структура проекта

Создание проекта

```
git clone -o gh git@github.com:user/lexprogram.git
cd lexprogram
touch src.src log.log \
      ypp.ypp lpp.lpp.hpp.hpp.cpp.cpp Makefile .gitignore
gvim -p src.src log.log ... Makefile .gitignore >> bat.bat
bat.bat
```

src.src	<i>bI</i>	текст программы в свободном синтаксисе
log.log	<i>bI</i>	лог интерпретатора
ypp.ypp	bison	парсер синтаксиса
lpp.lpp	flex	лексер
hpp.hpp	<i>C₊⁺</i>	хедеры
cpp.cpp	<i>C₊⁺</i>	ядро интерпретатора
Makefile	make	скрипты сборки проекта
.gitignore	git	маски файлов, не попадающие в git-проект
bat.bat	<i>Windows</i>	helper запуска (g)Vim

.gitignore

```
*~  
*.swp  
exe.exe  
log.log  
ypp.tab.?pp  
lex.yy.c
```

bat.bat

```
@start .
@gvim -p src.src log.log ypp.ypp lpp.lpp.hpp.hpp.cpp.cpp Makefile
```

Makefile

Makefile

```
MODULE = $(notdir $(CURDIR))
log.log: ./exe.exe src.src
    ./exe.exe < src.src > log.log && tail $(TAIL) log.log
C = cpp.cpp ypp.tab.cpp lex.yy.c
H = hpp.hpp ypp.tab.hpp
CXXFLAGS = -std=gnu++11 -DMODULE=\\"$(MODULE)\\"
./exe.exe: $(C) $(H)
$(CXX) $(CXXFLAGS) -o $@ $(C)
ypp.tab.cpp: ypp.ypp
bison $<
lex.yy.c: lpp.lpp
flex $<
```

MODULE имя программного модуля, в примере получается автоматически из имени каталога проекта; при компиляции интерпретатора добавляется как глобальная константа, и может быть использована в скриптах.

TAIL = -n7|-n17|<none> при успешном выполнении интерпретатора выводятся последние \$(TAIL) строк лога, при отладке скриптов удобно добавлять **в конец программы** вывод отладочной информации. Конкретное значение параметра команды **tail** выбирается в зависимости от настроек вашей IDE, для **eclipse** на старом 15" мониторе мне удобен TAIL=-n7, для **(g)Vim** и командной строки можно увеличить до TAIL=-n17.

CURDIR полный путь для текущего каталога

\$(notdir ...) функция выделяет из полного пути последний / элемент

ypp.ypp: синтаксический парсер

Весь код между %{...%} будет скопирован в выходной сгенерированный файл ypp.tab.cpp

Заголовочная часть с C_+^+ кодом

```
%{
#include "hpp.hpp"
%}
```

используем универсальный тип для хранения дерева разбора

```
%defines %union { Sym*; }
```

токены для скалярных типов

```
%token <o> SYM NUM STR /* symbol number 'string' */
```

правило для скалярных типов

scalar : SYM | NUM | STR ;

символ, число и строка — атомы информатики

токены для скобок

%token <o> LP RP LQ RQ LC RC /* () [] { } */

[L]eft/[R]ight [P]arens, [Q]uad, [C]url

пачка операторов V

%token <o> EQ AT TILD PERC PIPE /* = @ ~ % | */
%token <o> COLON DOT COMMA /* : . , */
%token <o> ADD SUB MUL DIV POW /* + - * / ^ */
%token <o> LL GG /* < > */

типы выражений

%type <o> ex scalar /* expression scalar */
%type <o> list lambda /* [list] { la:mbda } */

правила парсера помещаются между

%%

REPL-цикл интерпретатора

REPL : | REPL ex { cout << \$2->eval()->dump(); } ;

скаляры

scalar : SYM | NUM | STR ;

выражения

ex : scalar
| LP ex RP { \$\$=\$2; }
| LQ list RQ { \$\$=\$2; }
| LC lambda RC { \$\$=\$2; }
| ex COMMA ex { \$\$=new Cons(\$1,\$3); }
| TILD ex { \$\$=\$1; \$\$->push(\$2); }
;

списки

list : { \$\$= new List(); }
| list ex { \$\$=\$1; \$\$->push(\$2); }
;

лямбда-определения

```
lambda : { $$= new List(); }
| lambda SYM COLON { $$=$1; $$->par($2); }
| lambda ex { $$=$1; $$->push($2); }
;
```

lpp.lpp: лексер

Весь код между `%{...%}` будет скопирован в выходной сгенерированный файл `lex.yy.c`

Заголовочная часть с C_+^+ кодом

```
%{
#include "hpp.hpp"
string LexString;
%}
```

определенна дополнительная переменная `LexString`: буфер используемый при разборе строк.

опция

```
%option noyywrap
```

подавляет вывод сообщений об отсутствии функции `yywrap`

опция включения счетчика нумерации строк

```
%option yylineno
```

сохраняет в переменной `yylineno` номер текущей строки

правила лексера помещаются между

```
%%
```

строчные комментарии

```
#[^\n]* { }
```

разбор строк через специальное состояние лексера

```
%x lexstring
```

```
,
```

```
<lexstring> {BEGIN(lexstring); LexString="";}
```

```
{BEGIN(INITIAL);
```

```
yylval.o = new Str(LexString); return STR; }
```

```
<lexstring>\\t {LexString+='\\t';}
```

```
<lexstring>\\n      {LexString+='\n';}
<lexstring>\\n      {LexString+=yytext[0];}
<lexstring>.      {LexString+=yytext[0];}
```

распознавание чисел

```
S [\\+\\-]?
N [0-9]+
```

{S}{N}[eE]{S}{N}	TOC(Num,NUM)
{S}{N}\\.{N}	TOC(Num,NUM)
{S}{N}	TOC(Int ,NUM)
0x[0-9A-F]+	TOC(Hex ,NUM)
0b[01]+	TOC(Bin ,NUM)

односимвольные операторы

\=	TOC(Op ,EQ)
\@	TOC(Op ,AT)
\~	TOC(Op ,TILD)
\%	TOC(Op ,PERC)
\	TOC(Op ,PIPE)
\:	TOC(Op ,COLON)
\.	TOC(Op ,DOT)
\,	TOC(Op ,COMMA)
\+	TOC(Op ,ADD)
\-	TOC(Op ,SUB)
*	TOC(Op ,MUL)
\	TOC(Op ,DIV)
\^	TOC(Op ,POW)
\<	TOC(Op ,LL)
\!	TOC(Op ,EX)
\>	TOC(Op ,GG)

hpp.hpp: хедеры

```
#ifndef _H_SKELEX
#define _H_SKELEX
```

все остальное находится между препроцессорными “скобками”, блокирующими многократное включение кода

```
#endif // _H_SKELEX
```

```
#include
```

```
#include <iostream>
#include <sstream>
#include <cstdlib>
#include <vector>
#include <map>
using namespace std;
```

универсальный тип: Abstract Symbolic Type

```
struct Sym {
    string tag, val;
    Sym(string, string); Sym(string);
    vector<Sym*> nest; void push(Sym*);
    map<string, Sym*> pars; void par(Sym*);
    virtual string tagval(); string tagstr();
    virtual string dump(int=0); string pad(int);
    virtual Sym* eval();
    virtual Sym* eq(Sym*);
    virtual Sym* at(Sym*);
};
```

глобальная среда (таблица символов)

```
extern map<string, Sym*> env;
extern void env_init();
```

скаляры: строки

```
struct Str: Sym { Str(string); string tagval(); };
```

скаляры: числа

```
struct Int: Sym { Int(string); long val; string tagval(); };
struct Num: Sym { Num(string); double val; string tagval(); };
struct Hex: Sym { Hex(string); };
struct Bin: Sym { Bin(string); };
```

КОМПОЗИТЫ

```
struct List: Sym { List(); };
struct Cons: Sym { Cons(Sym*, Sym*); };
```

функционалы: оператор

```
struct Op: Sym { Op(string); };
```

встроенные функции

```
typedef Sym*(*FN)(Sym*);  
struct Fn: Sym { Fn(string ,FN); FN fn; };
```

лямбда-функции

```
struct Lambda: Sym { Lambda(); };
```

интерфейс к лексеру/парсеру

```
extern int yylex();  
extern int yylineno;  
extern char* yytext;  
#define TOC(C,X) { yyval.o = new C(yytext); return X; }  
extern int yyparse();  
extern void yyerror(string);  
#include "ypp.tab.hpp"
```

cpp.cpp: ядро интерпретатора

```
#include "hpp.hpp"
```

обработка ошибок синтаксического анализатора

```
#define YYERR "\n\n<<yylineno<<: "<<msg<< [ "<<yytext<<"]\n\n"  
void yyerror(string msg) { cout<<YYERR; cerr<<YYERR; exit(-1); }
```

функция main()

```
int main() { env_init(); return yyparse(); }
```

конструкторы AST

```
Sym::Sym(string T, string V) { tag=T; val=V; }  
Sym::Sym(string V):Sym("",V) {}
```

```
void Sym::push(Sym*o) { nest.push_back(o); }  
void Sym::par(Sym*o) { pars[o->val]=o; }
```

дамп AST

```
string Sym::tagval() { return "<" + tag + ":" + val + ">"; }  
string Sym::pad(int n) { string S; for (int i=0;i<n; i++) S+='\t'; ret  
string Sym::dump(int depth) { string S = "\n" + pad(depth)+tagval();  
for (auto it=nest.begin(), e=nest.end(); it!=e; it++)  
    S += (*it)->dump(depth+1);  
return S; }
```

```
Sym* Sym:: eval () {
    Sym*E = env[ val ]; if (E) return E;
    for (auto it=nest.begin(), e=nest.end(); it!=e; it++)
        (*it) = (*it)->eval ();
    return this; }
```

```
Sym* Sym:: eq (Sym*o) { env[ val ]=o; return o; }
Sym* Sym:: at (Sym*o) { push(o); return this; }
```

строки и Sym::tagstr()

```
Str:: Str (string V):Sym("str",V) {}
string Str:: tagval () { return tagstr (); }
string Sym:: tagstr () { string S = '"';
    for (int i=0,n=val.length(); i<n; i++) {
        char c=val[ i ]; switch (c) {
            case '\t': S+="\\t"; break;
            case '\n': S+="\\n"; break;
            default: S+=c;
        }
    }
return S+"\"; }
```

числа

```
Int:: Int (string V):Sym("int","");
string Int:: tagval () { ostringstream os;
    os << "<" << tag << ":" << val << ">" ; return os.str(); }

Num:: Num (string V):Sym("num","");
string Num:: tagval () { ostringstream os;
    os << "<" << tag << ":" << val << ">" ; return os.str(); }
```

```
Hex:: Hex (string V):Sym("hex",V) {}
Bin:: Bin (string V):Sym("bin",V) {}
```

КОМПОЗИТЫ

```
List:: List ():Sym("[","]") {}
```

функционалы: оператор

```
Op:: Op (string V):Sym("op",V) {}
```

встроенная функция

```
Fn:: Fn (string V, FN F):Sym("fn",V) { fn=F; }
```

```
Lambda :: Lambda () : Sym( "^", "^" ) { }
```

глобальная таблица символов

```
map<string ,Sym*> env;
void env_init() {
    env[ "MODULE" ] = new Sym(MODULE);
}
```

Тестирование интерпретатора

Комментарии

test/comment.src

```
# this is line comment from # till end of line
```

test/comment.log

Скаляры и базовые композиты

test/coretypes.src

```
# core scalar and composite types
```

```
[
    [                                     # numbers / nested list /
        [                                # integers / list /
            -01 , 00 , +002             # int's / linked cons/
            0x12AF                      # machine hex
            0b1101                      # binary string
        ]
        [                                     # floating numbers / cons/
            -01.230 ,                  # point
            -04e+05                     # exponential
        ]
    symbol 'string
can\tbe
multilined ,
```

test/coretypes.log

```
<[:]>
  <:>
    <int:-1>
    <:>
      <int:0>
      <int:2>
    <hex:0x12AF>
    <bin:0b1101>
  <:>
    <num:-1.23>
    <num:-400000>
<:symbol>
'string\ncan\tbe\n\tmultilined'
```

Операторы

A+B	add	сложение
A-B	sub	вычитание
A*B	mul	умножение
A/B	div	деление
A^B	pow	возведение в степень
A>>B	rsh	правый сдвиг
A<<B	lsh	левый сдвиг
<hr/>		
A>B	great	больше
A=>B	greateq	больше или равно
A<B	less	меньше
A<=B	lesseq	меньше или равно
A==B	eq	равно
A!=B	noteq	неравно
A&B	and	и
A B	or	или
A^B	xor	исключающее или
!A	not	не

A=B	assign	назначение/присвоение переменной <i>A предварительно вычисляется</i> ,
A@B	apply	результат является указателем на переменную применение (функции) <i>A</i> к (параметру) <i>B</i> применимо не только к функциям: в общем случае <i>A</i> может быть любым типом в том числе классом: в роли конструктора объекта
~A	quote	блокировка вычисления выражения <i>A</i>
A B	map	применить распределенно <i>A</i> к членам <i>B</i> функция <i>map</i> : <i>A</i> функция, вычислить список → список параллельное вычисление: <i>A</i> constant-функция $f(x) = x$ <i>A@B</i> вычисляются параллельно при наличии поддержки в ядре интерпретатора
A%B	member	вложить <i>B</i> как член <i>A</i> чаще всего используется в определении (добавлении) членов класса
A:B	inherit	наследовать <i>B</i> от <i>A</i> если <i>A</i> составное, выполняется множественное наследование в порядке итерации если <i>A</i> не класс , выполняется наследование копированием
A.B	index	доступ по индексу: <i>B</i> -ый член <i>A</i> <i>B</i> может быть именем или числовым индексом вложенного элемента из <i>A</i>
A<>B	symm	симметричное правило замены $A \leftrightarrow B$
A>>B	is	одностороннее правило замены $A \rightarrow B$
A<!>B	notsym	симметричный запрет замены $A \nleftrightarrow B$
A!>B	notis	односторонний запрет замены $A \not\rightarrow B$

Часть VI

emLinux для встраиваемых систем

Структура встраиваемого микро*Linux*

syslinux Загрузчик

em*Linux* поставляется в виде двух файлов:

1. ядро `(HW)(APP).kernel`
2. сжатый образ корневой файловой системы `(HW)(APP).rootfs`

Загрузчик считывает их с одного из носителей данных, который поддерживается загрузчиком², распаковывает в память, включив защищенный режим процессора, и передает управление ядру *Linux*.

Для работы em*Linux* не требуются какие-либо носители данных: вся (виртуальная) файловая система располагается в ОЗУ. При необходимости к любому из каталогов корневой ФС может быть *подмонтирована* любая существующая дисковая или сетевая файловая система (FAT,NTFS,Samba,NFS,...), причем можно явно запретить возможность записи на нее, защитив данные от разрушения.

Использование rootfs в ОЗУ позволяет гарантировать защиту базовой ОС и пользовательских исполняемых файлов от внезапных выключений питания и ошибочной записи на диск. Еще большую защиту даст отключение драйверов загрузочного носителя в ядре. Если отключить драйвера SATA/IDE и грузиться с USB флешки, практически невозможно испортить основную установку ОС и пользовательские файлы на чужом компьютере.

kernel Ядро *Linux* 3.13.xx

ulibc Базовая библиотека языка Си

busybox Ядро командной среды UNIX, базовые сетевые сервера

дополнительные библиотеки

zlib сжатие/распаковка gzip

???? библиотека помехозащищенного кодирования

png библиотека чтения/записи графического формата .png

freetype рендер векторных шрифтов (TTF)

SDL полноэкранная (игровая) графика, аудио, джойстик

кодеки аудио/видео форматов: ogg vorbis, mp3, mpeg, ffmpeg/gsm

² IDE/SATA HDD, CDROM, USB флешка, сетевая загрузка с BOOTP-сервера по Ethernet

К базовой системе добавляются пользовательские программы */usr/bin* и динамические библиотеки */usr/lib*.

Процедура сборки

Глава 14

clock: коридорные электронные
часы = контроллер умного
дурдома

Глава 15

gambox: игровая приставка

Часть VII

GNU Toolchain и C_+^+ для встраиваемых систем

Глава 16

Программирование встраиваемых систем с использованием GNU Toolchain [23]

© Vijay Kumar B.¹ перевод ²

16.1 Введение

Пакет компиляторов GNU toolchain широко используется при разработке программного обеспечения для встраиваемых систем. Этот тип разработки ПО также называют *низкоуровневым*, *standalone* или *bare metal* программированием (на Си и C_+^+). Написание низкоуровневого кода на Си добавляет программисту новых проблем, требующих глубокого понимания инструмента разработчика — **GNU Toolchain**. Руководства разработчика **GNU Toolchain** предоставляют отличную информацию по инструментарию, но с точки зрения самого **GNU Toolchain**, чем с точки зрения решаемой проблемы. Поэтому было написано это руководство, в котором будут описаны типичные проблемы, с которыми сталкивается начинающий разработчик.

Этот учебник стремится занять свое место, объясняя использование **GNU Toolchain** с точки зрения практического использования. Надеемся, что его будет достаточно для разработчиков, собирающихся освоить и использовать **GNU Toolchain** в их embedded проектах.

В иллюстративных целях была выбрана встроенная система на базе процессорного ядра ARM, которая эмулируется в пакете **Qemu**. С таким подходом вы сможете освоить **GNU Toolchain** с комфортом на вашем рабочем компьютере, без необходимости вкладываться в “физическое” железо, и бороться со сложностями с его запуском. Учебник не стремиться обучить работе с архитектурой

¹ © <http://bravegnu.org/gnu-eprog/>

² © <https://github.com/ponyatov/gnu-eprog/blob/ru/gnu-eprog.asciidoc>

ARM, для этого вам нужно будет воспользоваться дополнительными книгами или онлайн-учебниками типа:

- ARM Assembler <http://www.heyrick.co.uk/assembler/>
- ARM Assembly Language Programming <http://www.arm.com/miscPDFs/9658.pdf>

Но для удобства читателя, некоторое множество часто используемых ARM-инструкций описано в приложении [16.18](#).

16.2 Настройка тестового стенда

В этом разделе описано, как настроить на вашей рабочей станции простую среду разработки и тестирования ПО для платформы ARM, используя **Qemu** и **GNU Toolchain**. **Qemu** это программный³ эмулятор нескольких распространенных аппаратных платформ. Вы можете написать программу на ассемблере и C_+ , скомпилировать ее используя **GNU Toolchain** и отладить ее в эмуляторе **Qemu**.

16.2.1 Qemu ARM

Будем использовать **Qemu** для эмуляции отладочной платы **Gumstix connex** на базе процессора PXA255. Для работы с этим учебником у вас должен быть установлен **Qemu** версии не ниже 0.9.1.

Процессор⁴ PXA255 имеет ядро ARM с набором инструкций ARMv5TE. PXA255 также имеет в своем составе несколько блоков периферии. Некоторая периферия будет описана в этом курсе далее.

16.2.2 Инсталляция Qemu на *Debian GNU/Linux*

Этот учебник требует **Qemu** версии не ниже 0.9.1. Пакет **Qemu** доступный для современных дистрибутивов *Debian GNU/Linux*, вполне удовлетворяет этим условиям, и собирать свежий **Qemu** из исходников совсем не требуется⁵. Установим пакет командой:

```
$ sudo apt install qemu
```

16.2.3 Установка кросс-компилятора **GNU Toolchain** для ARM

Если вы предпочитаете простые пути, установите пакет кросс-компилятора командной

```
sudo apt install gcc-arm-none-eabi
```

или

³ для i386 — программно-аппаратный, использует средства виртуализации хост-компьютера

⁴ Точнее SoC: система-на-кристалле

⁵ хотя может быть и очень хочется

1. Годные чуваки из CodeSourcery⁶ уже давно запилили несколько вариантов **GNU Toolchain**ов для разных ходовых архитектур. Скачайте готовую бинарную бесплатную lite-сборку **GNU Toolchain-ARM**
2. Распакуйте tar-архив в каталог */toolchains*:

```
$ mkdir ~/toolchains  
$ cd ~/toolchains  
$ tar -jxf ~/downloads/arm-2008q1-126-arm-none-eabi-i686-pc-linux-gr
```

3. Добавьте bin-каталог тулчайна в переменную среды PATH.

```
$ PATH=$HOME/toolchains/arm-2008q1/bin:$PATH
```

4. Чтобы каждый раз не выполнять предыдущую команду, вы можете прописать ее в дот-файл **.bashrc**.

Для совсем упертых подойдет рецепт сборки полного комплекта кросс-компиляции из исходных текстов, описанный в 18.

16.3 Hello ARM

В этом разделе вы научитесь пользоваться arm-ассемблером, и тестировать вашу программу на голом железе — эмуляторе платы **connex** в **Qemu**.

Файл исходника ассемблерной программы состоит из последовательности инструкций, по одной на каждую строку. Каждая инструкция имеет формат (каждый компонент не обязателен):

<метка> : <инструкция> @ <комментарий>

метка — типичный способ пометить адрес инструкции в памяти. Метка может быть использована там, где требуется указать адрес, например как операнд в команде перехода. Метка может состоять из латинских букв, цифр⁷, символов _ и \$.

комментарий начинается с символа @ — все последующие символы игнорируются до конца строки

инструкция может быть инструкцией процессора или директивой ассемблера, начинаящейся с точки “.”

Вот пример простой ассемблерной программы 3 для процессора ARM, складывающей два числа:

⁶ подразделение Mentor Graphics

⁷ не может быть первым символом метки

Листинг 3: Сложение двух чисел

```
.text
start:
    mov    r0, #5          @ загрузить в регистр r0 значение 5
    mov    r1, #4          @ загрузить в регистр r1 значение 4
    add    r2, r1, r0      @ сложить r0+r1 и сохранить в r2 (справа налево)

stop:   b stop           @ пустой бесконечный цикл для останова выполнения
```

.text ассемблерная директива, указывающая что последующий код должен быть *ассемблирован* в *секцию кода .text* а не в секцию .data. *Секции* будут подробно описаны далее.

16.3.1 Сборка бинарника

Сохраните программу в файл **add.s**⁸. Для ассемблирования файла вызовите ассемблер **as**:

```
$ arm-none-eabi-as -o add.o add.s
```

Опция -o указывает выходной файл с *объектным кодом*, имеющий стандартное расширение .o⁹.

Команды кросс-тулчайна всегда имеют префикс целевой архитектуры (target triplet), для которой они были собраны, чтобы предотвратить конфликт имен с хост-тулчайном для вашего рабочего компьютера. Далее утилиты **GNU Toolchain** будут использоваться без префикса для лучшей читаемости. **не забывайте добавлять arm-none-eabi-, иначе получите множество странных ошибок типа “unexpected command”.**

```
$ (arm-none-eabi-)as -o add.o add.s
$ (arm-none-eabi-)objdump -x add.o
```

вывод команды **arm-none-eabi-objdump -x**: ELF-заголовки в файле объектного кода

```
add.o:      file format elf32-littlearm
add.o
architecture: armv4, flags 0x00000010:
HAS_SYMS
```

⁸ .S или .S стандартное расширение в мире OpenSource, указывает что это файл с программной на ассемблере

⁹ и внутренний формат ELF (как завещал великий *Linux*)

```
start address 0x00000000  
private flags = 5000000: [ Version5 EABI]
```

Sections:

Idx	Name	Size	VMA	LMA	File off	Algn
0	.text	00000010	00000000	00000000	00000034	2**2
		CONTENTS, ALLOC, LOAD, READONLY, CODE				
1	.data	00000000	00000000	00000000	00000044	2**0
		CONTENTS, ALLOC, LOAD, DATA				
2	.bss	00000000	00000000	00000000	00000044	2**0
		ALLOC				
3	.ARM.attributes	00000014	00000000	00000000	00000044	2**0
		CONTENTS, READONLY				

SYMBOL TABLE:

00000000	l	d	.text	00000000	.text
00000000	l	d	.data	00000000	.data
00000000	l	d	.bss	00000000	.bss
00000000	l		.text	00000000	start
0000000c	l		.text	00000000	stop
00000000	l	d	.ARM.attributes	00000000	.ARM.attributes

Секция .text имеет размер `Size=0x0010 =16` байт, и содержит [машинный код](#):

машинный код из секции .text: **objdump -d**

```
add.o:      file format elf32-littlearm
```

Disassembly of section .text:

```
00000000 <start>:
```

```
 0:   e3a00005    mov r0 , #5
  4:   e3a01004    mov r1 , #4
  8:   e0812000    add r2 , r1 , r0
```

```
0000000c <stop>:
```

```
 c:   eaffffff     b     c <stop>
```

Для генерации [исполняемого файла](#)¹⁰ вызовем [линкер ld](#):

```
$ arm-none-eabi-ld -Ttext=0x0 -o add.elf add.o
```

Опять, опция -o задает выходной файл. -Ttext=0x0 явно указывает адрес, от которого будут отсчитываться все метки, т.е. секция инструкций начинается с адреса 0x0000. Для просмотра адресов, назначенных меткам, можно использовать команду (`arm-none-eabi-)`nm¹¹:

¹⁰ обычно тот же формат ELF.o, сплеленный из одного или нескольких объектных файлов, с некоторыми модификациями см. опцию -T далее

¹¹ NaMes

```
ponyatov@bs:/tmp$ arm-none-eabi-nm add.elf
...
00000000 t start
0000000c t stop
```

* если вы забудете опцию `-T`, вы получите этот вывод с адресами `00008xxx` — эти адреса были заданы при компиляции **GNU Toolchain-ARM**, и могут не совпадать с необходимыми вам. Проверяйте ваши .elfы с помощью **nm** или **objdump**, если программы не запускаются, или **Qemu** ругается на ошибки (защиты) памяти.

Обратите внимание на **назначение адресов** для меток `start` и `stop`: адреса начинаются с `0x0`. Это адрес первой инструкции. Метка `stop` находится после третьей инструкции. Каждая инструкция занимает 4 байта¹², так что `stop` находится по адресу $0xC_{hex} = 12_{dec}$. **Линковка** с другим **базовым адресом** `-Ttext=nnnn` приведет к сдвигу адресов, назначенных меткам.

```
$ arm-none-eabi-ld -Ttext=0x20000000 -o add.elf add.o
$ arm-none-eabi-nm add.elf
... clip ...
20000000 t start
2000000c t stop
```

Выходной файл, созданный **ld** имеет формат, который называется **ELF**. Существует множество форматов, предназначенных для хранения выполняемого и объектного кода¹³. Формат ELF применяется для хранения машинного кода, если вы запускаете его в базовой ОС¹⁴, но поскольку мы собираемся запускать нашу программу на bare metal¹⁵, мы должны сконвертировать полученный .elf файл в более простой **бинарный формат**.

Файл в **бинарном формате** содержит последовательность байт, начинающуюся с определенного адреса памяти, поэтому бинарный файл еще называют **образом памяти**. Этот формат типичен для утилит программирования флеш-памяти микроконтроллеров, так как все что требуется сделать — последовательно скопировать каждый байт из файла в FlashROM-память микроконтроллера, начиная с определенного начального адреса.¹⁶

Команда **GNU Toolchain objcopy** используется для конвертирования машинного кода между разными объектными форматами. Типичное использование:

¹² в множестве команд ARM-32, если вы компилируете код для микроконтроллера Cortex-Mx в режиме команд Thumb или Thumb2, команды 16-битные, т.е. 2 байта

¹³ можно отдельно отметить Microsoft COFF (объектные файлы .obj) и PE (.exe)cutable

¹⁴ прежде всего “большой” или встраиваемый *Linux*

¹⁵ голом железе

¹⁶ Та же операция выполняется и для SoC-систем с NAND-флешем: записать бинарный образ начиная с некоторого аппаратно фиксированного адреса.

```
$ objcopy -O <выходной_формат> <входной_файл> <выходной_файл>
```

Конвертируем **add.elf** в бинарный формат:

```
$ objcopy -O binary add.elf add.bin
```

Проверим размер полученного бинарного файла, он должен быть равен тем же 16 байтам¹⁷:

```
$ ls -al add.bin  
-rw-r--r-- 1 vijaykumar vijaykumar 16 2008-10-03 23:56 add.bin
```

Если вы не доверяете **ls**, можно дизассемблировать бинарный файл:

```
ponyatov@bs:/tmp$ arm-none-eabi-objdump -b binary -m arm -D add.bin  
  
add.bin:      file format binary
```

Disassembly of section .data:

```
00000000 <.data>:  
 0:   e3a00005      mov    r0, #5  
 4:   e3a01004      mov    r1, #4  
 8:   e0812000      add    r2, r1, r0  
 c:   ea\xff\xfe      b      0xc  
ponyatov@bs:/tmp$
```

Опция **-b** задает формат файла, опция **-m** (machine) архитектуру процессора, получить полный список сочетаний **-b/-m** можно командной **arm-none-eabi-objdump**

16.3.2 Выполнение в **Qemu**

Когда ARM-процессор сбрасывается, он начинает выполнять команды с адресе 0x0. На плате Connex установлен флеш на 16 мегабайт, начинающийся с адрес 0x0. Таким образом, при сбросе будут выполняться инструкции с начала флеша.

Когда **Qemu** эмулирует плату connex, в командной строке должен быть указан файл, который будет считаться образом флеш-памяти. Формат флеша очень прост — это побайтный образ флеша без каких-либо полей или заголовков, т.е. это тот же самый **бинарный формат**, описанный выше.

Для тестирования программы в эмуляторе Gumstix connex, сначала мы создаем 16-мегабайтный файл флеша, копируя 16М нулей из файла **/dev/zero** с помощью команды **dd**. Данные копируются 4Кбайтными блоками¹⁸ (4096 x 4K):

¹⁷ 4 инструкции по 4 байта каждая

¹⁸ опция **bs=** (blocksize)

```
$ dd if=/dev/zero of=flash.bin bs=4K count=4K
4096+0 записей получено
4096+0 записей отправлено
скопировано 16777216 байт (17 MB), 0,0153502 с, 1,1 GB/c
```

```
$ du -h flash.bin
16M    flash.bin
```

Затем переписываем начало **flash.bin** копируя в него содержимое **add.bin**:

```
$ dd if=add.bin of=flash.bin bs=4K conv=notrunc
0+1 записей получено
0+1 записей отправлено
скопировано 16 байт (16 B), 0,000173038 с, 92,5 kB/c
```

После сброса процессор выполняет код с адреса 0x0, и будут выполняться инструкции нашей программы. Команда запуска **Qemu**:

```
$ qemu-system-arm -M connex -pflash flash.bin -nographic -serial /dev/null
```

```
QEMU 2.1.2 monitor - type 'help' for more information
(qemu)
```

Опция **-M connex** выбирает режим эмуляции: **Qemu** поддерживает эмуляцию нескольких десятков вариантов железа на базе ARM процессоров. Опция **-pflash** указывает файл образа флеша, который должен иметь определенный размер (16M). **-nographic** отключает эмуляцию графического дисплея (в отдельном окне). Самая важная опция **-serial /dev/null** подключает последовательный порт платы на **/dev/null**, при этом в терминале после запуска **Qemu** вы получите **консоль монитора**.

Qemu выполняет инструкции, и останавливается в бесконечном цикле на **stop**, выполняя команду **stop: b stop**. Для просмотра содержимого регистров процессора воспользуемся **монитором**. Монитор имеет интерфейс командной строки, который вы можете использовать для контроля работы эмулируемой системы. Если вы запустите **Qemu** как указано выше, монитор будет доступен через **stdio**.

Для просмотра регистров выполним команду **info registers**:

```
(qemu) info registers
R00=00000005 R01=00000004 R02=00000009 R03=00000000
R04=00000000 R05=00000000 R06=00000000 R07=00000000
R08=00000000 R09=00000000 R10=00000000 R11=00000000
R12=00000000 R13=00000000 R14=00000000 R15=0000000c
PSR=400001d3 -Z-- A svc32
FPSCR: 00000000
```

Обратите внимание на значения в регистрах r00..r02: 4, 5 и ожидаемый результат 9. Особое значение для ARM имеет регистр r15: он является указателем команд, и содержит адрес текущей выполняемой машинной команды, т.е. 0x000c: b stop.

16.3.3 Другие команды монитора

Несколько полезных команд монитора:

help	список доступных команд
quit	выход из эмулятора
xp /fmt addr	вывод содержимого физической памяти с адреса addr
system_reset	перезапуск

Команда xp требует некоторых пояснений. Аргумент /fmt указывает как будет выводиться содержимое памяти, и имеет синтаксис <счетчик><формат><размер>:

счетчик число элементов данных

size размер одного элемента в битах: b=8 бит, h=16, w=32, g=64

format определяет формат вывода:

- x** hex
- d** десятичные целые со знаком
- u** десятичные без знака
- o** 8ричные
- c** символ (char)
- i** инструкции ассемблера

Команда xp в формате i будет дизассемблировать инструкции из памяти. Выведем дамп с адреса 0x0 указав fmt=4iw: 4 — 4 , i — инструкции размером w=32 бита:

```
(qemu) xp /4wi 0x0
0x00000000: e3a00005      mov  r0, #5   ; 0x5
0x00000004: e3a01004      mov  r1, #4   ; 0x4
0x00000008: e0812000      add  r2, r1, r0
0x0000000c: ea\xff\fe    b    0xc
```

16.4 Директивы ассемблера

В этом разделе мы посмотрим несколько часто используемых директив ассемблера, используя в качестве примера пару простых программ.

16.4.1 Суммирование массива

Следующий код 4 вычисляет сумму массива байт и сохраняет результат в r3:

Листинг 4: Сумма массива

```
.text
entry: b start
arr:    .byte 10, 20, 25
eoa:           .align
start:
        ldr r0, =eoa      @ r0 = &eoa
        ldr r1, =arr      @ r1 = &arr
        mov r3, #0         @ r3 = 0
loop:   ldrb r2, [r1], #1    @ r2 = *r1++
        add r3, r2, r3    @ r3 += r2
        cmp r1, r0         @ if (r1 != r2)
        bne loop          @ goto loop
stop:   b stop
```

В коде используются две новых ассемблерных директивы, описанных далее:
.byte и .align.

.byte

Аргументы директивы .byte асSEMBлируются в последовательность байт в памяти. Также существуют аналогичные директивы .2byte и .4byte для асSEMBлирования 16- и 32-битных констант:

```
.byte  exp1, exp2, ...
.2byte exp1, exp2, ...
.4byte exp1, exp2, ...
```

Аргументом может быть целый числовой литерал: двоичный с префиксом 0b, 8-ричный префикс 0, десятичный и hex 0x. Также может использоваться строковая константа в одиночных кавычках, асSEMBлируемая в ASCII значения.

Также аргументом может быть Си-выражение из литералов и других символов, примеры:

```
pattern: .byte 0b01010101, 0b00110011, 0b00001111
npattern: .byte npattern - pattern
halpha:   .byte 'A', 'B', 'C', 'D', 'E', 'F'
dummy:    .4byte 0xDEADBEEF
nalpha:   .byte 'Z' - 'A' + 1
```

```
.align
```

Архитектура ARM требует чтобы инструкции находились в адресах памяти, выровненных по границам 32-битного слова, т.е. в адресах с нулями в 2х младших разрядах. Другими словами, адрес каждого первого байта из 4-байтной команды, должен быть кратен 4. Для обеспечения этого предназначена директива `.align`, которая вставляет выравнивающие байты до следующего выровненного адреса. Ее нужно использовать только если в код вставляются байты или неполные 32-битные слова.

16.4.2 Вычисление длины строки

Этот код вычисляет длину строки и помещает ее в `r1`:

Листинг 5: Длина строки

```
.text
b start

str:    .asciz "Hello World"
        .equ    nul, 0

        .align
start:   ldr    r0, =str          @ r0 = &str
        mov    r1, #0

loop:    ldrb   r2, [r0], #1      @ r2 = *(r0++)
        add    r1, r1, #1      @ r1 += 1
        cmp    r2, #nul         @ if (r1 != nul)
        bne    loop            @ goto loop

        sub    r1, r1, #1      @ r1 -= 1
stop:   b stop
```

Код иллюстрирует применение директив `.asciz` и `.equ`.

```
.asciz
```

Директива `.asciz` принимает аргумент: строковый литерал, последовательность символов в двойных кавычках. Строковые литералы ассемблируются в память последовательно, добавляя в конец нулевой символ \0 (признак конца строки в языке Си и стандарте POSIX).

Директива `.ascii` аналогична `.asciz`, но конец строки не добавляется. Все символы — 8-битные, кириллица может не поддерживаться.

.equ

Ассемблер при своей работе использует **таблицу символов**: она хранит соответствия меток и их адресов. Когда ассемблер встречает очередное определение метки, он добавляет в таблицу новую запись. Когда встречается упоминание метки, оно заменяется соответствующим адресом из таблицы.

Использование директивы `.equ` позволяет добавлять записи в таблицу символов вручную, для привязки к именам любых числовых значений, не обязательно адресов. Когда ассемблер встречает эти имена, они заменяются на их значения. Эти имена-константы, и имена меток, называются **символами**, а таблицы записанные в объектные файлы, или в отдельные `.sym` файлы — **таблицами символов**¹⁹.

Синтаксис директивы `.equ`:

```
.equ <имя>, <выражение>
```

Имя символа имеет те же ограничения по используемым символам, что и метка. Выражение может быть одиночным литералом или выражением как и в директиве `.byte`.

В отличие от `.byte`, директива `.equ` не выделяет никакой памяти под аргумент. Она только добавляет значение в таблицу символов.

16.5 Использование ОЗУ (адресного пространства процессора)

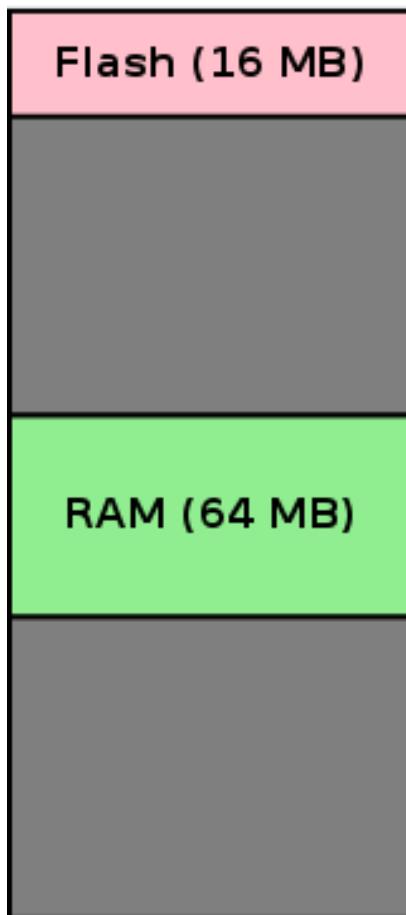
Flash-память описанная ранее, в которой хранится машинный код программы, один из видов EEPROM²⁰. Это вторичный носитель данных, как например жесткий диск, но в любом случае хранить данные и значения переменных во флашне неудобно как с точки зрения возможности перезаписи, так и прежде всего со скоростью чтения флаша, и кешированием.

В предыдущем примере мы использовали флаш как EEPROM для хранения константного массива байт, но вообще переменные должны храниться в максимально быстрой и неограниченно перезаписываемой RAM.

Плата connex имеет 64Мб ОЗУ начиная с адреса 0xA0000000, для хранения данных программы. Карта памяти может быть представлена в виде диаграммы:

¹⁹ также используются отладчиком, чтобы показывать адреса переходов в виде понятных программисту символов, а не мутных числовых констант

²⁰ Electrical Eraseable Programmable Read-Only Memory, электрически стираемая перепрограммируемая память только-для-чтения



0x0000_0000

0x0100_0000

0xA000_0000

0xA400_0000

Карта памяти Gumstix

21

Для настройки размещения переменных по нужным физическим адресам **нужна** некоторая **настройка адресного пространства**, особенно **если вы используете внешнюю память и переферийные устройства, подключаемые к внешнейшине**²².

Для этого нужно понять, какую роль в распределении памяти играют ассемблер и линкер.

²¹ здесь адреса считаются сверху вниз, что нетипично, обычно на диаграммах памяти адреса увеличиваются снизу вверх.

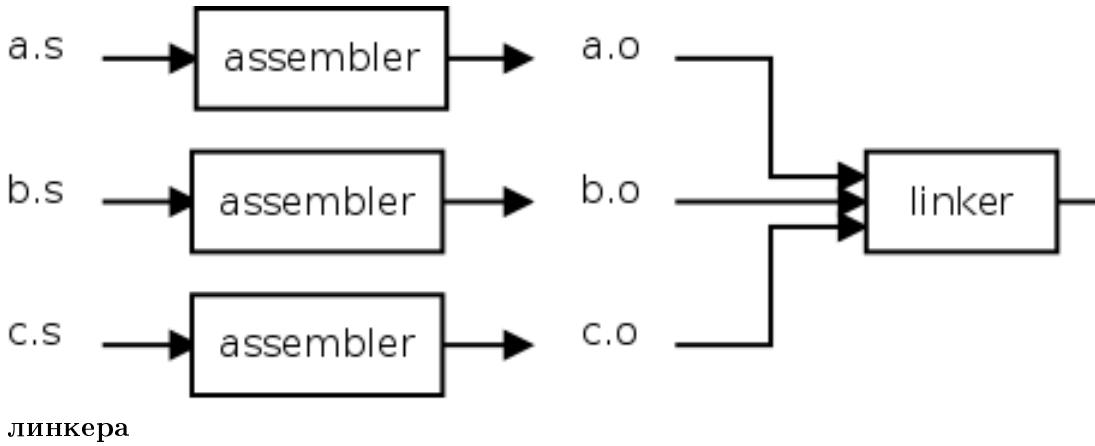
²² или используете малораспространенные клоны ARM-процессоров, типа Миландровского 1986BE9x “чернобыль”

16.6 Линкер

Линкер позволяет **скомпоновать** исполняемый файл программы из нескольких объектных файлов, поделив ее на части. Чаще всего это нужно при использовании нескольких компиляторов для разных языков программирования: ассемблер, компиляторы C^+ , Фортрана и Паскаля.

Например, очень известная библиотека численных вычислений на базе матриц BLAS/LAPACK написана на Фортране, и для ее использования с сишной программой нужно слинковать program.o, blas.a и lapack.a²³ в один исполняемый файл.

При написании многофайловой программы (еще это называют **инкрементной компоновкой**) каждый файл исходного кода ассемблируется в индивидуальный файл объектного кода. Линкер²⁴ собирает объектные файлы в финальный исполняемый бинарник.



При сборке объектных файлов, линкер выполняет следующие операции:

- symbol resolution (разрешение символов)
- relocation (релокация)

В этой секции мы детально рассмотрим эти операции.

16.6.1 Разрешение символов

В программе из одного файла при создании объектного файла все ссылки на метки заменяются их адресами непосредственно ассемблером. Но в программе из нескольких файлов существует множество ссылок на метки в других файлах, неизвестные на момент ассемблирования/компиляции, и ассемблер помечает

²³ .a — файлы архивов из пары сотен отдельных .o файлов каждый, по одному .o файлу на каждый возможный вариант функции линейной алгебры

²⁴ или компоновщик

их “unresolved” (неразрешённые). Когда эти объектные файлы обрабатываются линкером, он определяет адреса этих меток по информации из других объектных файлов, и корректирует код. Этот процесс называется **разрешением символов**.

Пример суммирования массива разделен на два файла для демонстрации разрешения символов, выполняемых линкером. Эти файлы ассемблируются отдельно, чтобы их таблицы символов содержали неразрешенные ссылки.

Файл **sumsub.s** содержит процедуру суммирования, а **summain.s** вызывает процедуру с требуемыми аргументами:

Листинг 6: summain.s вызов внешней процедуры

```
.text
b start          @ пропустить данные
arr: .byte 10, 20, 25      @ константный массив байт
eoa:             @ адрес массива + 1
.align
start:
    ldr r0, =arr        @ r0 = &arr
    ldr r1, =eoa        @ r1 = &eoa
    bl sum              @ (вложенный) вызов процедуры
stop:   b stop
```

Листинг 7: sumsub.s код процедуры

```
@ Аргументы (в регистрах)
@ r0: начальный адрес массива
@ r1: конечный адрес массива
@
@ Возврат результата
@ r3: сумма массива

.global sum

sum:   mov r3, #0           @ r3 = 0
loop:  ldrb r2, [r0], #1     @ r2 = *r0++ ; получить следующий элем
      add r3, r2, r3        @ r3 += r2       ; суммирование
      cmp r0, r1             @ if (r0 != r1) ; проверка на конец массива
      bne loop              @ goto loop      ; цикл
      mov pc, lr             @ pc = lr       ; возврат результата по lr
```

²⁵ в архитектуре ARM нет специальной команды возврата ret, ее функцию выполняет прямое присваивание регистра указателя команд mov pc,lr

Применение директивы `.global` обязательно. В Си все функции и переменные, определенные вне функций, считаются видимыми из других объектных файлов, если они не определены с модификатором `static`. В ассемблере наоборот все метки считаются *статическими*²⁶, если с помощью директивы `.global` специально не указано, что они должны быть видимы извне.

Ассемблируйте файлы, и посмотрите дамп их таблицы символов используя комманду `nm`:

```
$ arm-none-eabi-as -o main.o main.s
$ arm-none-eabi-as -o sum-sub.o sum-sub.s
$ arm-none-eabi-nm main.o
00000004 t arr
00000007 t eoa
00000008 t start
00000018 t stop
    U sum
$ arm-none-eabi-nm sum-sub.o
00000004 t loop
00000000 T sum
```

Теперь сфокусируемся на букве во втором столбце, который указывает тип символа: `t` указывает что символ определен в секции кода `.text`, `u` указывает что символ не определен. Буква в верхнем регистре указывает что символ глобальный и был указан в директиве `.global`.

Очевидно, что символ `sum` определ в `sum-sub.o` и еще не разрешен в `main.o`. Вызов линкера разрешает символьные ссылки, и создает исполняемый файл.

16.6.2 Релокация

Релокация — процесс изменения уже назначенных меткам адресов. Он также выполняет коррекцию всех ссылок, чтобы отразить заново назначенные адреса меток. В общем, релокация выполняется по двум основным причинам:

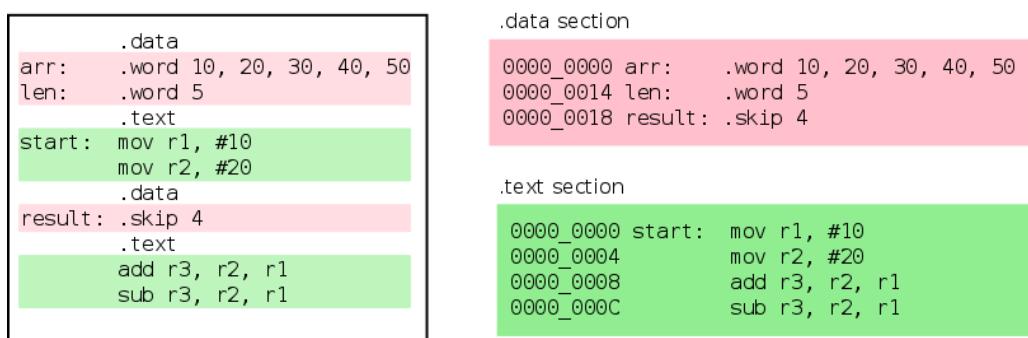
1. Объединение секций
2. Размещение секций

Для понимания процесса релокации, нужно понимание самой концепции секций.

Код и данные отличаются по требованиям при исполнении. Например код может размещаться в ROM-памяти, а данные требуют память открытую на запись. Очень хорошо, если **области кода и данных не пересекаются**. Для этого программы делятся на секции. Большинство программ имеют как минимум две секции: `.text` для кода и `.data` для данных. Ассемблерные директивы `.text` и `.data` ожидаются использовать для переключения между этими секциями.

²⁶ или локальными на уровне файла

Хорошо представить каждую секцию как ведро. Когда ассемблер натыкается на директиву секции, он начинает сливать код/данные в соответствующее ведро, так что они размещаются в смежных адресах. Эта диаграмма показывает как ассемблер упорядочивает данные в секциях:



Секции

Теперь, когда у нас есть общее понимание **секционирования** кода и данных, давайте посмотрим по каким причинам выполняется релокация.

Объединение секций

Когда вы имеете дело с многофайловыми программами, секции в каждом объектном файле имеют одинаковые имена ('.text',...), линкер отвечает за их объединение в выполняемом файле. По умолчанию секции с одинаковыми именами из каждого .o файла объединяются последовательно, и метки корректируются на новые адреса.

Эффекты объединения секций можно посмотреть, анализируя таблицы символов отдельно в объектных и исполняемом файлах. Многофайловая программа суммирования может иллюстрировать объединение секций. Дампы таблиц символов:

```
$ arm-none-eabi-nm main.o
00000004 t arr
00000007 t eoa
00000008 t start
00000018 t stop
    U sum
$ arm-none-eabi-nm sum-sub.o
00000004 t loop <1>
00000000 T sum
$ arm-none-eabi-ld -Ttext=0x0 -o sum.elf main.o sum-sub.o
$ arm-none-eabi-nm sum.elf
...
00000004 t arr
```

```
00000007 t eoa
00000008 t start
00000018 t stop
00000028 t loop <1>
00000024 T sum
```

1. символ `loop` имеет адрес `0x4` в `sum-sub.o`, и `0x28` в `sum.elf`, так как секция `.text` из `sum-sub.o` размещена сразу за секцией `.text` из `main.o`.

Размещение секций

Когда программа ассемблируется, каждой секции назначается стартовый адрес `0x0`. Поэтому всем переменным назначаются адреса относительно начала секции. Когда создается финальный исполняемый файл, секция размещаются по некоторому адресу `X`, и все адреса меток, назначенные в секции, увеличиваются на `X`, так что они указывают на новые адреса.

Размещение каждой секции по определенному месту в памяти и коррекцию всех ссылок на метки в секции, выполняет линкер.

Эффект размещения секций можно увидеть по тому же дампу символов, описанному выше. Для простоты используем объектный файл однофайловой программы суммирования `sum.o`. Для увеличения заметности искусственно разместим секцию `.text` по адресу `0x100`:

```
$ arm-none-eabi-as -o sum.o sum.s
$ arm-none-eabi-nm -n sum.o
00000000 t entry <1>
00000004 t arr
00000007 t eoa
00000008 t start
00000014 t loop
00000024 t stop
$ arm-none-eabi-ld -Ttext=0x100 -o sum.elf sum.o <2>
$ arm-none-eabi-nm -n sum.elf
00000100 t entry <3>
00000104 t arr
00000107 t eoa
00000108 t start
00000114 t loop
00000124 t stop
...
...
```

1. Адреса меток назначаются с `0` от начала секции.
2. Когда создается выполняемый файл, линкеру указано разместить секцию кода с адреса `0x100`.

3. Адреса меток в `.text` переназначаются начиная с 0x100, и все ссылки на метки корректируются.

Процесс объединения и размещения секций в общем показаны на диаграмме:

a.s (.text)

```
strcpy: ldrb r0, [r1], #1  
strb r0, [r2], #1  
cmp r0, 0  
bne strcpy  
mov pc, lr
```

b.s (.text)

```
strlen: ldrb r0, [r1]  
add r2, #1  
cmp r0, 0  
bne strlen  
mov pc, lr
```

Assembler

```
0000_0000 strcpy: ldrb r0, [r1], #1  
0000_0004 strb r0, [r2], #1  
0000_0008 cmp r0, 0  
0000_000C bne strcpy  
0000_0010 mov pc, lr
```

```
0000_0000 strlen: ldrb r0, [r1]  
0000_0004 add r2, #1  
0000_0008 cmp r0, 0  
0000_000C bne strlen  
0000_0010 mov pc, lr
```

Assembler

Merging .text sections from two files

```
0000_0000 strcpy: ldrb r0, [r1], #1  
0000_0004 strb r0, [r2], #1  
0000_0008 cmp r0, 0  
0000_000C bne strcpy  
0000_0010 mov pc, lr  
0000_0014 strlen: ldrb r0, [r1], #1  
0000_0018 add r2, #1  
0000_001C cmp r0, 0  
0000_0020 bne strlen  
0000_0024 mov pc, lr
```

Patch

New address
after merge

Placing .text section at 0x2000_0000

```
2000_0000 strcpy: ldrb r0, [r1], #1  
2000_0004 strb r0, [r2], #1  
2000_0008 cmp r0, 0  
2000_000C bne strcpy  
2000_0010 mov pc, lr  
2000_0014 strlen: ldrb r0, [r1], #1  
2000_0018 add r2, #1  
2000_001C cmp r0, 0  
2000_0020 bne strlen  
2000_0024 mov pc, lr
```

Patch

Объединение и размещение секций

16.7 Скрипт линкера

Как было описано в предыдущем разделе, объединение и размещение секций выполняется линкером. Программист может контролировать этот процесс через **скрипт линкера**. Очень простой пример скрипта линкера:

Листинг 8: Простой скрипт линкера

```
SECTIONS { <1>
. = 0x00000000; <2>
.text : { <3>
abc.o (.text);
def.o (.text);
} <3>
}
```

1. Команда **SECTIONS** наиболее важная команда, она указывает как секции объединяются, и по каким адресам они размещаются.
2. Внутри блока **SECTIONS** команда **.** (точка) представляет **указатель адреса размещения**. Указатель адреса всегда инициализируется **0x0**. Его можно модифицировать явно присваивая новое значение. Показанная явная установка на **0x0** на самом деле не нужна.
3. Этот блок скрипта определяет что секция **.text** выходного файла составляется из секций **.text** в файлах **abc.o** и **def.o**, причем именно в этом порядке.

Скрипт линкера может быть значительно упрощен и универсализирован с помощью использования символа шаблона ***** вместо индивидуального указания имен файлов:

Листинг 9: Шаблоны в скриптах линкера

```
SECTIONS {
. = 0x00000000;
.text : { * (.text); }
}
```

Если программа одновременно содержит секции **'.text'** и **'.data'**, объединение и размещение секций можно прописать вот так:

Листинг 10: Несколько секций

```
SECTIONS {
. = 0x00000000;
.text : { * (.text); }
```

```
. = 0x00000400;
.data : { * (.data); }
}
```

Здесь секция `.text` размещается по адресу `0x0`, а секция `.data` — `0x400`. Отметим, что если указателю размещения не привавивать значения, секции `.text` и `.data` будут размещены в смежных областях памяти.

16.7.1 Пример скрипта линкера

Для демонстрации использования скриптов линкера рассмотрим применение скрипта `??` для размещения секций `.text` и `.data`. Для этого используем немного измененный пример программы суммирования массива, разделив код и данные в отдельные секции:

Листинг 11: Программа суммирования массива

```
.data
arr: .byte 10, 20, 25 @ Read-only array of bytes
eoa: @ Address of end of array + 1
```

```
.text
start:
ldr r0, =eoa    @ r0 = &eoa
ldr r1, =arr @ r1 = &arr
mov r3, #0 @ r3 = 0
loop: ldrb r2, [r1], #1 @ r2 = *r1++
add r3, r2, r3 @ r3 += r2
cmp r1, r0 @ if (r1 != r2)
bne loop @ goto loop
stop: b stop
```

- Изменения заключаются в выделении массива в секцию `.data` и удалении директивы выравнивания `.align`.
- Также не требуется инструкция перехода на метку `start` для обхода данных, так как линкер разместит секции отдельно. В результате команды программы размещаются любым удобным способом, а скрипт линкера позаботится о правильном размещении сегментов в памяти.

При линковке программы в командной строке нужно указать использования скрипта:

```
$ arm-none-eabi-as -o sum-data.o sum-data.s
$ arm-none-eabi-ld -T sum-data.lds -o sum-data.elf sum-data.o
```

Опция `-T sum-data.lds` указывает что используется скрипт `sum-data.lds`. Выводим таблицу символов и видим размещение сегментов в памяти:

```
$ arm-none-eabi-nm -n sum-data.elf
00000000 t start
0000000c t loop
0000001c t stop
00000400 d arr
00000403 d eoa
```

Из таблицы символов видно что секция `.text` размещена с адреса 0x0, а секция `.data` с 0x400.

16.7.2 Анализ объектного/исполняемого файла утилитой `objdump`

Более подробную информацию даст утилита `objdump`:

```
$ arm-none-eabi-as -o sum-data.o sum-data.s
$ arm-none-eabi-ld -T sum-data.lds -o sum-data.elf sum-data.o
$ arm-none-eabi-objdump sum-data.elf
```

Листинг 12: `sum-data.objdump`

1. указание на архитектуру,
2. для которой предназначен исполняемый файл
3. стартовый адрес в секции `.text`, по умолчанию 0x0²⁷
4. ***ABI*** — соглашения о передаче
5. параметров в регистрах/стеке (для Си кода)
6. приведена подробная информация о секциях
7. `.text` секция кода
8. `.data` секция данных
9. служебная информация
10. столбец `Size` указывает размер секции в байтах (hex)
11. `VMA`²⁸ указывает адрес размещения сегмента

²⁷ обязателен и фиксирован для прошивок микроконтроллеров, т.к. на него перескакивает аппаратный сброс

²⁸ Virtual Memory Address

12. **Align** (Align) автоматическое выравнивание содержимого сегмента в памяти, в степени двойки 2^{**n} : код выравнивается кратно $2^{**2=4}$ байтам, данные не выравниваются $2^{**0=1}$
13. Флаг **ALLOC** (Allocate) указывает что при загрузке программы под этот сегмент должна быть выделена память.
14. **LOAD** указывает что содержимое сегмента должно загружаться из исполняемого файла в память при использовании ОС, а для микроконтроллеров указывает программатору что сегмент нужно прошивать.
15. **READONLY** сегмент с константными неизменяемыми данными, которые могут быть размещены в ROM, а при использовании ОС область памяти должна быть помечена в таблице системы защиты памяти как R/O. Отсутствие флага **READONLY** + наличие **LOAD** указывает что данные должны загружаться **только в ОЗУ**.
16. сегмент кода
17. сегмент данных
18. таблица символов
19. дизассемблированный код из секций, помеченных флагом **CODE**: `.text`

16.8 Данные в RAM, пример

Теперь мы знаем как писать скрипты линкера, и можем попытаться написать программу, разместив данные в секции `.data` в ОЗУ.

Программа сложения модифицирована для загрузки значений из ОЗУ, и записи результата обратно в ОЗУ: память для операндов и результат размещена в секции `.data`.

Листинг 13: Данные в ОЗУ

```
.data
val1: .4byte 10 @ First number
val2: .4byte 30 @ Second number
result: .4byte 0 @ 4 byte space for result
```

```
.text
.align
start:
ldr r0, =val1 @ r0 = &val1
ldr r1, =val2 @ r1 = &val2
```

```
ldr    r2, [r0] @ r2 = *r0
ldr    r3, [r1] @ r3 = *r1

add    r4, r2, r3 @ r4 = r2 + r3

ldr    r0, =result @ r0 = &result
str    r4, [r0] @ *r0 = r4

stop: b stop
```

Листинг 14: Скрипт для линковки

```
SECTIONS {
. = 0x00000000;
.text : { * (.text); }

. = 0xA0000000;
.data : { * (.data); }
}
```

Дамп таблицы символов:

```
$ arm-none-eabi-nm -n add-mem.elf
00000000 t start
00000001c t stop
a0000000 d val1
a0000001 d val2
a0000002 d result
```

Скрипт линкера решил проблему с размещением данных, но **проблема с использованием ОЗУ еще не решена !**

16.8.1 RAM энергозависима (volatile)!

ОЗУ стирается при отключении питания, поэтому для использования ОЗУ недостаточно разместить сегменты.

Во флеше должен храниться не только код, но **и данные**, чтобы при подаче питания специальный **startup код** выполнил **инициализацию ОЗУ**, копируя данные из флеша. Затем управление передается основной программе.

Поэтому секция `.data` имеет **два адреса размещения**: **адрес хранения** во флеше **LMA** и **адрес размещения** в ОЗУ **VMA**.

TIP: как видно из раздела ??, в терминах **ld** адрес хранения (загрузки) называется **LMA** (Load Memory Address), а адрес размещения (времени выполнения) **VMA** (Virtual Memory Address).

Нужно сделать следующие две модификации, чтобы программа работала корректно:

1. модифицировать .lds чтобы для секции .data в нем учитывались оба адреса: LMA и VMA.
2. написать небольшой кусочек кода, который будет **инициализировать память данных**, копируя образ секции .data из флеша (из адреса хранения LMA) в ОЗУ (по адресу исполнения, VMA).

16.8.2 Спецификация адреса загрузки LMA

VMA это адрес, который должен быть использован для вычисления адресов всех меток при исполнении программы. В предыдущем линк-скрипте мы задали VMA секции .data. LDA не указан, и по умолчанию равен VMA. Это нормально для сегментов, размещаемых в ROM. Но если используются инициализируемые сегменты в ОЗУ, нужно задать отдельно VMA и LMA.

Адрес загрузки LMA, отличающийся от адреса выполнения VMA, задается с помощью команды AT. Модифицированный скрипт показан ниже:

```
SECTIONS {  
    . = 0x00000000;  
    .text : { * (.text); }  
    etext = .; <1>  
  
    . = 0xA0000000;  
    .data : AT (etext) { * (.data); } <2>  
}
```

1. В блоках описания секций можно создавать символы, назначая им значения: числовой адрес или текущую позицию с помощью точки ". ". Символу **etext** назначается адрес флеша, следующий сразу за концом кода. Отметим что **etext** сам по себе не выделяет никакой памяти, а только помечает адрес LMA сегмента .data в таблице символов.
2. При настройке сегмента .data с помощью ключевого слова AT (**etext**) назначается LMA для хранения содержимого сегмента данных. Команде AT может быть передан любой адрес или символ²⁹. Так что в результате мы настроили адрес хранения .data на область флеша, помеченную символом **etext**.

²⁹ значением которого является валидный адрес

16.8.3 Копирование ‘.data’ в ОЗУ

Для копирования данных инициализации из флеши в ОЗУ требуется следующая информация:

1. Адрес данных во флеше (`flash_sdata`)
2. Адрес данных в ОЗУ (`ram_sdata`)
3. Размер секции `.data` (`data_size`)

Имея эту информацию, сегмент `.data` может быть инициализирован может быть скопирован следующим стартовым кодом:

```
ldr r0, =flash_sdata
ldr r1, =ram_sdata
ldr r2, =data_size
copy:
ldrb r4, [r0], #1
strb r4, [r1], #1
subs r2, r2, #1
bne copy
```

Для получения такой информации скрипт линкера нужно немного модифицировать:

Листинг 15: Скрипт линкера с символами для копирования секции `.data`

```
SECTIONS {
. = 0x00000000;
.text : { * (.text); }
flash_sdata = .; <1>

. = 0xA0000000;
ram_sdata = .; <2>
.data : AT(flash_sdata) { * (.data); }
ram_edata = .; <3>
data_size = ram_edata - ram_sdata; <3>
}
```

1. Начало данных во флеше сразу за секцией кода.
2. Начало данных — базовый адрес ОЗУ в адресном пространстве процессора.
3. Получение размера непросто: размер вычисляется вычитанием адресов метод начала и конца данных. Да, простые выражения тоже можно использовать в скрипте линкера.

Полный листинг программы с добавленной инициализацией данных:

Листинг 16: Инициализация ОЗУ

```
.data
val1: .4byte 10 @ First number
val2: .4byte 30 @ Second number
result: .space 4 @ 1 byte space for result

.text

;; Copy data to RAM.
start:
ldr r0, =flash_sdata
ldr r1, =ram_sdata
ldr r2, =data_size

copy:
ldrb r4, [r0], #1
strb r4, [r1], #1
subs r2, r2, #1
bne copy

;; Add and store result.
ldr r0, =val1 @ r0 = &val1
ldr r1, =val2 @ r1 = &val2

ldr r2, [r0] @ r2 = *r0
ldr r3, [r1] @ r3 = *r1

add r4, r2, r3 @ r4 = r2 + r3

ldr r0, =result @ r0 = &result
str r4, [r0] @ *r0 = r4

stop: b stop
```

Листинг 17: add-ram.objdump Программа была ассемблирована и скомпилирована используя .lds в ??.

Запуск и тестирование программы в Qemu:

```
qemu-system-arm -M connex -pflash flash.bin -nographic -serial /dev/null
(qemu) xp /4dw 0xA0000000
a0000000:          10            30            40            0
```

На реальной физической системе с SDRAM, память не может использована сразу. Сначала нужно инициализировать контроллер памяти, и только затем обращаться к ОЗУ. Наш код работает потому, что симулятор не требует инициализации контроллера.

16.9 Обработка аппаратных исключений

Все примеры программ, приведенные выше, содержат гигантский баг: **первые 8 машинных слов в адресном пространстве зарезервированы для векторов исключений**. Когда возникает исключение, выполняется аппаратный переход на один из этих жестко заданных меток. Исключения и их адреса приведены в следующей таблице:

Адреса векторов исключений

Исключение		Адрес
Сброс	Reset	0x00
Неопределенная инструкция	Undefined Instruction	0x04
Программное прерывание (SWI)	Software Interrupt (SWI)	0x08
Ошибка предвыборки	Prefetch Abort	0x0C
Ошибка данных	Data Abort	0x10
Резерв, не используется	Reserved, not used	0x14
Аппаратное прерывание	IRQ	0x18
Быстрое прерывание	FIQ	0x1C

Предполагается что по этим адресам находятся команды перехода, которые передадут управление на соответствующий произвольный адрес обработчика исключения. Во всех примерах ранее бы не вставляли таблицу обработчиков исключений, так как мы предполагали что эти исключения не случатся. Все эти программы можно скорректировать, слинковав их со следующим кодом:

```
.section "vectors"
reset: b      start
undef: b      undef
swi: b      swi
pabt: b     pabt
dabt: b     dabt
nop
irq: b      irq
fiq: b      fiq
```

Только обработчик `reset` векторизован на отдельный адрес `start`. Все остальные исключения векторизованы сами на себя. Таким образом если случится любое исключение, процессор зациклится на соответствующем векторе. В этом случае

возникшее исключение может быть идентифицировано в отладчике (мониторе Qemu, в нашем случае) по адресу указателя команд `pc=r15`.

В ассемблерном коде видно применение директивы `.section` которая позволяет создавать секции с произвольными именами, чтобы прописать для них отдельную обработку в скрипте линкера.

Чтобы обеспечить правильное размещение таблицы обработчиков, нужно скорректировать скрипт линкера:

```
SECTIONS {
. = 0x00000000;
.text : {
* (vectors);
* (.text);
...
}
...
}
```

Обратите внимание что секция `vectors` размещена сразу за инициализацией указателя размещения на первом месте, до всего остального кода, что гарантирует что таблица векторов будет находиться по жесткому адресу `0x0`.

16.10 Стартап-код на Си

Если процесс только что был сброшен, невозможно напрямую выполнить Си-код, так как в отличие от ассемблера, программы на Си требуют для себя некоторой предварительной настройки. В этом разделе описаны эти предварительные требования, и как их выполнить.

Мы возьмем пример Си-программы которая вычисляет сумму массива. И к концу раздела мы уже будем способны, сделав некоторые низкоуровневые настройки, передать управление и выполнить ее.

Листинг 18: Сумма массива на Си

```
static int arr[] = { 1, 10, 4, 5, 6, 7 };
static int sum;
static const int n = sizeof(arr) / sizeof(arr[0]);

int main()
{
int i;

for (i = 0; i < n; i++)
sum += arr[i];
```

}

Перед передачей управления Си-коду, нужно выполнить следующие настройки:

1. Стек
2. Глобальные переменные
 - (а) Инициализированные
 - (б) Неинициализированные
3. Константные данные

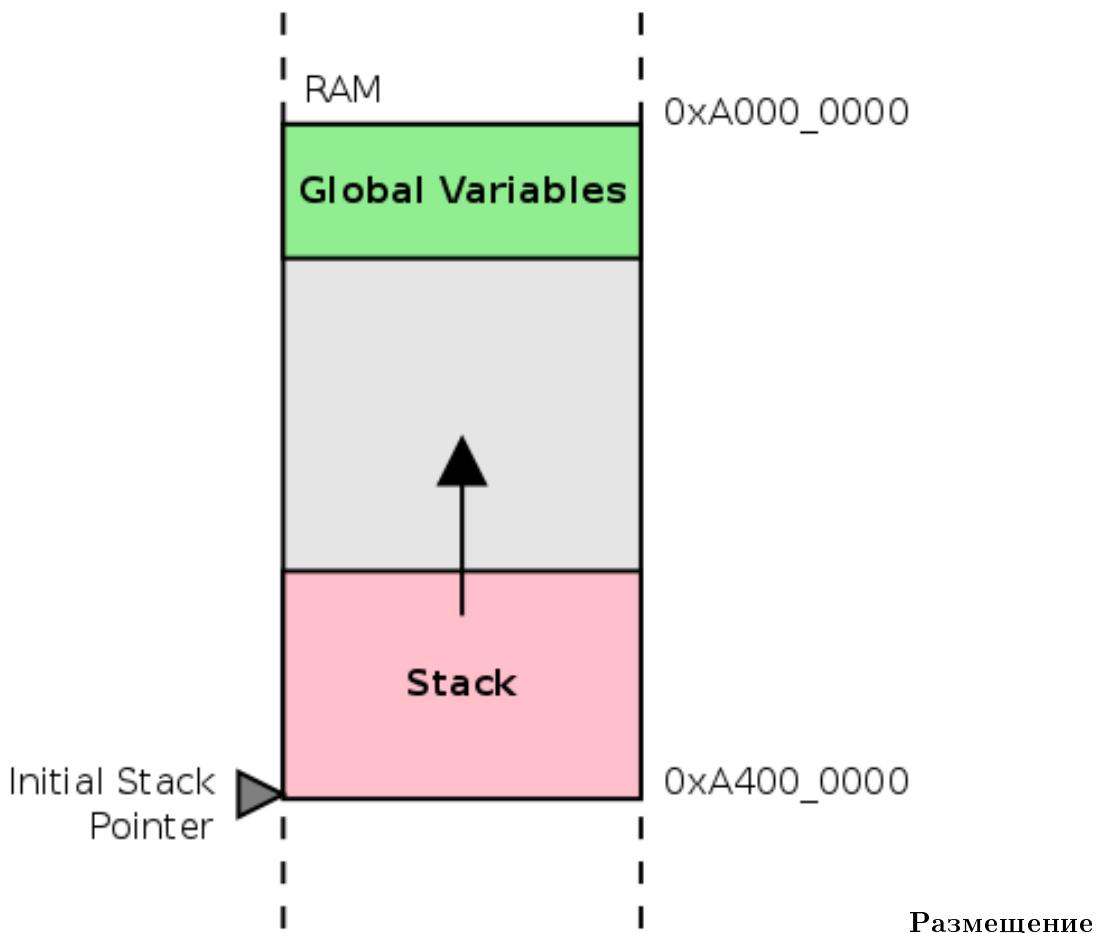
16.10.1 Стек

Си использует стек для хранения локальных (авто) переменных, передачи аргументов и результата функций, хранения адресов возврата из функций и т.д. Так что необходимо чтобы стек был настроен корректно перед передачей управление Си-коду.

На архитектуре ARM стеки очень гибкие, поэтому их реализация полностью ложиться на программное обеспечение. Для людей не знакомых с ARM, некоторый обзор приведен в ??.

Чтобы быть уверенным, что разные части кода, сгенерированного **разными** компиляторами, работали друг с другом, ARM создал [Стандарт вызова процедур для архитектуры ARM \(AAPCS\)](#). В нем описаны регистры которые должны быть использованы для работы со стеком и направление в котором растет стек. Согласно AAPCS, **регистр r13** должен быть использован для указателя стека. Также стек должен быть для указателя стека. Также стек должен быть **full-descending** (нисходящим).

Один из способов размещения глобальных переменных на стеке показан в диаграмме:



стека

Так что все, что нужно сделать в стартовом коде для стека — выставить `r13` на старший адрес ОЗУ, так что стек может расти вниз (в сторону младших адресов). Для платы `connex` это можно сделать командой

```
ldr sp, =0xA4000000
```

Обратите внимание что ассемблер предоставляет алиас `sp` для регистра `r13`.

Адрес 0xA4000000 сам по себе не указывает на ОЗУ. ОЗУ кончается адресом 0xA3FFFFFF. Но это нормально, так как стек **full-descending**, т.е. во время первого `push` в стек указатель **сначала уменьшится**, и только потом значение будет записано уже в ОЗУ.

16.10.2 Глобальные переменные

Когда компилируется Си-код, компилятор размещает инициализированные глобальные переменные в секцию `.data`. Как и для ассемблера, сегмент `.data` должен быть скопирован стартовым кодом в ОЗУ из флеша.

Язык Си гарантирует что все неинициализированные глобальные переменные будут инициализированы нулем³⁰. Когда Си-программа компилируется, создается отдельный сегмент `.bss` для неинициализированных переменных. Так как для всего сегмента должно быть выполнено обнуление, его не нужно хранить во флеше. Перед передачей управления на Си-код, содержимое `.bss` должно быть зачищено startup-кодом.

16.10.3 Константные данные

GCC размещает переменные, помеченные модификатором `const`, в отдельный сегмент `.rodata`. Также `.rodata` используется для хранения всех "строковых констант".

Так как содержимое `.rodata` не модифицируется, оно может быть размещено в Flash/ROM. Для этого нужно модифицировать `.lds`.

16.10.4 Секция `.eeprom` (AVR8)

При написании прошивок для Atmel AVR8, существует модификатор `EEMEM` определенный в `avr/eeprom.h`:

```
#define EEMEM __attribute__((section(".eeprom")))
```

который использует модификатор GCC `__attribute__((section("...")))`, который приписывает объект данных к любой указанной секции. В частности, секция `.eeprom` выделяется из финального объектного файла, и программируется в Atmel ATmega отдельным вызовом `avrdude` (ПО программатора).

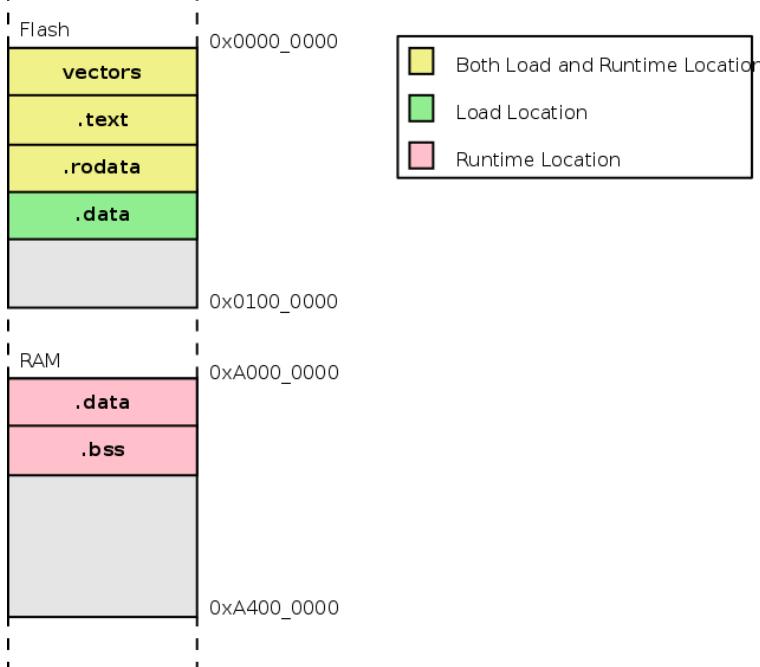
16.10.5 Стартовый код

Теперь все готово к написанию скрипта линкера и стартового кода. Скрипт ?? модифицируется с добавлением размещения секций:

1. `.bss`
2. `vectors`
3. `.rodata`

Секция `.bss` размещается сразу за секцией `.data` в ОЗУ. Также создаются символы маркирующие начало и конец секции `.bss`, которые будут использованы в startup-коде при ее очистке. `.rodata` размещается сразу за `.text` во флеше:

³⁰ старый стандарт Си не гарантировал



Размещение секций

Листинг 19: Скрипт линкера для Си кода

```

SECTIONS {
    . = 0x00000000;
    .text : {
        * (.vectors);
        * (.text);
    }
    .rodata : {
        * (.rodata);
    }
    flash_sdata = .;

    . = 0xA0000000;
    ram_sdata = .;
    .data : AT (flash_sdata) {
        * (.data);
    }
    ram_edata = .;
    data_size = ram_edata - ram_sdata;
}

```

```
sbss = .;
.bss : {
    * (.bss);
}
ebss = .;
bss_size = ebss - sbss;
}
```

Startup-код включает следующие части:

1. вектора исключений
2. код копирования `.data` из Flash в RAM
3. код обнуления `.bss`
4. код установки указателя стека
5. переход на `_main`

Листинг 20: Стартовый код для Си программы на ассемблере

```
.section "vectors"
reset: b      start
undef: b      undef
swi: b       swi
pabt: b      pabt
dabt: b      dabt
nop
irq: b      irq
fiq: b      fiq

.text
start:
@ Copy data to RAM.
ldr   r0, =flash_sdata
ldr   r1, =ram_sdata
ldr   r2, =data_size

@ Handle data_size == 0
cmp   r2, #0
beq   init_bss
copy:
ldrb  r4, [r0], #1
strb r4, [r1], #1
subs r2, r2, #1
bne  copy

init_bss:
```

```
@@ Initialize .bss
ldr    r0, =sbss
ldr    r1, =ebss
ldr    r2, =bss_size

@@ Handle bss_size == 0
cmp    r2, #0
beq    init_stack

        mov    r4, #0
zero:
strb   r4, [r0], #1
subs   r2, r2, #1
bne    zero

init_stack:
@@ Initialize the stack pointer
ldr    sp, =0xA4000000

bl    main

stop: b    stop
```

Для компиляции кода не требуется отдельно вызывать ассемблер, линкер и компилятор Си: программа **gcc** является оберткой, которая умеет делать это сама, автоматически вызывая ассемблер, компилятор и линкер в зависимости от типов файлов. Поэтому мы можем скомпилировать весь наш код одной командой:

```
$ arm-none-eabi-gcc -nostdlib -o csum.elf -T csum.lds csum.c startup.s
```

Опция **-nostdlib** используется для указания, что нам при компиляции не нужно подключать стандартную библиотеку Си (**newlib**). Эта библиотека крайне полезна, но для ее использования нужно выполнить некоторые дополнительные действия, описанные в разделе **??**.

Дамп таблицы символов даст больше информации о расположении объектов в памяти:

```
$ arm-none-eabi-nm -n csum.elf
00000000 t reset <1>
00000004 A bss_size
00000004 t undef
00000008 t swi
0000000c t pabt
00000010 t dabt
```

```
000000018 A data_size
000000018 t irq
00000001c t fiq
000000020 T main
000000090 t start <2>
000000a0 t copy
000000b0 t init_bss
000000c4 t zero
000000d0 t init_stack
000000d8 t stop
000000f4 r n <3>
000000f8 A flash_sdata
a0000000 d arr <4>
a0000000 A ram_sdata
a0000018 A ram_edata
a0000018 A sbss
a0000018 b sum <5>
a000001c A ebss
```

1. `reset` и остальные вектора исключений размещаются с `0x0`.
2. ассемблерный код находится сразу после 8 векторов исключений ($8 * 4 = 32$)
3. константные данные `n`, размещены во флеше после кода.
4. инициализированные данные `arr`, массив из 6 целых, размещен с начала ОЗУ `0xA0000000`.
5. неинициализированные данные `sum` размещен после массива из 6 целых. ($6 * 4 = 24 = 0x18$)

Для выполнения программы преобразуем ее в `.bin` формат, запустим в **Qemu**, и выведем дамп переменной `sum` по адресу `0xA0000018`:

```
$ arm-none-eabi-objcopy -O binary csum.elf csum.bin
$ dd if=/dev/zero of=flash.bin bs=4K count=4K
$ dd if=csum.bin of=flash.bin bs=4096 conv=notrunc
$ qemu-system-arm -M connex -pflash flash.bin -nographic -serial /dev/null
(qemu) xp /6dw 0xa0000000
a0000000:      1          10          4          5
a0000010:      6          7
(qemu) xp /1dw 0xa0000018
a0000018:     33
```

16.11 Использование библиотеки Си

FIXME: Эту секцию еще нужно написать.

16.12 Inline-ассемблер

FIXME: Эту секцию еще нужно написать.

16.13 Использование ‘make’ для автоматизации компиляции

Если вам надоело каждый раз вводить длинные команды, компилируя примеры из этого учебника, пришло время научиться пользоваться утилитой **make**. **make** это утилита, отслеживающая зависимости файлов, описанные в файле **Makefile**.

Умение читать и писать makeфайлы **must have** навык для программиста, особенно для больших многофайловых проектов, содержащих сотни и тысячи файлов, которые должны быть ассемблированы, откомпилированы или оттрансформированы в различные форматы.

Каждая зависимости между двумя или более файлами прописывается в **make-правиле**:

```
<цель> : <источник>
<tab><команда компиляции 1>
<tab><команда компиляции 2>
...
...
```

цель одно имя файла, или несколько имен, разделенных пробелами. Этот файл(ы) будут созданы или обновлены этим правилом.

источник 0+ имен файлов разделенных пробелами. Эти файл(ы) будут ‘проводиться на изменения’ используя метку времени последней модификации.

tab символ табуляции с ascii кодом 0x09, вы должны использовать текстовый редактор, который умеет работать с табуляциями, не заменяя их последовательностями пробелов.

команда компиляции любая команда, такая как вызов ассемблера или линкера, которая обновляет **цель**, выполняя некоторую полезную работу. **make-правило может не иметь команд компиляции**, если вам нужно прописать только зависимость файлов.

Основной принцип каждого make-правила: если один из файлов-источников **новее** чем один из целевых файлов, будет выполнено тело правила, которое обновит **цели**.

Давайте напишем простой **Makefile** для простейшей программы, описанной в разделе ??:

Листинг 21: Makefile

```
emulation: add.flash
    qemu-system-arm -M connex -pflash add.flash \
        -nographic -serial /dev/null
flash.bin: add.bin
    dd if=/dev/zero of=flash.bin bs=4K count=4K
    dd if=add.bin of=flash.bin bs=4K conv=notrunc
add.bin: add.elf
    arm-none-eabi-objcopy -O binary add.elf add.bin
add.elf: add.o
    arm-none-eabi-ld -o add.elf add.o
add.o: add.s
    arm-none-eabi-as -o add.o add.s
```

- обратите внимание на обратный слэш и следующую табулированную строку: вы можете делить длинные команды на несколько строк; каждая строка должна быть табулирована для следования синтаксису make-правила.

Введине в командной строке команду **make** без параметров, находясь в каталоге проета, в котором находится **Makefile** и исходные тексты программы, и вы сразу получите автоматически скомпилированные бинарные файлы и запущенный **Qemu**:

```
$ make
...
QEMU 2.1.2 monitor - type 'help' for more information
(qemu)
```

Если вы запускаете **make** без параметров, **первое правило** в **Makefile** будет обработано как **главная цель**, с обходом всех зависимостей в других правилах.

16.13.1 Выбор конкретной *цели*

Если вам нужно обновить только определенный файл-*цель*, поместите необходимое имя файла после команды **make**:

```
$ make add.o
make: 'add.o' is up to date.
```

Эта команда будет перекомпилировать только файл **add.o**, в том и только в том случае, если вы перед запуском команды изменили **add.s**. Если вы видите

сообщение типа **make**: **add.o is up to date.**, **исходные файлы не менялись**, и **make не будет запускать правило ассемблирования**.

Это очень полезно если у вас очень много файлов исходников³¹, и вы изменили несколько символов в одном файле. Без **make**³² каждое микроскопическое изменение потребует перекомпиляции всего проекта, которое может длиться **несколько часов (!)**. Использование **make** позволяет выполнить всего несколько вызовов компилятора и линкера, что будет намного намного быстрее.

Возарашаясь к нашему **add.o**, вы можете заставить ассемблер выполниться не изменяя файл **add.s**, через команду **touch**:

```
$ touch add.s  
$ make add.o  
arm-none-eabi-as -o add.o add.s
```

Команда **touch** изменяет только дату модификации исходного файла **add.s**, не меняя его содержимое, так что **make** увидит что этот файл обновился, и запустит ассемблер для указанной цели **add.o**.

По умолчанию **make** выводит каждую команду и ее вывод. Если у вас есть какие-то причины для "тихой" работы **make**, вы можете добавить префикс "-" (минус) к командам компиляции.

16.13.2 Переменные

16.14 13. Contributing

16.15 14. Credits

16.15.1 14.1. People

16.15.2 14.2. Tools

16.16 15. Tutorial Copyright

16.17 A. ARM Programmer's Model

16.18 B. ARM Instruction Set

16.19 C. ARM Stacks

³¹ например тысячи файлов, как у ядра *Linux*

³² используя простой .rc shell-скрипт или .batник

Глава 17

Embedded Systems Programming in C_+^+ [22]

1

Глава 18

Сборка кросс-компилятора **GNU Toolchain** из исходных текстов

Если вам по каким-то причинам не подходит одна из типовых сборок кросс-компиляторов, поставляемых в виде готовых бинарных пакетов из репозитория вашего дистрибутива *Linux*, **GNU Toolchain** можно легко скопилировать **из исходных текстов** и установить в систему, даже имея только пользовательские права доступа.

Сборка **GNU Toolchain** из исходников может понадобиться, если вы хотите:

- самую свежую или какую-то конкретную версию **GNU Toolchain**
- опции компиляции: малораспространенный **target**-процессор, **нетиповой формат файлов объектного кода**¹ или экспериментальные оптимизаторы, не включенные в бинарные пакеты из дистрибутива *Linux*
- полпроцента ускорения работы компилятора благодаря жесткой оптимизации его машинного кода точно под ваш рабочий компьютер (**-march=native** -

При сборке используется утилита **make 16.13**, которой можно передать набор переменных конфигурирования. В таблице перечислен набор переменных конфигурирования сборки с указанием их значения по умолчанию² и имя mk-файла, где оно задано:

¹ например для i386 может понадобится сборка кросс-компилятора с **-target=i486-none-elf** IX или **i686-linux-uclibc** вместо типовой компиляции для *Linux* типа **i486-linux-gnu**

² также приведены часто используемые варианты значения

APP	cross	Makefile	приложение: условное имя проекта (только латиница, буквы a-z)
HW	x86	Makefile	qemu vmware virtualpc x86 pc686 amd64 cortexM avr8
CPU	i386	hw/\$(HW).mk	
ARCH	i386	cpu/\$(CPU).mk	
TARGET	\$(CPU)-pc-elf	hw/\$(HW).mk	i686-linux-uclibc x86_64-linux i386-pc-elf arm-none-eabi avr-

APP/HW: приложение/платформа

Для сборки необходимо выбрать имя проекта³ и аппаратную платформу, для которой будет настраиваться пакет кросс-компилятора.

Особенно это важно для варианта сборки, когда собирается не только кросс-компилятор, но и базовая ОС — минимальная *Linux*-система из ядра, libc и дополнительных прикладных библиотек. В этом случае **APP/HW** используются для формирования имен файлов ядра **\$(APP)\$(HW).kernel**, названия и состава загрузочного образа **\$(APP)\$(HW).rootfs**, и внутренних настроек.

Подготовка BUILD-системы: необходимое ПО

Для сборки необходимо установить следующие пакеты:

```
sudo apt install gcc g++ make flex bison m4 bc bzip2 xz-utils libncurses-
```

dirs: создание структуры каталогов

```
user@bs:~/boox/cross$ make dirs
mkdir -p
/home/user/boox/cross/gz /home/user/boox/cross/src /home/user/boox/cross/toolchain /home/user/boox/cross/root
```

Командой **make dirs** создается набор вспомогательных каталогов:

TC	\$(CWD)/\$(APP)\$(ROOT).cross	каталог установки кросс-компилятора
ROOT	\$(CWD)/\$(APP)\$(ROOT)	каталог файловой системы для целевого
CWD	\$(CURDIR)	текущий каталог
GZ	\$(CWD)/gz	архивы исходных текстов GNU Toolchain, загрузчика, и библиотек
SRC	\$(CWD)/src	каталог для распаковки исходников
TMP	\$(CWD)/tmp	каталог для out-of-tree сборки GNU toolchain

³ только латиница, буквы a-z

```
CWD = $(CURDIR)
```

```
GZ = $(CWD) / gz
```

```
SRC = $(CWD) / src
```

```
TMP = $(CWD) / tmp
```

```
ROOT = $(CWD) / $(APP) $(HW)
```

```
TC = $(CWD) / $(APP) $(HW).cross
```

```
DIRS = $(GZ) $(SRC) $(TMP) $(TC) $(ROOT)
```

```
.PHONY: dirs
```

```
dirs:
```

```
    mkdir -p $(DIRS)
```

Сборка в ОЗУ на ramdiske

Если у вас есть админские права и достаточный объем RAM, после выполнения `make dirs` рекомендуется примонтировать на каталоги `SRC` и `TMP` файловую систему `tmpfs` — это значительно ускорит компиляцию, т.к. все временные файлы будут храниться только в ОЗУ:

```
/etc/fstab
```

```
tmpfs /home/user/src tmpfs auto,uid=yourlogin,gid=yourgroup 0 0
tmpfs /home/user/tmp tmpfs auto,uid=yourlogin,gid=yourgroup 0 0
```

Если вы прописали монтирование `ramdisk`ов в `/etc/fstab`, или сделали `mount -t` вручную, может оказаться нужным запускать `make` с явным указанием значений переменных `SRC/TMP`:

```
make blablabla SRC=/home/user/src TMP=/home/user/tmp
```

Пакеты системы кросс-компиляции

GNU Toolchain

```
1 # bintools: assembler, linker, objfile tools
2 BINUTILS_VER= 2.24
3 # 2.25 build error
4
5 # gcc: C/C++ cross-compiler
6 GCC_VER      = 4.9.2
7 # 4.9.2 used: bug arm/62098 fixed
```

```

8
9 # gcc support libraries
10 ## required for GCC build
11 GMP_VER      = 5.1.3
12 MPFR_VER     = 3.1.3
13 MPC_VER      = 1.0.2
14 ## loop optimisation
15 ISL_VER       = 0.11.1
16 # 0.11 need for binutils build
17 CLOOG_VER     = 0.18.1
18
19 # standard C/POSIX library libc (newlib)
20 NEWLIB_VER    = 2.3.0.20160226
21
22 # loader for i386 target
23 SYSLINUX_VER  = 6.03
24
25 # packages
26 BINUTILS      = binutils-$(BINUTILS_VER)
27 GCC            = gcc-$(GCC_VER)
28 GMP            = gmp-$(GMP_VER)
29 MPFR           = mpfr-$(MPFR_VER)
30 MPC            = mpc-$(MPC_VER)
31 ISL            = isl-$(ISL_VER)
32 CLOOG          = cloog-$(CLOOG_VER)
33 NEWLIB         = newlib-$(NEWLIB_VER)
34 SYSLINUX       = syslinux-$(SYSLINUX_VER)

```

make

newlib стандартная библиотека **libc**

gz: загрузка исходного кода для пакетов

```
user@bs$ make APP=cross HW=x86 GZ=/home/user/gz gz
```

В примере команды показано два обязательных параметра **APP/HW⁴** и необязательный **GZ**: поскольку я собираю кросс-компиляторы для нескольких целевых платформ, я создал каталог **\$(HOME)/gz** и загружаю туда архивы исходников **для всех проектов сразу⁵**. Более простой способ – просто сделать симлинк **ln -s ~/gz project/gz** и не переопределять переменную **GZ** явно.

⁴ по ним могут закачиваться дополнительные файлы исходников, зависящие от платформы — например исходник загрузчика или бинарные файлы (блöбы) драйверов от производителя железки

⁵ а не в **/gz** каждого проекта, нет смысла дублировать исходники **GNU Toolchain** одной и той же версии

mk/gz.mk

```
WGET = -wget -N -P $(GZ) -t2 -T2
```

```
.PHONY: gz
```

```
gz: gz_cross gz_libs gz_$(ARCH)
```

```
.PHONY: gz_cross
```

```
gz_cross:
```

```
$(WGET) ftp://ftp.gnu.org/pub/gmp/$(GMP).tar.bz2
```

```
$(WGET) http://www.mpfr.org/mpfr-current/$(MPFR).tar.bz2
```

```
$(WGET) http://www.multiprecision.org/mpc/download/$(MPC).tar.gz
```

```
$(WGET) ftp://gcc.gnu.org/pub/gcc/infrastructure/$(ISL).tar.bz2
```

```
$(WGET) ftp://gcc.gnu.org/pub/gcc/infrastructure/$(CLOOG).tar.gz
```

```
$(WGET) http://ftp.gnu.org/gnu/binutils/$(BINUTILS).tar.bz2
```

```
$(WGET) http://gcc.skazkaforyou.com/releases/$(GCC)/$(GCC).tar.bz2
```

```
.PHONY: gz_libs
```

```
gz_libs:
```

```
$(WGET) ftp://sourceware.org/pub/newlib/$(NEWLIB).tar.gz
```

```
.PHONY: gz_i386
```

```
gz_i386:
```

```
$(WGET) https://www.kernel.org/pub/linux/utils/boot/syslinux/$(SY
```

Макро-правила для автоматической распаковки исходников

mk/src.mk

```
$(SRC)/%/README: $(GZ)%.tar.gz
```

```
    cd $(SRC) && zcat $< | tar x && touch $@
```

```
$(SRC)/%/README: $(GZ)%.tar.bz2
```

```
    cd $(SRC) && bzcat $< | tar x && touch $@
```

```
$(SRC)/%/README: $(GZ)%.tar.xz
```

```
    cd $(SRC) && xzcat $< | tar x && touch $@
```

Общие параметры для .configure

mk/cfg.mk

```
# configure parameters for all packages
```

```
CFG_ALL = --disable-nls --disable-werror \
```

```
    --docdir=$(TMP)/doc --mandir=$(TMP)/man --infodir=$(TMP)/info
```

```
# [B]uild host configure
```

```
BCFG = configure $(CFG_ALL) --prefix=$(TC)
XPATH = PATH=$(TC)/bin:$PATH
# [T]arget configure
TCFG = configure $(CFG_ALL) --prefix=$(ROOT) CC=$(TARGET)-gcc
# get cpu cores
CPU_CORES ?= $(shell grep processor /proc/cpuinfo | wc -l)
# run make with -j flag or make CPU_CORES=<none> for one thread build
MAKE = make -j$(CPU_CORES)
INSTALL = make install
```

18.1 Сборка кросс-компилятора

Для пакетов кросс-компилятора существуют два варианта сборки пакетов:

Пакеты с 0 в конце имени задают сборку программ, которые будут выполняться на BUILD-компьютере, и компилировать код для TARGET-системы, т.е. это простейший вариант кросс-компиляции.

Пакеты без 0, которые могут появиться в будущем — **относятся только к сборке emLinux**, собирают кросс-компилятор **канадским крестом**:

- пакет собирается на BUILD-системе — ваш рабочий компьютер,
- выполняется на HOST-системе — например PC104 или роутер с emLinux,
- и компилирует код для TARGET-микропроцессора — модуль ввода/вывода на USB, подключенный к PC104)

18.1.1 cclibs0: библиотеки поддержки gcc

Для сборки **GNU Toolchain** необходим набор нескольких библиотек, причем **успешность сборки сильно зависит от их версий**, поэтому библиотеки **нужно собрать из исходников**, а не использовать девелоперские пакеты из дистрибутива BUILD-Linux.

Библиотеки чисел произвольной точности:

gmp0 целых

gmfr0 с плавающей точкой

gmc0 комплексных

Библиотеки для работы с графами (нужны для компилятора оптимизатора **Graphite**)

cloog0 polyhedral оптимизации

isl0 манипуляция наборами целочисленных точек

```

WITH_CCLIBS0 = --with-gmp=$(TC) --with-mpfr=$(TC) --with-mpc=$(TC) \
    --without-ppl --without-cloog
# --with-isl=$(TC) --with-cloog=$(TC)

CFG_CCLIBS0 = $(WITH_CCLIBS0) --disable-shared
.PHONY: cclibs0
cclibs0: gmp0 mpfr0 mpc0
# cloog0 isl0

CFG_GMP0 = $(CFG_CCLIBS0)
.PHONY: gmp0
gmp0: $(SRC) $(GMP) / README
    rm -rf $(TMP) $(GMP) && mkdir -p $(TMP) $(GMP) && cd $(TMP) $(GMP)
        $(SRC) $(BCFG) $(CFG_GMP0) && $(MAKE) && $(INSTALL)-strip

CFG_MPFR0 = $(CFG_CCLIBS0)
.PHONY: mpfr0
mpfr0: $(SRC) $(MPFR) / README
    rm -rf $(TMP) $(MPFR) && mkdir -p $(TMP) $(MPFR) && cd $(TMP) $(MPFR)
        $(SRC) $(BCFG) $(CFG_MPFR0) && $(MAKE) && $(INSTALL)-strip

CFG_MPC0 = $(CFG_CCLIBS0)
.PHONY: mpc0
mpc0: $(SRC) $(MPC) / README
    rm -rf $(TMP) $(MPC) && mkdir -p $(TMP) $(MPC) && cd $(TMP) $(MPC)
        $(SRC) $(BCFG) $(CFG_MPC0) && $(MAKE) && $(INSTALL)-strip

CFG_CLOOG0 = --with-gmp-prefix=$(TC) $(CFG_CCLIBS0)
.PHONY: cloog0
cloog0: $(SRC) $(CLOOG) / README
    rm -rf $(TMP) $(CLOOG) && mkdir $(TMP) $(CLOOG) && cd $(TMP) $(CLOOG)
        $(SRC) $(BCFG) $(CFG_CLOOG0) && $(MAKE) && $(INSTALL)-strip

CFG_ISL0 = --with-gmp-prefix=$(TC) $(CFG_CCLIBS0)
.PHONY: isl0
isl0: $(SRC) $(ISL) / README
    rm -rf $(TMP) $(ISL) && mkdir $(TMP) $(ISL) && cd $(TMP) $(ISL)
        $(SRC) $(BCFG) $(CFG_ISL0) && $(MAKE) && $(INSTALL)-strip

```

18.1.2 binutils0: ассемблер и линкер

Чтобы побыстрее получить результат, который можно сразу потестировать, соберем сначала кросс-**binutils**, а потом все что относится к Сициальному компилятору⁶.

-target триплет целевой системы, например **i386-pc-elf**

⁶ на самом деле **binutils0** надо собирать после **cclibs0**, так как есть зависимость от библиотек **isl0** и **cloog0**

CFG_ARCH CFG_CPU задаются в файлах `arch/$(ARCH).mk` и `cpu/$(CPU).mk`,
и определяют опции сборки `binutils/gcc` для конкретного процессора⁷

-`with-sysroot` каталог где должны храниться файлы для целевой системы: откомпилированные библиотеки и каталог `include`

-`with-native-system-header-dir` имя каталога с `include`-файлами, относительно `sysroot`

arch/i386.mk

CFG_ARCH =

cpu/i386.mk

ARCH = i386

CFG_CPU = --with-cpu=i386 --with-tune=i386

mk/bintools.mk

CFG_BINUTILS0 = --target=\$(TARGET) \$(CFG_ARCH) \$(CFG_CPU) \
--with-sysroot=\$(ROOT) --with-native-system-header-dir=/include \
--enable-lto --disable-multilib \$(WITH_CCLIBS0) \
--disable-target-libiberty --disable-target-zlib \
--disable-bootstrap --disable-decimal-float \
--disable-libmudflap --disable-libssp \
--disable-libgomp --disable-libquadmath

.PHONY: binutils0

binutils0: \$(SRC)/\$(BINUTILS)/README

rm -rf \$(TMP)/\$(BINUTILS) && mkdir -p \$(TMP)/\$(BINUTILS) && cd \$(SRC)/\$(BINUTILS)/\$(BCFG) \$(CFG_BINUTILS0) && \$(MAKE) && \$(INSTA

Файлы `binutils0` с TARGET- префиксами и типовые скрипты линкера

crossx86 . cross/bin/i386-pc-elf-readelf
crossx86 . cross/bin/i386-pc-elf-addr2line
crossx86 . cross/bin/i386-pc-elf-size
crossx86 . cross/bin/i386-pc-elf-objdump
crossx86 . cross/bin/i386-pc-elf-objcopy
crossx86 . cross/bin/i386-pc-elf-nm
crossx86 . cross/bin/i386-pc-elf-ld.bfd
crossx86 . cross/bin/i386-pc-elf-elfedit
crossx86 . cross/bin/i386-pc-elf-as
crossx86 . cross/bin/i386-pc-elf-ranlib
crossx86 . cross/bin/i386-pc-elf-c++filt
crossx86 . cross/bin/i386-pc-elf-gprof

⁷ например `-without-fpu` для `cpu/i486sx.mk`

```
crossx86.cross/bin/i386-pc-elf-ar
crossx86.cross/bin/i386-pc-elf-strip
crossx86.cross/bin/i386-pc-elf-strings

crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xr
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xsc
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xdc
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xu
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xc
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.x
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xbn
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xsw
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xs
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xw
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xn
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xdw
crossx86.cross/i386-pc-elf/lib/ldscripts/elf_i386.xd
```

18.1.3 **gcc00**: сборка stand-alone компилятора Си

Сборка кросс-компилятора Си выполняется в два этапа

gcc00 минимальный **gcc** необходимый для сборки libc ??

newlib сборка стандартной библиотеки Си

gcc0 пересборка полного кросс-компилятора Си/ C_+^+

mk/gcc.mk

CFG_GCC_DISABLE =

CFG_GCC00 = \$(CFG_BINUTILS0) \$(CFG_GCC_DISABLE) \
--disable-threads --disable-shared --without-headers --with-newl
--enable-languages="c"

CFG_GCC0 = \$(CFG_BINUTILS0) \$(CFG_GCC_DISABLE) \
--with-newlib \
--enable-languages="c , c++"

.PHONY: gcc00

gcc00: \$(SRC) / \$(GCC) / README

```
rm -rf $(TMP) / $(GCC) && mkdir -p $(TMP) / $(GCC) && cd $(TMP) / $(GCC)
$(SRC) / $(GCC) / $(BCFG) $(CFG_GCC00)
cd $(TMP) / $(GCC) && $(MAKE) all-gcc && $(INSTALL)-gcc
cd $(TMP) / $(GCC) && $(MAKE) all-target-libgcc && $(INSTALL)-target-libgcc
```

18.1.4 newlib: сборка стандартной библиотеки libc

Стандартная библиотека **libc**⁸ обеспечивает слой совместимости со стандартом POSIX для ваших программ. Это удобно при адаптации чужих программ под вашу ОС, и при написании собственного **мультиплатформенного** кода.

mk/libc.mk

```
CFG_NEolib = --host=$(TARGET)  
.PHONY: newlib  
newlib : $(SRC) / $(NEWLIB) / README  
        rm -rf $(TMP) / $(NEWLIB) && mkdir -p $(TMP) / $(NEWLIB) && cd $(TMP)  
        $(XPATH) $(SRC) / $(NEWLIB) / $(TCFG) $(CFG_NEolib)  
#      && $(MAKE) && $(INSTALL)-strip
```

18.1.5 gcc0: пересборка компилятора Си/ C_+

18.2 Поддерживаемые платформы

18.2.1 i386: ПК и промышленные PC104

arch/i386.mk

```
CFG_ARCH =
```

18.2.2 x86_64: серверные системы

arch/x86_64.mk

18.2.3 AVR: Atmel AVR Mega

arch/avr.mk

18.2.4 arm: процессоры ARM Cortex-Mx

arch/arm.mk

18.2.5 armhf: SoCи Cortex-A, PXA270,..

arch/armhf.mk

⁸ для микроконтроллерных систем — обрезанная версия, **newlib**

18.3 Целевые аппаратные системы

18.3.1 **x86**: типовой компьютер на процессоре i386+

hw/x86.mk

CPU = i386

TARGET = \$(CPU)-pc-elf

Глава 19

Porting The GNU Tools To Embedded Systems

Embed With GNU

Porting The GNU Tools To Embedded Systems

Spring 1995

Very *Rough* Draft

Rob Savoye - Cygnus Support

http://ieee.uwaterloo.ca/coldfire/gcc-doc/docs/porting_toc.html

Глава 20

Оптимизация кода

20.1 PGO оптимизация

¹

Часть VIII

Микроконтроллеры Cortex-Mx

Часть IX

**os86: низкоуровневое
программирование i386**

Если вам по каким-то причинам не подходит одна из типовых распространенных ОС, например требуется сделать систему управления жесткого реального времени², информация в этом разделе поможет сделать ОС-поделку для типового WinInt ПК.

Специализированный GNU Toolchain для i386-pc-gnu

Для компиляции кода вам потребуется специально собранный из исходников кросс-**GNU Toolchain** для целевой архитектуры i386 — *triplet* TARGET=i386-pc-elf. Процесс сборки подробно описан в отдельном разделе [18](#).

Для упрощения не будем завязываться на особенности конкретного ПК или эмулятора **Qemu**³, все они вполне аппаратно совместимы с любым i386 компьютером в базовой конфигурации, для которого мы и будем рассматривать примеры кода:

- APP=bare metal программирование, без базовой ОС
- HW=x86 типовой минимальный i386 компьютер

os86/Makefile

```
APP = bare
HW = x86
TARGET = i386-pc-elf

TODO = gz dirs cclibs0 binutils0 gcc00 newlib
.PHONY: toolchain
toolchain: $(APP) $(HW).cross /bin /$(TARGET)-g++
$(APP) $(HW).cross /bin /$(TARGET)-g++:
    cd .. / cross; $(MAKE) $(TODO) \
        CWD=$(CURDIR) GZ=$(HOME) /L/gz SRC=$(HOME) /L/src TMP=$(HOME) /L/tmp
    APP=$(APP) HW=$(HW)
```

MultiBoot-загрузчик

Благодаря усилиям сообщества разработчиков OpenSource была успешно решена одна из проблем начинающего системного программиста — было создано несколько универсальных **загрузчиков**, берущих на себя заботу о чтении ядра ОС или bare metal программы, начальную инициализацию оборудования, включении защищенного режима, и передачу управления вашей ОС.

Чтобы ваша bare metal программа была успешно загружена, она должна удовлетворять требованиям **спецификации MultiBoot X** быть слинкована в формат ELF и включать заголовок multiboot.

² или вы любитель гадить из прикладного ПО в аппаратные порты в обход всех соглашений и средств защиты ОС

³ VMWare, VirtualPC

Часть X

Спецификация MultiBoot

Этот файл документирует ***Спецификацию Multiboot***, проект стандарта на последовательность загрузки. Этот документ имеет редакцию 0.6.96.

Copyright © 1995,96 Bryan Ford <baford@cs.utah.edu>

Copyright © 1995,96 Erich Stefan Boleyn <erich@uruk.org>

Copyright © 1999,2000,2001,2002,2005,2006,2009 Free Software Foundation, Inc.

Permission is granted to make and distribute verbatim copies of this manual provided the copyright notice and this permission notice are preserved on all copies.

Permission is granted to copy and distribute modified versions of this manual under the conditions for verbatim copying, provided also that the entire resulting derived work is distributed under the terms of a permission notice identical to this one.

Permission is granted to copy and distribute translations of this manual into another language, under the above conditions for modified versions.

Глава 21

Introduction to Multiboot Specification

This chapter describes some rough information on the Multiboot Specification. Note that this is not a part of the specification itself.

21.1 The background of Multiboot Specification

Every operating system ever created tends to have its own boot loader. Installing a new operating system on a machine generally involves installing a whole new set of boot mechanisms, each with completely different install-time and boot-time user interfaces. Getting multiple operating systems to coexist reliably on one machine through typical chaining mechanisms can be a nightmare. There is little or no choice of boot loaders for a particular operating system — if the one that comes with the operating system doesn't do exactly what you want, or doesn't work on your machine, you're screwed.

While we may not be able to fix this problem in existing proprietary operating systems, it shouldn't be too difficult for a few people in the free operating system communities to put their heads together and solve this problem for the popular free operating systems. That's what this specification aims for. Basically, it specifies an interface between a boot loader and a operating system, such that any complying boot loader should be able to load any complying operating system. This specification does not specify how boot loaders should work — only how they must interface with the operating system being loaded.

21.2 The target architecture

This specification is primarily targeted at i386 PC, since they are the most common and have the largest variety of operating systems and boot loaders. However, to the extent that certain other architectures may need a boot specification and do not have one already, a variation of this specification, stripped of the x86-specific details, could

be adopted for them as well.

21.3 The target operating systems

This specification is targeted toward free 32-bit operating systems that can be fairly easily modified to support the specification without going through lots of bureaucratic rigmarole. The particular free operating systems that this specification is being primarily designed for are Linux, the kernels of FreeBSD and NetBSD, Mach, and VSTa. It is hoped that other emerging free operating systems will adopt it from the start, and thus immediately be able to take advantage of existing boot loaders. It would be nice if proprietary operating system vendors eventually adopted this specification as well, but that's probably a pipe dream.

21.4 Boot sources

It should be possible to write compliant boot loaders that load the OS image from a variety of sources, including floppy disk, hard disk, and across a network.

Disk-based boot loaders may use a variety of techniques to find the relevant OS image and boot module data on disk, such as by interpretation of specific file systems¹, using precalculated *blocklists*², loading from a special *boot partition*³, or even loading from within another operating system⁴. Similarly, network-based boot loaders could use a variety of network hardware and protocols.

It is hoped that boot loaders will be created that support multiple loading mechanism increasing their portability, robustness, and user-friendliness.

21.5 Configure an operating system at boot-time

It is often necessary for one reason or another for the user to be able to provide some configuration information to an operating system dynamically at boot time. While this specification should not dictate how this configuration information is obtained by the boot loader, it should provide a standard means for the boot loader to pass such information to the operating system.

21.6 How to make OS development easier

OS images should be easy to generate. Ideally, an OS image should simply be an ordinary 32-bit executable file in whatever file format the operating system normally uses. It should be possible to **nm** or disassemble OS images just like normal executables.

¹ e.g. the BSD/Mach boot loader

² e.g. LILO

³ e.g. OS/2

⁴ e.g. the VSTa boot code, which loads from DOS

Specialized tools should not be required to create OS images in a **special** file format. If this means shifting some work from the operating system to a boot loader, that is probably appropriate, because all the memory consumed by the boot loader will typically be made available again after the boot process is created, whereas every bit of code in the OS image typically has to remain in memory forever. The operating system should not have to worry about getting into 32-bit mode initially, because mode switching code generally needs to be in the boot loader anyway in order to load operating system data above the 1MB boundary, and forcing the operating system to do this makes creation of OS images much more difficult.

Unfortunately, there is a horrendous variety of executable file formats even among free Unix-like pc-based operating systems — generally a different format for each operating system. Most of the relevant free operating systems use some variant of a.out format, but some are moving to elf. It is highly desirable for boot loaders not to have to be able to interpret all the different types of executable file formats in existence in order to load the OS image — otherwise the boot loader effectively becomes operating system specific again.

This specification adopts a compromise solution to this problem. Multiboot-compliant OS images always contain a magic *Multiboot header* (see OS image format ??), which allows the boot loader to load the image without having to understand numerous a.out variants or other executable formats. This magic header does not need to be at the very beginning of the executable file, so kernel images can still conform to the local a.out format variant in addition to being Multiboot-compliant.

21.7 Boot modules

Many modern operating system kernels, such as Mach and the microkernel in VSta, do not by themselves contain enough mechanism to get the system fully operational: they require the presence of additional software modules at boot time in order to access devices, mount file systems, etc. While these additional modules could be embedded in the main OS image along with the kernel itself, and the resulting image be split apart manually by the operating system when it receives control, it is often more flexible, more space-efficient, and more convenient to the operating system and user if the boot loader can load these additional modules independently in the first place.

Thus, this specification should provide a standard method for a boot loader to indicate to the operating system what auxiliary boot modules were loaded, and where they can be found. Boot loaders don't have to support multiple boot modules, but they are strongly encouraged to, because some operating systems will be unable to boot without them.

The definitions of terms used through the specification

must We use the term must, when any boot loader or OS image needs to follow a rule — otherwise, the boot loader or OS image is not Multiboot-compliant.

should We use the term should, when any boot loader or OS image is recommended to follow a rule, but it doesn't need to follow the rule.

may We use the term may, when any boot loader or OS image is allowed to follow a rule.

boot loader Whatever program or set of programs loads the image of the final operating system to be run on the machine. The boot loader may itself consist of several stages, but that is an implementation detail not relevant to this specification. Only the final stage of the boot loader — the stage that eventually transfers control to an operating system — must follow the rules specified in this document in order to be Multiboot-compliant; earlier boot loader stages may be designed in whatever way is most convenient.

OS image The initial binary image that a boot loader loads into memory and transfers control to start an operating system. The OS image is typically an executable containing the operating system kernel.

boot module Other auxiliary files that a boot loader loads into memory along with an OS image, but does not interpret in any way other than passing their locations to the operating system when it is invoked.

Multiboot-compliant A boot loader or an OS image which follows the rules defined as must is Multiboot-compliant. When this specification specifies a rule as should or may, a Multiboot-compliant boot loader/OS image doesn't need to follow the rule.

u8 The type of unsigned 8-bit data.

u16 The type of unsigned 16-bit data. Because the target architecture is little-endian, **u16** is coded in **little-endian**.

u32 The type of unsigned 32-bit data. Because the target architecture is little-endian, **u32** is coded in **little-endian**.

u64 The type of unsigned 64-bit data. Because the target architecture is little-endian, **u64** is coded in little-endian.

Глава 22

The exact definitions of Multiboot Specification

There are three main aspects of a boot loader/OS image interface:

1. The format of an OS image as seen by a boot loader.
2. The state of a machine when a boot loader starts an operating system.
3. The format of information passed by a boot loader to an operating system.

22.1 OS image format

An OS image may be an ordinary 32-bit executable file in the standard format for that particular operating system, except that it may be linked at a non-default load address to avoid loading on top of the pc's I/O region or other reserved areas, and of course it should not use shared libraries or other fancy features.

An OS image must contain an additional header called *Multiboot header*, besides the headers of the format used by the OS image. The Multiboot header must be contained completely within the first 8192 bytes of the OS image, and must be longword (32-bit) aligned. In general, it should come **as early as possible**, and may be embedded in the beginning of the text segment after the real executable header.

22.1.1 The layout of Multiboot header

The layout of the Multiboot header must be as follows:

Offset	Type	Field Name	Note
0	u32	magic	required
4	u32	flags	required
8	u32	checksum	required
12	u32	header_addr	if flags[16] is set
16	u32	load_addr	if flags[16] is set
20	u32	load_end_addr	if flags[16] is set
24	u32	bss_end_addr	if flags[16] is set
28	u32	entry_addr	if flags[16] is set
32	u32	mode_type	if flags[2] is set
36	u32	width	if flags[2] is set
40	u32	height	if flags[2] is set
44	u32	depth	if flags[2] is set

The fields ‘magic’, ‘flags’ and ‘checksum’ are defined in Header magic fields 22.1.2, the fields ‘header_addr’, ‘load_addr’, ‘load_end_addr’, ‘bss_end_addr’ and ‘entry_addr’ are defined in Header address fields 22.1.1, and the fields ‘mode_type’, ‘width’, ‘height’ and ‘depth’ are defined in Header graphics fields 22.1.4.

22.1.2 The magic fields of Multiboot header

‘magic’ The field ‘magic’ is the magic number identifying the header, which must be the hexadecimal value 0x1BADB002.

‘flags’ The field ‘flags’ specifies features that the OS image requests or requires of an boot loader. Bits 0-15 indicate requirements; if the boot loader sees any of these bits set but doesn’t understand the flag or can’t fulfill the requirements it indicates for some reason, it must notify the user and fail to load the OS image. Bits 16-31 indicate optional features; if any bits in this range are set but the boot loader doesn’t understand them, it may simply ignore them and proceed as usual. Naturally, all as-yet-undefined bits in the ‘flags’ word must be set to zero in OS images. This way, the ‘flags’ fields serves for version control as well as simple feature selection.

If bit 0 in the ‘flags’ word is set, then all boot modules loaded along with the operating system must be aligned on page (4KB) boundaries. Some operating systems expect to be able to map the pages containing boot modules directly into a paged address space during startup, and thus need the boot modules to be page-aligned.

If bit 1 in the ‘flags’ word is set, then information on available memory via at least the ‘mem_*’ fields of the Multiboot information structure (see Boot information format ??) must be included. If the boot loader is capable of passing a memory map (the ‘mmap_*’ fields) and one exists, then it may be included as well.

If bit 2 in the ‘flags’ word is set, information about the video mode table (see Boot information format ??) must be available to the kernel.

If bit 16 in the ‘flags’ word is set, then the fields at offsets 12-28 in the Multiboot header are valid, and the boot loader should use them instead of the fields in the actual executable header to calculate where to load the OS image. This information does not need to be provided if the kernel image is in elf format, but it must be provided if the images is in a.out format or in some other format. Compliant boot loaders must be able to load images that either are in elf format or contain the load address information embedded in the Multiboot header; they may also directly support other executable formats, such as particular a.out variants, but are not required to.

‘checksum’ The field ‘checksum’ is a 32-bit unsigned value which, when added to the other magic fields (i.e. ‘magic’ and ‘flags’), must have a 32-bit unsigned sum of zero.

22.1.3 The address fields of Multiboot header

All of the address fields enabled by flag bit 16 are physical addresses. The meaning of each is as follows:

header_addr Contains the address corresponding to the beginning of the Multiboot header — the physical memory location at which the magic value is supposed to be loaded. This field serves to **synchronize** the mapping between OS image offsets and physical memory addresses.

load_addr Contains the physical address of the beginning of the text segment. The offset in the OS image file at which to start loading is defined by the offset at which the header was found, minus (header_addr - load_addr). load_addr must be less than or equal to header_addr.

load_end_addr Contains the physical address of the end of the data segment. (load_end_addr - load_addr) specifies how much data to load. This implies that the text and data segments must be consecutive in the OS image; this is true for existing a.out executable formats. If this field is zero, the boot loader assumes that the text and data segments occupy the whole OS image file.

bss_end_addr Contains the physical address of the end of the bss segment. The boot loader initializes this area to zero, and reserves the memory it occupies to avoid placing boot modules and other data relevant to the operating system in that area. If this field is zero, the boot loader assumes that no bss segment is present.

entry_addr The physical address to which the boot loader should jump in order to start running the operating system.

22.1.4 The graphics fields of Multiboot header

All of the graphics fields are enabled by flag bit 2. They specify the preferred graphics mode. Note that that is only a recommended mode by the OS image. If the mode exists, the boot loader should set it, when the user doesn't specify a mode explicitly. Otherwise, the boot loader should fall back to a similar mode, if available.

The meaning of each is as follows:

mode_type Contains ‘0’ for linear graphics mode or ‘1’ for EGA-standard text mode.

Everything else is reserved for future expansion. Note that the boot loader may set a text mode, even if this field contains ‘0’.

width Contains the number of the columns. This is specified in pixels in a graphics mode, and in characters in a text mode. The value zero indicates that the OS image has no preference.

height Contains the number of the lines. This is specified in pixels in a graphics mode, and in characters in a text mode. The value zero indicates that the OS image has no preference.

depth Contains the number of bits per pixel in a graphics mode, and zero in a text mode. The value zero indicates that the OS image has no preference.

22.2 Machine state

When the boot loader invokes the 32-bit operating system, the machine must have the following state:

‘EAX’ Must contain the magic value ‘0x2BADB002’; the presence of this value indicate to the operating system that it was loaded by a Multiboot-compliant boot loader (e.g. as opposed to another type of boot loader that the operating system can also be loaded from).

‘EBX’ Must contain the 32-bit physical address of the Multiboot information structure provided by the boot loader (see Boot information format).

‘CS’ Must be a 32-bit read/execute code segment with an offset of ‘0’ and a limit of ‘0xFFFFFFFF’. The exact value is undefined.

‘DS’

‘ES’

‘FS’

‘GS’

'SS' Must be a 32-bit read/write data segment with an offset of '0' and a limit of '0xFFFFFFFF'. The exact values are all undefined.

'A20 gate' Must be enabled.

'CR0' Bit 31 (PG) must be cleared. Bit 0 (PE) must be set. Other bits are all undefined.

'EFLAGS' Bit 17 (VM) must be cleared. Bit 9 (IF) must be cleared. Other bits are all undefined.

All other processor registers and flag bits are undefined. This includes, in particular:

'ESP' The OS image must create its own stack as soon as it needs one.

'GDTR' Even though the segment registers are set up as described above, the 'GDTR' may be invalid, so the OS image must not load any segment registers (even just reloading the same values!) until it sets up its own 'GDT'.

'IDTR' The OS image must leave interrupts disabled until it sets up its own IDT.

However, other machine state should be left by the boot loader in normal working order, i.e. as initialized by the bios (or DOS, if that's what the boot loader runs from). In other words, the operating system should be able to make bios calls and such after being loaded, as long as it does not overwrite the bios data structures before doing so. Also, the boot loader must leave the pic programmed with the normal bios/DOS values, even if it changed them during the switch to 32-bit mode.

22.3 Boot information format

FIXME: Split this chapter like the chapter "OS image format".

Upon entry to the operating system, the EBX register contains the physical address of a Multiboot information data structure, through which the boot loader communicates vital information to the operating system. The operating system can use or ignore any parts of the structure as it chooses; all information passed by the boot loader is advisory only.

The Multiboot information structure and its related substructures may be placed anywhere in memory by the boot loader (with the exception of the memory reserved for the kernel and boot modules, of course). It is the operating system's responsibility to avoid overwriting this memory until it is done using it.

The format of the Multiboot information structure (as defined so far) follows:

0	+-----+ flags (required) +-----+
---	--

4	mem_lower	(present if flags[0] is set)
8	mem_upper	(present if flags[0] is set)
	+-----+	
12	boot_device	(present if flags[1] is set)
	+-----+	
16	cmdline	(present if flags[2] is set)
	+-----+	
20	mods_count	(present if flags[3] is set)
24	mods_addr	(present if flags[3] is set)
	+-----+	
28 - 40	syms	(present if flags[4] or flags[5] is set)
	+-----+	
44	mmap_length	(present if flags[6] is set)
48	mmap_addr	(present if flags[6] is set)
	+-----+	
52	drives_length	(present if flags[7] is set)
56	drives_addr	(present if flags[7] is set)
	+-----+	
60	config_table	(present if flags[8] is set)
	+-----+	
64	boot_loader_name	(present if flags[9] is set)
	+-----+	
68	apm_table	(present if flags[10] is set)
	+-----+	
72	vbe_control_info	(present if flags[11] is set)
76	vbe_mode_info	
80	vbe_mode	
82	vbe_interface_seg	
84	vbe_interface_off	
86	vbe_interface_len	
	+-----+	

The first longword indicates the presence and validity of other fields in the Multiboot information structure. All as-yet-undefined bits must be set to zero by the boot loader. Any set bits that the operating system does not understand should be ignored. Thus, the ‘flags’ field also functions as a version indicator, allowing the Multiboot information structure to be expanded in the future without breaking anything.

If bit 0 in the ‘flags’ word is set, then the ‘mem_*’ fields are valid. ‘mem_lower’ and ‘mem_upper’ indicate the amount of lower and upper memory, respectively, in kilobytes. Lower memory starts at address 0, and upper memory starts at address 1 megabyte. The maximum possible value for lower memory is 640 kilobytes. The value returned for upper memory is maximally the address of the first upper memory hole minus 1 megabyte. It is not guaranteed to be this value.

If bit 1 in the ‘flags’ word is set, then the ‘boot_device’ field is valid, and indicates which bios disk device the boot loader loaded the OS image from. If the OS image was not loaded from a bios disk, then this field must not be present (bit 3 must be clear). The operating system may use this field as a hint for determining its own root device, but is not required to. The ‘boot_device’ field is laid out in four one-byte subfields as follows:

+	-	-	-	-	-	+					
	part3		part2		part1		drive				
+	-	-	-	-	-	+	-	-	+	-	+

The first byte contains the bios drive number as understood by the bios INT 0x13 low-level disk interface: e.g. 0x00 for the first floppy disk or 0x80 for the first hard disk.

The three remaining bytes specify the boot partition. ‘part1’ specifies the top-level partition number, ‘part2’ specifies a sub-partition in the top-level partition, etc. Partition numbers always start from zero. Unused partition bytes must be set to 0xFF. For example, if the disk is partitioned using a simple one-level DOS partitioning scheme, then ‘part1’ contains the DOS partition number, and ‘part2’ and ‘part3’ are both 0xFF. As another example, if a disk is partitioned first into DOS partitions, and then one of those DOS partitions is subdivided into several BSD partitions using BSD’s disklabel strategy, then ‘part1’ contains the DOS partition number, ‘part2’ contains the BSD sub-partition within that DOS partition, and ‘part3’ is 0xFF.

DOS extended partitions are indicated as partition numbers starting from 4 and increasing, rather than as nested sub-partitions, even though the underlying disk layout of extended partitions is hierarchical in nature. For example, if the boot loader boots from the second extended partition on a disk partitioned in conventional DOS style, then ‘part1’ will be 5, and ‘part2’ and ‘part3’ will both be 0xFF.

If bit 2 of the ‘flags’ longword is set, the ‘cmdline’ field is valid, and contains the physical address of the command line to be passed to the kernel. The command line is a normal C-style zero-terminated string.

If bit 3 of the ‘flags’ is set, then the ‘mods’ fields indicate to the kernel what boot modules were loaded along with the kernel image, and where they can be found. ‘mods_count’ contains the number of modules loaded; ‘mods_addr’ contains the physical address of the first module structure. ‘mods_count’ may be zero, indicating no boot modules were loaded, even if bit 1 of ‘flags’ is set. Each module structure is formatted as follows:

0		mod_start				
4		mod_end				
8		string				
12		reserved (0)				
	+	-	-	-	-	+

The first two fields contain the start and end addresses of the boot module itself. The ‘string’ field provides an arbitrary string to be associated with that particular boot module; it is a zero-terminated ASCII string, just like the kernel command line. The ‘string’ field may be 0 if there is no string associated with the module. Typically the string might be a command line (e.g. if the operating system treats boot modules as executable programs), or a pathname (e.g. if the operating system treats boot modules as files in a file system), but its exact use is specific to the operating system. The ‘reserved’ field must be set to 0 by the boot loader and ignored by the operating system.

Caution: Bits 4 & 5 are mutually exclusive.

If bit 4 in the ‘flags’ word is set, then the following fields in the Multiboot information structure starting at byte 28 are valid:

	+-----+
28	tabsz
32	strsz
36	addr
40	reserved (0)
	+-----+

These indicate where the symbol table from an a.out kernel image can be found. ‘addr’ is the physical address of the size (4-byte unsigned long) of an array of a.out format nlist structures, followed immediately by the array itself, then the size (4-byte unsigned long) of a set of zero-terminated ascii strings (plus sizeof(unsigned long) in this case), and finally the set of strings itself. ‘tabsz’ is equal to its size parameter (found at the beginning of the symbol section), and ‘strsz’ is equal to its size parameter (found at the beginning of the string section) of the following string table to which the symbol table refers. Note that ‘tabsz’ may be 0, indicating no symbols, even if bit 4 in the ‘flags’ word is set.

If bit 5 in the ‘flags’ word is set, then the following fields in the Multiboot information structure starting at byte 28 are valid:

	+-----+
28	num
32	size
36	addr
40	shndx
	+-----+

These indicate where the section header table from an ELF kernel is, the size of each entry, number of entries, and the string table used as the index of names. They correspond to the ‘shdr_*’ entries (‘shdr_num’, etc.) in the Executable and Linkable Format (elf) specification in the program header. All sections are loaded, and the physical address fields of the elf section header then refer to where the sections are in

memory (refer to the i386 elf documentation for details as to how to read the section header(s)). Note that ‘shdr_num’ may be 0, indicating no symbols, even if bit 5 in the ‘flags’ word is set.

If bit 6 in the ‘flags’ word is set, then the ‘mmap_*’ fields are valid, and indicate the address and length of a buffer containing a memory map of the machine provided by the bios. ‘mmap_addr’ is the address, and ‘mmap_length’ is the total size of the buffer. The buffer consists of one or more of the following size/structure pairs (‘size’ is really used for skipping to the next pair):

-4	size	
0	base_addr	
8	length	
16	type	
		+-----+

where ‘size’ is the size of the associated structure in bytes, which can be greater than the minimum of 20 bytes. ‘base_addr’ is the starting address. ‘length’ is the size of the memory region in bytes. ‘type’ is the variety of address range represented, where a value of 1 indicates available ram, and all other values currently indicated a reserved area.

The map provided is guaranteed to list all standard ram that should be available for normal use.

If bit 7 in the ‘flags’ is set, then the ‘drives_*’ fields are valid, and indicate the address of the physical address of the first drive structure and the size of drive structures. ‘drives_addr’ is the address, and ‘drives_length’ is the total size of drive structures. Note that ‘drives_length’ may be zero. Each drive structure is formatted as follows:

0	size	
4	drive_number	
5	drive_mode	
6	drive_cylinders	
8	drive_heads	
9	drive_sectors	
10 - xx	drive_ports	
		+-----+

The ‘size’ field specifies the size of this structure. The size varies, depending on the number of ports. Note that the size may not be equal to $(10 + 2 * \text{the number of ports})$, because of an alignment.

The ‘drive_number’ field contains the BIOS drive number. The ‘drive_mode’ field represents the access mode used by the boot loader. Currently, the following modes are defined:

‘0’ CHS mode (traditional cylinder/head/sector addressing mode).

‘1’ LBA mode (Logical Block Addressing mode).

The three fields, ‘drive_cylinders’, ‘drive_heads’ and ‘drive_sectors’, indicate the geometry of the drive detected by the bios. ‘drive_cylinders’ contains the number of the cylinders. ‘drive_heads’ contains the number of the heads. ‘drive_sectors’ contains the number of the sectors per track.

The ‘drive_ports’ field contains the array of the I/O ports used for the drive in the bios code. The array consists of zero or more unsigned two-bytes integers, and is terminated with zero. Note that the array may contain any number of I/O ports that are not related to the drive actually (such as dma controller’s ports).

If bit 8 in the ‘flags’ is set, then the ‘config_table’ field is valid, and indicates the address of the rom configuration table returned by the GET CONFIGURATION bios call. If the bios call fails, then the size of the table must be zero.

If bit 9 in the ‘flags’ is set, the ‘boot_loader_name’ field is valid, and contains the physical address of the name of a boot loader booting the kernel. The name is a normal C-style zero-terminated string.

If bit 10 in the ‘flags’ is set, the ‘apm_table’ field is valid, and contains the physical address of an apm table defined as below:

Examples

History

Index

Часть XI

Технологии

Часть XII

Сетевое обучение

Часть XIII

Базовая теоретическая подготовка

Глава 23

Математика

23.1 Высшая математика в упражнениях и задачах [68]

В этом разделе будут размещены решения некоторых задач из [68] в “техническом” стиле: главное быстрый результат, а не точное аналитическое решение, поэтому будем использовать системы компьютерной математики. Будут рассмотрены приемы применения OpenSource пакетов:

Maxima [19] символьная математика, аналог **MathCAD**, on-line <http://maxima.org/>

Octave [21] численная математика, аналог **MATLAB**, on-line <http://octave-online.net/>

GNUPLOT [?] простейшее средство построения 3D/3D графиков

WolframAlpha <http://www.wolframalpha.com/> бесплатная on-line система символьной математики и база знаний, функционал и интерфейс очень ограничены, но вполне полезна в качестве **символьного калькулятора**

Python скриптовый язык программирования, в последнее время получил широкое применение в области численных методов, анализа данных и автоматизации, чаще всего применяется в комплекте с библиотеками:

NumPy поддержка многомерных массивов (включая матрицы) и высокочувственных математических функций, предназначенных для работы с ними

SciPy библиотека предназначенная для выполнения научных и инженерных расчётов: поиск минимумов и максимумов функций, вычисление интегралов функций, поддержка специальных функций, обработка сигналов, обработка изображений, работа с генетическими алгоритмами, решение обыкновенных дифференциальных уравнений,...

Sympy библиотека символьной математики <https://en.wikipedia.org/wiki/Sympy>

Matplotlib библиотека на языке программирования Python для 2D/3D визуализации данных. Получаемые изображения могут быть использованы в качестве иллюстраций в публикациях.

Подробно с применением *Python* при обработке данных можно ознакомиться в <http://scipy-cookbook.readthedocs.org/>

Также этот раздел можно использовать как пример использования системы верстки L^AT_EX для научных публикаций —смотрите **исходные тексты** файла <https://github.com/ponyatov/boox/tree/master/math/danko/danko.tex>.

Запуск **Maxima** и **Octave** в пакетном режиме

При запуске **Maxima**/**Octave** выводится информация о программе и license disclaim.
При их использовании в автоматическом режиме¹ требуется блокировать лишний вывод опцией -q. Как пример можно привести набор правил для **make**:

```
% .pdf: %.plot
    gnuplot $<
%.pdf: %.mac
    maxima -q < $<
%.log: %.mac
    maxima -q < $< > $@
%.pdf: %.m Makefile
    octave -q $< && pdfcrop o$@ $@
%.log: %.m Makefile
    octave -q $< > $@
```

\$@ **левая** часть make-правила

\$< **первый элемент** правой части правила

&& выполнить следующую команду только если предыдущая вернула код успешного выполнения `exit(0)`

pdfcrop <in> <out> **octave** выводит графики в полный лист А4, **pdfcrop** выполняет обрезку

23.1.1 Аналитическая геометрия на плоскости

Прямоугольные и полярные координаты

1. Координаты на прямой. Деление отрезка в данном отношении. Точку M координатной оси Ox , имеющую **абсциссу** x , обозначают через $M(x)$.

Расстояние d между точками $M_1(x_1)$ и $M_2(x_2)$ оси при любом расположении точек на оси находятся по формуле:

$$d = |x_2 - x_1| \quad (23.1)$$

¹ например в файлах Makefile 16.13

Пусть на произвольной прямой задан отрезок AB (A — начало отрезка, B — конец), тогда всякая третья точка C этой прямой делить отрезок AB в некотором отношении λ , где $\lambda = \frac{AC}{CB}$. Если отрезки AC и CB направлены в одну сторону, то λ приписывают знак “плюс”; если же отрезки AC и CB направлены в противоположные стороны, то λ приписывают знак “минус”. Иными словами, $\lambda > 0$ если точка C лежит между точками A и B ; $\lambda < 0$ если точка C лежит вне отрезка AB .

Пусть точки A и B лежат на оси Ox , тогда **координата точки $C(\bar{x})$** , делящей отрезок между точками $A(x_1)$ и $B(x_2)$ в отношении λ , находится по формуле:

$$\bar{x} = \frac{x_1 + \lambda x_2}{1 + \lambda} \quad (23.2)$$

В частности, при $\lambda = 1$ получается формула для координаты середины отрезка:

$$\bar{x} = \frac{x_1 + x_2}{2} \quad (23.3)$$

Формула 23.2 легко выводится из системы

$$\begin{cases} |A(x_1)C(\bar{x})| = \bar{x} - x_1 = a > 0 \Leftrightarrow \bar{x} > x_1 \\ |C(\bar{x})B(x_2)| = x_2 - \bar{x} = b > 0 \Leftrightarrow x_2 > \bar{x} \\ |A(x_1)B(x_2)| = x_2 - x_1 = a + b; \\ \lambda = a/b; \end{cases}$$

WolframAlpha

```
solve x-x1=a ; x2-x=b ; x2-x1=a+b ; lambda=a/b for x
Reduce[{ x-x1==a, x2-x==b, x2-x1==a+b, lambda==a/b },{x}]
```

- Построить на прямой точки $A(3)$, $B(-2)$, $C(0)$, $D(\sqrt{2})$, $E(-3.5)$.

WolframAlpha

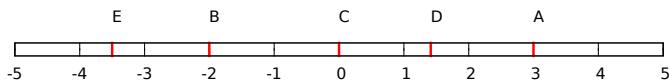


```
number line 3,-2,0,sqrt(2),-3.5 • 3 | • -2 | • 0 | • √2 | • -3.5
```

Листинг
GNUPLOT

22:

```
set terminal pdf
set output 'g_1_1_1.pdf'
set size ratio .02
unset key
unset ytics
set xtics 1
set label "A" at 3,3
set label "B" at -2,3
set label "C" at 0,3
set label "D" at sqrt(2),3
set label "E" at -3.5,3
plot [-5:+5][0:1] '-' u 1:2 w i lw 5
3 1
-2 1
0 1
1.4142 1
-3.5 1
e
```

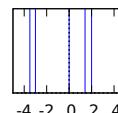


2

² $\sqrt{2}$ пришлось указать численно, значение функции не подставилось

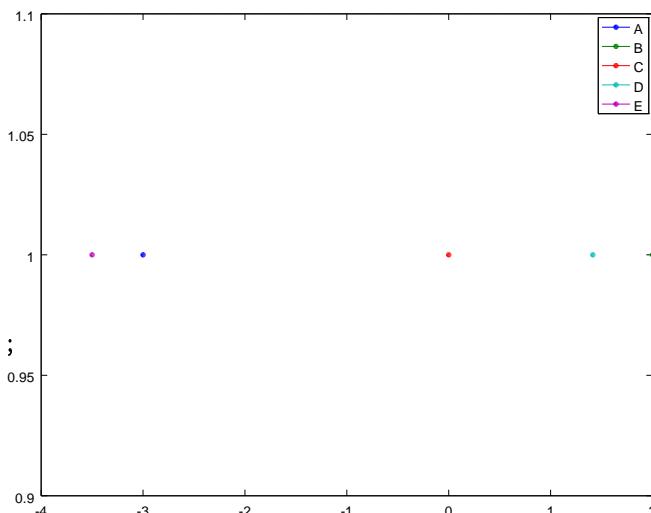
Листинг 23: Maxima

```
A:-3;  
B:2;  
C:0;  
D:sqrt(2);  
E:-3.5;  
  
dat: [[A,1],[B,1],[C,1],[D,1],[E,1]];  
  
plot2d([discrete,dat],\  
 [x,-5,+5],[y,0,1],\  
 [style,impulses],[yticks,false],\  
 [xlabel,false],[ylabel,false],\  
 [gnuplot_term,"pdf size 5,1"],\  
 [gnuplot_out_file,"./m_1_1_1.pdf"]);
```



Листинг 24: Octave

```
A=-3;  
B=2;  
C=0;  
D=sqrt(2);  
E=-3.5;  
  
plot(A,1,B,1,C,1,D,1,E,1)  
legend('A','B','C','D','E');  
print o_1_1_1.pdf
```



2. Отрезок AB четырьмя точками разделен на пять равных частей. Найти координату ближайшей к A точки деления, если $A(-3)$, $B(7)$.

Пусть $C(\bar{x})$ — искомая точка, тогда $\lambda = \frac{AC}{CB} = \frac{1}{4}$. Следовательно, по формуле 23.2 находим

$$C(\bar{x}) = \frac{x_1 + \lambda x_2}{1 + \lambda} = \frac{-3 + \frac{1}{4} \cdot 7}{1 + \frac{1}{4}} = C(-1)$$

Maxima

```
m_1_1_2 (x1 ,x2 ,lambda) := (x1+lambda*x2)/(1+lambda);  
A : -3 ;  
B : 7 ;  
lambda : 1/4 ;  
  
C = m_1_1_2(A,B,lambda);
```

Определяем функцию `m(maxima)` <глава> <параграф> <задача> (по нумерации задач в [68]), и вычисляем функцию с подстановкой числовых значений.

```
(%i1)  
(%o1)      m_1_1_2(x1 , x2 , lambda) := 
$$\frac{x1 + \text{lambda} \cdot x2}{1 + \text{lambda}}$$
  
(%i2) (%o2)  
(%i3) (%o3)  
(%i4)      - 3  
(%o4)      7  
           1  
           -  
           4  
(%i5) (%o5)      C = - 1  
(%i6)
```

В **Octave** **файлы с расширением .m** могут содержать не только последовательность команд, но и **выполнять роль определения библиотечной функции**. В этом случае имя функции должно совпадать с именем файла, где прописано ее определение.

Листинг 25: шаблон определения функции

```
function [<результат_1>,<результат_2>,..] = <имя> [<параметр_1>,<параметр_2>...  
    оператор_1;  
    ...  
    оператор_N;  
end;
```

Octave – o_1_1_2.m

```
function [xn] = o_1_1_2 (x1 ,x2 ,lambda)  
    xn = (x1+lambda*x2)/(1+lambda);  
end  
A = -3  
B = 7  
lambda = 1/4  
  
o_1_1_2(A,B,lambda)
```

Определяем функцию `o(ctave)_<глава>_<параграф>_<задача>` (по нумерации задач в [68]), и вычисляем функцию с подстановкой числовых значений.

```
A = -3
B = 7
lambda = 0.25000
ans = -1
```

3. Известны точки $A(1)$, $B(5)$ — концы отрезка AB ; вне этого отрезка расположена точка C , причем ее расстояние от точки A в 3 раза больше расстояния от точки B . Найти координату точки C .

Нетрудно установить что $\lambda = -\frac{AC}{BC} = -3$, таким образом

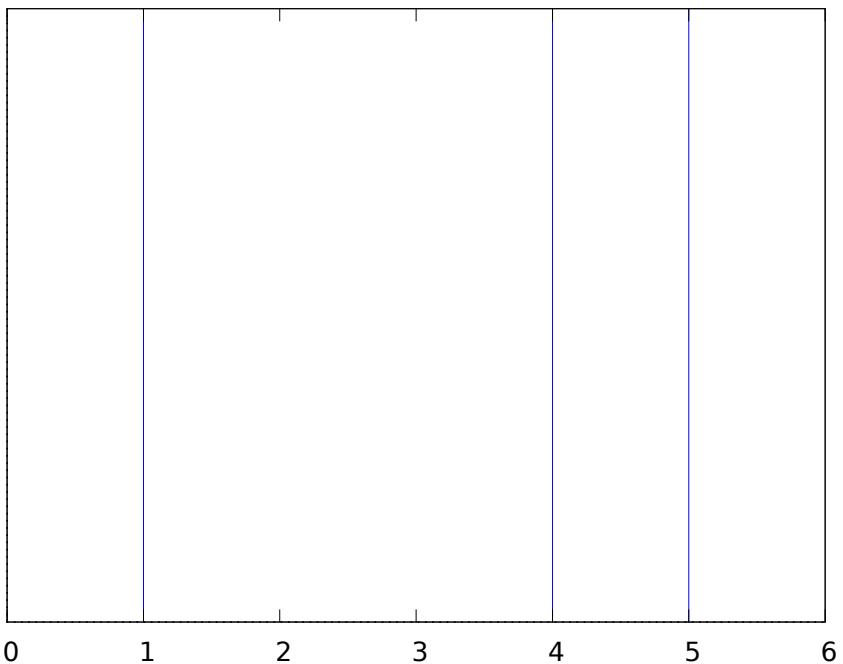
$$C(\bar{x}) = \frac{1 - 3 \cdot 5}{1 - 3} = C(7) \quad (23.4)$$

Maxima

```
A:1;
B:5;
lambda:3;
C:(A+lambda*B)/(A+lambda);

dat: [[A,1],[B,1],[C,1]];

plot2d ([ discrete ,dat ], \
[x,A-1,B+1],[y,0,1], \
[ style , impulses ] , \
[ xlabel , false ] , [ ylabel , false ] , [ ytics , false ] , \
[ gnuplot _term , pdf ] , \
[ gnuplot _out _file , "./m_1_1_3.pdf "]);
```



4. Найти расстояние между точками

1. $M(3) N(-5)$

Python

```
abs( (-5) - (3) )
8
```

2. $P(-5.5) Q(-2.5)$

Python

```
def distance(a,b)
    return abs(a-b)
```

```
distance( -5.5 , -2.5 )
3.0
```

5. Найти координаты середины отрезка, если известны его концы³:

1. $A(-6) B(7)$

2. $C(-5) D(0.5)$

³ используем формулу 23.3

```
% [danko3] equation:
function midpoint = danko3 (x1 , x2)
    midpoint = (x1+x2)/2;
end

danko3( -6 , 7 )
danko3( -5 , 0.5 )
```

Листинг 26:

```
ans = 0.50000
ans = -2.2500
```

6. Найти точку M , симметричную точке $N(-3)$ относительно точки $P(2)$.

$$N(x_1)P(\bar{x}) = P(\bar{x})M(x_2)$$

Из 23.3:

$$2 = \frac{(-3) + x_2}{2}$$

WolframAlpha solve N=-3;P=2;P=(N+M)/2 for M \Rightarrow 7

Maxima

```
N: -3;
P: 2;
solve (P=(N+M)/2 ,M);
```

m_1_1_6.log

(%i1) (%o1)	- 3
(%i2) (%o2)	2
(%i3) (%o3)	[M = 7]
(%i4)	

Часть XIV

Прочее

Ф.И.Атауллаханов об учебниках США и России

© Доктор биологических наук Фазли Иноятович Атауллаханов.
МГУ им. М. В. Ломоносова, Университет Пенсильвании, США

<http://www.nkj.ru/archive/articles/19054/>

...

У необходимости рекламировать науку есть важная обратная сторона: каждый американский учёный непрерывно, с первых шагов и всегда, учится излагать свои мысли внятно и популярно. В России традиции быть понятными у учёных нет. Как пример я люблю приводить двух великих физиков: русского Ландау и американца Фейнмана. Каждый написал многотомный учебник по физике. Первый — знаменитый “Ландау-Лифшиц”, второй — “Лекции по физике”. Так вот, “Ландау-Лифшиц” прекрасный справочник, но представляет собой полное издательство над читателем. Это типичный памятник автору, который был, мягко говоря, малоприятным человеком. Он излагает то, что излагает, абсолютно пре-небрегая своим читателем и даже издеваясь над ним. А у нас целые поколения выросли на этой книге, и считается, что всё нормально, кто справился, тот младец. Когда я столкнулся с “Лекциями по физике” Фейнмана, я просто обалдел: оказывается, можно по-человечески разговаривать со своими коллегами, со студентами, с аспирантами. Учебник Ландау — пример того, как устроена у нас вся наука. Берёшь текст русской статьи, читаешь с самого начала и ничего не можешь понять, а иногда сомневаешься, понимает ли автор сам себя. Конечно, крупицы осмысленного и разумного и оттуда можно вынуть. Но автор явно считает, что это твоя работа — их оттуда извлечь. Не потому, что он не хочет быть понятым, а потому, что его не научили правильно писать. Не учат у нас человека ни писать, ни говорить внятно, это считается неважным.

...

Думаю, американская наука в целом устроена именно так: она продаёт не просто себя, а всю свою страну. Сегодня американцы дороги не метут, сапоги не тачают, даже телевизоры не собирают, за них это делает весь остальной мир. А что же делают американцы? Самая богатая страна в мире? Они объяснили, в первую очередь самим себе, а заодно и всему миру, что они — мозг планеты. Они изобретают. “Мы придумываем продукты, а вы их делайте. В том числе и для нас”. Это прекрасно работает, поэтому они очень ценят науку.

...

Глава 24

Настройка редактора/IDE **(g)Vim**

При использовании редактора/IDE **(g)Vim** удобно настроить сочетания клавиш и подсветку синтаксиса языков, которые вы используете так, как вам удобно.

24.1 для вашего собственного скриптового языка

Через какое-то время практики FSP у вас выработается один диалект скриптов для всех программ, соответствующий именно вашим вкусам в синтаксисе, и в этом случае его нужно будет описать только в файлах `/.vim/(ftdetect|syntax).vim`, и привязать их к расширениям через dot-файлы **(g)Vim** в вашем домашнем каталоге:

<code>filetype.vim</code>	(g)Vim	привязка расширений файлов (.src .lo
<code>syntax.vim</code>	(g)Vim	синтаксическая подсветка для скрипт
<code>/.vimrc</code>	<i>Linux</i>	настройки для пользователя
<code>/vimrc</code>	<i>Windows</i>	
<code>/.vim/ftdetect/src.vim</code>	<i>Linux</i>	привязка команд к расширению .src
<code>/vimfiles/ftdetect/src.vim</code>	<i>Windows</i>	
<code>/.vim/syntax/src.vim</code>	<i>Linux</i>	синтаксис к расширению .src
<code>/vimfiles/syntax/src.vim</code>	<i>Windows</i>	

Книги must have любому техническому специалисту

Математика, физика, химия

- Бермант **Математический анализ** [35]
- Тихонов, Самарский **Математическая физика** [44, 69]
- Демидович, Марон **Численные методы** [49, 50]
- Кремер **Теория вероятностей и матстатистика** [40]
- Ван дер Варден **Математическая статистика** [36]
- Кострикин **Введение в алгебру** [38, 39]

- Ван дер Варден **Алгебра** [37]

- Демидович **Сборник задач по математике для втузов. В 4 частях** [70, ?, ?, ?]
- Будак, Самарский, Тихонов **Сборник задач по математической физике** [69]

Фейнмановские лекции по физике

1. Современная наука о природе. Законы механики. [55]
2. Пространство. Время. Движение. [56]
3. Излучение. Волны. Кванты. [57]
4. Кинетика. Теплота. Звук. [58]
5. Электричество и магнетизм [59]
6. Электродинамика. [60]
7. Физика сплошных сред. [61]
8. Квантовая механика 1. [62]
9. Квантовая механика 2. [63]

- Цирельсон **Квантовая химия** [65]
- Розенброк **Вычислительные методы для инженеров-химиков** [66]
- Шрайвер Эткинс **Неорганическая химия** [67]

Обработка экспериментальных данных и метрология

- Смит **Цифровая обработка сигналов** [41]
- Князев, Черкасский **Начала обработки экспериментальных данных** [42]

Программирование

- **Система контроля версий Git и git-хостинга GitHub**
хранение наработок с полной историей редактирования, правок, релизов для разных заказчиков или вариантов использования
- **Язык Python** [26]
написание скриптов обработки данных, автоматизации, графических оболочек и т.п. утилит
- **JavaScript** [24] + **HTML**
генерация отчетов и ввод исходных данных, интерфейс к сетевым расчетным серверам на *Python*, простые браузерные граф.интерфейсы и расчетки
- **Реляционные (и объектные) базы данных** /MySQL, Postgres (,ZODB,GC)
хранение и простая черновая обработка табличных (объектных) данных экспериментов, справочников, настроек, пользователей.

- Язык C_+ , утилиты GNU toolchain [22, 23] (gcc/g++, make, ld)
базовый Си, ООП очень кратко¹, без излишеств профессионального программирования², чисто вспомогательная роль для написания вычислительных блоков и критичных к скорости/памяти секций, использовать в связке с Python.
Знание базового Си **критично при использовании микроконтроллеров**, из C_+ необходимо владение особенностями использования ООП и управления крайне ограниченной памятью: пользовательские менеджеры памяти, статические классы.
- Использование утилит **flex/bison**
обработка текстовых форматов данных, часто необходимая вещь.

САПР, пакеты математики, моделирования, визуализации

- **Maxima** символьная математика [19]
- **Octave** численные методы [21]
- **GNUPLOT** простой вывод графиков
- **ParaView/VTK** навороченнейший пакет/библиотека визуализации всех видов
- **LATEX** верстка научных публикаций и генерация отчетов
- **KiCAD + ng-spice** электроника: расчет схем и проектирование печатных плат
- **FreeCAD** САПР общего назначения
- **Elmer, OpenFOAM** расчетные пакеты метода конечных элементов (мультифизика, сопротивление материалов, конструкционная устойчивость, газовые и жидкостные потоки, теплопроводность)
- **CodeAster + Salome** пакет МКЭ, особо заточенный под сопромат и расчет конструкций
- **OpenModelica** симуляция моделей со средоточенными параметрами³ (электроника, электротехника, механика, гидропневмоавтоматика и системы управления)
- **V-REP** робототехнический симулятор
- **SimChemistry**⁴ интересный демонстрационный симулятор химической кинетики молекул на микроуровне (обсчитывается движение и столкновение отдельных молекул)
- **Avogadro** 3D редактор молекул

¹ наследование, полиморфизм, операторы для пользовательских типов, использование библиотеки STL

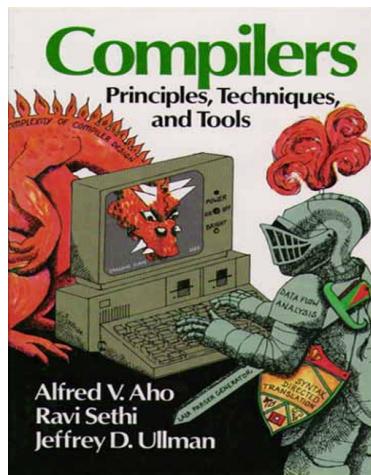
² мегабиблиотека Boost, написание своих библиотек шаблонов и т.п.

³ для описания моделей элементов использует ООП-язык Modelica

⁴ Windows

Литература

Разработка языков программирования и компиляторов



[1] **Dragon Book**

Компиляторы. Принципы, технологии, инструменты.

Альфред Ахо, Рави Сети, Джейфри Ульман.

Издательство Вильямс, 2003.

ISBN 5-8459-0189-8

[2] **Compilers: Principles, Techniques, and Tools**

Aho, Sethi, Ullman

Addison-Wesley, 1986.

ISBN 0-201-10088-6

**Structure and
Interpretation
of Computer
Programs**

Second Edition



Harold Abelson and
Gerald Jay Sussman
with Julie Sussman

SICP

[3]

Структура и интерпретация компьютерных программ

Харольд Абельсон, Джеральд Сассман

ISBN 5-98227-191-8

EN: web.mit.edu/alexmv/6.037/sicp.pdf



[4]

Функциональное программирование

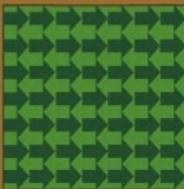
Филд А., Харрисон П.

М.: Мир, 1993

ISBN 5-03-001870-0

МАТЕМАТИЧЕСКОЕ
ОБЕСПЕЧЕНИЕ
ЭВМ

П.Хендерсон
ФУНКЦИОНАЛЬНОЕ
ПРОГРАММИРОВАНИЕ
Применение
и реализация



[5]

Функциональное программирование: применение и реализация

П.Хендерсон

М.: Мир, 1983



LLVM: инфраструктура
для разработки компиляторов

Бруно Кардос Лопес

Рафаэль Аулер



[6]

LLVM. Инфраструктура для разра-

ботки компиляторов

Бруно Кардос Лопес, Рафаэль Аулер

Lisp/Sheme

Haskell

ML

[7] <http://homepages.inf.ed.ac.uk/mfourman/teaching/mlCourse/notes/L01.pdf>

Basics of Standard ML

© Michael P. Fourman

перевод 1

[8] <http://www.soc.napier.ac.uk/course-notes/sml/manual.html>

A Gentle Introduction to ML

© Andrew Cumming, Computer Studies, Napier University, Edinburgh

[9] <http://www.cs.cmu.edu/~rwh/smlbook/book.pdf>

Programming in Standard ML

© Robert Harper, Carnegie Mellon University

Электроника и цифровая техника



[10]

An Introduction to Practical Electronics, Microcontrollers and Software Design

Bill Collis

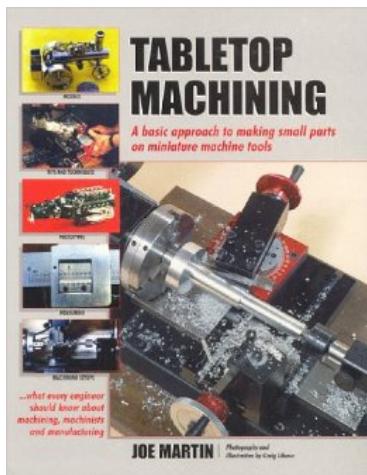
2 edition, May 2014

<http://www.techideas.co.nz/>

Конструирование и технология

Приемы ручной обработки материалов

Механообработка



[11]

Tabletop Machining

Martin, Joe and Libuse, Craig
Sherline Products, 2000

[12] Home Machinists Handbook

Briney, Doug, 2000

[13] Маленькие станки

Евгений Васильев

Псков, 2007

<http://www.coilgun.ru/stanki/index.htm>

Использование OpenSource программного обеспечения

LATEX



[14]

Набор и вёрстка в системе LATEX

С.М. Львовский

3-е издание, исправленное и дополненное, 2003

<http://www.mccme.ru/free-books/llang/newllang.pdf>



[15]

LATEX 2ε по-русски И. Котельников, П. Чеботаев

ISBN: 5-87550-195-2

[16] e-Readers and LATEX

Alan Wetmore

<https://www.tug.org/TUGboat/tb32-3/tb102wetmore.pdf>

[17] How to cite a standard (ISO, etc.) in BibLATEX ?
<http://tex.stackexchange.com/questions/65637/>

Математическое ПО: Maxima, Octave, GNUPLOT,..

[18] Система аналитических вычислений Maxima для физиков-теоретиков
Б.А. Ильина, П.К.Силаев
<http://tex.bog.msu.ru/numtask/max07.ps>



[19] Компьютерная математика с Maxima
Евгений Чичкарев

[20] Graphics with Maxima
Wilhelm Haager



[21] Введение в Octave для инженеров и математиков

САПР, электроника, проектирование печатных плат

Программирование

GNU Toolchain

- [22] **Embedded Systems Programming in C₊⁺**
© <http://www.bogotobogo.com/>
<http://www.bogotobogo.com/cplusplus/embeddedSystemsProgramming.php>
- [23] **Embedded Programming with the GNU Toolchain**
Vijay Kumar B.
<http://bravegnu.org/gnu-eprog/>

JavaScript, HTML, CSS, Web-технологии:

- [24] **On-line пошаговый учебник JavaScript** на английском, поддерживает множество языков и ИТ-технологий, курс очень удобен и прост для совсем начинающих <https://www.codecademy.com>
- [25] On-line учебник *JavaScript* на русском <http://learn.javascript.ru/>

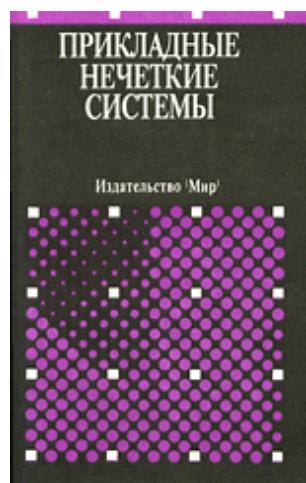
Python

- [26] **Язык программирования Python**
Россум, Г., Дрейк, Ф.Л.Дж., Откидач, Д.С., Задка, М., Левис, М., Монтаро, С., Реймонд, Э.С., Кучлинг, А.М., Лембург, М.-А., Йи, К.-П., Ксиллаг, Д., Петрилли, Х.Г., Варсав, Б.А., Ахлстром, Дж.К., Роскинд, Дж., Шеменор, Н., Муландер, С.
© Stichting Mathematisch Centrum, 1990–1995 and Corporation for National Research Initiatives, 1995–2000 and BeOpen.com, 2000 and Откидач, Д.С., 2001
<http://rus-linux.net/MyLDP/BOOKS/python.pdf>

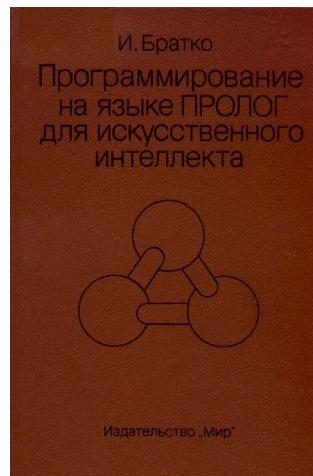
Python является простым и, в то же время, мощным интерпретируемым объектно-ориентированным языком программирования. Он предоставляет структуры данных высокого уровня, имеет изящный синтаксис и использует

динамический контроль типов, что делает его идеальным языком для быстрого написания различных приложений, работающих на большинстве распространенных платформ. Книга содержит вводное руководство, которое может служить учебником для начинающих, и справочный материал с подробным описанием грамматики языка, встроенных возможностей и возможностей, предоставляемых модулями стандартной библиотеки. Описание охватывает наиболее распространенные версии Python: от 1.5.2 до 2.0.

Prolog и логическое программирование

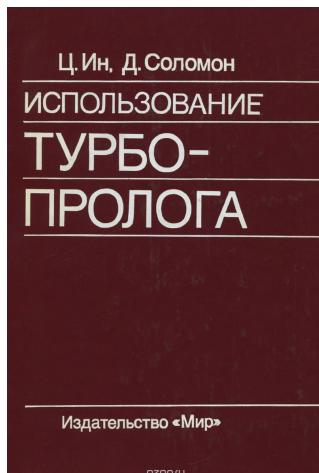


[27] Прикладные нечеткие системы
Тэрано Т., Асай К., Сугэно М. [djvu](#)

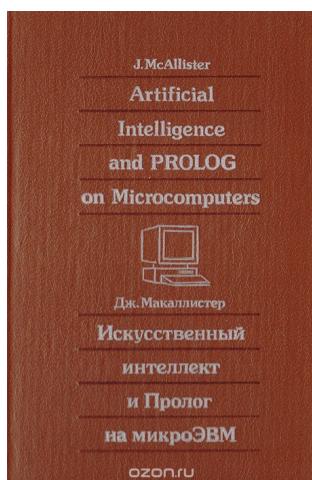


[28] Программирование на языке Пролог для искусственного интеллекта
Иван Братко

Мир, 1990
ISBN 5-03-001425-X, 0-201-14224-4



- [29] **Использование Турбо-Пролога**
Чин Маун Ин, Дэвид Соломон
Мир, 1993
ISBN 5-03-001181-1



- [30] **Искусственный интеллект и Пролог на микроЭВМ**
Дж. Макаллистер
Машиностроение, 1990
ISBN 5-217-00973-X

- [31] **Программирование на языке Пролог**
Клоксин У., Меллиш К.
Мир, 1987

- [32] **Искусство программирования на языке Пролог**
Л. Стерлинг, Э. Шапиро

Мир 1990

ISBN: 5-0300-0406-8

- [33] Интеллектуальные информационные системы. PROLOG- язык разработки интеллектуальных и экспертных систем: учебное пособие Хабаров С.П.
СПб. СПбГЛТУ, 2013.- 138 с. [pdf](#)

Разработка операционных систем и низкоуровневого ПО

- [34] OSDev Wiki
<http://wiki.osdev.org>

Базовые науки

Математика



- [35] Краткий курс математического анализа для ВТУЗов
Бермант А.Ф., Араманович И.Г.
М.: Наука, 1967
<https://drive.google.com/file/d/0B0u4WeMj0894U1Y1dEJ6cnCxU28/view?usp=sharing>

Пятое издание известного учебника, охватывает большинство вопросов программы по высшей математике для инженерно-технических специальностей вузов, в том числе дифференциальное исчисление функций одной переменной

и его применение к исследованию функций; дифференциальное исчисление функций нескольких переменных; интегральное исчисление; двойные, тройные и криволинейные интегралы; теорию поля; дифференциальные уравнения; степенные ряды и ряды Фурье. Разобрано много примеров и задач из различных разделов механики и физики. **Отличается крайней доходчивостью и отсутвием филонианов и “легко догадаться”.**

- [36] Математическая статистика Б.Л. Ван дер Варден
- [37] Алгебра Б.Л. Ван дер Варден
- [38] Введение в алгебру. В 3 частях. Часть 1. Основы алгебры А.И. Ко стрикин
- [39] Введение в алгебру. В 3 частях. Линейная алгебра. Часть 2 А.И. Кострикин



- [40] Теория вероятностей и математическая статистика
Наум Кремер
М.: Юнити, 2010



[41]

Цифровая обработка сигналов. Практическое руководство для инженеров и научных работников

Стивен Смит

Додэка XXI, 2008

ISBN 978-5-94120-145-7

В книге изложены основы теории цифровой обработки сигналов. Акцент сделан на доступности изложения материала и объяснении методов и алгоритмов так, как они понимаются при практическом использовании. Цель книги - практический подход к цифровой обработке сигналов, позволяющий преодолеть барьер сложной математики и абстрактной теории, характерных для традиционных учебников. Изложение материала сопровождается большим количеством примеров, иллюстраций и текстов программ

[42] Начала обработки экспериментальных данных

Б.А.Князев, В.С.Черкасский

Новосибирский государственный университет, кафедра общей физики, Новосибирск, 1996

http://www.phys.nsu.ru/cherk/Metodizm_old.PDF

Учебное пособие предназначено для студентов естественно-научных специальностей, выполняющих лабораторные работы в учебных практикумах. Для его чтения достаточно знаний математики в объеме средней школы, но оно может быть полезно и тем, кто уже изучил математическую статистику, поскольку исходным моментом в нем является не математика, а эксперимент. Во второй части пособия подробно описан реальный эксперимент — от появления идеи и проблем постановки эксперимента до получения результатов и обработки данных, что позволяет получить менее формализованное представление о применении математической статистики. Пособие дополнено обучающей программой, которая позволяет как углубить и уточнить знания, полученные в методическом пособии, так и проводить собственно обработку результатов лабораторных работ. Приведен список литературы для желаю-

щих углубить свои знания в области математической статистики и обработки данных.



[43]

Принципы современной математической физики Р. Рихтмайер

[44] Уравнения математической физики А.Н. Тихонов, А.А. Самарский

Символьная алгебра

[45] Компьютерная алгебра

Панкратьев Евгений Васильевич
МГУ, 2007

Настоящее пособие составлено на основе спецкурсов, читавшихся автором на механико-математическом факультете в течение более 10 лет. Выбор материала в значительной мере определялся пристрастиями автора. Наряду с классическими результатами компьютерной алгебры в этих спецкурсах (и в настоящем пособии) нашли отражение исследования нашего коллектива. Прежде всего, это относится к теории дифференциальной размерности.

Е. В. ПАНКРАТЬЕВ

ЭЛЕМЕНТЫ
КОМПЬЮТЕРНОЙ
АЛГЕБРЫ



ozon.ru

[46]

Элементы компьютерной алгебры

Евгений Панкратьев

Год выпуска 2007

ISBN 978-5-94774-655-6, 978-5-9556-0099-4

Учебник посвящен описанию основных структур данных и алгоритмов, применяемых в символьных вычислениях на ЭВМ. В книге затрагивается широкий круг вопросов, связанных с вычислениями в кольцах целых чисел, многочленов и дифференциальных многочленов.



[47]

Элементы абстрактной и компьютерной алгебры

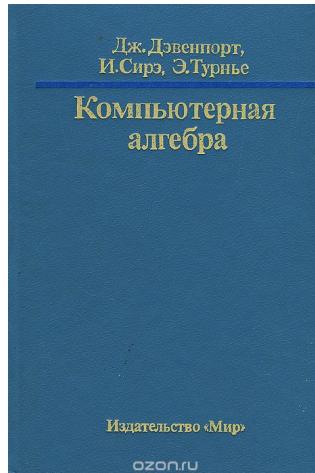
Дмитрий Матрос, Галина Поднебесова

2004

ISBN 5-7695-1601-1

В книгу включены следующие главы: алгебры, введение в системы компьютерной алгебры, кольцо целых чисел, полиномы от одной переменной, полиномы от нескольких переменных, формальное интегрирование, кодирование. Разбор доказательств утверждений и выполнение упражнений, приведенных

в учебном пособии, позволяют студентам овладеть методами решения практических задач, навыками конструирования алгоритмов.



[48]

Компьютерная алгебра

Дж.Дэвенпорт, И.Сирэ, Э.Турнье

Книга французских специалистов, охватывающая различные вопросы компьютерной алгебры: проблему представления данных, полиномиальное упрощение, современные алгоритмы вычисления НОД полиномов и разложения полиномов на множители, формальное интегрирование, применение систем компьютерной алгебры. Первый автор знаком читателю по переводу его книги "Интегрирование алгебраических функций"

(М.: Мир, 1985).

Численные методы

[49] **Основы вычислительной математики**

Борис Демидович, Исаак Марон

Книга посвящена изложению важнейших методов и приемов вычислительной математики на базе общего вузовского курса высшей математики. Основная часть книги является учебным пособием по курсу приближенных вычислений для вузов.

[50] **Численные методы анализа. Приближение функций, дифференциальные и интегральные уравнения**

Б. П. Демидович, И. А. Марон, Э. З. Шувалова

В книге излагаются избранные вопросы вычислительной математики, и по содержанию она является продолжением учебного пособия [49]. Настоящее, третье издание отличается от предыдущего более доходчивым изложением. Добавлены новые примеры.

Теория игр

[51] Теория игр

Петросян Л. А. Зенкевич Н.А., Семина Е.А.

Учеб. пособие для ун-тов. — М.: Высш. шк., Книжный дом «Университет», 1998.

ISBN 5-06-001005-8, 5-8013-0007-4.

[52] Математическая теория игр и приложения

Мазалов В.В.

Санкт-Петербург - Москва - Краснодар: Лань, 2010.

ISBN 978-5-8114-1025-5.

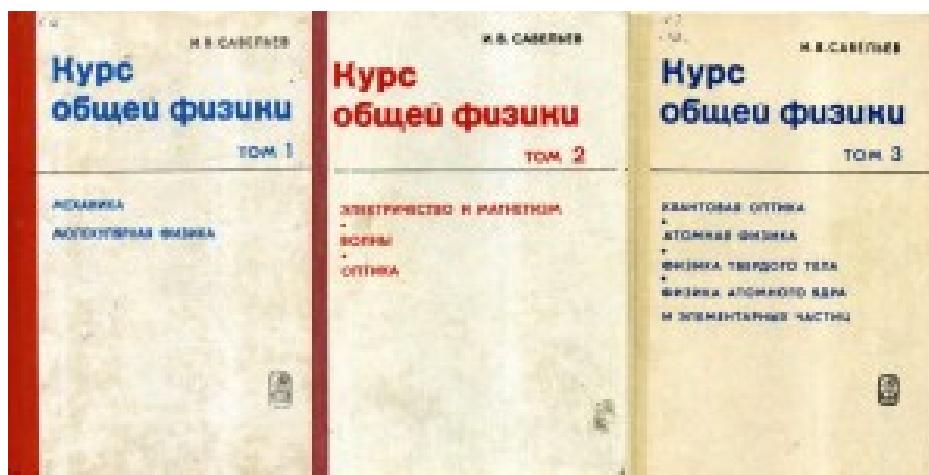
[53] Теория игр

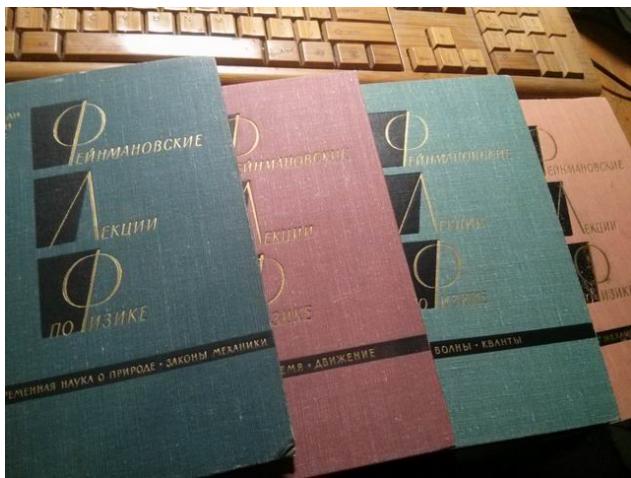
Оуэн Г.

Книга представляет собой краткое и сравнительно элементарное учебное пособие, пригодное как для первоначального, так и для углубленного изучения теории игр. Для ее чтения достаточно знания элементов математического анализа и теории вероятностей.

Книга естественно делится на две части, первая из которых посвящена играм двух лиц, а вторая — играм N лиц. Она охватывает большинство направлений теории игр, включая наиболее современные. В частности, рассмотрены антагонистические игры, игры двух лиц с ненулевой суммой и основы классической кооперативной теории. Часть материала в монографическом изложении появляется впервые. Каждая глава снабжена задачами разной степени сложности.

Физика





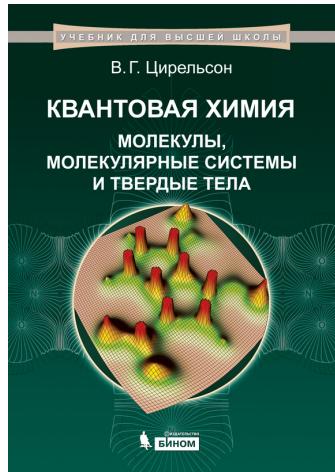
Фейнмановские лекции

по физике

Ричард Фейнман, Роберт Лейтон, Мэттью Сэндс

- [55] Современная наука о природе. Законы механики.
- [56] Пространство. Время. Движение.
- [57] Излучение. Волны. Кванты.
- [58] Кинетика. Теплота. Звук.
- [59] Электричество и магнетизм.
- [60] Электродинамика.
- [61] Физика сплошных сред.
- [62] Квантовая механика 1.
- [63] Квантовая механика 2.
- [64] Основы квантовой механики Д.И. Блохинцев

Химия



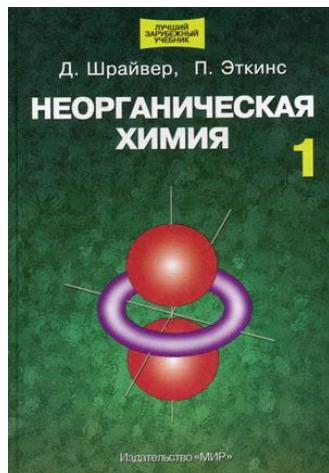
[65]

Квантовая химия. Молекулы, молекулярные системы и твердые тела. Учебное пособие Владимир Цирельсон



[66]

Вычислительные методы для инженеров-химиков X. Розенброк, С. Стори



[67]

Неорганическая химия В 2 томах
Д. Шрайвер, П. Эткинс

Задачники

Математика



[68]

Высшая математика в упражнениях и задачах
П.Е. Данко, А.Г.Попов, Т.Я. Кожевникова, С.П. Данко

[69] **Сборник задач по математической физике** Будак Б.М., Самарский А.А., Тихонов А.Н.

[70] **Сборник задач по математике для втузов. В 4 частях. Часть 1. Линейная алгебра и основы математического анализа**
Демидович

Стандарты и ГОСТы

- [71] 2.701-2008 Схемы. Виды и типы. Общие требования к выполнению
http://rtu.samgtu.ru/sites/rtu.samgtu.ru/files/GOST_ESKD_2.701-2008.pdf

Предметный указатель

- Prolog*,
 IeC функтор, 84
 IeC тэг, 85
арность, 84
константа, 84
куча, 85
несвязанная переменная, 85
переменная, 84
регистр, 87
сплющенная форма, 88
структурная ячейка, 86
субтерм, 84
терм, 84
терм программы, 85
терм запроса, 85
ячейка функтора, 86
ячейка переменной, 85
„, 147
:, 112
абсцисса, 221
адрес хранения, 168
адрес размещения, 168
анонимная переменная, 28
базовый адрес, 149
бинарный формат, 149
биндинг логической переменной, 28
цель (Пролог), 34
дерево вывода, 36, 46
дерево заключений, 32
факт, 31, 34
граф смежности, 29
грамматика, 113
инкрементная компоновка, 157
канадский крест, 190
компоновка, 157
конъюнктивная цель, 28
консеквенция, 32
координата точки, 222
линкер, 148
линковка, 149
логическая переменная, 28
монитор *Qemu*, 151
назначение адресов, 149
низкоуровневое программирование, 144
объектный код, 147
оператор, 106
переменная цели, 34
правило, 34
привязка логической переменной, 28
разрешение символов, 158
релокация символов, 159
секционирование, 160
семантическое дерево, 29
символ, 106, 155
символьный тип, 105
синтаксическое дерево, 113
скрипт линкера, 164
состояние лексера, 115
спецификация MultiBoot, 200, 202
строчный комментарий, 114
таблица символов, 155
токен, 113
трассировка, 36
указатель адреса размещения, 164
унификация, 39, 46
вывод *Prolog*-программы, 32
заголовок правила, 34
загрузчик, 200
ABI, 166

application term, 68

backtracking, 47

bare metal, 144

closure, 69

cut, 49

ELF, 149

lambda term, 68

LMA, 168

make-правило, 181

standalone, 144

startup код, 168

VMA, 168